

An Evaluation of Non-Gaussian Data Assimilation Methods in Moist Convective Regimes

JOSHUA MCCURRY,^a JONATHAN POTERJOY,^a KENT KNOPFMEIER,^b AND LOUIS WICKER^c

^a *University of Maryland, College Park, College Park, Maryland*

^b *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*

^c *NOAA/National Severe Storms Laboratory, Norman, Oklahoma*

(Manuscript received 28 September 2022, in final form 14 March 2023, accepted 14 March 2023)

ABSTRACT: Obtaining a faithful probabilistic depiction of moist convection is complicated by unknown errors in subgrid-scale physical parameterization schemes, invalid assumptions made by data assimilation (DA) techniques, and high system dimensionality. As an initial step toward untangling sources of uncertainty in convective weather regimes, we evaluate a novel Bayesian data assimilation methodology based on particle filtering within a WRF ensemble analysis and forecasting system. Unlike most geophysical DA methods, the particle filter (PF) represents prior and posterior error distributions nonparametrically rather than assuming a Gaussian distribution and can accept any type of likelihood function. This approach is known to reduce bias introduced by Gaussian approximations in low-dimensional and idealized contexts. The form of PF used in this research adopts a dimension-reduction strategy, making it affordable for typical weather applications. The present study examines posterior ensemble members and forecasts for select severe weather events between 2019 and 2020, comparing results from the PF with those from an ensemble Kalman filter (EnKF). We find that assimilating with a PF produces posterior quantities for microphysical variables that are more consistent with model climatology than comparable quantities from an EnKF, which we attribute to a reduction in DA bias. These differences are significant enough to impact the dynamic evolution of convective systems via cold pool strength and propagation, with impacts to forecast verification scores depending on the particular microphysics scheme. Our findings have broad implications for future approaches to the selection of physical parameterization schemes and parameter estimation within preexisting data assimilation frameworks.

SIGNIFICANCE STATEMENT: The accurate prediction of severe storms using numerical weather models depends on effective parameterization schemes for small-scale processes and the assimilation of incomplete observational data in a manner that faithfully represents the probabilistic state of the atmosphere. Current generation methods for data assimilation typically assume a standard form for the error distributions of relevant quantities, which can introduce bias that not only hinders numerical prediction, but that can also confound the characterization of errors from the model itself. The current study performs data assimilation using a novel method that does not make such assumptions and explores characteristics of resulting model fields and forecasts that might make such a method useful for improving model parameterization schemes.

KEYWORDS: Bayesian methods; Severe storms; Kalman filters; Ensembles; Cloud parameterizations; Data assimilation

1. Introduction

As ensemble sizes become large, particle filter (PF) approaches to data assimilation (e.g., Doucet et al. 2001) present theoretical advantages for weather prediction, mainly by avoiding the Gaussian assumptions inherent to contemporary parametric data assimilation methods. Imbalances in forecast initial conditions caused by such assumptions commonly lead to transient adjustment effects and degradation of forecast performance at extended lead times and small spatial scales, acting in a way that is difficult to distinguish from other sources of error (e.g., Poterjoy et al. 2017; Poterjoy 2022a). As nonparametric methods, PFs are capable of quantifying complex non-Gaussian prior probability distributions such as those characteristic of modeled physical relationships between hydrometeor mixing ratios and radar reflectivity.

Standard PFs have long been impractical for use with high-dimensional geophysical models due to the ensemble sizes required for moderate- to high-dimensional applications (e.g., Bengtsson et al. 2008; Bickel et al. 2008; Snyder et al. 2008). However, increased affordability of PFs has recently followed from approximations such as localization, which directly modulate dependence between state variables updated by data assimilation; see Farchi and Bocquet (2018) and van Leeuwen et al. (2019) for a review. The localized particle filter (LPF) (Poterjoy 2016; Poterjoy et al. 2019; Poterjoy 2022b) adopted for this study limits the spatial influence of update steps in a manner similar to that of covariance localization in Gaussian data assimilation methods, such as the ensemble Kalman filter (EnKF), which has proven to be a tractable way of overcoming steep computational costs for alternative ensemble filters (e.g., 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).

Corresponding author: Joshua McCurry, jmc Curry@umd.edu

DOI: 10.1175/MWR-D-22-0260.1

© 2023 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

In the context of moist processes in atmospheric models, past idealized model simulations have shown that the LPF more frequently preserves positive hydrometeor mixing ratios during filter updates, when compared to typical EnKF updates. Nevertheless, the extent of consistency between LPF updates and non-Gaussian, model constrained posterior densities for moist convection has not been demonstrated outside of idealized simulations. To this point, covariance localization techniques with the EnKF are known to cause adjustments to wind and mass fields that scale inversely with localization radii (Kepert 2009; Greybush et al. 2011), and localization in the LPF likewise constitutes a residual source of imbalance in posterior fields that must be addressed through the tuning of localization radii. The primary goal of the current study and of future research is to gauge the ability of the LPF to meet expectations afforded to standard PF methods in regard to minimizing data assimilation-related bias for non-Gaussian posteriors. Establishing a high level of consistency between LPF posteriors and non-Gaussian error structures imposed by nonlinearities in model physics would allow those investigating model bias to apply data assimilation without introducing a significantly confounding source of error and could therefore provide a basis for characterizing biases introduced by microphysical schemes and other model physics components.

The current paper adopts a Weather Research and Forecasting (WRF) Model ensemble framework for investigating systematic differences between posterior quantities generated by the LPF and corresponding quantities from parametric data assimilation methods that assume Gaussian error structures. The modeling system resembles the configuration designed for past iterations of the National Severe Storms Laboratory (NSSL) Warn-on-Forecast system (WoFs; Wheatley et al. 2015; Jones et al. 2016; Lawson et al. 2018), which targets forecast improvements at short lead times. As a proof-of-concept, the present manuscript discusses results from small (64-member) ensemble sizes affordable for most research and operational weather prediction systems, focusing on microphysical variables that relate nonlinearly to observations of radar reflectivity. Future studies will build on the comparison methodology established here to evaluate filter solutions for large ensemble runs of $\mathcal{O}(10^3)$ members. Larger ensembles will reduce sampling error and allow for the application of the LPF using relaxed heuristic features, including larger localization radii. While certain aspects of this study focus on forecast performance, the primary goal is not to assess the LPF's utility for operational data assimilation, but rather to evaluate their potential for future work contingent on the quality of posterior representations afforded under nonidealized conditions.

2. Modeling system

a. Model configuration

Experiments were performed with the Advanced Research WRF (V4.2) (Skamarock et al. 2019), using a single $900 \times 900 \text{ km}^2$ domain with 50 vertical levels and a 3-km horizontal grid spacing, which is sufficient to explicitly permit convection in numerical experiments. This domain

configuration was kept constant for all experiments but moved to appropriate geographical locations to accommodate events of interest.

Choices for physical parameterization schemes were motivated by past experiments performed at NSSL for the WoFs (Jones et al. 2018; Potvin et al. 2020), which served as the benchmark for our modeling system. For microphysics parameterization, we chose the NSSL two-moment variable-density scheme (NVD). The NVD scheme is notable for its fully double-moment representation of all hydrometeor classes, and the inclusion of graupel density as a diagnosable parameter (Mansell 2010). We selected the Rapid Radiative Transfer Model (RRTM) (Iacono et al. 2008) and the Dudhia scheme (Chen and Dudhia 2001) as our respective longwave and shortwave radiative transfer schemes. Additional physics options included the RUC Land Surface Model for surface physics (Smirnova et al. 2016), the Monin–Obukhov Similarity scheme for surface-layer physics (Jiménez et al. 2012), and the YSU scheme for PBL physics (Ghoniya et al. 2017). Since our model configuration is convective allowing, we did not specify a cumulus parameterization scheme. In addition to our default configuration using NSSL microphysics, we performed a limited set of experiments using the double-moment Thompson microphysics scheme (Thompson et al. 2008), which is commonly used in research and operational contexts.

b. Data assimilation methods

The LPF adopted in this study was first introduced in Poterjoy (2016; hereafter P16) and revised in Poterjoy et al. (2019). The filter operates by assimilating observations with independent errors sequentially and combining sampled particles and prior particles for each observation. The LPF satisfies the bootstrap PF solution (Gordon et al. 1993) for state variables located in close geographical proximity to observations in the sequence but maintains the prior particles for state variables located far from observations. A smooth correlation function that tapers to zero at a finite user-specified distance controls the spatial influence observations have on posterior estimates, which greatly reduces the number of particles needed for geophysical data assimilation. The current version of the LPF incorporates improvements introduced by Poterjoy (2022b) to further reduce sampling error for limited ensemble sizes. These include tempering, which uses a factorization of the Bayesian observation likelihood to iteratively approach a posterior solution. Tempering can be extended to include a “hybrid” methodology that combines the LPF with parametric filters, such as ensemble Kalman filters (EnKFs).

For experiments that employed the hybrid variant of the LPF, we specified a “minimum residual” parameter value of 0.8, indicating that tempering steps were structured so as to perform an LPF update using 20% of the likelihood and an EnKF update using the remaining 80% of the likelihood; see Poterjoy (2022b) for a detailed description of this factorization and Kurosawa and Poterjoy (2022) for potential strategies for choosing the factorization adaptively. This choice followed from a series of sensitivity tests (not shown), which were aimed at reducing prior mean squared errors while retaining most of the positive effects of using general approximations

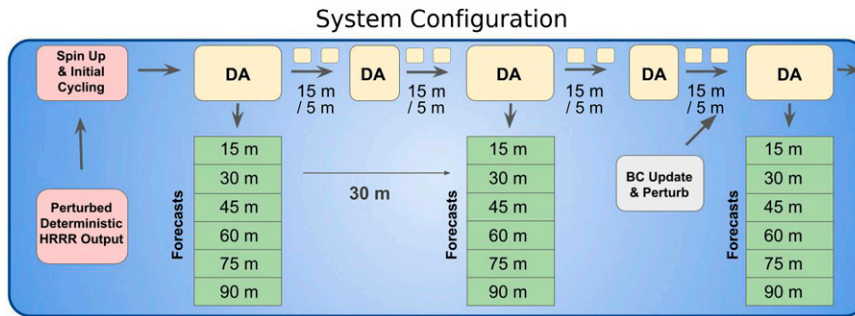


FIG. 1. Configuration diagram for experimental modeling system used for sequential data assimilation experiments.

for error distributions. As we will demonstrate, this particular setup does not necessarily produce updates dominated by the succeeding EnKF step, since the innovation term between observations and modeled quantities can be substantially reduced after the initial application of the LPF. Following the LPF update with an EnKF update, however, greatly reduces challenges associated with sampling bias in the standard LPF, as will be demonstrated in the following section. To further cope with sampling error, we set the particle mixing parameter (Poterjov 2022b) to 0.5, which helped maintain diversity in posterior model states in the absence of localization. As a form of posterior covariance inflation, we applied relaxation-to-prior-spread (RTPS) (Whitaker and Hamill 2012) on the EnKF portion of the hybrid update with a relaxation coefficient of 0.8, but retained the full update performed by the LPF. Experiments with the nonhybrid variant of the LPF used the same mixing parameter, but with localization radii multiplied by a factor of 0.75.

For comparison with a Gaussian data assimilation method, we performed identical data assimilation experiments with an ensemble adjustment Kalman filter (EAKF), which is a deterministic square root variant of the EnKF (Anderson 2001). Square root EnKFs, like the EAKF, represent the most commonly used ensemble data assimilation method in atmospheric science (Houtekamer and Zhang 2016). To correspond with the hybrid particle filter variant, we applied RTPS with a coefficient of 0.8. Localization radii were identical to those used with the hybrid variant LPF.

c. Experiment design

We performed sequential data assimilation experiments for a select number of severe weather outbreaks during the 2019–20 warm seasons. These experiments generated posterior ensemble members at various times along the life cycle of each event, which are the primary subject of this paper. The modeling system constructed for this research was adapted from the NSSL experimental WoFS (Wheatley et al. 2015; Jones et al. 2016; Lawson et al. 2018).¹ It incorporates the

Data Assimilation Research Test bed (DART), which supports a host of ensemble filters, including the EAKF and the LPF variant adopted for this study.

To support a flexible ensemble size in future experiments, our modeling system generates initial and lateral boundary conditions by perturbing an analysis from the High Resolution Rapid Refresh (HRRR) (James et al. 2022) forecast with Gaussian noise for model horizontal wind, potential temperature, and moisture fields. Hourly updates were applied to lateral boundary conditions for each member during data assimilation cycling, using information from the posterior ensemble and HRRR forecasts. After each hourly update, additional Gaussian noise was added to boundary conditions to maintain spread. All cases used 64 ensemble members, which is roughly twice the size used by the WoFS. Members were initially generated from 1-h forecasts run from a perturbed HRRR analysis, which were used to initiate sets of experiments that consist of 14 h of sequential DA. To cope with under-dispersion in the ensemble that cannot be treated by RTPS alone, we also used additive inflation at 15-min intervals near areas of high observed reflectivity (Dowell and Wicker 2009). We retained only the last 9 h of posterior ensemble members for comparison in order to reduce the dependence of model solutions on the initial HRRR analysis. We also initialized 90-min ensemble forecasts every 30 min from these solutions, giving a total of 19 ensemble forecasts per case study. We generated two sets of cycling data assimilation experiments and forecasts using 15- and 5-min cycling frequencies. The 15-min cycling frequencies are characteristic of experimental real-time prediction systems such as the WoFS. Five-minute cycling frequencies allowed us to investigate posterior characteristics when generated from priors evolved over a shorter integration time, which reduced deviations from Gaussianity caused by nonlinear model processes. Furthermore, all experiments used WRF “restart” files to carry forward all prognostic variables and tendencies. This choice further isolated the impact of data assimilation on forecast evolution at short lead times, by avoiding contributions to transient adjustments from the equilibration period associated with cold-starts from history files. An overview of our modeling system configuration is shown in Fig. 1.

To improve the relevance of our results to operational conditions, we assimilated as many observational types as feasible

¹ The WoFS was implemented by NSSL as part of the NOAA Warn-on-Forecast program to spur improvements in forecast lead time and accuracy for severe threat events using convection-allowing ensembles and advanced data assimilation techniques (Stensrud et al. 2009, 2013).

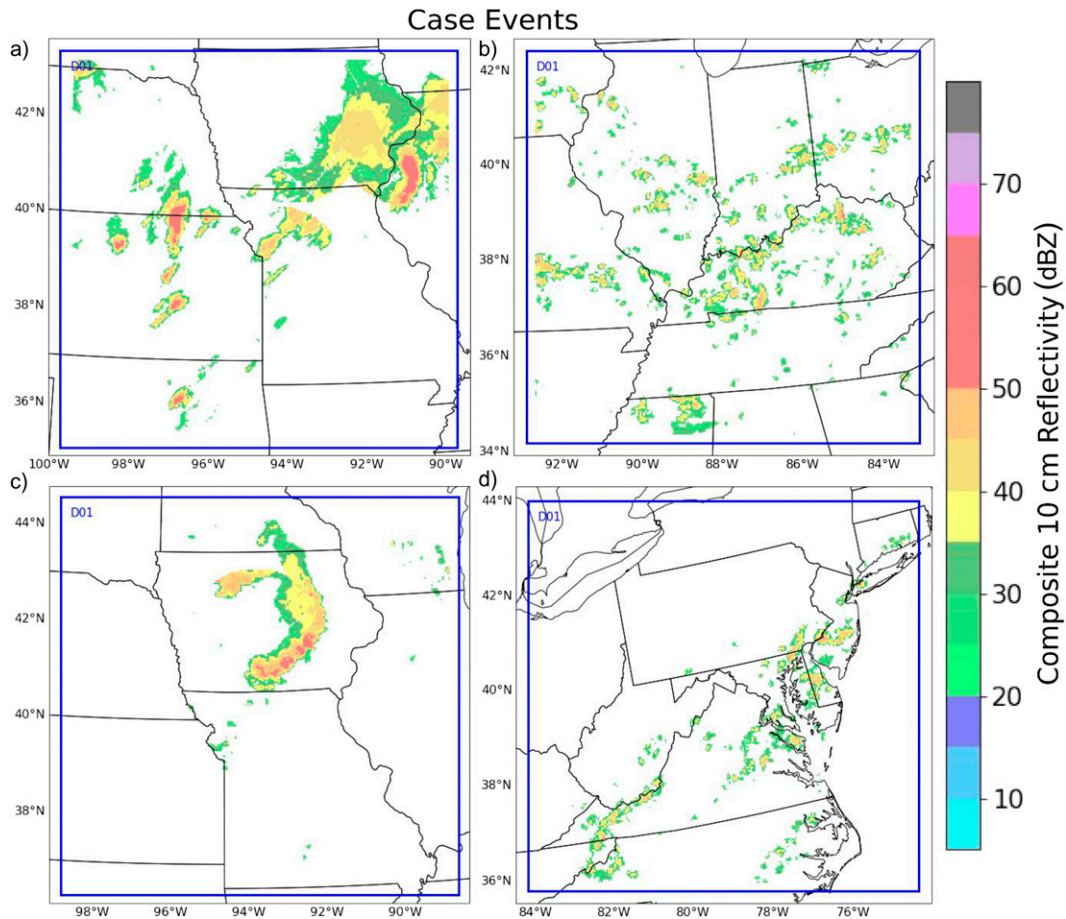


FIG. 2. Observed 10-cm composite reflectivity plotted at representative hourly times for test cases used in data assimilation experiments. Sample reflectivity output shown for experiments starting at (a) 1200 UTC 28 May 2019, (b) 1000 UTC 3 Jul 2019, (c) 0700 UTC 17 Jul 2019, and (d) 1000 UTC 12 Aug 2020.

based on data availability and compatibility with the forward operators in DART. The Meteorological Assimilation Data Ingest System (MADIS) from NOAA provided surface observations of temperature and wind from mesonet and Meteorological Aerodrome Reports (METAR) data, as well as observations of temperature and wind in the lower troposphere from the Aircraft Communications Addressing and Reporting System (ACARS). Although temporally sparse, we also assimilated derived tropospheric winds from the Geostationary Operational Environmental Satellite, made available through MADIS. From the Next Generation Radar (NEXRAD) network, we assimilated observations of radar reflectivity and radial velocity in areas of precipitation, as well as Multi-Radar Multi-Sensor (MRMS) clear-air estimates of reflectivity in areas not experiencing storm activity. The latter are useful in removing areas of predicted reflectivity and associated convective activity in the prior that do not coincide with observed storm features.

d. Case studies

In the interest of sampling a set of events sufficiently representative of deep moist convective regimes, we chose four test cases to cover several distinct dynamical setups (Fig. 2). The

defining difference between these cases is in convective mode, ranging from single-cell convection to organized convection within a mesoscale convective system. Each case is covered by an observational network of comparable density and observation type composition.

Our first test case covers the period from 1200 UTC 28 May to 0300 UTC 29 May 2019, within a domain centered on northern Missouri (Fig. 2a). A mesoscale convective system initially associated with a shortwave trough propagates across the domain in a northeastward direction. The complex synoptic-scale setup with weakening convective inhibition and marginal directional shear made for difficult forecasting conditions at the time of the event (Leitman and Thompson 2019). The system was ultimately associated with a diverse array of convective modes dominated by multicellular storms but including discrete supercells and simple single-cell convection.

The second test case covers the period from 1000 UTC 3 July to 0100 UTC 4 July 2019 over the Ohio Valley region (Fig. 2b). Ordinary single-cell convection appears beginning after 1600 UTC under strong mean-layer CAPE and weak directional shear, becoming multicellular with new cells evident at outflow boundaries, but without widespread organization

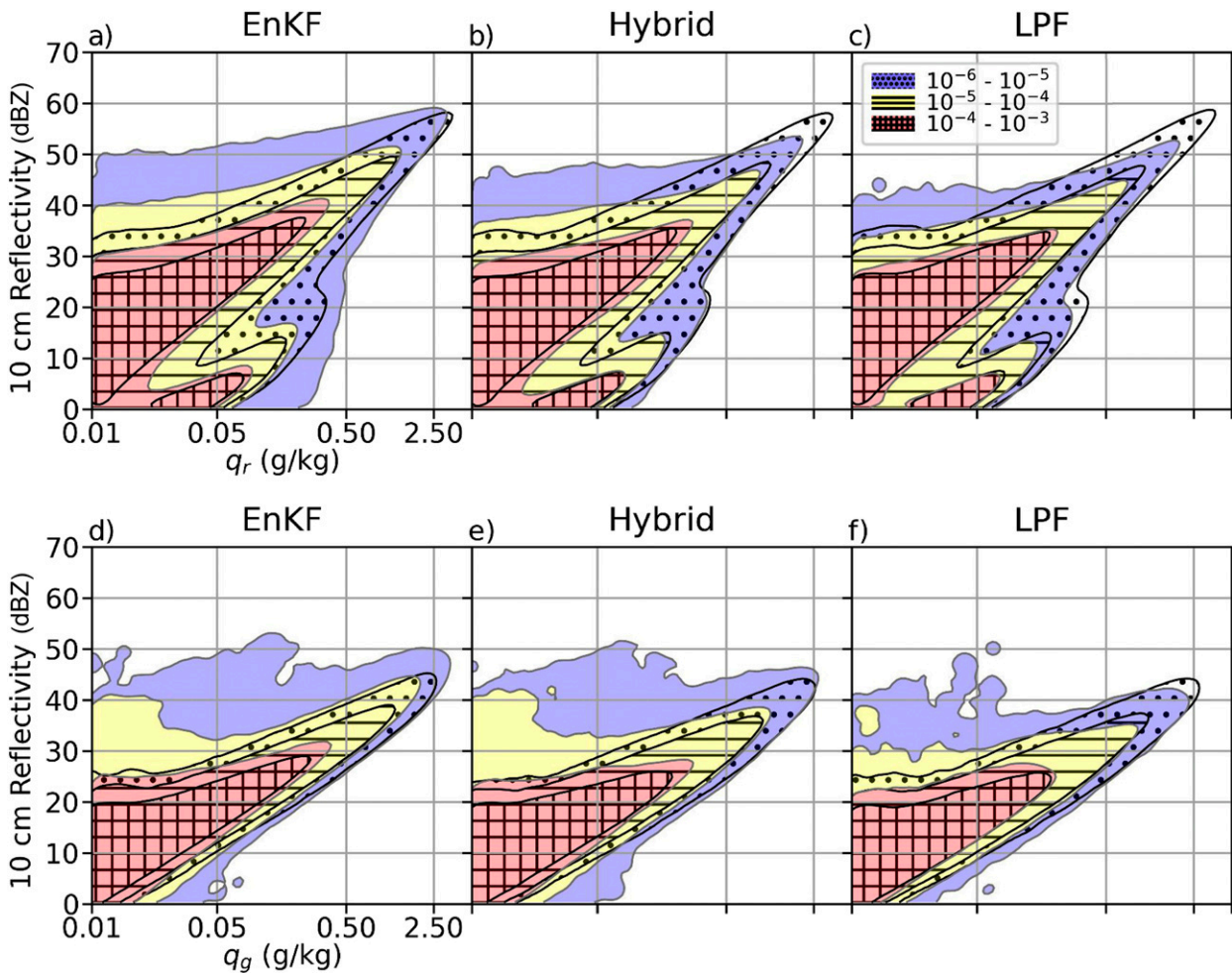


FIG. 3. Marginal bivariate probability density function (PDF) plots for 15-min cycling frequency experiments. Prior (hatched) and posterior (shaded) probability density for 10-cm radar reflectivity and rainwater mixing ratio for sequential assimilation with (a) EnKF, (b) hybrid particle filter, and (c) localized particle filter. PDFs for 10-cm radar reflectivity and graupel mixing ratio for sequential assimilation with (d) EnKF, (e) hybrid particle filter, and (f) localized particle filter. Axis scaling is maintained across panel rows.

into quasi-linear structures. The transition from single to multicellular convection is quite rapid in this case, presenting a notable challenge for data assimilation.

The third case extends from 0700 UTC 17 July to 0100 UTC 18 July 2019 over a domain nearly collocated with that of the first test case (Fig. 2c). A quasi-linear convective system with embedded bow-echo-type segments propagates from the northwest to southeast of the domain, beginning to dissipate after nightfall. This case features highly organized multicellular convection with discernible inflow and outflow structures.

The fourth and final case begins at 1000 UTC 12 August and ends at 0100 UTC 13 August 2020, in a domain centered on the mid-Atlantic seaboard (Fig. 2d). It is a classic pulse-type event that does not show any obvious signs of multicellular organization. Convection initiates nearly uniformly at 1600 UTC, continuing under strong surface heating until nightfall.

3. Characterizing posterior ensemble members

a. Analysis of prior and posterior climatology

To investigate systematic differences in EnKF and LPF state estimates we approximated climatological bivariate marginal probability density functions (PDFs) of radar reflectivity and hydrometeor mixing ratios for model prior and posterior states. This choice followed the substantial influence of radar reflectivity measurements on analyses and forecasts produced over these events. The samples used to generate these PDFs came from WRF gridpoint values at all times within the duration of sequential cycling and across all four case study events. We present PDFs generated in this manner for both 15-min (Fig. 3) and 5-min cycling frequencies (Fig. 4). Gridpoint sampling was confined to the eighth vertical WRF coordinate level of our domain when examining rainwater mixing ratio (q_r) and the sixteenth level when considering graupel mixing ratio (q_g) in order to ensure that the population from which

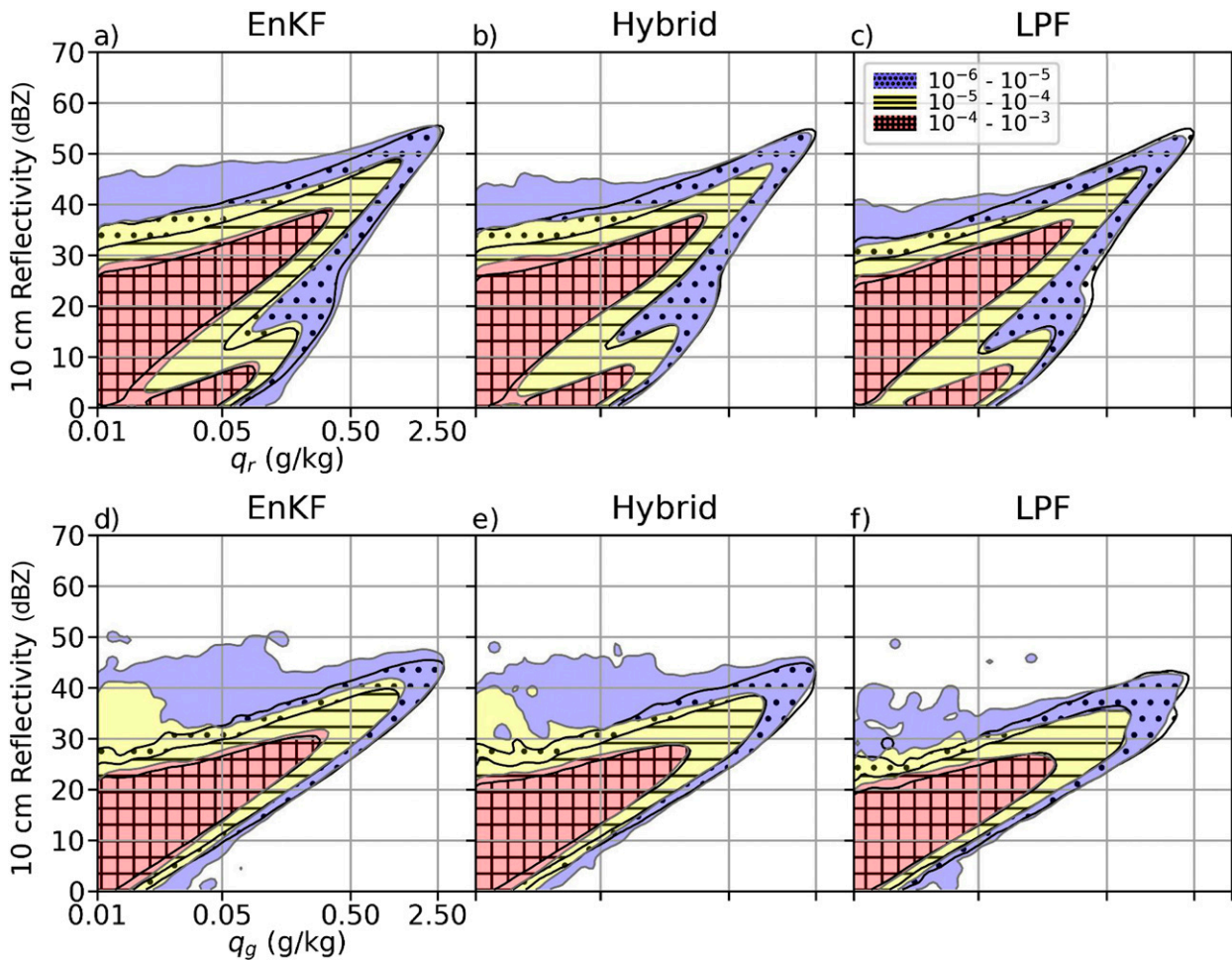


FIG. 4. As in Fig. 3, but for 5-min cycling frequency.

grid cells were sampled consisted of predominately unfrozen and frozen species, respectively. The eighth and sixteenth vertical levels roughly correspond to pressure levels of 850 and 550 hPa, respectively, and will be referred to as such for the remainder of the text. Sampled grid points from these levels were additionally filtered to ensure that the relevant species was the only contributor to reflectivity. Posterior output for radar reflectivity was recalculated from updated mixing ratios using the relations provided in the NSSL two-moment scheme. The NSSL microphysics configuration used in this study forms Z - q relationships for rainwater that are solely functions of mixing ratio and number concentration, with number concentration in the expression denominator. Therefore, in the case of rainwater, we may fully characterize the relation between reflectivity updates and updates to q_r by examining additional PDFs for hydrometeor mass mixing ratio and number concentration n_r . Although the relation between q_g and reflectivity is additionally complicated by diagnosed ice density, we may also gain insight by examining PDFs with graupel number concentration, n_g . For these PDFs we again show figures for 15-min (Fig. 5) and 5-min (Fig. 6) cycling,

taken at the 850- and 550-hPa pressure levels for rain and graupel, respectively.

The 15-min EnKF posterior PDF for q_r and radar reflectivity is notable since it indicates a significant presence of high reflectivity grid points above 50 dBZ (Fig. 3a). Some of these high reflectivity posterior points are collocated with elevated q_r above 0.5 g kg^{-1} , falling within regions of the state-space supported by the corresponding EnKF prior. However, significant probability density at high reflectivity values also occurs with q_r between 0.05 and 0.5 g kg^{-1} , indicating that the EnKF commonly produces state-space solutions not supported by the model. Looking at the corresponding 15-min PDF between q_r and n_r (Fig. 5a), the EnKF updates show a shift of probability density toward lower number concentration from prior to posterior at all mixing ratios above 0.05 g kg^{-1} , presumably the cause of elevated reflectivity at lower q_r . In contrast, the LPF configurations at the same pressure level produce posterior densities for q_r and radar reflectivity that are more consistent with the support of their priors, with the pure LPF producing the least disagreement (Figs. 3b,c). Posterior members generated by both LPF methods feature few

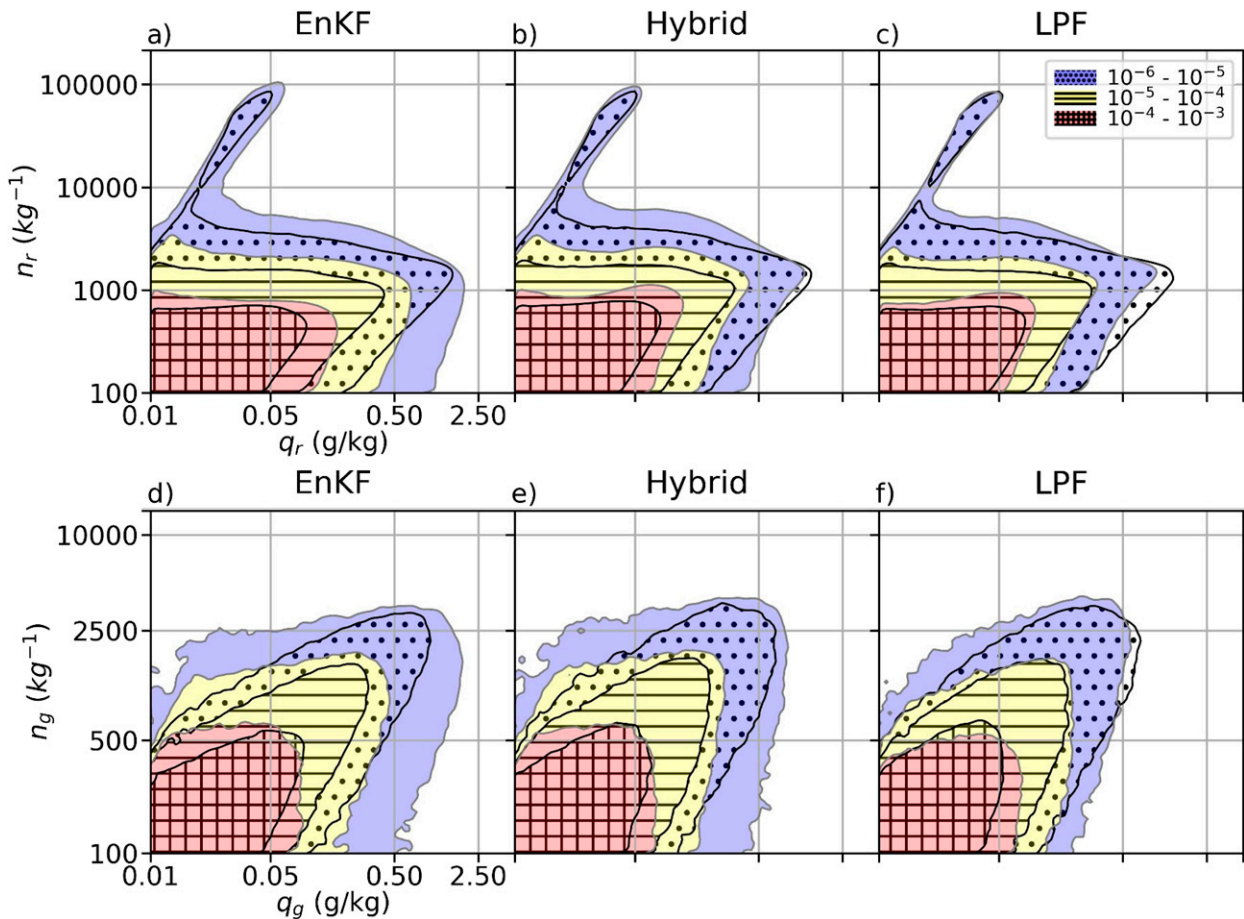


FIG. 5. Marginal bivariate probability density function (PDF) plots for 15-min cycling frequency experiments. Prior (hatched) and posterior (shaded) probability density for rainwater mixing ratio and rainwater number concentration with (a) EnKF, (b) hybrid particle filter, and (c) localized particle filter. PDFs for graupel mixing ratio and graupel number concentration for sequential assimilation with (d) EnKF, (e) hybrid particle filter, and (f) localized particle filter. Axis scaling is maintained across panel rows.

grid points with 50+ dBZ reflectivity at q_r below 0.5 g kg^{-1} , and also indicate a reduced occurrence of points with q_r above 0.5 g kg^{-1} when compared to their respective priors. Again looking at the 15-min PDF for q_r and n_r (Figs. 6b,c), while the PF posteriors reduce areas of high probability density for elevated q_r , the pure LPF does not substantially change the nature of the relationship between q_r and n_r from that indicated by the prior PDF, and the hybrid does so only to a degree intermediate between EnKF and LPF. The behavior here is consistent with a persistent removal of large hydrometeors by the LPF methods in favor of lower reflectivity, lower q_r model solutions during resampling, while preserving the physical relationships between q_r , n_r and reflectivity output in model priors.

When comparing experiments with 15 and 5-min cycling frequencies, notable differences are apparent in the span of prior support toward higher reflectivity and mixing ratios. The 5-min cycling reduces support in state space regions of reflectivity above 50 dBZ and largely excludes q_r above 2.5 g kg^{-1} . EnKF posteriors produced from 5-min cycling

priors still generate points of high reflectivity outside the prior, but in this case, they occur over a more limited region of state space and compose a smaller portion of posterior probability density (Fig. 4a). As before, these anomalous solutions are mediated by additional posterior density at low values of n_r as shown in the 5-min PDF for q_r and n_r (Fig. 6a). The LPF configurations for 5-min cycling again produce posterior PDFs that more consistently remain within the support of priors compared to the EnKF for both sets of marginals, but here the removal of probability density at high mixing ratios between prior and posterior is virtually absent due to the already reduced density of high mixing ratio solutions with 5-min PF priors (Figs. 4b,c). Increased cycling frequency allows the prior to remain more Gaussian by limiting the duration of the model advance and associated nonlinear error growth and should therefore be more amenable to assimilation with a Gaussian filter such as the EnKF. The apparent differences between 5- and 15-min EnKF posteriors indicate, at the least, a strong dependence for filter behavior on cycling frequency. Changes to LPF posteriors are more muted, with

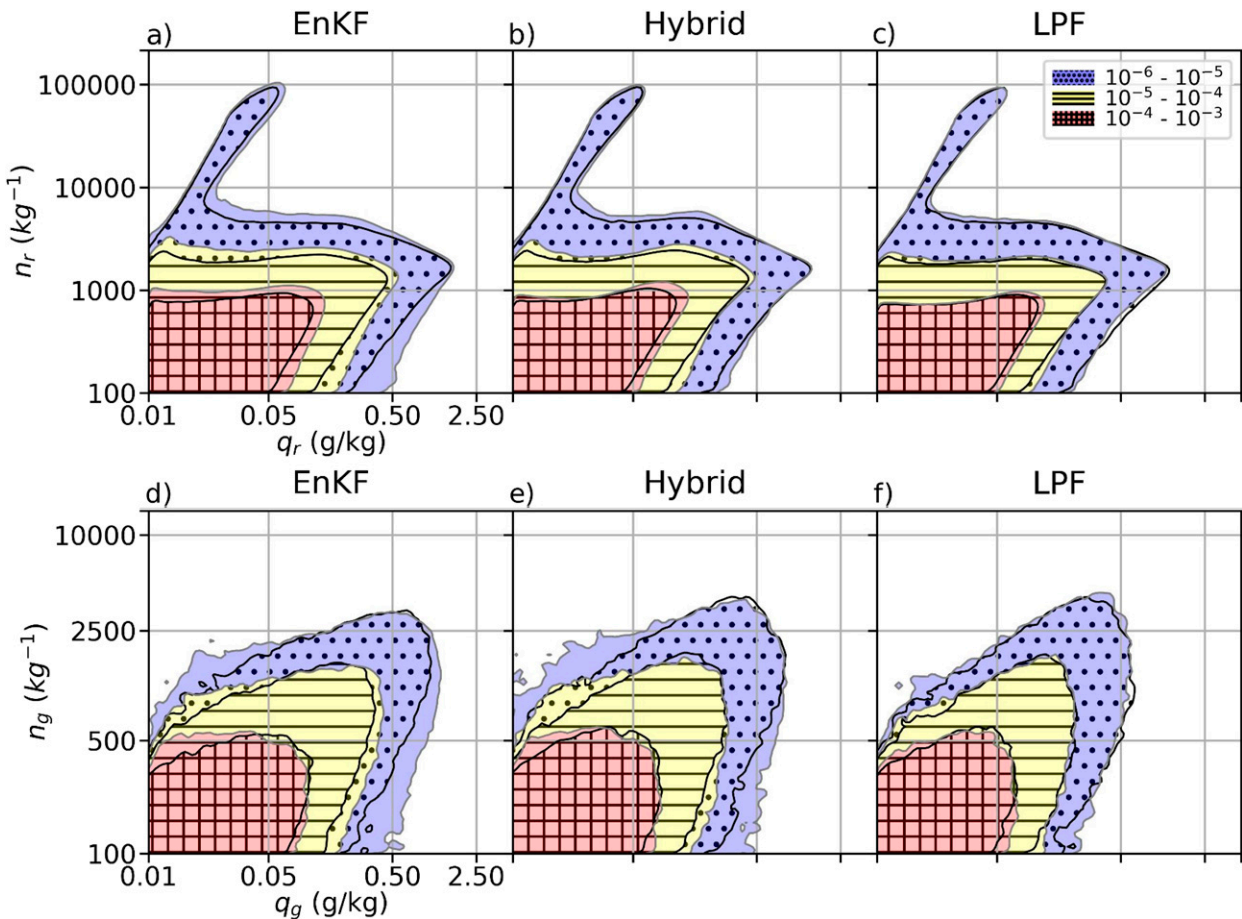


FIG. 6. As in Fig. 5, but for 5-min cycling frequency.

both cycling frequencies producing posteriors similar to the 5-min cycling frequency EnKF. This could suggest that the presence of high reflectivity, high q_r points in the 15-min EnKF posterior is mediated by Gaussian assumptions inherent to the EnKF. The decrease in probability density for state-space regions of high reflectivity and high mixing ratio between 15- and 5-min priors is equally notable and may indicate a model bias toward strong convection or another source of high reflectivity that is ameliorated by all three filters when subject to frequent data assimilation. Indeed, several studies lend support to the idea that NSSL microphysics overpredicts regions of high reflectivity (Johnson 2019; Potvin et al. 2020; Skinner et al. 2018).

The corresponding marginal PDFs taken at the 550-hPa pressure level for q_g and reflectivity show similarly filter-specific behavior, but differ significantly in trends between 5- and 15-min cycling frequencies. The 15-min EnKF posterior PDF indicates elevated probability densities for reflectivity values above 40 dBZ at a broad range of mixing ratios, in contrast to the prior PDF where such solutions are sparse and limited to mixing ratios above 0.5 g kg^{-1} (Fig. 3d). As with the marginal for q_r and reflectivity, this is again mediated by a density shift toward smaller number concentrations (Fig. 5d).

The PDF for the pure LPF, in contrast, features little density for high reflectivity solutions of 40+ dBZ and is the most consistent with its prior, with the PDF for the hybrid taking a middle ground between both methods (Figs. 3e,f). The 5-min cycling frequency retains this trend, but reduces the magnitude of differences between priors and posteriors for the EnKF and hybrid particle filter (Figs. 4e,f). Unlike the marginal PDFs for prior q_r and reflectivity, the prior PDFs for q_g and reflectivity do not lose density in high mixing ratio areas of state-space between 15- and 5-min cycling (Fig. 6). The bivariate marginal PDFs for prior q_g and n_g also remain distinct between filtering methods at 15-min cycling frequency, with the support of the LPF methods spanning slightly higher number concentrations than that of the EnKF (Figs. 5d-f). Interestingly, this distinction is maintained from 15- to 5-min cycling frequency, in contrast to the trend with rainwater mixing ratios and number concentrations where the univariate marginal distribution for number concentration evens out between filtering methods. This behavior is driven by strong EnKF adjustments to the marginal PDF that are maintained from the posterior into the prior with the effect of moving density toward lower n_g even at short cycling frequencies (Fig. 6d). LPF adjustments on the other hand largely maintain

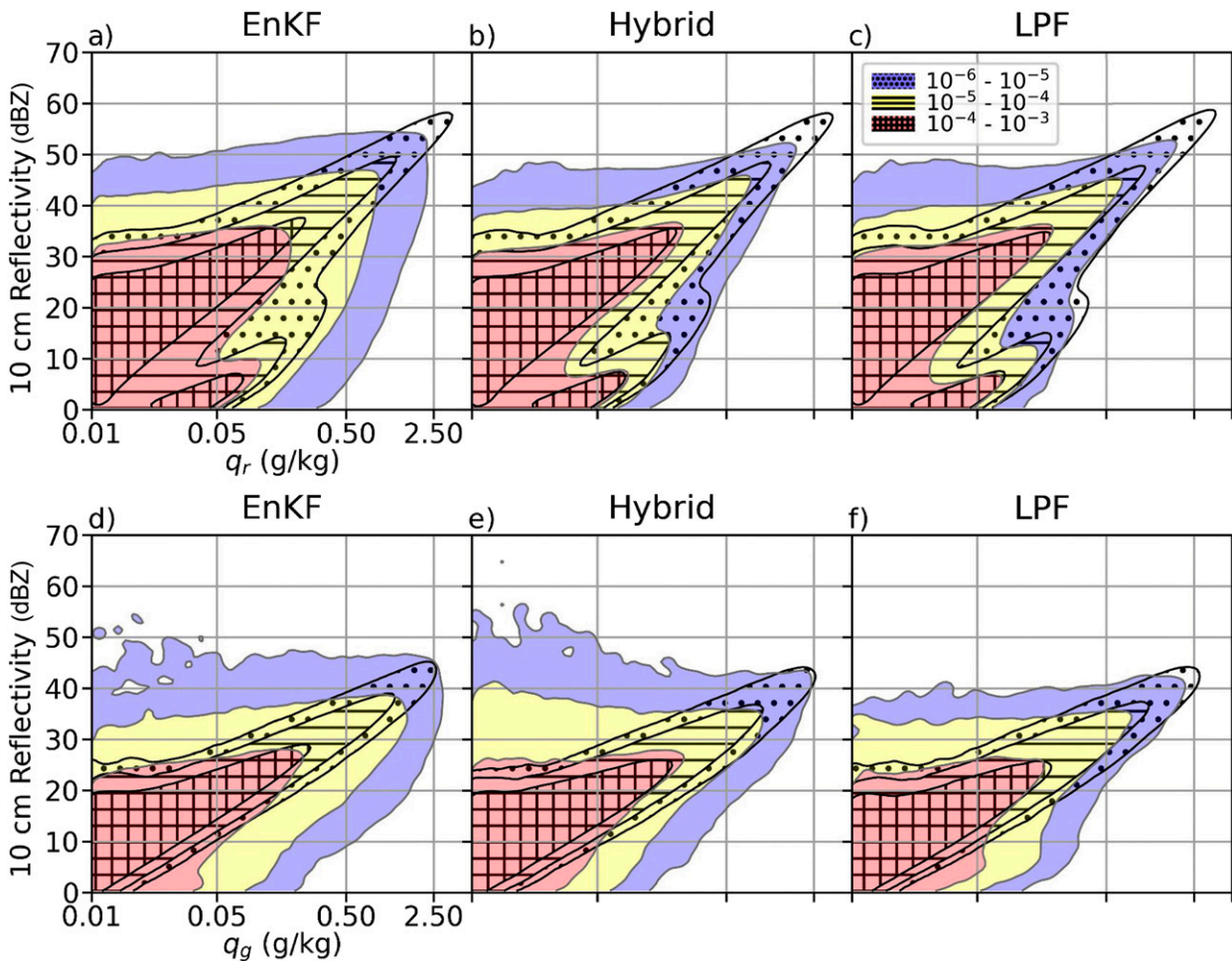


FIG. 7. Marginal bivariate probability density function (PDF) plots for 15-min cycling frequency experiments. Prior (hatched) and posterior (shaded) probability density for DART posterior output 10-cm radar reflectivity and rainwater mixing ratio for sequential assimilation with (a) EnKF, (b) hybrid particle filter, and (c) localized particle filter. PDFs for 10-cm radar reflectivity and graupel mixing ratio for sequential assimilation with (d) EnKF, (e) hybrid particle filter, and (f) localized particle filter. Axis scaling is maintained across panel rows.

the prior marginal distribution of n_g and q_g (Figs. 6e,f). Although adjustments to the n_g marginal from prior to posterior are not necessarily nonphysical when associated with changes to the dynamical evolution of storms, the fact that downward adjustments to n_g are seemingly independent of q_g above 0.05 g kg^{-1} would suggest that it is instead an artifact of Gaussian assumptions made by the EnKF.

Since the DART implementation of the EnKF updates observation-space priors during sequential assimilation according to Anderson and Collins (2007), posterior diagnosed reflectivity may not reflect values consistent with hydrometeor quantities given a nonlinear forward operator. Though the pure LPF is not directly affected by nonlinearity in measurement operators, similar discrepancies can be produced by the method’s kernel density distribution mapping (KDDM) step. The DART output reflectivity is informative as the working variable “seen” by the filter algorithm during assimilation and before recalculation of diagnostic reflectivity during the WRF advance step. With this in mind, we examined bivariate PDFs

of mixing ratio and DART output reflectivity. We again produced PDFs for 15- and 5-min cycling frequencies (Figs. 7 and 8). The 15-min EnKF posteriors for both species extend probability density at high mixing ratios toward lower values of reflectivity than supported by the prior PDF, and extend probability density at low mixing ratios toward higher values of reflectivity (Fig. 7a). The PDFs for 5-min cycling show the same, albeit muted, extensions of probability density (Fig. 8a). As with recalculated reflectivity, the posterior PDFs for the LPF methods are more consistent with the support of their respective priors, especially in the case of the pure LPF (Figs. 7b,c and 8b,c). These results together show that nonlinear forward operators affect the EnKF algorithm to a greater degree than either LPF method when adopting the parallel filter configuration (Anderson and Collins 2007). The low reflectivity states found in the posterior EnKF PDF and absent in the prior further suggest that linear approximations for measurement operators induce a bias that may account for the recalculated reflectivity estimates being higher than in the LPF methods.

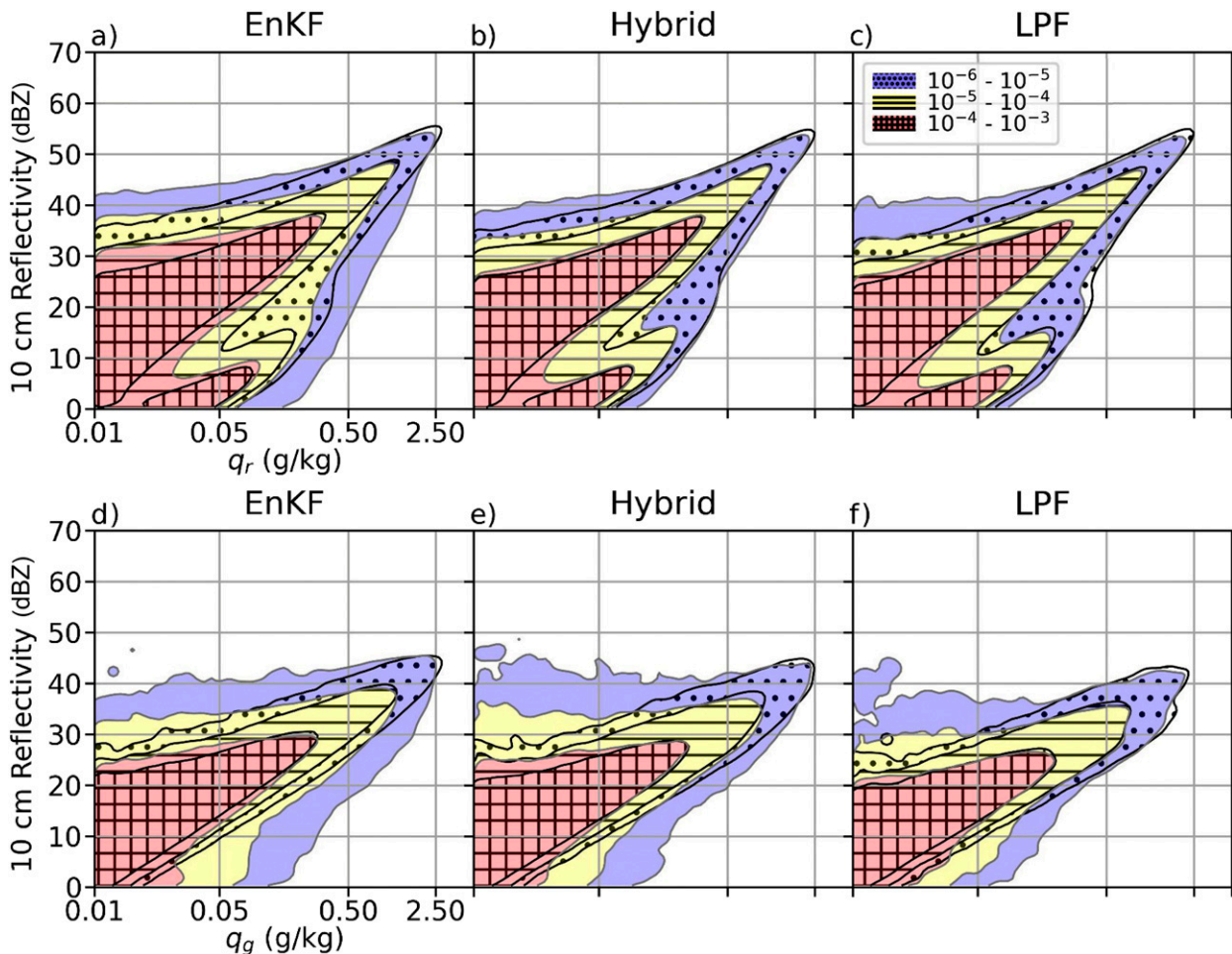


FIG. 8. As in Fig. 7, but for 5-min cycling frequency.

As an addendum to the preceding analysis and to provide specific contexts for the aforementioned marginal PDFs, we plotted prior and posterior ensemble member values for q_r and NSSL estimated reflectivity, as well as for n_r and reflectivity at two characteristic grid point locations. We generated posteriors here from a single filtering step applied to identical priors dated 1800 UTC 3 July from our sequential data assimilation experiment with the LPF. In this way, we show the distinct effects of each filtering method that lead to the aggregated behavior shown above for longer cycling intervals. Grid points presented here were identified according to the ensemble prior and posterior means for reflectivity, so that distinct behavior could be characterized according to the direction of reflectivity adjustment, and feature minimal contributions to reflectivity from graupel, hail or snow. We further required that sampled grid points be collocated with reflectivity observations. We show ensemble member states for a grid cell with a negative analysis increment for reflectivity in Fig. 9 and those for a grid cell with a positive analysis increment in Fig. 10. Results at this single grid cell location highlight the comparative advantages and disadvantages offered by LPF methods with relatively small ensemble sizes. Downward adjustments with

the LPF and hybrid methods sample new ensemble members effectively from physical states that conform to the $Z-Q$ relations governing reflectivity generated by rainfall in the NSSL microphysical scheme (Figs. 9b,c), whereas the EnKF samples members in such a way as to create mixing ratios that wildly diverge from the reflectivity indicated by observations (Fig. 9a), accompanied by a proliferation of unexpectedly high number concentrations for low mixing ratio solutions (Fig. 9d). Upward adjustments by the LPF are hindered by sampling issues related to insufficient ensemble size; with only a single prior member featuring reflectivity close to that of the nearby observation, the LPF posterior solutions cluster around prior members with moderate reflectivity and fail to sample from regions of higher mixing ratio (Fig. 10c). In contrast, the EnKF and hybrid move their posterior ensemble members away from the prior toward appropriate regions of state space (Figs. 10a,b). In this case, the EnKF adjustment produces spurious number concentrations in the posterior that are well below those suggested as plausible by prior ensemble members, while number concentrations in the hybrid posterior remain close to those for high mixing ratio solutions in the prior (Figs. 10d,e).

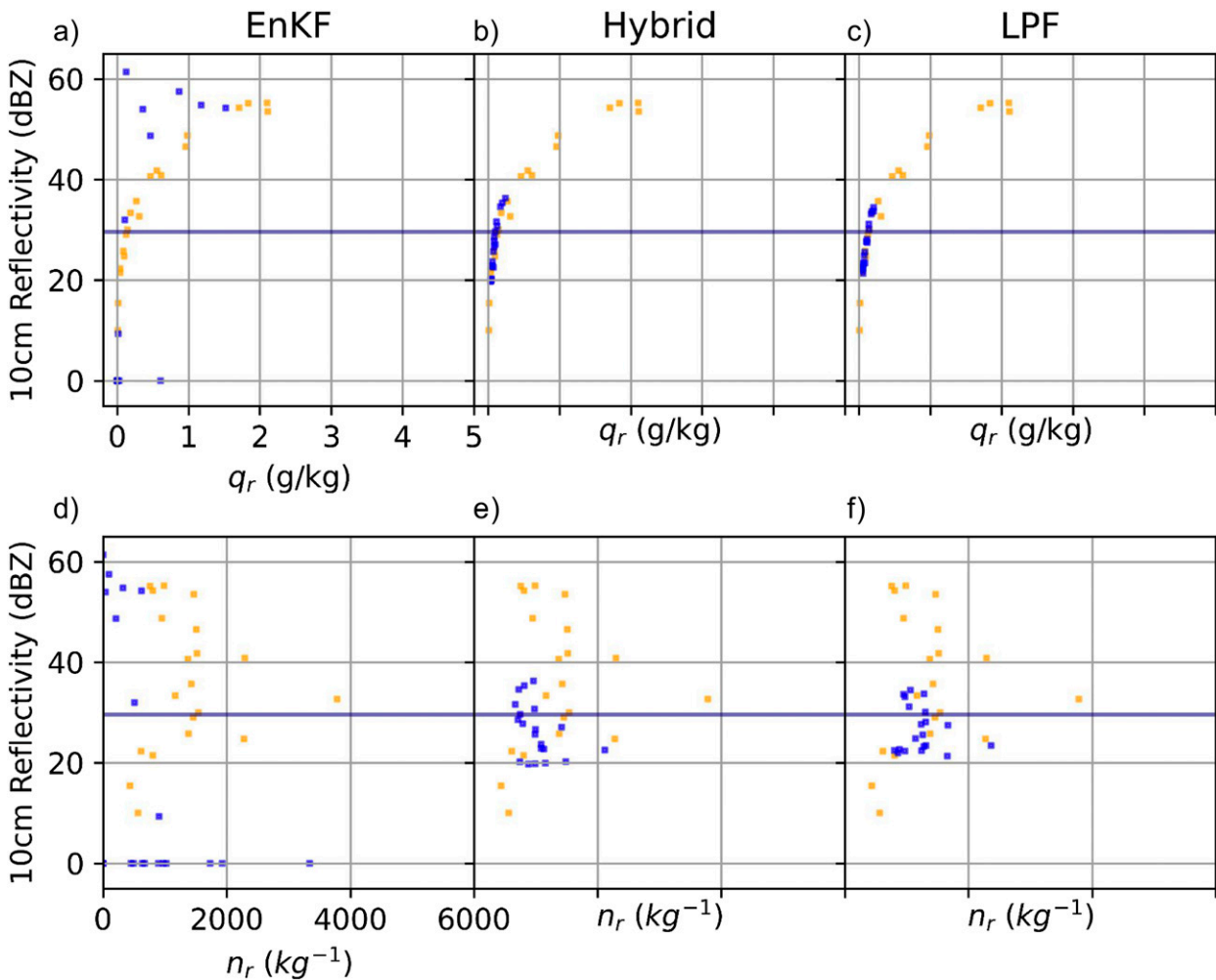


FIG. 9. 10-cm radar reflectivity and rainwater mixing ratio for prior (orange) and posterior (blue) ensemble members at a sampled grid-point location for a single adjustment by (a) the EnKF, (b) hybrid LPF, and (c) LPF. 10-cm radar reflectivity and rainwater number concentration for prior (orange) and posterior (blue) ensemble members at a sampled gridpoint location for a single adjustment by (d) the EnKF, (e) hybrid LPF, and (f) LPF. Average reflectivity value for observations within the corresponding gridpoint location is indicated by horizontal line in blue.

b. Cycling period verification using root mean departure from observations

Although the LPF methods appear to produce posterior members that more faithfully represent model solutions produced by the WRF model, this does not guarantee that they will produce the most accurate forecasts. Model error can be especially challenging for particle filtering methods if prior model states do not overlap with the true prior (Poterjoy et al. 2017). To gauge forecast accuracy from each method, we performed a simple verification over the period of sequential data assimilation by calculating the root-mean-square departure from observations (RMSD). For this purpose, we verified ensemble mean quantities over observation locations with valid entries for prior mean, and discounted observations rejected by all three filters. A squared error type metric applied to ensemble mean fields provides confirmation of fit to

observed quantities, albeit without considering higher-order moments. RMSD for 15-min cycling is presented for radar reflectivity (Fig. 11) and radial velocity (Fig. 12) measurements for the duration of sequential cycling for each test event. Results for reflectivity and radial velocity verifications show a clear advantage for the hybrid and EnKF over the pure LPF across events, with the EnKF slightly outperforming the hybrid in terms of prior RMSDs. Note that RMSD scores show strong filter independent variability between events that can be ascribed to event dynamics, such as relatively high RMSD scores for radial velocity when considering the quasi-linear convective system of 17–18 July 2019 that featured strong directional winds (Fig. 12c). Strong results for the hybrid correspond with findings from perfect model experiments with the Lorentz 96 model (Poterjoy 2022b), which showed that the tempering iterations of the hybrid method allowed it to outperform both a standard LPF and an EnKF in terms of

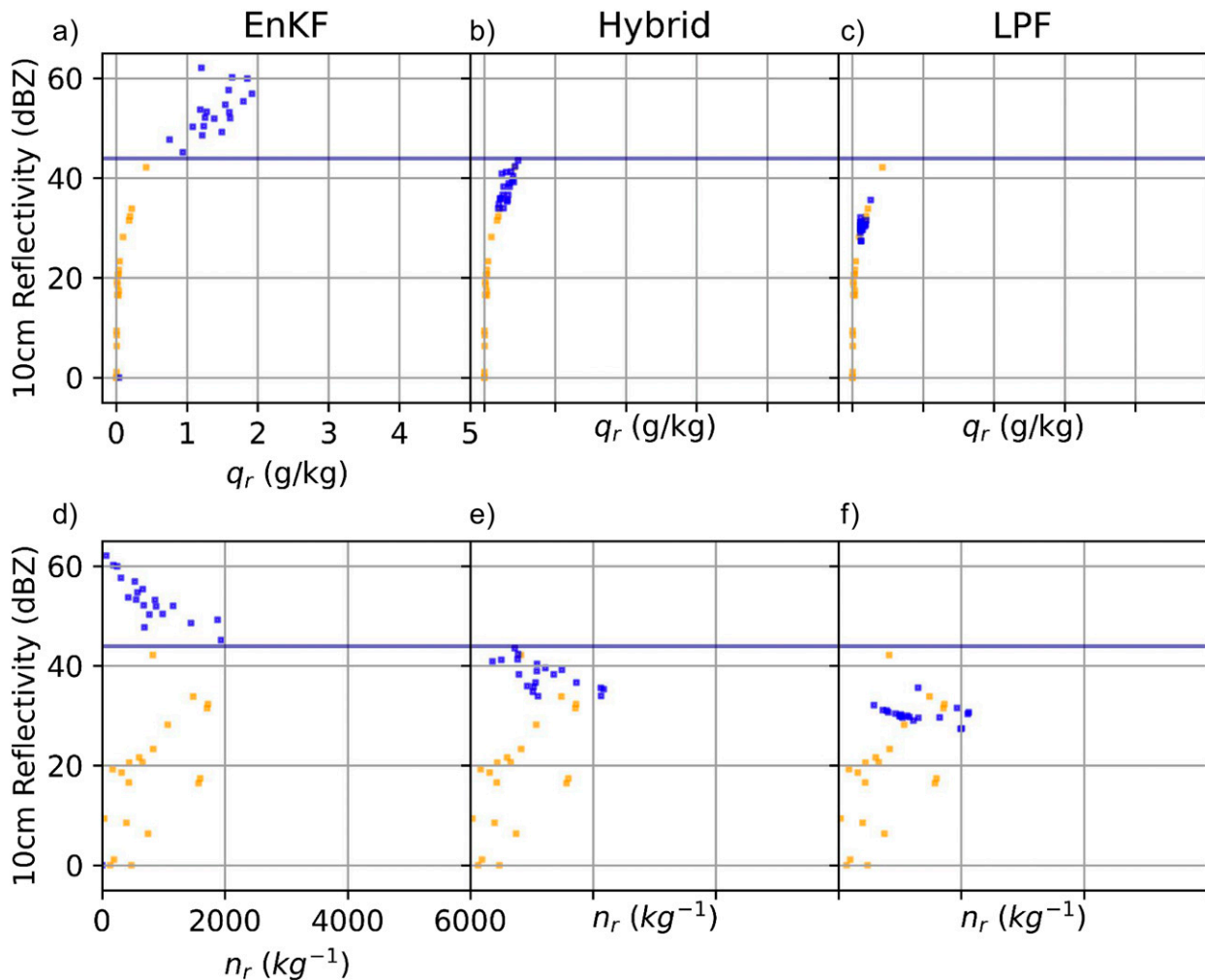


FIG. 10. As in Fig. 9, but for a sampled grid point with a positive analysis increment.

posterior RMSE in the presence of model error. When taken along with the previous characterization of prior and posterior densities, the verification results suggest compelling advantages to the hybrid LPF as a method that retains posterior fidelity to model solutions while producing a better forecast fit to observations than the pure LPF. For this reason, and to simplify the presentation of proceeding figures, we chose the hybrid as a representative non-Gaussian method for further comparison against the EnKF. Although not presented, we note that the aforementioned differences in filter performance persisted with 5-min cycling.

4. Qualitative comparison of storm evolution

a. Domain-averaged quantities during continuous cycling data assimilation

Modeling systems that represent mesoscale processes face significant challenges from data assimilation bias mediated by Gaussian assumptions and associated with spurious corrections to small-scale features (Poterjoy 2022a). We wish to

connect the data assimilation-induced biases for microphysical quantities indicated in our posterior PDFs with concrete impacts on storm evolution and forecast verification. To this end, we first examined a proxy for cold pool strength during the course of sequential data assimilation with a 15-min cycling frequency for all four of our test case events. Cold pool intensity mediates dynamical processes responsible for convective organization and can be indicative of how data assimilation is handling convective initiation or modifying the strength of preexisting moist convection.

To gauge cold-pool activity we plotted the difference between EnKF and hybrid surface virtual potential temperatures (θ_v), averaged over prior ensemble member grid cells featuring a composite reflectivity above 25 dBZ (Fig. 13). The reflectivity threshold serves as a generalized indicator of convective activity where cold pools are likely to form. For all four events, sequential data assimilation with the EnKF quickly results in lower θ_v than assimilation with the hybrid near areas of convection. This trend continues to strengthen over the course of sequential assimilation. Lower θ_v produced

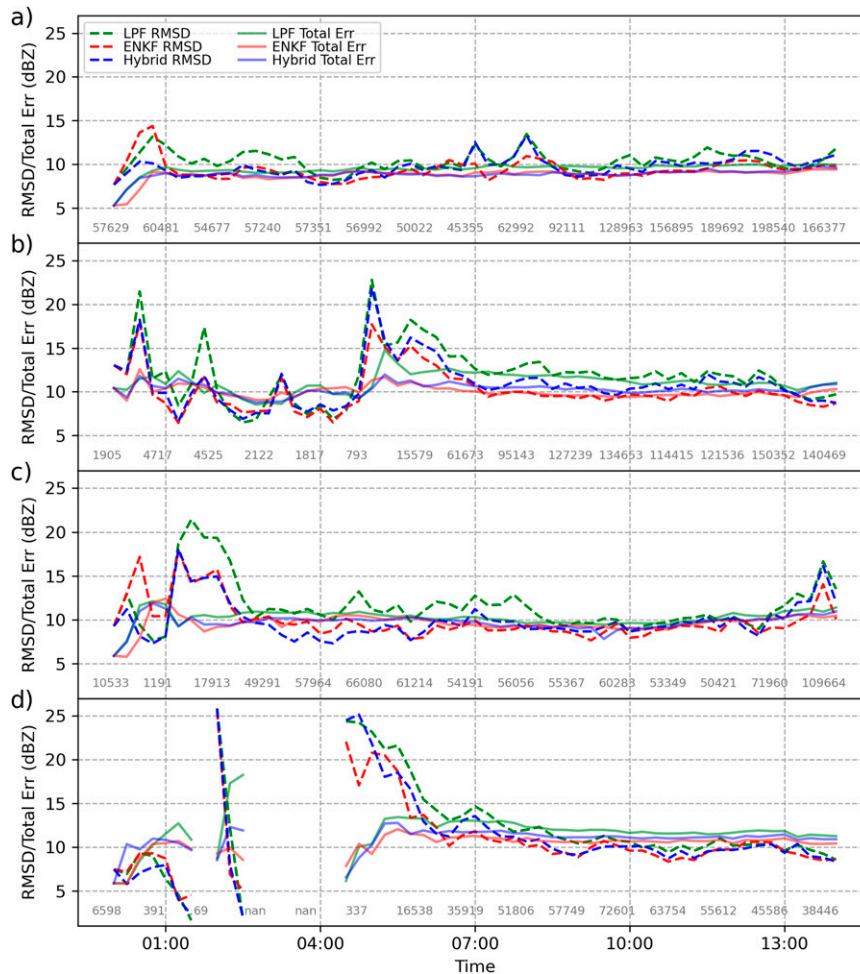


FIG. 11. Prior root-mean departure from observations (dashed) and total error (solid) for 10-cm radar reflectivity, plotted for the duration of 15-min frequency data assimilation for (a) 28–29 May 2019, (b) 3–4 Jul 2019, (c) 17–18 Jul 2019, and (d) 12–13 Aug 2020 events. Number of observations considered for RMSD calculations annotated in gray at hourly intervals.

during EnKF cycling is indicative of stronger or more widespread convection that produces more intense cold-pool development and confirms that the two assimilation methods are producing measurably different outcomes for system evolution.

b. Forecast evolution from identical priors

To more closely examine how differences in posterior quantities from a filtering methodology impact storm evolution, we generated two sets of forecast initial conditions from an identical prior ensemble by applying a single filtering step with a 15-min assimilation window for the EnKF and hybrid methods. We chose the initialization time as 1930 UTC 3 July for the case study covering this date due to its temporal placement after the start of convective initiation, but before peak development of the MCS. We took prior members from the corresponding time stamp during sequential data assimilation

with the pure LPF, then used posterior members to initialize 45-min ensemble forecasts output at 1-min intervals. This approach produced posterior fields and forecast output that differed only in the filtering method used for analyses and evolution from the resulting initial conditions. We supplemented our tests using an additional set of experiments that omitted hydrometeor updates during data assimilation, with the goal of isolating impacts attributable to microphysical variables. We also present results from forecasts initialized without data assimilation for reference.

We again examined cold pool activity as a proxy for the presence and strength of convection. Because we considered only a short forecast period with minimal input from the diurnal cycle and synoptic-scale forcing, we could conveniently define cold pool extent using a fixed threshold of surface-level virtual potential temperature (Fig. 14a). We set this threshold at 306.5 K to achieve the closest fit to manually identified cold pool structures while avoiding contamination from other

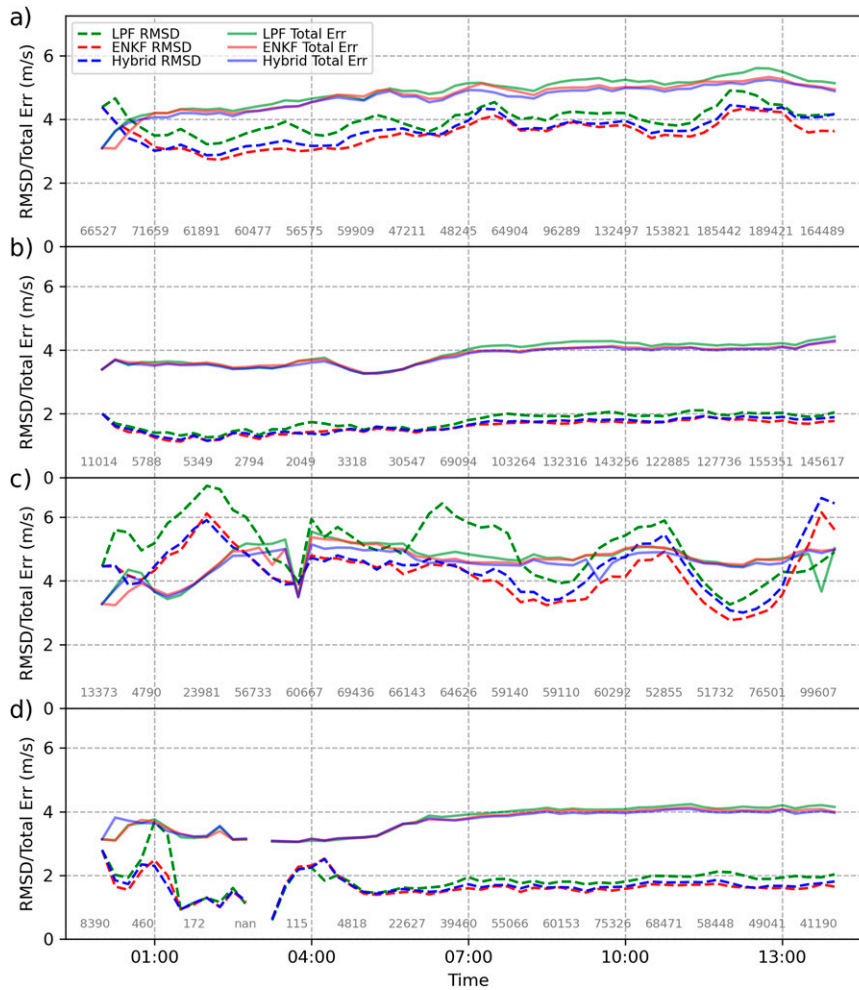


FIG. 12. As in Fig. 11, but for radar radial velocity.

features such as bodies of water. We additionally plotted domain-averaged accumulated gridcell precipitation for the same filter configurations (Fig. 14b). Forecasts initialized from standard EnKF initial conditions quickly produce greater cold pool extent than any other filter configuration, including no assimilation. Notably, the increased cold pool extent is not created directly in the posterior ensemble but appears rapidly over the first 15 min of the forecast period, suggesting that standard EnKF adjustments create conditions for enhanced convective development. Results for accumulated precipitation show similarly filter-specific behavior, with the standard hybrid forecasts producing significantly less precipitation over the forecast period than either the standard EnKF or no assimilation cases.

Forecasts initialized from filter updates that do not consider hydrometeor quantities show quite similar progressions of cold pool extent, with both filters producing slightly more prolific cold pools than forecasts initialized from the prior ensemble. The same EnKF and hybrid configurations also result in closely matched reductions to accumulated precipitation over the forecast period when compared to forecasts from the prior ensemble.

Together, these results suggest that filter-specific behavior between the EnKF and hybrid methods is strongly mediated by updates to microphysical quantities during data assimilation that temporarily enhance (weaken) convective activity.

To establish bulk trends in the direction of filter updates to hydrometeors, we plotted the evolution of rainwater column mass averaged across the domain (Fig. 14c). Filter adjustments for standard configurations of the EnKF and hybrid act in opposite directions, with the EnKF increasing column mass and the hybrid decreasing mass from the prior to initial conditions. There is relatively little adjustment to column mass for configurations that do not update hydrometeors. Although column mass returns to similar values for both standard filter configurations by the end of the 45-min forecast period, it is notable that significant differences in column mass remain by 15-min post-initialization, which implies that the effects of filter adjustment to bulk microphysical quantities could compound during sequential cycling. The persistence of transient adjustments from data assimilation is the likely cause for differing trends in cold pool intensity between EnKF and hybrid during our sequential runs.

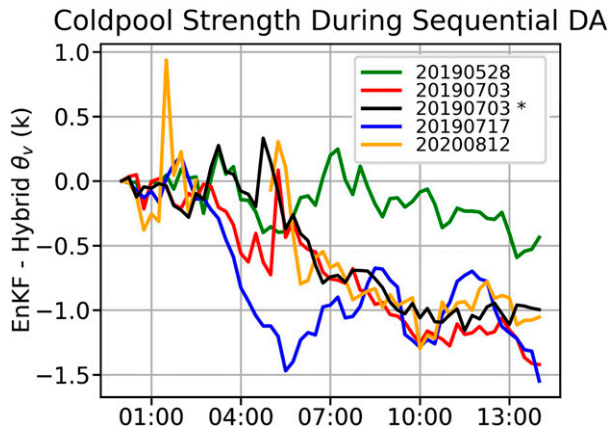


FIG. 13. Ensemble-averaged surface virtual potential temperature difference between EnKF and hybrid priors over the duration of sequential cycling for each test case event. Calculated for grid cells featuring composite reflectivity above 25 dBZ. The x axis denotes hours since the start of sequential assimilation. The asterisk (*) in legend indicates the experiment using Thompson microphysics.

We supplement the preceding time series of spatial average quantities with snapshots of cross sections covering a $300 \times 300 \text{ km}^2$ region of concentrated storm development near the center of the domain. The horizontal cross sections in Fig. 15 indicate areas of ensemble-averaged surface θ_v below 306.5 K at initialization (Fig. 15a) and 15 min after assimilation with either the EnKF (Fig. 15b) or hybrid (Fig. 15c) method. For the latter two cases, we also show shaded areas of ensemble probability for exceeding 2-mm accumulated surface precipitation. Areas of surface θ_v below 306.5 K roughly correspond to cold pool extent for the domain of interest. The EnKF ensemble produces large regions with accumulated precipitation 15-min post-initialization, while the hybrid ensemble shows such accumulation only for more limited regions and for a smaller subset of ensemble members. Regions of high probability of

accumulation in the EnKF largely correspond to areas where the cold pool is present in the EnKF ensemble average, but not in the hybrid. This result provides further evidence that the domain-averaged tendency of the hybrid method toward higher surface θ_v during sequential cycling is most likely the result of the accumulated effects of filter updates which restrict the areal extent of moderate to heavy precipitation in comparison to assimilation with an EnKF.

c. Spatial verification of forecasts

We verified the performance of forecasts initialized during our sequential data assimilation experiments using the fractions skill score (FSS) forecast verification technique given in Roberts and Lean (2008). This is a spatial verification method that uses a neighborhood approach to account for displacement errors and gauge forecast skill for varying degrees of spatial specificity. For our verifying event, we chose the occurrence of composite radar reflectivity over 25 dBZ to correspond with areas of active storm activity. FSS statistics for 5- and 15-min cycling shown in Fig. 16 took as a sample all 20 forecast ensemble members for every forecast initialization time within the scope of our four events, for a total of 1520 scored forecasts. Results for median, 25th- and 75th-percentile scores therefore represent a generalized assessment of forecast performance under variable constraints approximating real-time forecasting conditions. In addition to fractions skill score we also show related neighborhood scores using false alarm rate (FAR) and probability of detection (POD), which are given by the following relations:

$$\text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \tag{1}$$

$$\text{POD} = \frac{\text{TP}}{\text{FN} + \text{TP}}, \tag{2}$$

where FN and FP are, respectively, the numbers of false negatives and false positives given by the model with respect to

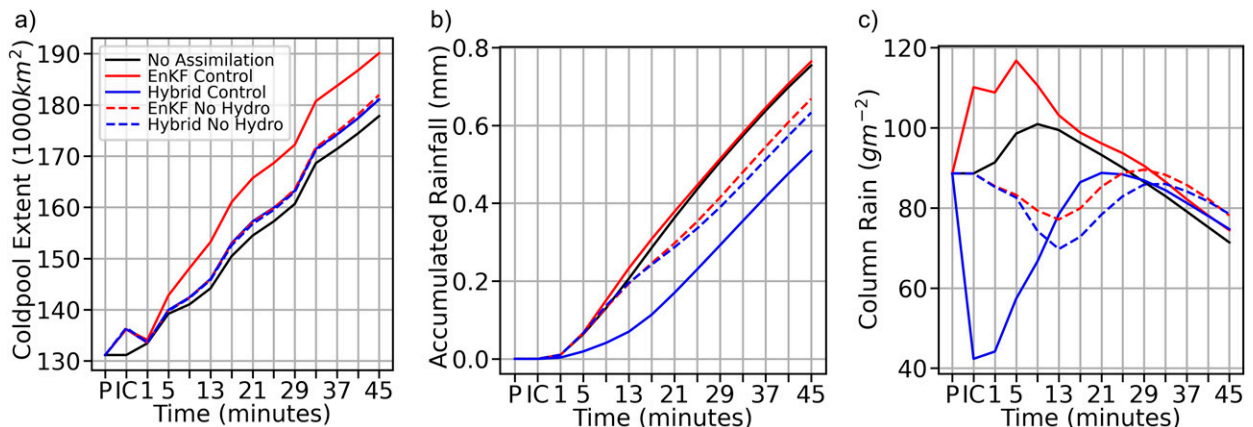


FIG. 14. (a) Cold pool extent, (b) accumulated rainfall, and (c) column rainwater mass for prior, initial conditions, and subsequent forecast times for identical prior experiments initialized at 1730 UTC 3 Jul, averaged over forecast members. Values are shown for (black) no assimilation, (red) assimilation with standard EnKF and (blue) hybrid configurations, and (dashed lines) assimilation with EnKF and hybrid configurations that do not update microphysical variables.

Coldpool Extent and Accumulated Precipitation

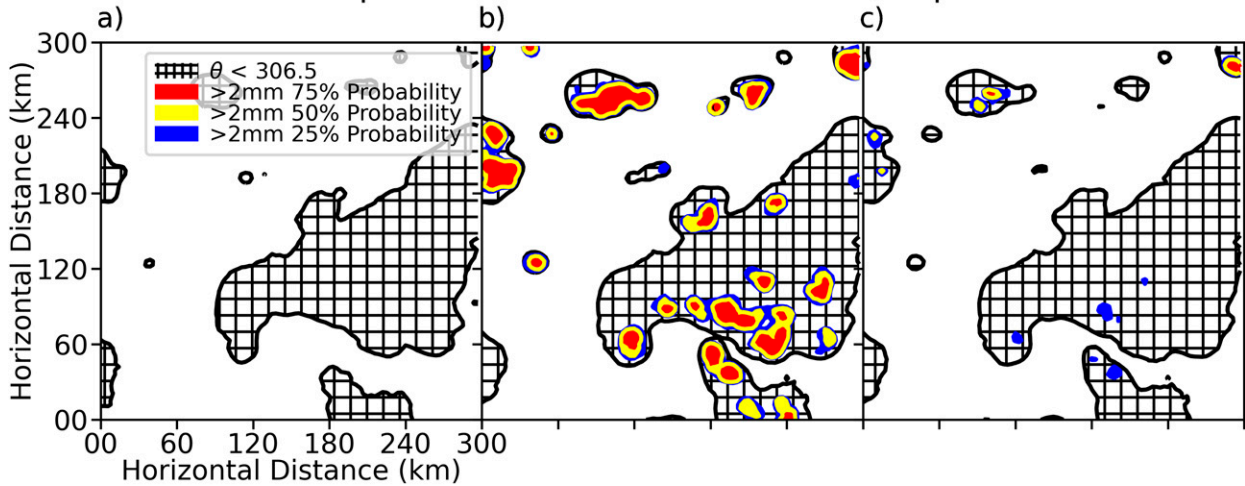


FIG. 15. (a) Cold pool extent (hatched area) at initialization. Cold pool extent and ensemble probabilities of exceeding 2-mm accumulated surface precipitation 15 min after single assimilation cycle with (b) EnKF and (c) hybrid.

the 25-dBZ threshold as verified by radar observations. TN and TP are likewise true positives and true negatives. All preceding calculations consider only grid points within a neighborhood window centered on a single gridpoint location.

Domain averaged values were then found by centering a window at each grid point, with the exception of points less than 100 km from the domain boundary. In the specific case of FSS, the number of windows with at least a single observed

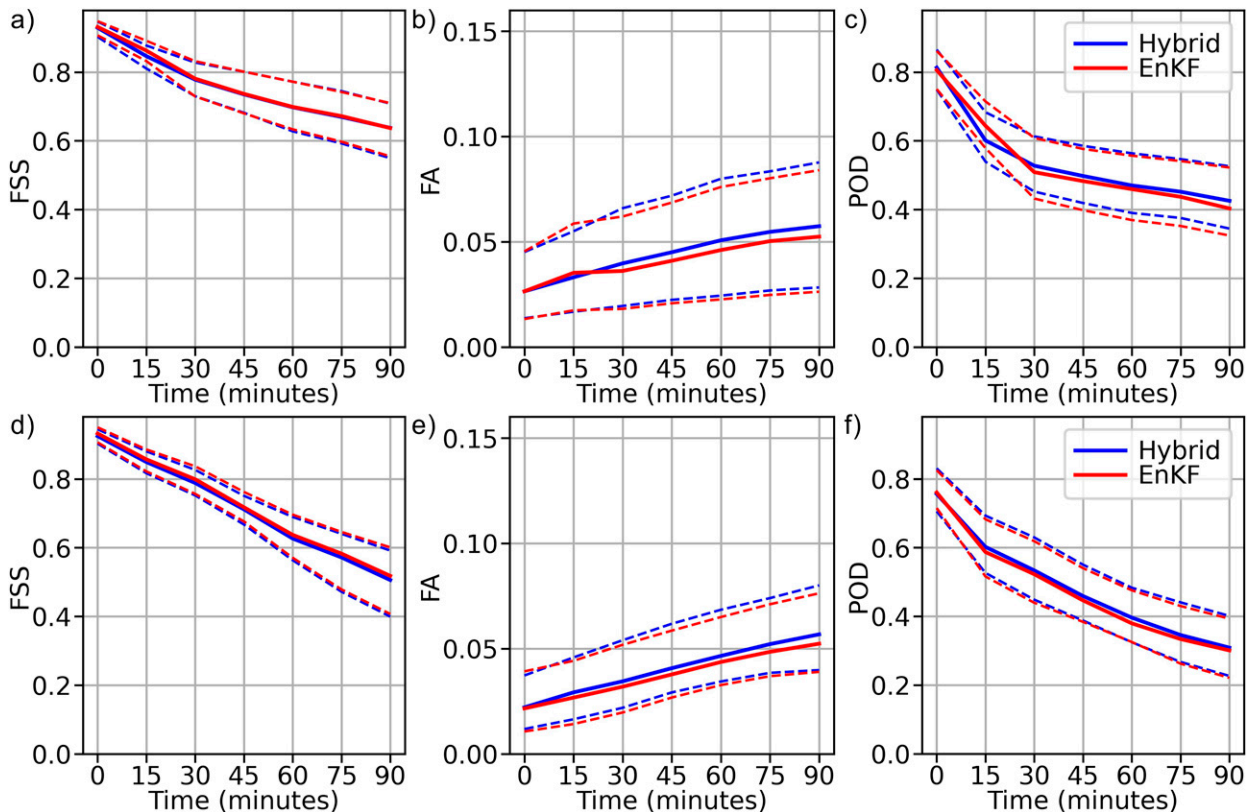


FIG. 16. Fractions skill score for (a) 15- and (d) 5-min cycling. False alarm rate for (b) 15- and (e) 5-min cycling. Probability of detection for (c) 15- and (f) 5-min cycling. Metrics are aggregated across all events based on time after initialization. Solid lines indicate median values and dashed lines indicate 25th- and 75th-percentile values.

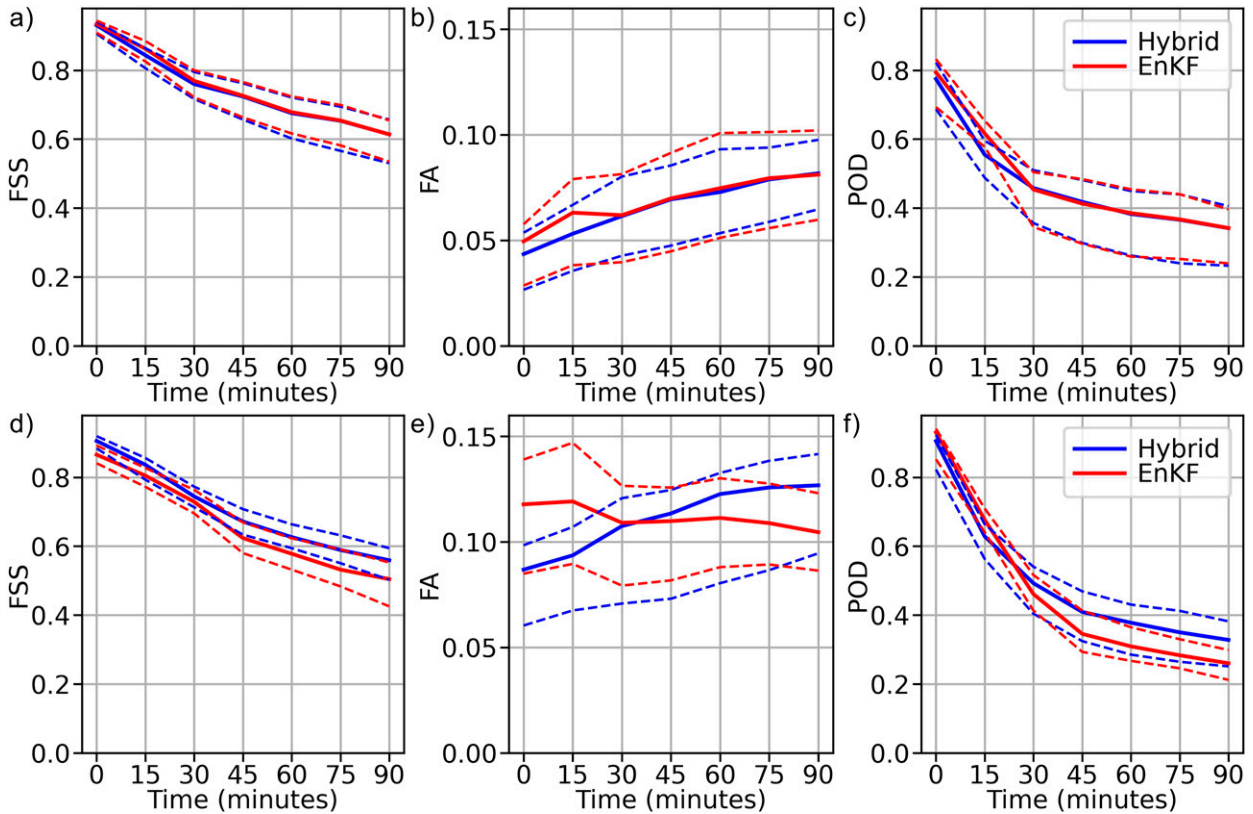


FIG. 17. Fractions skill score for forecasts from sequential data assimilation with (a) NSSL and (d) Thompson microphysics. False alarm rate with (b) NSSL and (e) Thompson microphysics. Probability of detection with (c) NSSL and (f) Thompson microphysics. Metrics are aggregated for the 3 Jul 2019 test event based on time after initialization. Solid lines indicate median values and dashed lines indicate 25th- and 75th-percentile values.

event acted as a normalizing factor for the calculation of the domain averaged score, whereas for all other metrics the domain average was taken as the simple numerical mean of window scores. We chose a neighborhood length scale of 30 km for all preceding calculations, which provided a balanced metric informed by both the resolution of discrete storm structures as well as more general performance characteristics such as precipitation bias.

For 15-min cycling, the EnKF shows a slight advantage in terms of FSS for the first 30-min post-initialization, with both filters achieving nearly identical scores for time stamps from 30- to 90-min post-initialization. Looking at scores for POD and FAR, the EnKF maintains an advantage for POD from initialization to 30 min, from which point the hybrid achieves higher scores. Hybrid forecasts produce a slightly lower FAR than EnKF forecasts from initialization to 30-min time stamps, with longer forecasts showing a higher FAR for the hybrid. As FSS reflects an award for predicted events as well as a penalization for false alarms, it is apparent that the slight false alarm advantage of the hybrid during early forecast time stamps is canceled out by the enhanced detection of the EnKF. Later time stamps reflect a balance between a higher false alarm rate in hybrid initialized forecasts and a lower probability of detection in forecasts initialized with the EnKF.

d. Sequential data assimilation cycling with the Thompson microphysical scheme

Having established the relevance of microphysical quantities for filter-specific behavior indicated by forecast evolutions, we expanded our analysis by performing an additional sequential data assimilation experiment using the Thompson microphysical scheme in place of the NSSL microphysical scheme used in our other case studies. For this experiment, we performed sequential assimilation with a 15-min frequency for the 3 July 2019 test case event featuring mixed-mode convection and introduce no changes in experiment parameters other than a different microphysical scheme. From our set of forecasts, we again present verification metrics in the form of FSS, POD, and FAR. We present these along with corresponding metrics for the same test case event using NSSL microphysics in Fig. 17. In contrast to results presented in the preceding subsection, forecast verification with Thompson microphysics shows an unambiguous advantage for cycling with hybrid assimilation, which is associated with relatively high false alarm rates for the EnKF at early forecast time stamps and a strong advantage in POD for the hybrid at later time stamps. As for our experiments with NSSL microphysics, we also produced estimates for cold pool strength over the course of sequential cycling. These results are shown as the black

line in Fig. 13. For the 3 July 2019 case study, sequential assimilation with Thompson microphysics shows a highly similar filter specific difference in cold pool intensity compared to the same experiment performed with NSSL microphysics.

5. Discussion

a. Posterior characteristics of the hybrid particle filter

The benefits of particle filtering for state estimation lie in the ability to correctly sample from underlying non-Gaussian distributions, which comes at the expense of increased sampling error compared to parametric methods. This trade-off motivates the use of a hybrid method that applies EnKF assimilation after tempering iterations with a localized particle filter have transformed the prior sampling distribution into Gaussian form, which can occur for Gaussian likelihoods (Poterjoy 2022b). Our sequential data assimilation experiments show that the hybrid method is able to maintain desirable posterior characteristics when sampling from non-Gaussian marginals for microphysical variables while outperforming the pure LPF in terms of RMSD for reflectivity and radial velocity observations. Crucially, our results demonstrate that 15-min frequency cycling with the hybrid method produces posterior PDFs for microphysical variables that largely remain within the region of state space support indicated by the prior ensemble. This contrasts with EnKF assimilation, which produces posterior ensemble members containing state-space solutions outside of the model attractor basin. Since solutions outside of climatological model priors can only be produced by data assimilation, we know that the hybrid is at the very least producing posterior members that more accurately depict physical relationships represented by the WRF Model.

One of the key advantages in reducing bias induced by data assimilation is the reduction of confounding factors for identifying residual uncertainty. In the context of our sequential data assimilation experiments, we see a tendency for 5-min cycling to maintain lower mixing ratios and reflectivity in prior members than for 15-min cycling, regardless of the assimilation method. 5-min prior solutions reflect model states with less time to deviate from observational constraint and are potentially closer to the “true” system state barring consideration of observational uncertainty which is relatively low for our dense network of reflectivity observations. The discrepancy may be the result of model bias toward high reflectivity introduced into prior ensemble members with longer model runs between assimilation steps. It is especially notable that the posterior PDF generated by the hybrid at a 15-min cycling frequency reduces probability density among high reflectivity, high mixing ratio solutions in a way that closely resembles the prior PDFs for 5-min cycling. The fact that these solutions maintain high posterior probability densities during 15-min cycling with the EnKF suggests that data assimilation bias is strong enough to cover up signatures from model bias under conditions of significantly non-Gaussian prior probability.

b. Impact of data assimilation bias on forecast evolution

Although the presence of data assimilation bias has been acknowledged and studied in more theoretical contexts (e.g., Posselt 2016), our modeling framework, along with the hybrid LPF as a reference method, allows for more direct insight into how these biases impact forecast evolution and prior model states during sequential assimilation with Gaussian filters. With regards to the EnKF, data assimilation bias consistently drives the development of heavier surface precipitation and cold pools that are more intense and widespread than those of the hybrid. We note that the filter specific trend in cold pool intensity is present for sequential assimilation with both NSSL and Thompson microphysics.

We speculate that for experiments with NSSL microphysics, the comparable performance of the hybrid LPF and EnKF methods in terms of root-mean-square error during sequential data assimilation and fractions skill score for forecasts could be the result of a compensatory effect from strong cold pools toward making up for sluggish convective initiation by the model. Such an effect would explain the relatively high POD present in EnKF forecasts up to the end of the 90-min forecast period. As previously noted, experiments with Thompson microphysics show a clear verification advantage for the hybrid filter and low POD for EnKF forecasts past 30 min, which suggests that compensatory advantages from DA bias are not a factor in such cases. This may indicate that the structure of data assimilation bias alleviated by the hybrid filter varies significantly between schemes, or that these schemes produce differential storm evolutions from similarly structured bias.

Interestingly, results obtained from data assimilation experiments using identical priors suggest that filter updates to wind, temperature, and moisture are mediated almost entirely by updates to microphysical quantities for time scales of about an hour. This finding is consistent with the cumulative effects of hydrometeor updates on cold pool strength that occurs during cycling data assimilation and is not unprecedented given the nonlinear relationships between microphysical variables and consequent non-Gaussianity of their marginal prior PDFs. Updates to microphysical variables may therefore produce more significant data assimilation bias than other state variables. Bias introduced to microphysical variables may also be more impactful for convective organization than bias in other variables. Further investigation in this respect is warranted due to the limited nature of our experiments that considered only a single initialization time within a single event.

6. Conclusions

a. Summary of results

The current study presents results obtained from a set of comparative real data assimilation experiments using the localized particle filter (LPF), a hybrid LPF, and an ensemble Kalman filter (EnKF). For each filter, we examined posterior characteristics from sequential assimilation by constructing prior and posterior marginal PDFs for microphysical variables with defined nonlinear physical relationships. Posterior EnKF

ensemble members show a significant departure from the support of climatological PDFs indicated by prior ensemble members in accordance with model physics over the duration of our test events, providing evidence for strong data assimilation bias in updates to microphysical quantities. In contrast, both hybrid and LPF assimilation produce posterior PDFs that more closely match climatological prior PDFs. We calculated verification statistics to quantify prior mean errors during sequential data assimilation, showing that the hybrid method is able to outperform the pure LPF and achieve similar verification scores to assimilation with an EnKF. We then compared the impact of hybrid and EnKF assimilation on forecast evolution by calculating a proxy for cold pool strength over the course of sequential assimilation, demonstrating that cycling with the EnKF produces significantly stronger cold pools than cycling with a hybrid particle filter. To understand these findings, we compared data assimilation experiments performed using identical priors for the pair of data assimilation schemes. Our results suggest that even a single filter update produces initial conditions that are distinct enough to cause major differences in cold pool propagation. Crucially, these experiments also show that updating only non-microphysical variables is insignificant for reproducing filter-specific differences, which we attribute to strong non-Gaussianity for microphysical quantities. Having established contrasting behavior in forecast evolution with EnKF and hybrid assimilation, we used neighborhood verification metrics to examine the performance of ensemble forecasts initialized during sequential data assimilation experiments. Fractions skill scores calculated for simulated reflectivity are comparable between filters and are mediated by distinct behavior in probability of detection and false alarm rate. Nevertheless, repeating experiments using Thompson microphysics yielded significant changes in relative forecast skill, with the hybrid showing advantages over the EnKF. This finding underscores the importance of uncovering bias induced by data assimilation assumptions to validate choices for subgrid-scale parameterization schemes. As a whole, our results suggest that the hybrid particle filter introduces less data assimilation bias during updates to microphysical and other quantities than Gaussian methods, which translates to quantifiable differences in forecast verification and in the dynamic evolution of modeled convective storms.

b. The hybrid particle filter as an emerging tool for scheme selection and parameter estimation

The unique properties offered by the hybrid particle filter make it well suited for parameterization scheme selection and parameter estimation within the context of a given scheme. The latter could be done through a joint state and parameter space estimation using an augmented state vector approach. Joint state-parameter estimation has been performed with EnKFs but can suffer degraded performance from bias induced by Gaussian assumptions. This bias arises from non-Gaussianity in the state and is further aggravated by the nonlinear coupling of parameters to model responses (Ruckstuhl and Janjić 2018). As demonstrated by our current findings, the hybrid particle

filter is able to significantly reduce bias introduced when sampling non-Gaussian PDFs, and therefore may be more amenable to such an approach. The high-dimensional, nonlinear optimization problems posed in scheme selection and parameter estimation may also be conveniently approached using genetic algorithm-type optimization methods that avoid expensive and often intractable efforts at quantifying bias across the parameter space (Wang 1997). In the geosciences, these methods have been applied to parameter optimization for runoff and mineral deposit models (Siriwardene and Perera 2006; Abdelazeem et al. 2019). Applying this method for atmospheric models would involve the optimization of schemes and parameters by a natural selection type process that uses verification statistics to modify, add, or remove discrete ensembles with varying parameter space configurations over the course of sequential data assimilation for a variety of nonidealized case events. Although this approach is technically feasible using any ensemble filtering method, our results have shown that data assimilation bias introduced through Gaussian assumptions can quickly translate to large impacts on forecast evolution and verification statistics that vary between parameterization schemes, and which could overwhelm selection pressures imposed by model behavior on its own. Using the hybrid LPF as our filtering method opens the door to a genetic algorithm approach that selects parameter configurations based on the performance of model configurations informed by observations in a manner consistent with their own attractor basin. This approach, therefore, allows for optimization in nonidealized contexts without significant hindrance by initial condition uncertainty.

Acknowledgments. This study was supported by a National Science Foundation CAREER Award AGS1848363. We would like to also acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

Data availability statement. HRRR data are publicly available from NOAA at archives hosted by Amazon Web Services (<https://registry.opendata.aws/noaa-hrrr-pds/>) and Google Cloud Platform (<https://console.cloud.google.com/marketplace/product/noaa-public/hrrr?project=python-232920&pli=1>). NEXRAD data are made openly available by NOAA via multiple sources (<https://doi.org/10.7289/V5W9574V>). Code used by the authors to generate ensemble analyses and forecasts is available on a Github repository (<https://github.com/synopticscale/WRF-DART>).

REFERENCES

- Abdelazeem, M., M. Gobashy, M. H. Khalil, and M. Abdrabou, 2019: A complete model parameter optimization from self-potential data using Whale algorithm. *J. Appl. Geophys.*, **170**, 103825, <https://doi.org/10.1016/j.jappgeo.2019.103825>.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2).

- , and N. Collins, 2007: Scaleable implementation of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463, <https://doi.org/10.1175/JTECH2049.1>.
- Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.*, **108**, 8775, <https://doi.org/10.1029/2002JD002900>.
- , P. Bickel, and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman*, D. Nolan and T. Speed, Eds., Vol. 2, Institute of Mathematical Statistics, 316–334.
- Bickel, P., B. Li, and T. Bengtsson, 2008: Sharp failure rates for the bootstrap particle filter in high dimensions. *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, B. Clarke and S. Ghosal, Eds., Vol. 3, Institute of Mathematical Statistics, 318–329.
- Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2).
- Chustagulprom, N., S. Reich, and M. Reinhardt, 2016: A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems. *SIAM/ASA J. Uncertainty Quantif.*, **4**, 592–608, <https://doi.org/10.1137/15M1040967>.
- Doucet, A., N. de Freitas, and N. Gordon, Eds., 2001: An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2–14.
- Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927, <https://doi.org/10.1175/2008JTECHA1156.1>.
- Farchi, A., and M. Bocquet, 2018: Review article: Comparison of local particle filters and new implementations. *Nonlinear Processes Geophys.*, **25**, 765–807, <https://doi.org/10.5194/npg-25-765-2018>.
- Ghonima, M. S., H. Yang, C. K. Kim, T. Heus, and J. Kleissl, 2017: Evaluation of WRF SCM simulations of stratocumulus-topped marine and coastal boundary layers and improvements to turbulence and entrainment parameterizations. *J. Adv. Model. Earth Syst.*, **9**, 2635–2653, <https://doi.org/10.1002/2017MS001092>.
- Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian state estimation. *IEE Proc.*, **140**, 107–113, <https://doi.org/10.1049/ip-f-2.1993.0015>.
- Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522, <https://doi.org/10.1175/2010MWR3328.1>.
- 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>.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shepard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, <https://doi.org/10.1029/2008JD009944>.
- James, E. P., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, <https://doi.org/10.1175/WAF-D-21-0130.1>.
- Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, and J. P. Montávez, 2012: A revised scheme for the WRF surface layer formulation. *Mon. Wea. Rev.*, **140**, 898–918, <https://doi.org/10.1175/MWR-D-11-00056.1>.
- Johnson, R. M., 2019: *Assessment of One-Moment and Two-Moment Bulk Microphysics and Spectral Bin Microphysics Schemes Using Idealized Supercell Simulations and Real Data Convective-Scale Predictions*. University of Oklahoma, 206 pp.
- Jones, T. A., K. H. Knopfmeier, D. M. Wheatley, and G. J. Creager, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, <https://doi.org/10.1175/WAF-D-15-0043.1>.
- , P. Skinner, K. Knopfmeier, E. Mansell, P. Minnis, R. Palikonda, and W. Smith Jr., 2018: Comparison of cloud microphysics schemes in a Warn-on-Forecast System using synthetic satellite objects. *Wea. Forecasting*, **33**, 1681–1708, <https://doi.org/10.1175/WAF-D-18-0112.1>.
- 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>.
- Kurosawa, K., and J. Poterjoy, 2022: A statistical hypothesis testing strategy for adaptively blending particle filters and ensemble Kalman filters for data assimilation. *Mon. Wea. Rev.*, **151**, 105–125, <https://doi.org/10.1175/MWR-D-22-0108.1>.
- Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, <https://doi.org/10.1175/WAF-D-17-0145.1>.
- Lee, Y., and A. J. Majda, 2016: State estimation and prediction using clustered particle filters. *Proc. Natl. Acad. Sci. USA*, **113**, 14 609–14 614, <https://doi.org/10.1073/pnas.1617398113>.
- Leitman, E., and R. Thompson, 2019: Mesoscale discussion 851. NOAA/Storm Prediction Center, accessed 12 October 2020, <https://www.sp.noaa.gov/products/md/2019/md0851.html>.
- Mansell, E. R., 2010: On sedimentation and advection in multimoment bulk microphysics. *J. Atmos. Sci.*, **67**, 3084–3094, <https://doi.org/10.1175/2010JAS3341.1>.
- Penny, S. G., and T. Miyoshi, 2016: A local particle filter for high dimensional geophysical systems. *Nonlinear Processes Geophys.*, **23**, 391–405, <https://doi.org/10.5194/npg-23-391-2016>.
- Posselt, D. J., 2016: A Bayesian examination of deep convective squall-line sensitivity to changes in cloud microphysical parameters. *J. Atmos. Sci.*, **73**, 637–665, <https://doi.org/10.1175/JAS-D-15-0159.1>.
- 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>.
- , 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>.
- , 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>.
- , 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>.
- , 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>.
- , L. Wicker, and 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>.
- Potvin, C., and Coauthors, 2020: Assessing systematic impacts of PBL schemes on storm evolution in the NOAA Warn-on-Forecast System. *Mon. Wea. Rev.*, **148**, 2567–2590, <https://doi.org/10.1175/MWR-D-19-0389.1>.
- Robert, S., and H. R. Künsch, 2017: Localizing the ensemble Kalman particle filter. *Tellus*, **69A**, 1282016, <https://doi.org/10.1080/16000870.2017.1282016>.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- Ruckstuhl, Y. M., and T. Janjić, 2018: Parameter and state estimation with ensemble Kalman filter based algorithms for convective-scale applications. *Quart. J. Roy. Meteor. Soc.*, **144**, 826–841, <https://doi.org/10.1002/qj.3257>.
- Siriwardene, N. R., and B. J. C. Perera, 2006: Selection of genetic algorithm operators for urban drainage model parameter optimisation. *Math. Comput. Modell.*, **44**, 415–429, <https://doi.org/10.1016/j.mcm.2006.01.002>.
- Skamarock, W. C., and Coauthors, 2019: A description of the Advanced Research WRF Model version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp., <https://doi.org/10.5065/1dfh-6p97>.
- Skinner, P., and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast System. *Wea. Forecasting*, **33**, 1225–1250, <https://doi.org/10.1175/WAF-D-18-0020.1>.
- Smirnova, T. G., J. M. Brown, S. G. Benjamin, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, <https://doi.org/10.1175/MWR-D-15-0198.1>.
- Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640, <https://doi.org/10.1175/2008MWR2529.1>.
- Stensrud, D. J., and Coauthors, 2009: Convective-scale Warn-on-Forecast System: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, <https://doi.org/10.1175/2009BAMS2795.1>.
- , and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, <https://doi.org/10.1016/j.atmosres.2012.04.004>.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>.
- van Leeuwen, P. J., H. R. Künsch, L. Nerger, R. Potthast, and S. Reich, 2019: Particle filters for high-dimensional geoscience applications: A review. *Quart. J. Roy. Meteor. Soc.*, **145**, 2335–2365, <https://doi.org/10.1002/qj.3551>.
- Wang, Q. J., 1997: Using genetic algorithms to optimise model parameters. *Environ. Modell. Software*, **12**, 27–34, [https://doi.org/10.1016/S1364-8152\(96\)00030-8](https://doi.org/10.1016/S1364-8152(96)00030-8).
- Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast System. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, <https://doi.org/10.1175/WAF-D-15-0043.1>.
- 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>.