ABSTRACT

Title of Dissertation: IMPLICATIONS OF NEW OBSERVING AND DATA ASSIMILATION STRATEGIES FOR MOIST CONVECTION IN THE ATMOSPHERE

Joshua McCurry
Doctor of Philosophy, 2024

Dissertation Directed by: Professor Jonathan Poterjoy
Department of Atmospheric & Oceanic Sciences

Obtaining a faithful probabilistic depiction of moist convection is challenging because of intrinsic predictability limits of geophysical systems at small scales as well as errors present within models and observations. Understanding each of these challenges is further complicated by assumptions made in data assimilation (DA) techniques. In order to untangle sources of uncertainty in convective weather regimes and ultimately improve state estimates for severe convective storms, we evaluate novel data assimilation and observing strategies using a WRF modeling system adapted and expanded from the National Severe Storm Laboratories Warn on Forecast System. As a core component of our vision for moist-convective state estimation, we evaluate particle filtering (PF) methods as an alternative to DA methods based on linear estimation theory. From a Bayesian perspective, PF's represent prior and posterior error distributions non-parametrically rather than assuming a Gaussian distribution and can accept any type of likelihood function. This approach is known to reduce bias introduced by data assimilation methods that rely on only the

first two statistical moments to perform state estimation. The form of PF used in this research implements a localization strategy that makes it affordable for large geophysical applications, such as convection-allowing weather models. Beyond the methods used for DA, our work also explores fundamental aspects of how mesoscale convective systems (MCS's) are constrained by observations in analyses and forecasts, with implications for the configuration of observing systems and choices made in parameterization schemes. The first part of this dissertation examines posterior ensembles and their forecasts for selected severe weather events between 2019 — 2020, comparing results from the PF with those from an Ensemble Kalman Filter (EnKF). We find that PF-based DA with 64-member ensembles 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-induced bias. These differences are significant enough to impact the dynamic evolution of convective systems via cold pool strength and propagation. The second phase of this dissertation leverages the same mesoscale analysis and forecasting system to examine observation collection strategies for the evolving Maryland Mesonet. This portion of research introduces a framework for conducting observing system simulation experiments (OSSEs) and establishes limitations on the marginal utility of surface observing system density for EnKF assimilation in moist-convective contexts. The final portion of this dissertation expands upon earlier findings with the PF by evaluating a set of techniques intended to improve state estimation with the PF by addressing deficiencies caused by model error and pre-processing constraints placed on observations, as well as relaxing assumptions of regime-invariance and Gaussianity made for likelihood functions. We find that modifications to key parameters within the cloud microphysics scheme adopted for this research benefit precipitation and hydrometeor estimates produced by the PF—in a manner that is difficult to discern from similarly configured EnKF ex-

periments. Additional benefits are found in observing configurations for reflectivity that remove the typical distinction between reflectivity returns above and below a cut-off threshold, which we again hypothesize to be a design choice that has a great impact on PF-based DA. Lastly, we find that the removal of Gaussian assumptions does not improve the resulting state estimates, but note that potential improvements may be achieved with observation uncertainty quantification strategies that better account for analysis error when forming non-parametric representations of observation errors within the DA framework.

IMPLICATIONS OF NEW OBSERVING AND DATA ASSIMILATION STRATEGIES FOR MOIST CONVECTION IN THE ATMOSPHERE

by

Joshua McCurry

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Professor Jonathan Poterjoy, Chair/Advisor
Professor Kayo Ide
Professor Brian Hunt
Professor Maria Molina
Professor Alfredo Ruiz-Barradas

Preface

This dissertation contains material reproduced or modified from two first-authored publications. Chapter one was originally published in 2023 in the Monthly Weather Review McCurry et al. (2023). Material in chapter two was published was originally published in 2024 in Weather and Forecasting McCurry and Poterjoy (2024) . Chapter three presents additional findings from sequential data assimilation experiments intended to extend concepts developed in chapter one. Copyright information for work originally published in AMS journals is included below:

Acknowledgments

I thank all of my committee members: Jon Poterjoy, Kayo Ide, Brian Hunt, Maria Molina, and Alfredo Ruiz-Barradas for their invaluable assistance and feedback in crafting my dissertation.

I further acknowledge my advisor, Jon Poterjoy, for guiding my academic and professional development over the past six years. Jon's enthusiasm for data assimilation and genuine desire to impart knowledge, combined with his unwavering dependability and patience over my time here were exactly what I needed to succeed. I could not have asked for a better advisor.

I would also like to acknowledge my fellow labmates in ATL 4339, especially Kenta Kurosawa, Henry Santer and Joseph Knisely who've been my comrades in the struggle against background error. In this effort I was also greatly helped by friendly chats with Benjamin Woods, Hannah Daley, Henry Hausmann, Sarah Loughran and others that made my time here more enjoyable.

I would like to send a general thank you to the AOSC department, for giving me the rare and priceless opportunity to develop a lifelong passion for weather into a career, and for supporting me throughout my time here.

For their special help in my academic journey, I thank my senior thesis advisor at JHU: professor Anand Gnanadesikan, who assisted me in escaping the world of chemical engineering, and Debra Baker at ESSIC who first put me in touch with research opportunities at UMD.

Of course, I could not have done any of this without my people at home: my friends Eric and Max, my parents, and most of all my loving wife Whitney who is the joyful, outgoing yin to my yang.

And finally, I dedicate this paper to the city of Baltimore: the greatest city in America, my home from 2013 to 2023. Most of what I know of atmospheric science and data assimilation was learned first at AOSC but digested on a walk past rowhomes on Calvert, Charles or St. Paul.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACARS	Aircraft Communications Addressing and Reporting System
CSI	Critical Success Index
DA	Data assimilation
DART	Data Assimilation Research Toolkit
EAKF	Ensemble Adjustment Kalman Filter
EQD	equidistant
EQD+	modified-equidistant
EnKF	Ensemble Kalman filter
FSOI	Forecast Sensitivity to Observation Impact
FAR	False alarm rate
FP	False positive
FN	False negative
FSS	Fractions Skill Score
HRRR	High Resolution Rapid Refresh
HRRRE	High Resolution Rapid Refresh Ensemble
IC	Initial conditions
IPF	Iterative particle filter
LBC	Lateral boundary conditions
LPF	Localized Particle Filter
MADIS	Meteorological Assimilation Data Ingest System
MCS	Mesoscale convective system
METAR	METeorological Aerodrome Report
MM	Maryland Mesonet
MRMS	Multi-Radar Multi-Sensor
NEXRAD	Next Generation Radar
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical weather prediction
NSSL	National Severe Storms Laboratory
NVD	NSSL 2-moment variable-density scheme
OSE	Observing system experiment
OSSE	Observing system simulation experiment
PDF	Probability density function
POD	Probability of detection
POP	Population-centered

QLCS	quasi-linear convective system	
RAP	Rapid Refresh RMSD	root-mean-square departure from observations
RMSE	root-mean-square error	
RRFS	Rapid Refresh Forecast System	
RRTM	Rapid Radiative Transfer Model	
TP	True positive	
TN	True negative	
WEST	West-centered	
WoFS	Warn on Forecast System	
WRF	Weather Research & Forecasting Model	

Chapter 1: Motivation and background for experiments implementing novel data assimilation and observing strategies

1.1 Introduction

Although geophysical data assimilation (DA) is commonly understood as the use of observational data to inform numerical models and their predictions, it is less commonly appreciated as the combined action of two forms of model–one statistical, one physical–through which the statistical model draws initial conditions by updating a probabilistic representation of quantities from an underlying deterministic physical model. In this context, a perfect physical model would exactly capture all processes acting upon an input system across all scales, whereas a perfect statistical model would exactly capture all ways in which observations impact uncertainty in our state estimates via Bayes' Theorem. As with physical models, a perfect statistical model is not achievable and their practical implementation requires various forms of approximation. These approximations vary significantly between modeling systems and include the approximation of dynamically evolving covariances with Gaussian ensemble statistics in the case of the ensemble Kalman filter, as well as the compulsory use of inflation and localization to cope with model error and sampling deficiency.

Approaches to DA relying on such approximations and assumptions have proven quite ver-

satile to an array of domain applications–from the medium-range prediction of global synoptic-scale weather patterns to the short-range prediction of severe weather impacts. It is often the case that settings for inflation, localization, and fixed observational error covariances are heavily tuned for operational contexts, such that they reflect an optimized configuration that may outperform systems featuring weaker assumptions. At the same time, evolving forecast requirements and the desire for state estimates that more closely reflect solutions to Bayes' theorem may require the implementation of strategies that—among other things-–do not depend on Gaussian assumptions for prior errors. While Gaussian-assuming methods like the EnKF are inherently rooted in Bayesian estimation theory, and produce solutions to Bayes theorem that minimize posterior errors in a root-mean-squared sense regardless of the actual shape of prior distributions, they are not capable of resolving higher order moments in posterior solutions found when the true prior or likelihood function is non-Gaussian.

To expand upon this idea, we note that improving the expected level of forecast accuracy—or physical consistency—from a given modeling system inherently changes the demands placed on DA by tasking it with the production of state estimates that fall within a smaller, more accurate subset of solutions than would previously have satisfied requirements. Achieving this level of accuracy may not be feasible through simple tuning, but may instead require substantial modifications to more fundamental modeling system components, such as observing strategies, physical parameterization schemes, or DA algorithms.

For the application of DA to moist-convective regimes, there are strong motivations in favor of revisiting common assumptions in the interest of enhancing predictive capabilities. The short timescales and spatial gradients characteristic of severe convective events mean that there is immense social value in even a marginal improvement to skillful lead times or to the spatial

resolution at which potentially hazardous weather features are skillfully predicted. At the same time, skillful forecast lead times for the prediction of organized convective systems do not appear to have reached the upper limits imposed on them by inherent predictability Ramanathan et al. (2019); Ramanathan and Satyanarayana (2021), which implies that improvement is at the very least theoretically—if not simply-–achievable. We may also add–more subjectively–that there does not seem to be a unifying vision of what exactly future convection-allowing numerical prediction systems should look like, in contrast to efforts towards global models that have predominantly centered around coupled DA and observing strategies for satellite radiances.

In this dissertation, we propose an approach to DA that accounts for a broad assortment of errors constraining the accuracy of state estimates and forecasts in convective regimes, including those introduced by DA itself. The introduction of Bayesian strategies like particle filtering that minimize assumptions for prior distributions may allow for the ultimate reduction of forecast error by removing errors of approximation that conceal sources of model and observational uncertainty. These error sources may in turn be addressed with targeted observing strategies and model configurations.

1.2 Localized particle filtering for geophysical systems

Particle filter (PF) methods in DA refer to a class of sequential Monte Carlo methods that approximate probability distributions using delta functions corresponding to ensemble or 'particle' states. Perhaps the simplest form of PF for state estimation is the sequential-importance-resampling or 'bootstrap' PF. This form of particle filter applies Bayes Theorem by weighting prior particles according to observation likelihoods before applying a resampling procedure that

removes particles unlikely to account for observations while duplicating those with high likelihood. Bootstrap PFs are very simple to implement and--unlike the EnKF and other methods based on linear estimation theory—do not assume a parametric form for prior distributions. They may therefore produce posterior state estimates closer to the true Bayesian solution. Unfortunately, bootstrap PFs are limited by the occurrence of weight collapse with observations that produce large variance in likelihood across ensemble membersSnyder et al. (2008). To avoid weight-collapse, bootstrap PFs require an ensemble size that increases exponentially with the state dimension Bickel et al. (2008); Bengtsson et al. (2008)—making them unaffordable for use with high-dimensional geophysical models. This phenomenon indicates that a 'curse of dimensionality' will likely always limit the effectiveness of PFs that do not incorporate additional approximations.

Particle filtering may be made more practical by applying localization strategies that replace the direct solution of Bayes' theorem for high-dimensional systems with the solution of Bayes' theorem for lower-dimensional, localized subsystems. The LPF considered in this dissertation was first introduced in Poterjoy (2016). 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 DA. The specific version of LPF implemented for DA experiments incorporates improvements introduced by Poterjoy (2022b) to further reduce sampling error for limited ensemble sizes. These

include tempering, which uses a factorization of the 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).

1.3 Moist-convective dynamics and implications for observing strategies

The assimilation and observing strategies described in this dissertation address problems that follow directly from the non-linearity of moist-convective dynamics, and are applied to real cases featuring severe weather. Although a thorough description of moist convection is beyond the scope of this dissertation, we briefly introduce key concepts that are further developed by individual chapters and comment on factors influencing the observational constraint of important structures.

Moist convection encompasses all types of convective instability resulting in the release of latent heat via condensation. Depending on their particular form, moist-convective events may be classified as organized or unorganized. For unorganized events, the factors governing convective instability remain limited to the immediate environment of discrete convective cells. Unorganized convection typically features relatively short-lived updrafts that are eventually off from heat and moisture by evaporatively-cooled downdrafts. By contrast, organized convection is distinguished by its self-reinforcing and re-distributing nature, which acts to extend the lifespan of convection and expand the spatial extent of convective activity. A discrete formation of organized, multi-cellular convective activity is known as a mesoscale convective system (MCS). Whether a particular synoptic-scale environment supports organized or unorganized convection depends heavily on prevailing conditions of wind shear and convective-available-potential-energy

(CAPE).

Individual MCS's may be seen as complete physical systems associated with convective activity on relatively large scales (meso-β or higher) while being internally driven by processes and structures that can be significantly smaller. Key among these features are precipitation-induced cold pools and associated outflow boundaries that act to initiate secondary convection near the periphery of the MCS. The strength and spatial orientation of secondary convection are in turn influenced by the interaction of horizontal vorticity from gust fronts with that arising from prevailing environmental wind shear Rotunno et al. (1988). This can have special significance for squall-line events-–also known as quasi-linear-convective-systems (QLCSs)–which may be distinguished as an important subtype of MCS based on their unique dynamical evolution and spatial profile. In these systems, the development of an upshear-sloping convective line leads to the formation of a trailing-stratiform region from convective outflow. Stratiform precipitation, in turn, contributes to the development of a rear-inflow jet that directs evaporatively-cooled outflow towards the convective line, which can substantially prolong MCS life-cycle, and which may itself be associated with dangerous straight-line wind conditions. A highly simplified schematic of key features for a squall-line type MCS is shown below in Figure 1.1.

In contrast to micro-scale phenomena like turbulent eddies and entrainment processes, meso-γ scale MCS structures such as outflow boundaries and updrafts are resolvable at grid spacings affordable to convection-allowing ensemble systems. The proliferation of double-moment microphysical parameterization schemes–that diagnose number concentrations for hydrometeor species in addition to mixing ratios–means that such systems may also reasonably describe hydrometeor populations within convective cloud structures and their subsequent impact on latent heating. Uncertainty at meso-γ scales is therefore a prime target for constraint through DA. The

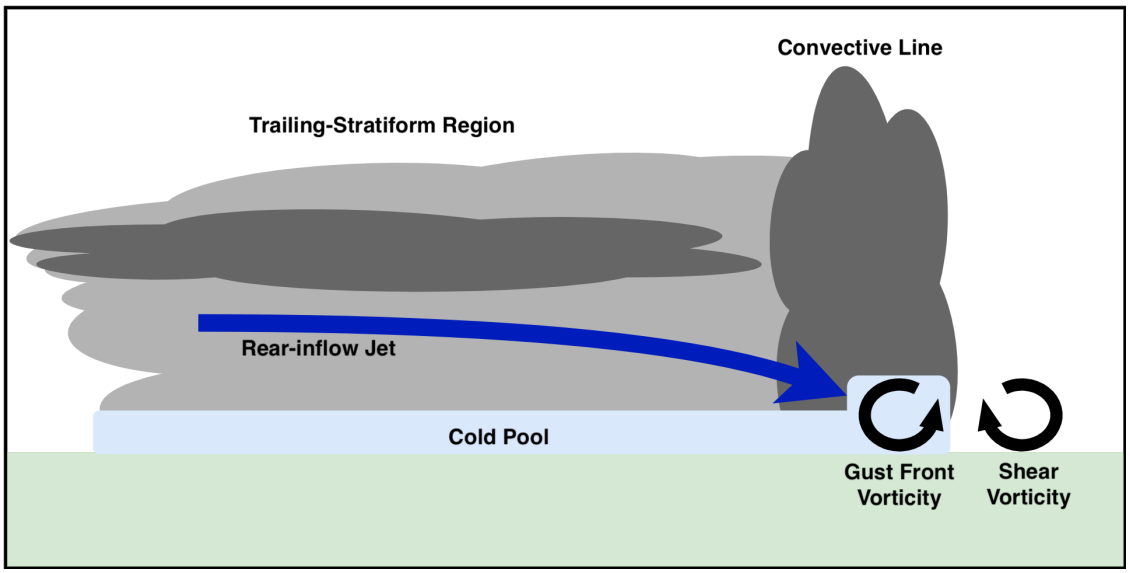

Figure 1.1: Idealized vertical cross-section showing key features for a mature squall-line type MCS. Shading indicates areas of elevated (dark grey) and moderate (light grey) radar reflectivity returns. Dark shading in the trailing-stratiform region is associated with the radar 'bright band' caused by aggregation and melting of frozen hydrometeors. Vorticity for cold pool gust front and environmental wind shear pertains to theory introduced by Rotunno et al. (1988).

extent to which this constraint is possible depends not only on how DA is performed algorithmically but also on the density and information content of observing systems. In this sense, sparse observation types like radiosonde measurements are less informative for the constraint of discrete MCS structures than denser, more frequently assimilated observation types such as those provided by weather radar systems.

1.4 WoFS-derived modeling system used for DA experiments

Although state estimation and prediction for moist-convective regimes do not inherently exclude global modeling systems of sufficient resolution, the DA and observing strategies presented in this dissertation are specifically targeted toward regional systems specialized for convective-scale prediction. To allow for simple interpretation of results in probabilistic terms, we focus on

sequential Bayesian filtering with convection-allowing ensembles while avoiding the extension of variational approaches that seek modal solutions.

Within these bounds, we may consider the National Severe Storm Laboratory (NSSL) Warn-on-Forecast System (WoFS) as an example of a state-of-the-art approach to analysis and prediction for moist-convective events. 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 DA techniques [(Stensrud et al., 2009; Stensrud and Coauthors, 2013)]. The WoFS reflects a number of design decisions that build on the capabilities of its parent modeling system, the NOAA High-Resolution-Rapid-Refresh Ensemble (HRRRE). These include the implementation of 15-minute intervals between DA updates-–significantly more frequent than the hourly assimilation intervals used to update the HRRRE. Frequent updates are intended to constrain the growth of background error associated with rapidly evolving convective features. Both conventional and radar observations are assimilated using an ensemble Kalman filter (EnKF). The direct assimilation of radar reflectivity differs from the use of latent-heat initialization techniques in the HRRRE and other convection-allowing prediction systems Dowell et al. (2022), and allows observations and ensemble covariances to inform microphysical quantities in state estimates without intermediate steps.

Microphysical processes are represented in the WoFS using the NSSL 2-moment variable-density (NVD) scheme. 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). In comparative studies against other double-moment microphysical schemes, the NVD scheme has generally been shown to perform well in representing the distribution of

hydrometeors in forecasts for convective storms (Gallo and coauthors, 2017; Choate et al., 2018), although other studies have noted biases in reflectivity, precipitation, and cloud-extent (Johnson, 2019; Li et al., 2023).

Experiments presented in subsequent chapters are performed using an analysis and forecast system substantially derived from the WoFS, but implemented with modifications suiting the needs of each study. Accordingly, experiments are performed with the Advanced Research WRF (V4.2) (Skamarock et al., 2019), and implement DA with the Data Assimilation Research Toolkit (DART). Domain size is set at 900 $\times$ 900 km^2, with 50 vertical levels and a 3 km horizontal grid spacing, which is sufficient to explicitly permit convection in numerical experiments. This domain configuration is kept constant for all experiments but moved to appropriate geographical locations as needed. Key configuration settings for individual dissertation chapters are noted below in table 1.1, including specifications for physical parameterization schemes in the WRF model. With certain exceptions noted in the following chapters, choices for physical parameterization schemes were motivated by past experiments performed at NSSL for the WoFS (Jones et al., 2018; Potvin et al., 2020). In addition to experiments with the LPF and its hybrid variant, the work presented in this dissertation includes DA experiments performed with an Ensemble Adjustment Kalman Filter (EAKF). The EAKF is a deterministic square root variant of the EnKF (Anderson, 2001).

To support a flexible ensemble size, we generate initial and lateral boundary conditions by perturbing an analysis from the operational NOAA High Resolution Rapid Refresh (HRRR) forecast with Gaussian noise for model horizontal wind, potential temperature, and moisture fields. Hourly updates are applied to lateral boundary conditions for each member during DA cycling, using information from the posterior ensemble and HRRR forecasts. Members are initially gen-

Table 1.1: Physical parameterization settings used for sequential DA experiments in dissertation chapters

	Chapter Two	**Chapter Three**	**Chapter Four**
Ensemble size	64	40	40
Assimilation Frq.	15/5 minutes	15 minutes	15 minutes
DA method	EAKF/hybrid-LPF/LPF	EAKF	LPF
Microphysics	NVD	NVD/Thompson (Thompson et al., 2008)	NVD/modified-NVD
Pl. Boundary Layer	YSU (Ghonima et al., 2017)	YSU	YSU
Land Surface Model	RUC (Smirnova et al., 2016)	RUC	RUC
Radiation	RRTM/Dudhia (Chen and Dudhia, 2001; Iacono et al., 2008)	RRTM/Dudhia	RRTM/Dudhia

erated from 1-h forecasts run from a perturbed HRRR analysis, which are 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 use additive inflation at 15-minute intervals near areas of high observed reflectivity (Dowell and Wicker, 2009). We retain only the last 9 h of posterior ensembles for further analysis in order to reduce the dependence of model solutions on the initial HRRR analysis. We also initialize 90-minute ensemble forecasts every 30 min from these solutions, giving a total of 19 ensemble forecasts per experiment. Our default cycling frequency is set at 15 minutes, since this is characteristic of real-time prediction systems such as the WoFS, although we also perform a limited set of experiments in chapter one using a 5-minute cycling frequency. All experiments use WRF "restart" files to carry forward all prognostic variables and tendencies. This choice further isolates the impact of DA on forecast evolution at short lead times. A graphical overview of our modeling system configuration is shown in Fig. 1.2.

To improve the relevance of our results to operational conditions, we assimilate as many observational types as feasible based on data availability and compatibility with forward operators in DART. From the Next Generation Radar (NEXRAD) network, we assimilate 3-km spaced observations of radar reflectivity and radial velocity where reflectivity returns are at or above 25 dBZ. We additionally assimilate Multi-Radar Multi-Sensor (MRMS) clear-air estimates of reflec-

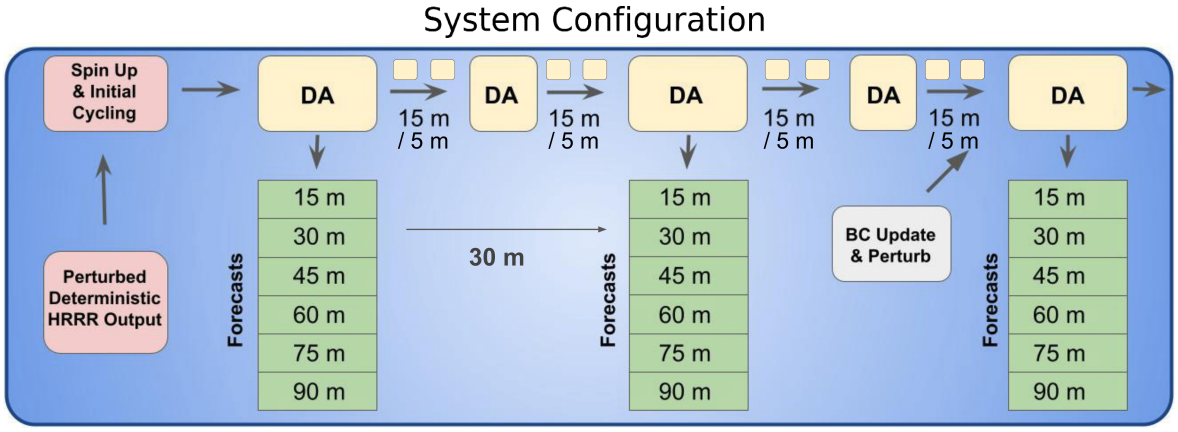

Figure 1.2: Configuration diagram for experimental modeling system used for sequential DA experiments.

tivity in areas not experiencing storm activity. Clear-air observations are useful in removing areas of predicted reflectivity and associated convective activity in the prior that do not coincide with otherwise observed storm features. Two levels of clear-air observations–at 1000 and 6000 m–are created during observation processing in regions with observed composite reflectivity below 25 dBZ, and are gridded at a slightly sparser 6-km spacing. The NOAA Meteorological Assimilation Data Ingest System (MADIS) provides 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, for experiments performed in chapters three and five we also assimilate derived tropospheric winds from the Geostationary Operational Environmental Satellite, made available through MADIS.

1.5 Metrics used for verification of model output

Skill associated with DA and forecast experiments is verified with a few primary metrics, the simplest of which are root-mean-square error (RMSE)–and closely related root-mean-

departure-from-observations (RMSD). The former is given when errors are known relative to the truth, while the latter is given for errors taken as differences compared to observations. As another primary form of verification, we frequently present values for Fractions skill score (FSS) (Roberts and Lean, 2008), a spatial verification metric that may better represent the qualitative performance of forecasts. FSS considers the ratio of modeled and observed occurrence of events for defined 'neighborhoods' of grid points. In doing so, it does not overly penalize small displacement errors in forecasts of discrete weather features, including convective cells and precipitation maxima. Depending on the specific context, verifying events are chosen as the occurrence of composite radar reflectivity or precipitation rate above a given threshold, usually 25 or 35 dBZ in the case of radar reflectivity. Calculations for FSS consider only grid points within a neighborhood window centered on a single grid point location. The size of this window is set by the neighborhood radius, which we specify separately in each chapter. Spatially averaged values are then found by centering a window at each grid point within a verifying region and normalizing the resulting sum. Unless otherwise mentioned, we take this verification region as the entirety of the domain excluding a boundary equal in width to the neighborhood radius. In addition to RMSE and FSS, we make occasional use of three other event-based metrics, probability of detection (POD), false alarm rate (FA), and critical success index (CSI), which are given by the following relations:

$$FA = \frac{FP}{FP + TN} \tag{1.1}$$

$$POD = \frac{TP}{FN + TP} \tag{1.2}$$

$$CSI = \frac{TP}{FN + FP + TP} \tag{1.3}$$

with TP and FP indicating the number of true and false positives within the verification region. FN and TN likewise indicate the number of false negatives and true positives. These metrics provide qualitative information on forecast behavior beyond that available from our primary metrics. POD and FA focus, respectively, on the verification of event occurrence and non-occurrence, while CSI may be seen as a more generalized indicator of skill.

Chapter 2: An evaluation of non-Gaussian data assimilation methods in moist convective regimes

2.1 Introduction

The primary goal of work presented in this chapter is to gauge the ability of the LPF to meet expectations afforded to PF methods in regard to minimizing DA-related bias for non-Gaussian posteriors. Such bias can come about due to non-physical multivariate state updates that may occur considering only the first two moments of errors during DA. 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 DA without introducing a significantly confounding source of error.

Using the ensemble modeling system framework introduced in the first chapter, we perform an investigation of systematic differences between posterior quantities generated by the LPF and corresponding quantities from parametric DA methods that assume Gaussian error structures. As a proof-of-concept, the present chapter discusses results from small (64-member) ensemble sizes affordable for most research and operational weather prediction systems, focusing on microphysical variables that relate non-linearly to observations of radar reflectivity. While certain aspects of this work focus on forecast performance, the primary goal is not to assess the LPF's utility for

operational DA but rather to evaluate the potential for future work contingent on the quality of posterior representations of moist convection.

2.2 Methodology

2.2.1 Note on experiment design

We perform sequential DA for a select number of severe weather outbreaks from the 2019 and 2020 warm seasons, using the LPF, a hybrid variant of the LPF (through mixed filter updates during tempering), and an EnKF. All experiments are performed using the modeling system and methodology established in the first chapter. Using prior and posterior member output from sequential DA, we comparatively evaluate the properties of analyses produced using each considered filtering method. Further insight is gained from analysis of verification results from the cycling period and from generated forecasts.

Subsequent experiments are performed to further isolate factors relevant to the performance of filtering methods. The first is intended to isolate the influence of background error on filter estimates by applying DA to an identical set of priors. The second investigated the role of parameterization error by replacing the NVD scheme used in our standardized modeling system with the Thompson scheme. These experiments are less comprehensive than our primary set of sequential DA experiments and considered only a single case event.

2.2.2 Case studies

In the interest of sampling a set of events sufficiently representative of deep moist convective regimes, we choose four test cases to cover several distinct dynamical setups (Fig. 2.1). 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 of 2019, within a domain centered on Northern Missouri (Fig. 2.1a). 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 multi-cellular storms but including discrete super-cells 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. 2.1b). Ordinary single-cell convection appears beginning after 1600 UTC under strong mean-layer CAPE and weak directional shear, becoming multi-cellular with new cells evident at outflow boundaries, but without widespread organization into quasi-linear structures. The transition from single to multi-cellular convection is quite rapid in this case, presenting a notable challenge for DA.

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. 2.1c). 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 multi-cellular 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,

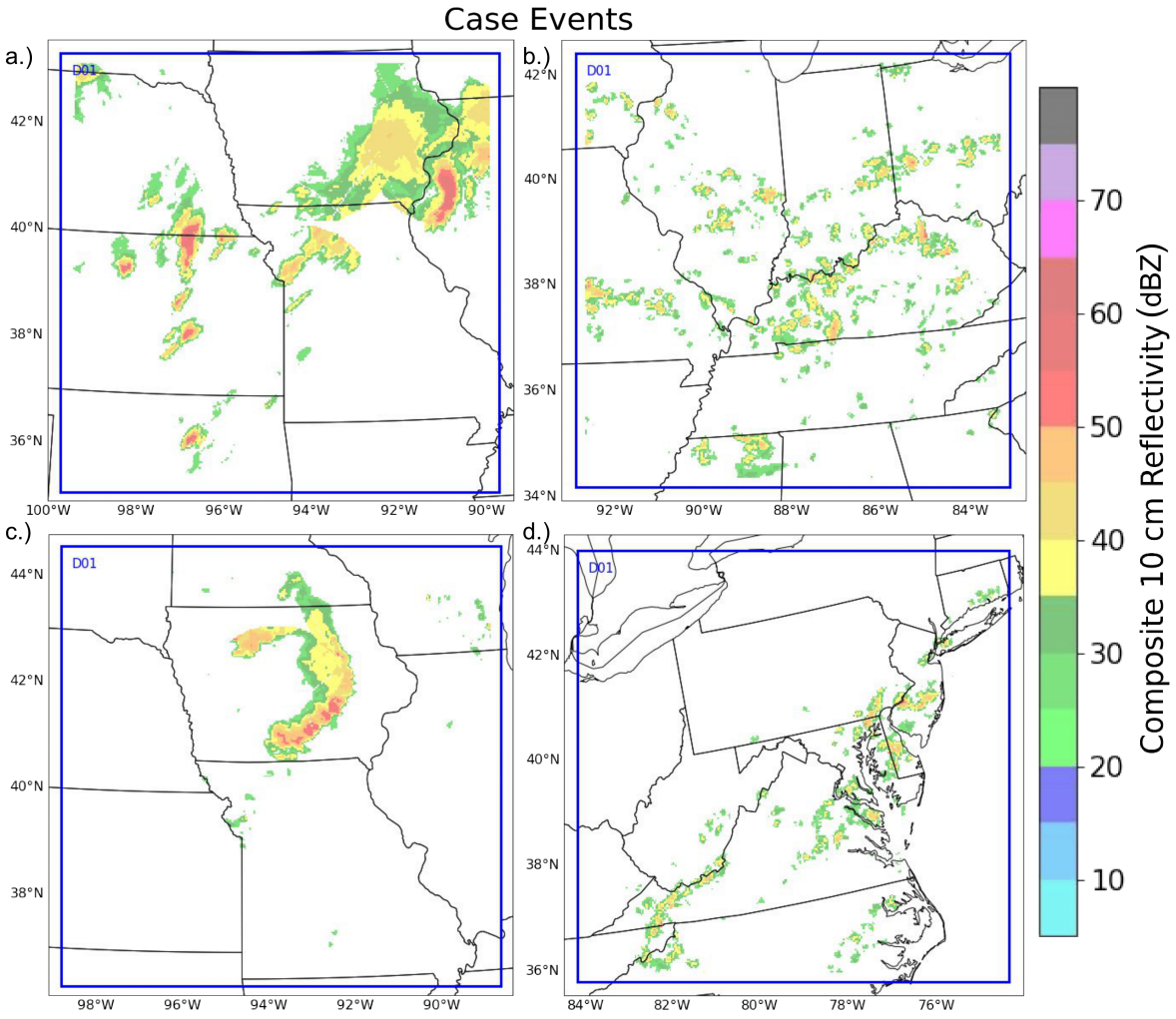

Figure 2.1: Observed 10 cm composite reflectivity plotted at representative hourly times for test cases used in DA experiments. Sample reflectivity output shown for experiments starting (a) 1200 UTC 28 May 2019, (b) 1000 UTC 3 July 2019, (c) 0700 UTC 17 July 2019, and (d) 1000 UTC 12 August 2020.

2020, in a domain centered on the Mid-Atlantic seaboard (Fig. 2.1d). It is a classic pulse-type event that does not show any obvious signs of multi-cellular organization. Convection initiates nearly uniformly at 1600 UTC, continuing under strong surface heating until nightfall.

2.3 Characterizing posterior ensemble members

2.3.1 Analysis of prior and posterior climatology

To investigate systematic differences in EnKF and LPF state estimates we approximated climatological bi-variate marginal probability density functions (PDFs) of radar reflectivity and hydrometeor mixing ratios for model prior and posterior states. This choice followed from 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 grid-point 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-minute (Fig. 2.2) and 5-minute cycling frequencies (Fig. 2.3). Grid-point sampling was confined to the 8th vertical WRF coordinate level of our domain when examining rainwater mixing ratio (q_r) and the 16th level when considering graupel mixing ratio (q_g) in order to ensure that the population from which grid cells were sampled consisted of predominately unfrozen and frozen species respectively. The 8th and 16th vertical levels roughly correspond to pressure levels of 850 mb and 550 mb 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 re-calculated from updated mixing ratios using the relations provided in the NVD scheme. The NVD microphysics scheme 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-minute (Fig. 2.4) and 5-minute (Fig. 2.5) cycling, taken at the 850 mb and 550 mb pressure levels for rain and graupel respectively.

The 15-minute 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. 2.2a). Some of these high reflectivity posterior points are co-located with elevated q_r above 0.5 g/kg, 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 0.05-0.5 g/kg, indicating that the EnKF commonly produces state-space solutions not supported by the model. Looking at the corresponding 15-minute PDF between q_r and n_r (Fig. 2.4a), the EnKF updates show a shift of probability density towards lower number concentration from prior to posterior at all mixing ratios above 0.05 g/kg, 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 (Fig. 2.2b,c). Posterior members generated by both LPF methods feature few grid points with 50+ dBZ reflectivity at q_r below 0.5 g/kg, and also indicate a reduced occurrence of points with q_r above 0.5 g/kg when compared to their respective priors. Again looking at the 15-minute PDF for q_r and n_r (Fig. 2.5b,c), while the PF posteriors reduce areas

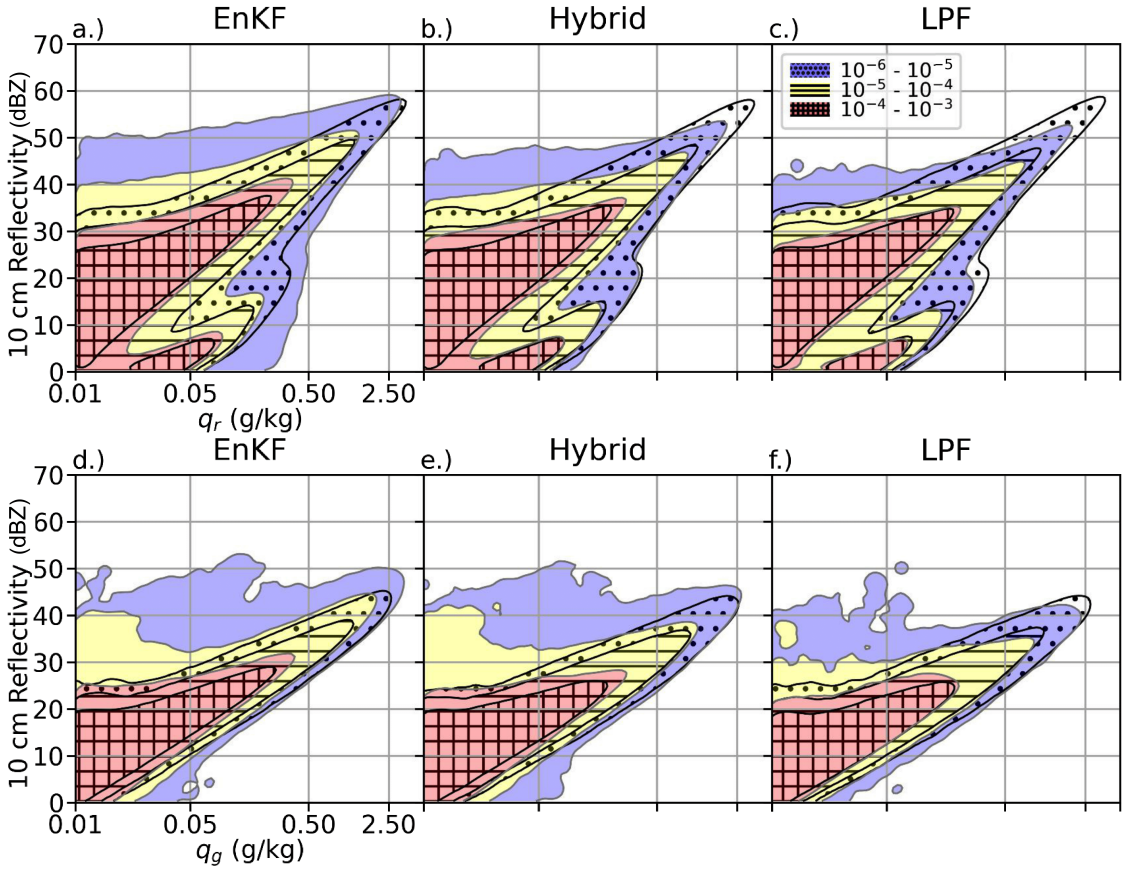

Figure 2.2: Marginal bi-variate probability density function (PDF) plots for 15 minute cycling frequency experiments. Prior (hatched) and posterior (shaded) probability density shown for 10 cm radar reflectivity and rainwater mixing ratio for sequential assimilation with (a) EnKF (b) hybrid particle filter and (c) localized particle filter. PDFs shown 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.

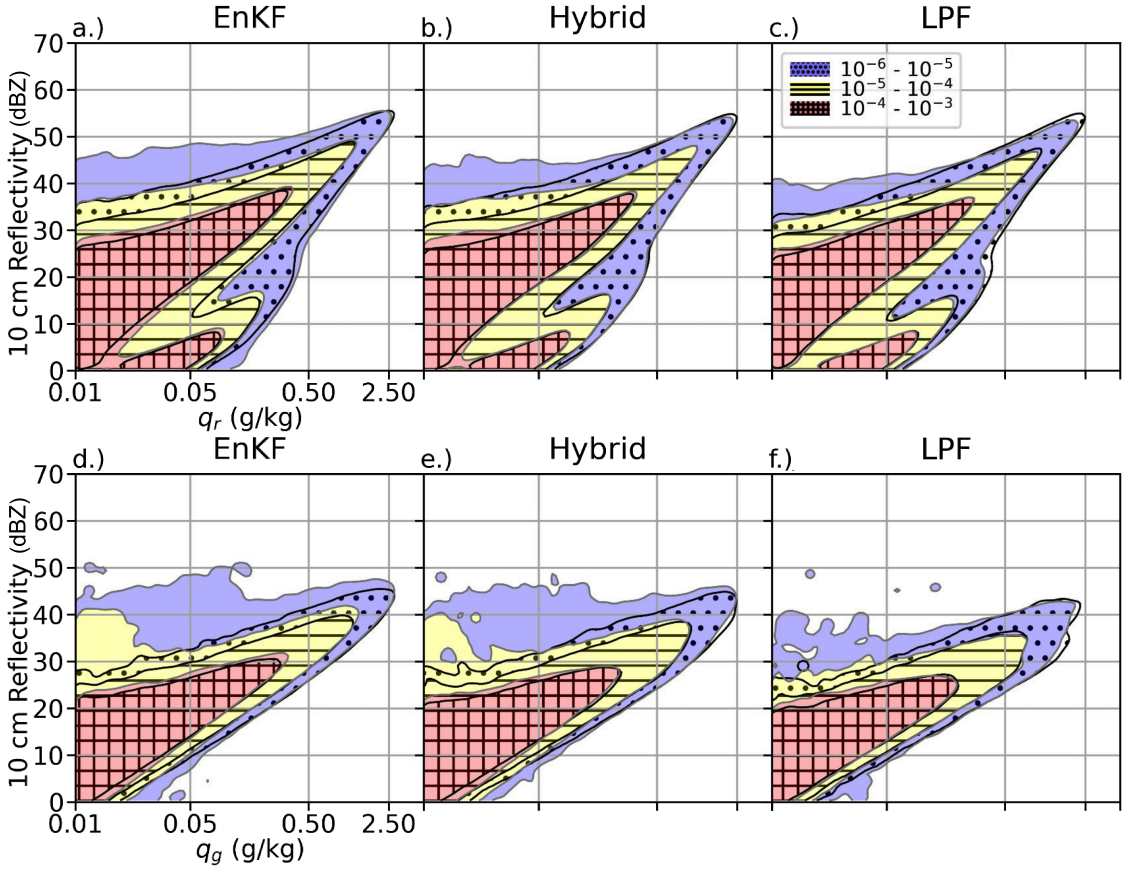

Figure 2.3: Same as Fig. 2.2, but for 5-minute cycling frequency.

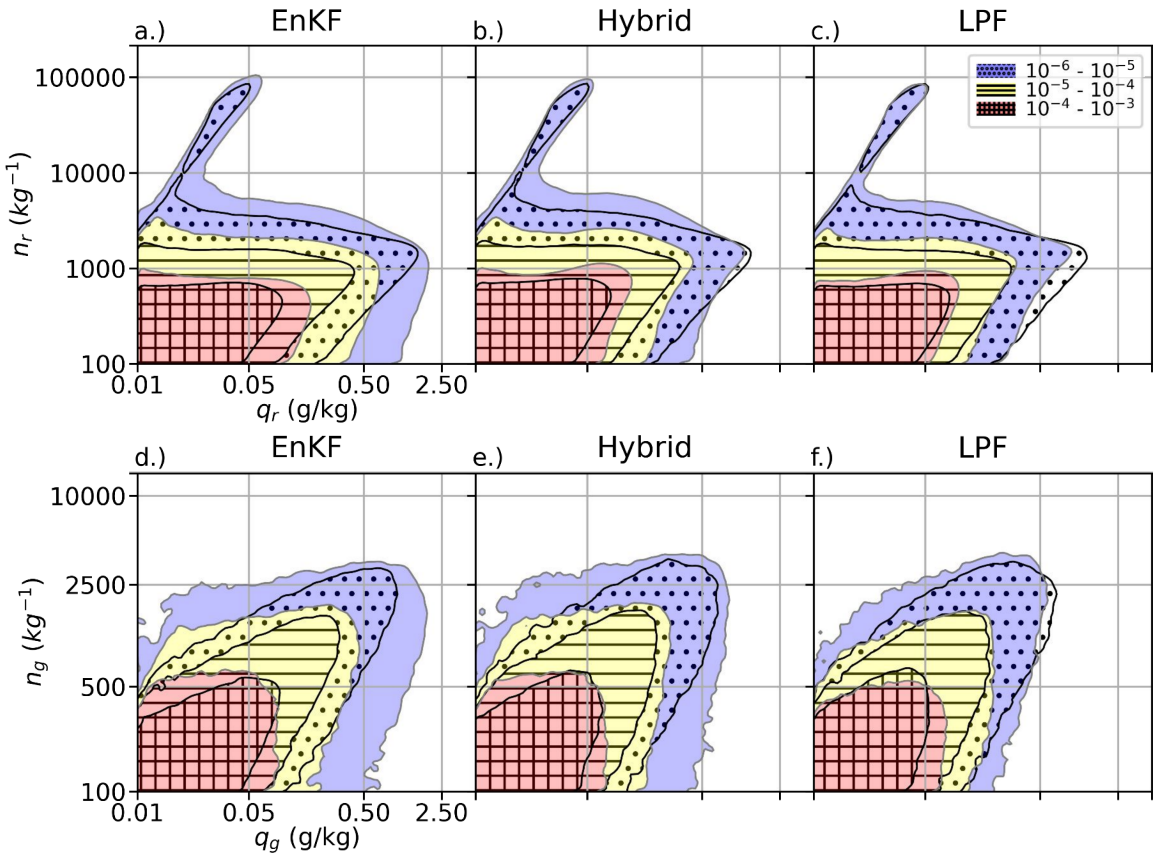

Figure 2.4: Marginal bi-variate probability density function (PDF) plots for 15 minute cycling frequency experiments. Prior (hatched) and posterior (shaded) probability density shown for rainwater mixing ratio and rainwater number concentration with (a) EnKF (b) hybrid particle filter and (c) localized particle filter. PDFs shown 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.

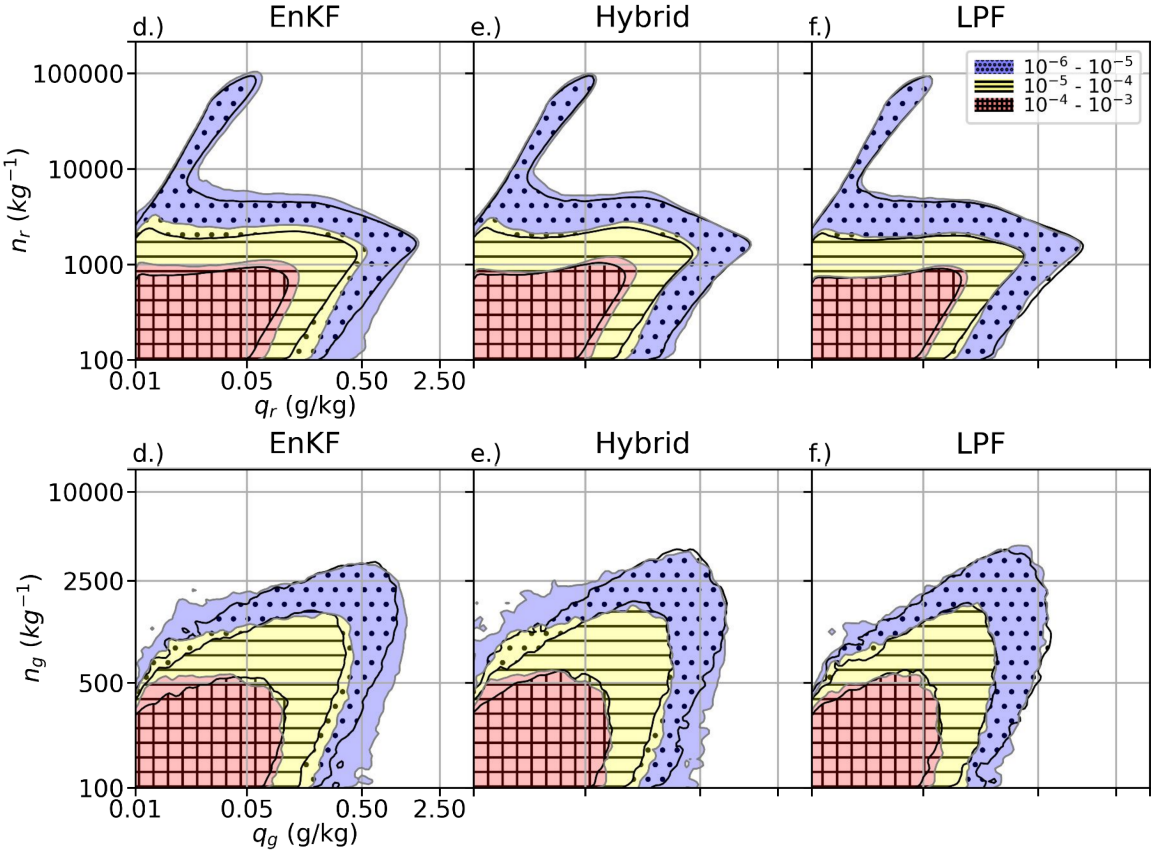

Figure 2.5: Same as Fig. 2.4, but for 5-minute cycling frequency.

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 re-sampling, while preserving the physical relationships between q_r, n_r and reflectivity output in model priors.

When comparing experiments with 15 and 5-minute cycling frequencies, notable differences are apparent in the span of prior support towards higher reflectivity and mixing ratios. 5 minute cycling reduces support in state space regions of reflectivity above 50 dBZ and largely excludes q_r above 2.5 g/kg. EnKF posteriors produced from 5-minute 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. 2.3a). As before, these anomalous solutions are mediated by additional posterior density at low values of n_r as shown in the 5-minute PDF for q_r and n_r (Fig. 2.5a). The LPF configurations for 5-minute 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 minute PF priors (Fig. 2.3b,c). Increased cycling frequency allows the prior to remain more Gaussian by limiting the duration of the model advance and associated non-linear 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-minute EnKF posteriors indicate, at the least, a strong dependence for filter behavior on cycling frequency. Changes to LPF posteriors are more muted, with both cycling frequencies producing

posteriors similar to the 5-minute cycling frequency EnKF. This could suggest that the presence of high reflectivity, high q_r points in the 15-minute 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-minute priors is equally notable and may indicate a model bias towards strong convection or another source of high reflectivity that is ameliorated by all three filters when subject to frequent DA. Indeed, several studies lend support to the idea that NVD microphysics over-predicts regions of high reflectivity (Johnson, 2019; Potvin et al., 2020; Skinner et al., 2018).

The corresponding marginal PDFs taken at the 550 mb pressure level for q_g and reflectivity show similarly filter-specific behavior, but differ significantly in trends between 5- and 15-minute cycling frequencies. The 15-minute 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 (Fig. 2.2d). As with the marginal for q_r and reflectivity, this is again mediated by a density shift towards smaller number concentrations (Fig. 2.4d). 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 (Fig. 2.2e,f). The 5-minute cycling frequency retains this trend, but reduces the magnitude of differences between priors and posteriors for the EnKF and hybrid particle filter (Fig. 2.3e,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-minute cycling (Fig. 2.5). The bivariate marginal PDFs for prior q_g and n_g also remain distinct between filtering methods at 15 minute cycling frequency, with the support of the LPF methods spanning slightly higher number

concentrations than that of the EnKF (Fig. 2.4d,e,f). Interestingly, this distinction is maintained from 15 to 5-minute cycling frequency, in contrast to the trend with rain water 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 towards lower n_g even at short cycling frequencies (Fig. 2.5d). LPF adjustments on the other hand largely maintain the prior marginal distribution of n_g and q_g (Fig. 2.5e,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 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 non-linear forward operator. Though the pure LPF is not directly affected by non-linearity 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 re-calculation of diagnostic reflectivity during the WRF advance step. With this in mind, we examined bi-variate PDFs of mixing ratio and DART output reflectivity. We again produced PDFs for 15-minute and 5-minute cycling frequencies (Fig. 2.6;Fig. 2.7). The 15-minute EnKF posteriors for both species extend probability density at high mixing ratios towards lower values of reflectivity than supported by the prior PDF, and extend probability density at low mixing ratios towards higher values of reflectivity (Fig. 2.6a). The

PDFs for 5-minute cycling show the same, albeit muted, extensions of probability density (Fig. 2.7a). As with re-calculated 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 (Fig. 2.6b,c; Fig. 2.7b,c). These results together show that non-linear 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 re-calculated reflectivity estimates being higher than in the LPF methods.

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 NVD 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 DA 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 co-located with reflectivity observations. We show ensemble member states for a grid cell with a negative analysis increment for reflectivity in Fig. 2.8 and those for a grid cell with a positive analysis increment in Fig. 2.9. Results at this single grid cell location highlight the comparative advantages and disadvantages offered by LPF methods with relatively small ensemble sizes. Downward adjustments

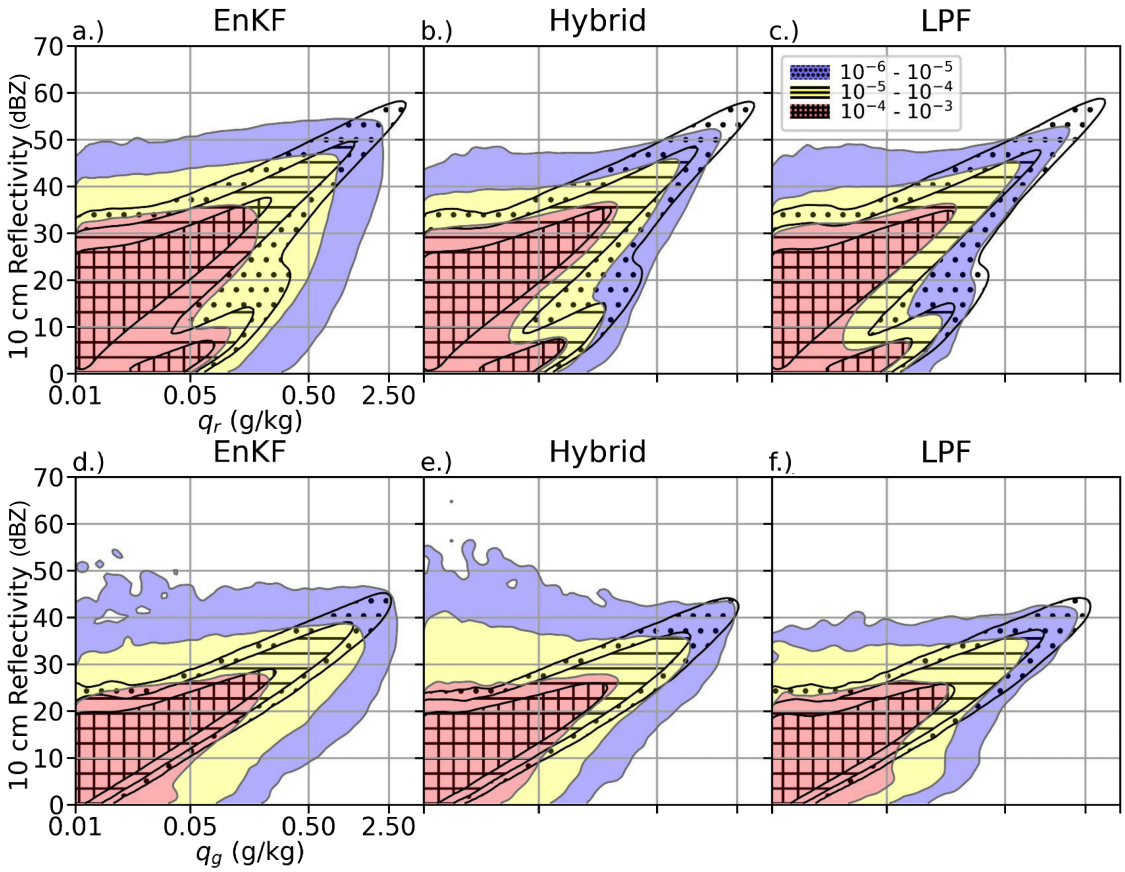

Figure 2.6: Marginal bi-variate probability density function (PDF) plots for 15 minute cycling frequency experiments. Prior(hatched) and posterior (shaded) probability density shown for DART posterior output 10 cm radar reflectivity and rainwater mixing ratio for sequential assimilation with (a) EnKF (b) hybrid particle filter and (c) localized particle filter. PDFs shown 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.

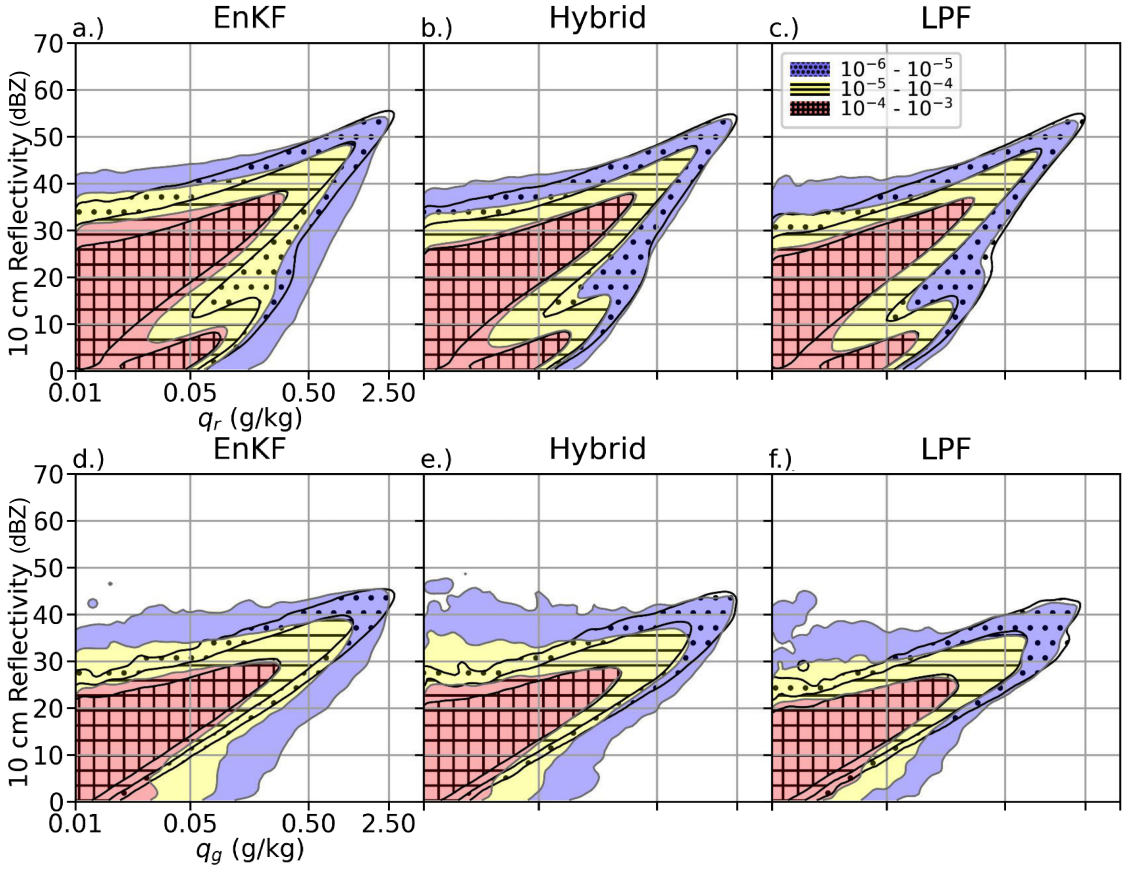

Figure 2.7: Same as Fig. 2.6, but for 5-minute cycling frequency.

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 NVD microphysics scheme (Fig. 2.8b,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. 2.8a), accompanied by a proliferation of unexpectedly high number concentrations for low mixing ratio solutions (Fig. 2.8d). 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. 2.9c). In contrast, the EnKF and hybrid move their posterior ensemble members away from the prior towards appropriate regions of state space (Fig. 2.9a,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 (Fig. 2.9d,e).

2.3.2 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 DA by calculating the root mean square departure from observations (RMSD). For this

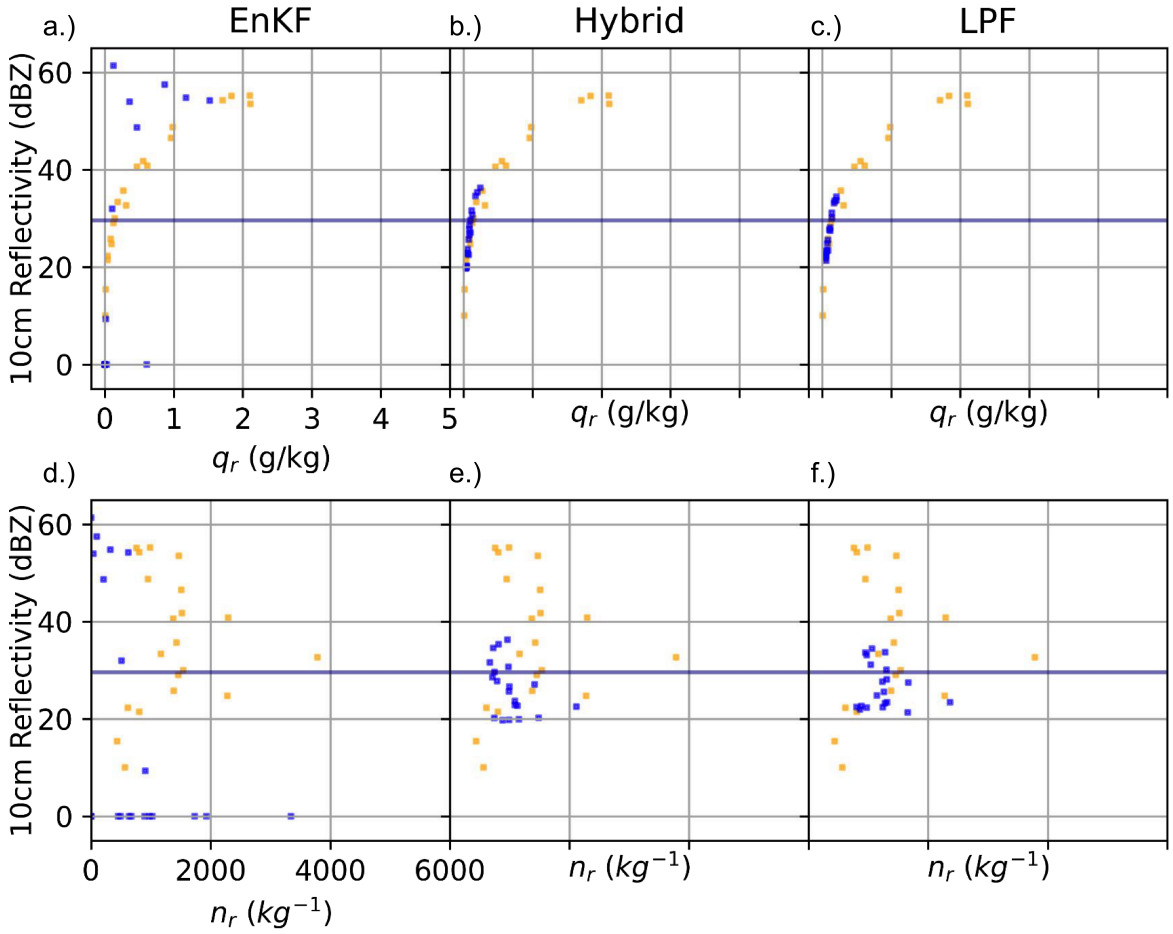

Figure 2.8: 10 cm radar reflectivity and rainwater mixing ratio for prior (orange) and posterior (blue) ensemble members at sampled grid point location for a single adjustment by (a) the EnKF, (b) hybrid LPF, and (c) LPF. 10 cm radar reflectivity and rainwater number concentration for prior (orange) and posterior (blue) ensemble members at sampled grid point location for a single adjustment by (d) the EnKF, (e) hybrid LPF, and (f) LPF. Average reflectivity value for observations within corresponding grid point location is indicated by horizontal line in blue.

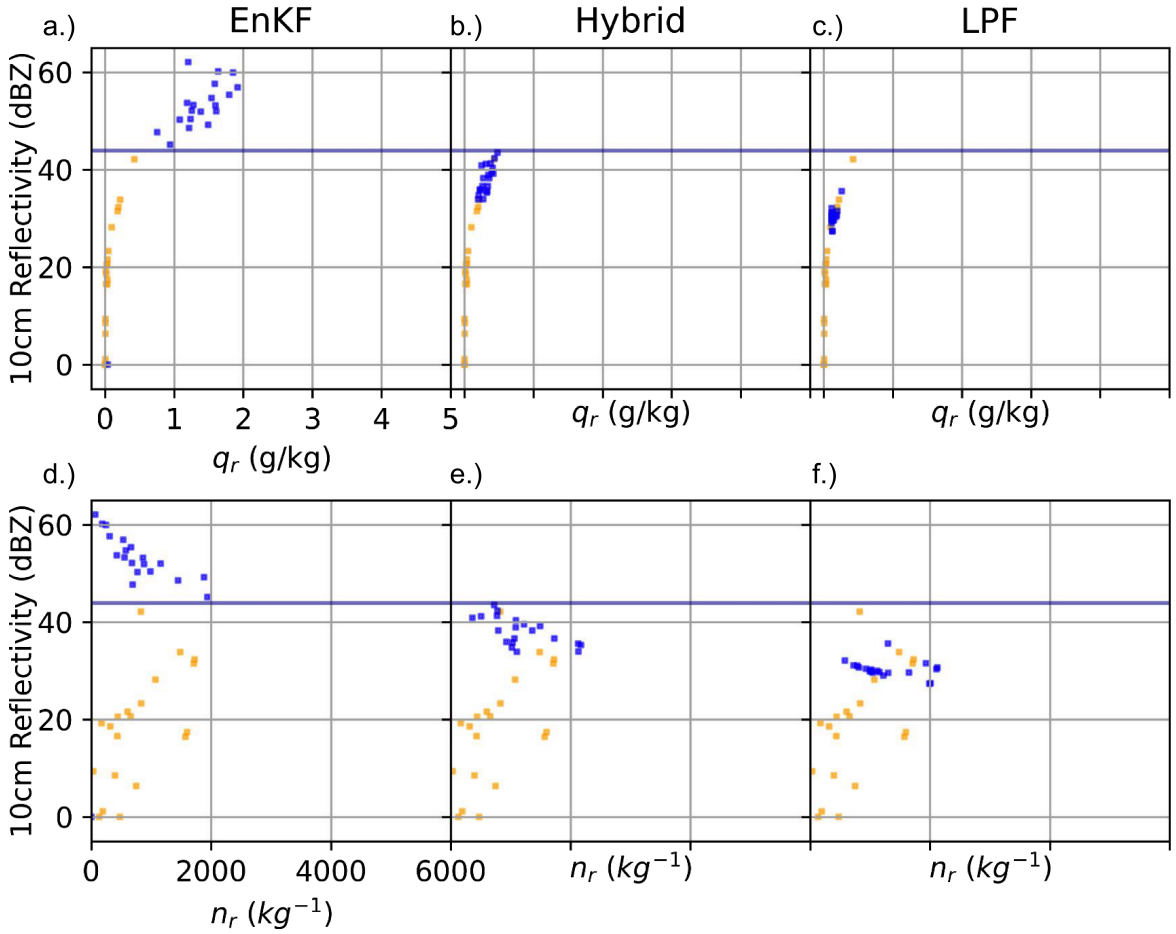

Figure 2.9: Same as Fig. 2.8, but for sampled grid point with positive analysis increment.

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-minute cycling is presented for radar reflectivity (Fig. 2.10) and radial velocity (Fig. 2.11) 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. 2.11c). 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 standalone LPF and an EnKF in terms of 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 minute cycling.

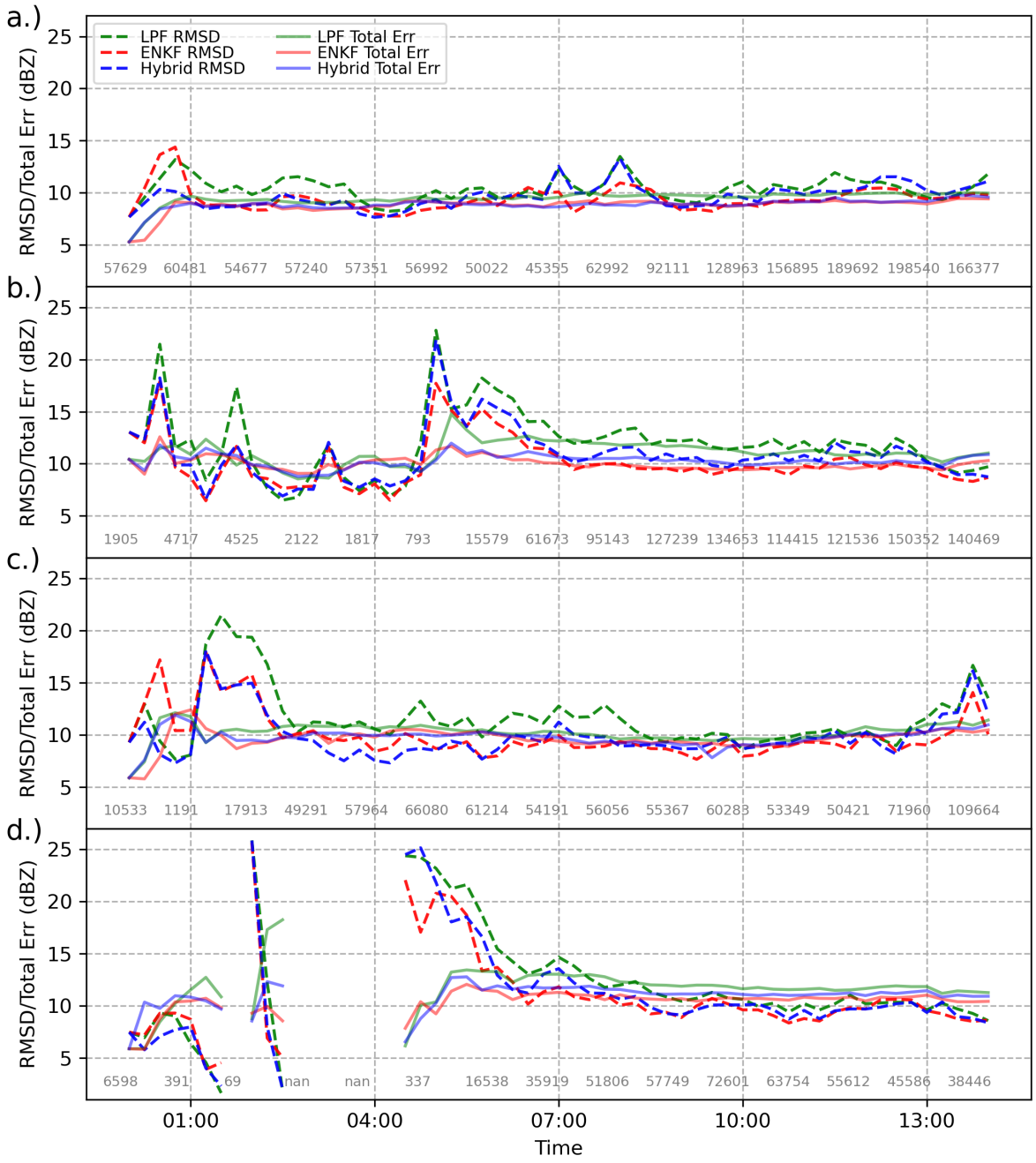

Figure 2.10: Prior root mean departure from observations (dashed) and total error (solid) for 10 cm radar reflectivity, plotted for the duration of 15 minute frequency DA for (a) 28-29 May 2019, (b) 3-4 July 2019, (c) 17-18 July 2019, and (d) 12-13 August 2020 events. Number of observations considered for RMSD calculations annotated in grey at hourly intervals.

2.4 Qualitative comparison of storm evolution

2.4.1 Domain-averaged quantities during continuous cycling DA

Modeling systems that represent mesoscale processes face significant challenges from DA bias mediated by Gaussian assumptions and associated with spurious corrections to small-scale

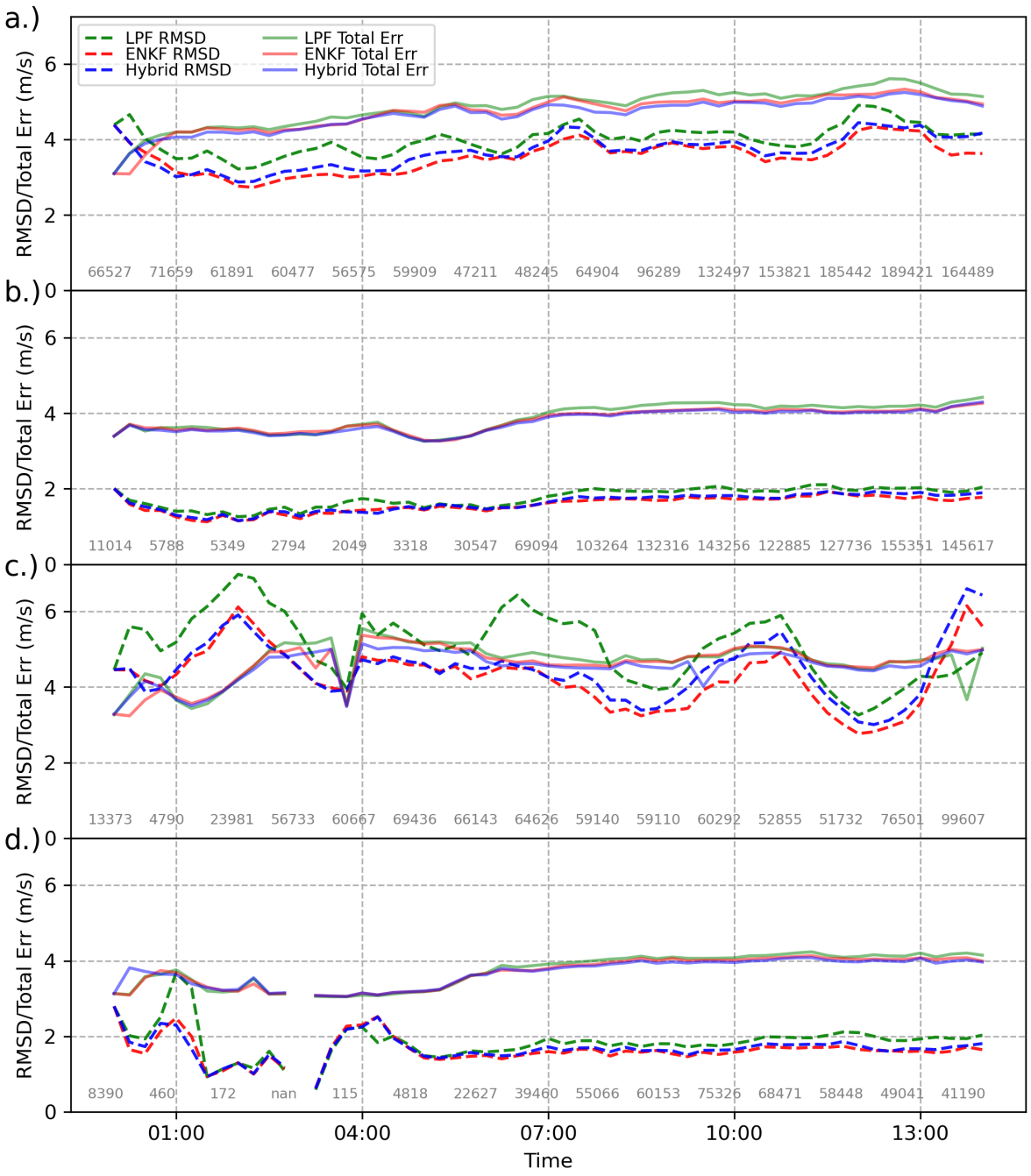

Figure 2.11: Same as Fig. 2.10, but for radar radial velocity.

features (Poterjoy, 2022a). We wish to connect the DA-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 DA with a 15-minute 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 DA is handling convective initiation or modifying the strength of pre-existing 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. 2.12). The reflectivity threshold serves as a generalized indicator of convective activity where cold pools are likely to form. For all four events, sequential DA 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 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.

2.4.2 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-minute assimilation window for the EnKF and hybrid methods. We chose the initialization time as 1930 UTC 3 July for the case

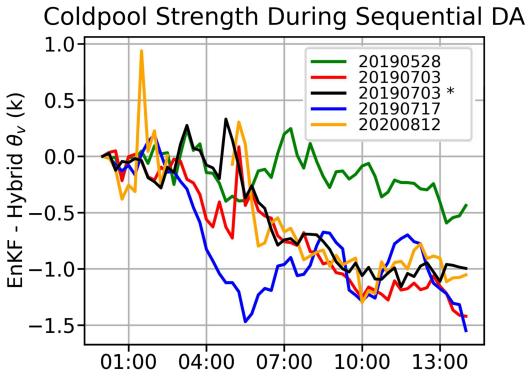

Figure 2.12: Ensemble-averaged surface virtual potential temperature difference between EnKF and hybrid priors over the duration of sequential cycling for each test case event. Calculated for grid cells featuring composite reflectivity above 25 dBZ. X axis denotes hours since start of sequential assimilation. Star (*) in legend indicates experiment using Thompson microphysics.

study covering this date due to it's temporal placement after the start of convective initiation, but before peak development of the MCS. We took prior members from the corresponding timestamp during sequential DA with the pure LPF, then used posterior members to initialize 45-minute ensemble forecasts output at one-minute 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 DA, with the goal of isolating impacts attributable to microphysical variables. We also present results from forecasts initialized without DA 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. 2.13a). We set this threshold at 306.5 K to achieve the closest fit to manually identified cold pool structures while avoiding contamination from other features such as bodies of water. We additionally plotted domain-averaged accumulated grid cell precipitation for the same filter configurations (Fig. 2.13b). Forecasts ini-

tialized from 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 minutes of the forecast period, suggesting that EnKF adjustments create conditions for enhanced convective development. Results for accumulated precipitation show similarly filter-specific behavior, with the hybrid forecasts producing significantly less precipitation over the forecast period than either the 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 DA 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. 2.13c). Filter adjustments produced by 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 filter configurations by the end of the 45-minute forecast period, it is notable that significant differences in column mass remain by 15 minutes 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 DA is

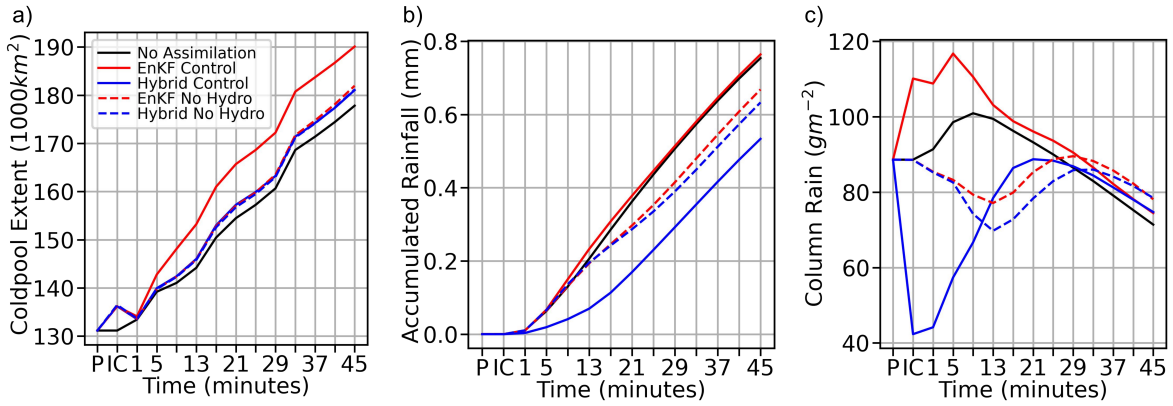

Figure 2.13: Cold pool extent (a), accumulated rainfall (b) and column rainwater mass (c) for prior, initial conditions, and subsequent forecast times for identical prior experiments initialized at 1730 UTC 3 July, averaged over forecast members. Values are shown for (black) no assimilation, (red) assimilation with EnKF and (blue) hybrid configurations, and (dashed lines) assimilation with EnKF and hybrid configurations that do not update microphysical variables.

the likely cause for differing trends in cold pool intensity between EnKF and hybrid during our sequential runs.

We supplement the preceding time series of spatial average quantities with snapshots of cross sections covering a 300 $\times$ 300 km region of concentrated storm development near the center of the domain. The horizontal cross sections in Fig. 2.14 indicate areas of ensemble-averaged surface θ_v below 306.5 K at initialization (Fig. 2.14a) and 15 minutes after assimilation with either the EnKF (Fig. 2.14b) or hybrid (Fig. 2.14c) 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 minutes 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 towards 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.

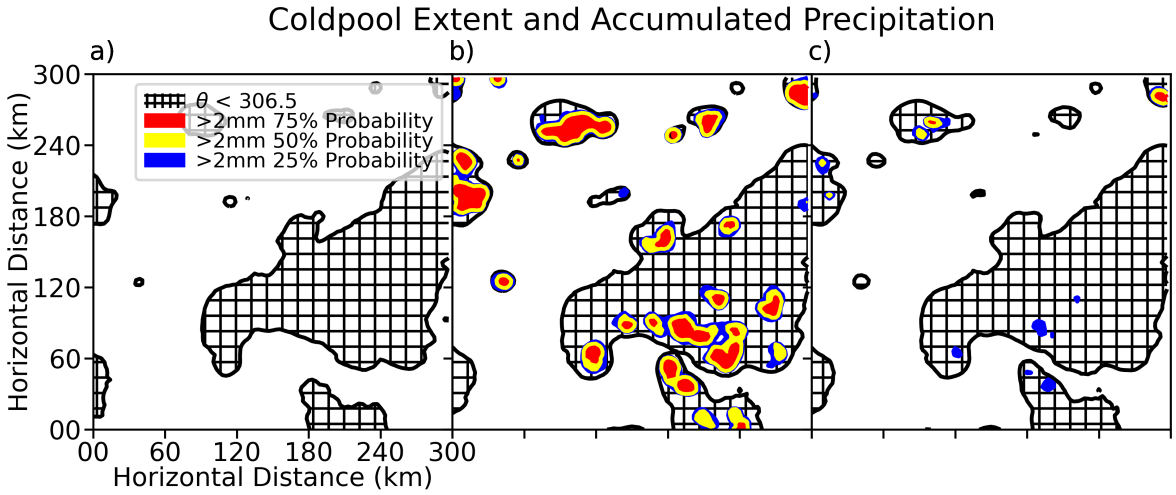

Figure 2.14: Cold pool extent (hatched area) at initialization (a). Cold pool extent and ensemble probabilities of exceeding 2mm accumulated surface precipitation 15 minutes after single assimilation cycle with EnKF (b) and hybrid (c).

2.4.3 Spatial verification of forecasts

We verified the performance of forecasts initialized from our sequential DA experiments using the fractions skill score (FSS) technique given in Roberts and Lean (2008) and introduced in chapter one of this dissertation. 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-minute cycling shown in Fig. 2.15 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. We chose a neighborhood length scale of 30 km for FSS 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. In addition to fractions skill score we also show related categorical scores using false alarm rate (FAR) and probability of detection (POD).

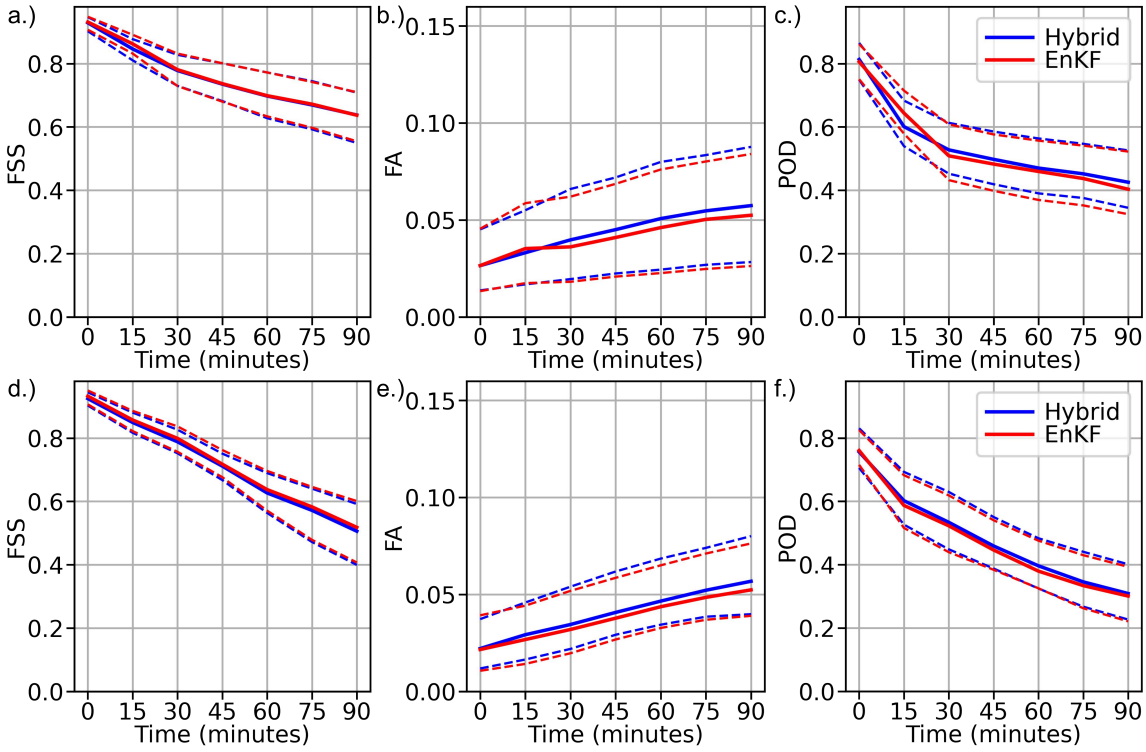

Figure 2.15: Fractions skill score for (a) 15-minute and (d) 5-minute cycling. False alarm rate for (b) 15-minute and (e) 5-minute cycling. Probability of detection for (c) 15-minute and (f) 5-minute cycling. Metrics are aggregated across all events based on time after initialization. Solid lines indicate median values, dashed lines indicate 25th and 75th percentile values.

For 15-minute cycling, the EnKF shows a slight advantage in terms of FSS for the first 30 minutes post initialization, with both filters achieving nearly identical scores for timestamps from 30 to 90 minutes post initialization. Looking at scores for POD and FAR, the EnKF maintains an advantage for POD from initialization to 30 minutes, from which point the hybrid achieves higher scores. Hybrid forecasts produce a slightly lower FAR than EnKF forecasts from initialization to 30-minute timestamps, 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 timestamps is canceled out by the enhanced detection of the EnKF. Later timestamps 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.

2.4.4 Sequential DA 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 DA experiment using the Thompson microphysical scheme in place of the NVD microphysics scheme used in our other case studies. For this experiment, we performed sequential assimilation with a 15 minute 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 NVD microphysics in Fig. 2.16. 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 timestamps and a strong advantage in POD for the hybrid at later timestamps. As for our experiments with NVD 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. 2.12. For the 3 July 2019 case study, sequential assimilation with Thompson microphysics shows a

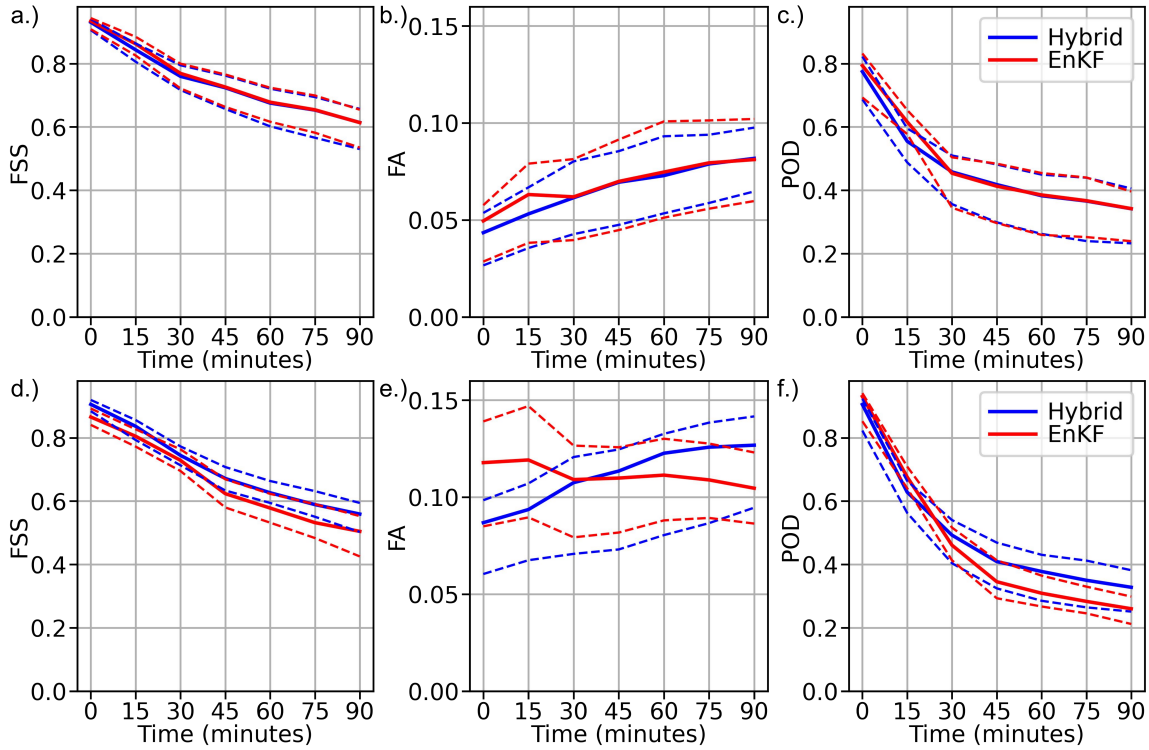

Figure 2.16: Fractions skill score for forecasts from sequential DA with (a) NVD and (d) Thompson microphysics. False alarm rate with (b) NVD and (e) Thompson microphysics. Probability of detection with (c) NVD and (f) Thompson microphysics. Metrics are aggregated for the 3 July 2019 test event based on time after initialization. Solid lines indicate median values, dashed lines indicate 25th and 75th percentile values.

highly similar filter specific difference in cold pool intensity compared to the same experiment performed with NVD microphysics.

2.5 Discussion

2.5.1 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 DA 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-minute 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 DA, 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 DA is the reduction of confounding factors for identifying residual uncertainty. In the context of our sequential DA experiments, we see a tendency for 5-minute cycling to maintain lower mixing ratios and reflectivity in prior members than for 15-minute cycling, regardless of the assimilation method. 5-minute 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 towards 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-minute cycling frequency reduces probability density among high reflectivity, high mixing ratio solutions in a way that closely resembles the prior PDFs for 5-minute cycling. The

fact that these solutions maintain high posterior probability densities during 15-minute cycling with the EnKF suggests that DA bias is strong enough to cover up signatures from model bias under conditions of significantly non-Gaussian prior probability.

2.5.2 Impact of DA bias on forecast evolution

Although the presence of DA 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, DA 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 NVD and Thompson microphysics.

We speculate that for experiments with NVD microphysics, the comparable performance of the hybrid LPF and EnKF methods in terms of root mean square error during sequential DA and fractions skill score for forecasts could be the result of a compensatory effect from strong cold pools towards 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 minute 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 minutes, which suggests that compensatory advantages from DA bias are not a factor in such cases. This may indicate that the structure of DA 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 DA 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 DA, and is not unprecedented given the non-linear relationships between microphysical variables and consequent non-Gaussianity of their marginal prior PDFs. Updates to microphysical variables may therefore produce more significant DA 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.

2.6 Conclusion

2.6.1 Summary of results

The current chapter presents results obtained from a set of comparative real DA 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 non-linear physical relationships. Posterior EnKF ensemble members show a significant departure from the support of climatological PDFs indicated by prior ensemble members in accordance to model physics over the duration of our test events, providing evidence for strong DA 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 DA, 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 DA experiments performed using identical priors for the pair of DA 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 DA 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 DA assumptions to validate choices for subgrid-scale parameterization schemes. As a whole, our results suggest that the hybrid particle filter introduces less DA bias during updates to microphysical and other quantities than Gaussian-assuming methods, which translates to quantifiable differences in forecast verification and in the dynamic evolution of modeled convective storms.

2.6.2 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 non-linear coupling of parameters to model responses (Ruckstuhl and Janjic, 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, non-linear 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 run-off and mineral deposit models (Siriwardene and Perera, 2006; M. 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 DA for a variety of non-idealized case events. Although this approach is technically feasible using any ensemble filtering method, our results have shown that DA 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 non-idealized contexts without significant hindrance by initial condition uncertainty.

Chapter 3: Optimizing numerical weather prediction utility of the Maryland Mesonet with observing system simulation experiments

3.1 Introduction

3.1.1 Maryland Mesonet Project

Carefully structured observing system simulation experiments (OSSEs) may be used to evaluate–and potentially optimize–the NWP utility that can be expected from an observing system given existing NWP systems, while providing insight on how afforded utility may evolve under future NWP systems that feature improvements such as more accurate model parameterization schemes, better constrained initial and boundary conditions, and alternative DA strategies. The current chapter explores an application of such an approach for a new mesonet network to be established in the State of Maryland. The new network aims to improve analyses and numerical predictions for high-impact weather events, with a specific emphasis on hazards associated with severe convective storms.

The Maryland Mesonet (MM) Project will oversee the placement of 75 state-of-the-art, in-situ surface observing stations, with the goal of improving climate records, producing more accurate regional analyses and forecasts, and mitigating the impact of severe weather events across the state. It will join other similarly motivated mesonet systems that exist across the

country, such as the pioneering Oklahoma mesonet established in 1994. As the definition of a mesonet has historically been contingent on local requirements (Fiebrich et al., 2020), for our study we define a mesonet as a network of observing stations that collects surface wind velocity, temperature, humidity, etc. with a spatial and temporal frequency sufficient to capture mesoscale processes and the development of severe convective storms. Such observing systems typically provide information at a granularity beyond the resolution available from pre-existing surface observing stations. The MM will be constructed in two phases. The first 28 stations will be distributed on a roughly county-by-county basis with specific locations contingent on the input of local officials and emergency managers. The second phase of the network design will place the remaining 47 stations based on an optimization strategy that maximizes network utility. The current paper outlines the optimization strategy for the second phase of this project and discusses findings from numerical experiments that were performed for this purpose.

3.1.2 Challenges and limitations of assimilating surface observations for NWP

The effective use of in-situ surface data from mesonets in atmospheric modeling systems is a nuanced problem that requires consideration of error sources in convective scale NWP systems. Subgrid-scale physical parameterization schemes, for components such as land surface, cloud microphysics, and planetary boundary layer in numerical models may not accurately reflect physical processes involved in moist convection. Model process error can often be highly correlated, difficult to estimate, and treated heuristically by inflating background error covariance during DA. As such, model error can produce large biases in estimates of model states and the respective uncertainty in the state estimate. Furthermore, true error correlation length scales across

point locations for state variables are non-stationary. They may, for instance, decrease as large-scale synoptic flow gives way to convective regimes and vice versa. Multiple factors, including the grid spacing of the modeling system, ensemble size, model error, and choices of heuristic DA parameters, such as localization and inflation can determine the accuracy with which cross-covariances are dynamically estimated using ensembles (Pannekoucke et al., 2008; Hodyss and Morzfeld, 2023). The meso-γ and micro-scale correlation structures that characterize organized moist convection could require finer grid spacing and larger ensemble sizes than presently available in operational contexts (Necker et al., 2019). Importantly, sampling error associated with unresolved variability can translate to biased analysis increments (Hodyss and Morzfeld, 2023) that degrade or fail to improve state estimates and resulting forecasts. Even where uncertainty is accurately estimated with ensembles, Gaussian assumptions made by current DA methods such as Ensemble Kalman Filters (EnKFs) and variational methods may introduce errors in state estimates due to the nonlinearity of moist convective dynamics and resulting non-Gaussianity of prior probability distributions (McCurry et al., 2023).

The assimilation of mesonet observations has been the subject of multiple observing system sensitivity studies, although the diversity of observing system configurations, modeling system parameters, case studies, and broader experiment methodology within this body of research makes it difficult to draw firm conclusions on their practical utility for improving analyses and forecasts for severe events in the context of operational, limited-area NWP systems. Results from Strensud et al. (2009) demonstrate the utility of mesonet observations in terms of improving the depiction of cold pools and storm wind structures in state estimates. Ha and Snyder (2014) likewise showed reduced low-level moisture and temperature biases and improved estimates for the location and structure of frontal boundaries. For the latter study, error reductions for wind and

temperature were seen to extend vertically into the mid-troposphere, and forecasts initialized with mesonet data outperformed verification scores from a control run on several metrics–including accumulated precipitation–even 6 hours post initialization. We note that both of the preceding studies used coarse ($>$ 10 km) grid-spacing and infrequent assimilation intervals ($>$ 1 h), with no assimilation of Doppler wind observations in areas of storm activity. Studies using higher resolution modeling systems that are perhaps more relevant for convective-scale DA have shown mixed results for utility gained from assimilating mesonet observations. When performing hourly DA for a convective-allowing modeling system using conventional surface and airborne observations from NOAA Global Systems Division Meteorological Analysis and Data Ingest System (MADIS; Miller et al. 2005), Knopfmeier and Stensrud (2013) found no significant difference between verification scores of state estimates produced using all possible mesonet observations and those from data denial experiments that assimilated only 25% of available mesonet observations. Sobash and Stensrud (2015) assimilated MADIS mesonet observations along with other real surface observation types in a modeling system with 3-km grid spacing and found that the addition of mesonet observations improved model timing of convective initiation by reducing surface biases for temperature and dew point, but only when observations were assimilated at a 5-min frequency. Sobash and Stensrud (2015) also noted limitations in the ability of surface observations to improve state estimates of quantities above the boundary layer, which contributed to limited mesonet utility for an event featuring a strong capping inversion. This issue also appears in Marquis et al. (2014), which evaluated a mobile mesonet system for real cases using a modeling system with 500-m grid spacing that also assimilated Doppler radar observations. The authors noted that surface observations provided by the mobile mesonet did not produce significant change to state estimates above a near-surface cold pool layer 2-3 km in depth. Al-

though both Knopfmeier and Stensrud (2013) and Sobash and Stensrud (2015) examined the impact of mesonet observations separately from METAR and other conventional surface observations, no study so far has comparatively evaluated the impact associated with distinct spatial configurations of mesonet systems. Due to a relative lack of comprehensive OSSE experiments involving modeling system uncertainty, it is also not well understood how factors like model process error and boundary condition error mediate information available to modeling systems from surface observing networks. It is conceivable, for instance, that reductions in uncertainty from improved model physics parameterizations and boundary conditions that remove synoptic-scale errors may isolate error sources associated with mesoscale dynamics to the point where surface observing systems can help inform state estimates for the free atmosphere in the vicinity of severe storms. Alternatively, improved error covariance estimates for surface-based mesoscale features like cold-pools and horizontal convergence zones may allow surface observations to inform model representations of such phenomena at finer scales and therefore produce more accurate forecasts even without reducing analysis error above the near-surface environment.

3.1.3 Observing system simulation experiments

OSSEs are a commonly used method for numerically estimating the impact that may be expected from changes applied to any component of an NWP system, including changes to the type and density of observations ingested during the DA process. All OSSEs involve the creation of a nature run—a single model integration defined as the benchmark 'truth' from which errors may be unambiguously defined. 'Synthetic' observations are then generated from the nature run, and assimilated by a modeling system that produces initial conditions and forecasts. OSSEs provide

a natural framework for investigating the impact of non-existent observing systems such as the Maryland Mesonet and also offer flexibility in tuning modeling system error structures against a reference state. Our study makes use of the latter property of OSSEs to comparatively evaluate the utility of four candidate networks for a 'standard-uncertainty' modeling system configuration representative of contemporary limited-area, convective-permitting NWP systems, and a 'constrained-uncertainty' modeling system configuration imitating future NWP systems with improved model physics and well-constrained boundary conditions. Although alternative methods exist for optimizing observing networks, such as the adaptive observation framework mentioned in Khare and Anderson (2006), our choice of a simple comparative evaluation of pre-generated mesonet configurations avoided assumptions for sampling and model error that may complicate such approaches.

3.2 Methodology

3.2.1 Case event selection

Our experiments simulated seven 18-h real-event cases featuring warm-season moist convection within the state of Maryland. Cases were chosen from within the archival range for forecast data from the High Resolution Rapid Refresh (HRRR) model (James et al., 2022), spanning 2018-2022 at the time of our experiments. Our selection process emphasized events featuring multiple SPC storm reports within state boundaries, with further filtering based on storm severity within the densely populated Baltimore-Washington metropolitan area. Severity in this context was judged by available observations of composite reflectivity and accumulated event precipitation. Limiting events to those that impacted densely populated areas allows human impact factors

to play an implicit role in the evaluation of mesonet configurations without dividing our verification region into smaller sub-domains that may suffer from issues of sampling error related to the spatial distribution of impact from a limited number of test cases. We chose initial conditions for our test case events so that the peak of convective activity occurs after the beginning of forecast initiation following a five-hour spin-up period of sequential DA.

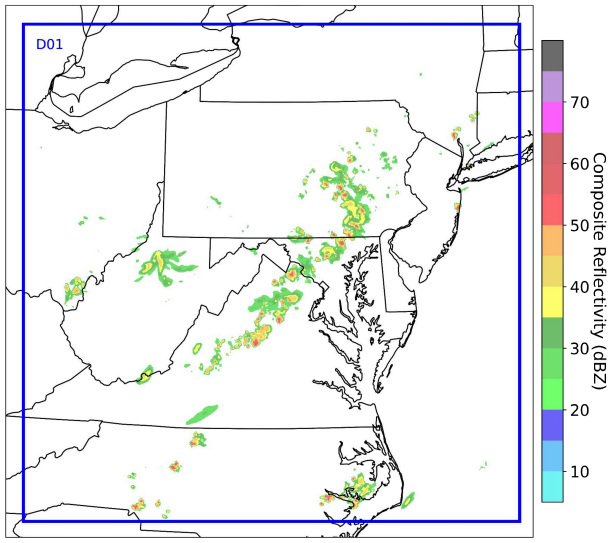

Figure 3.1: Mid-Atlantic domain used for OSSE experiments. Colored shading shows the nature-run composite reflectivity from 5 – 80 dBZ at 1800 UTC for one of the selected events on July 16th 2022.

3.2.2 Modeling system and DA

For the analysis and forecasting component of our OSSE framework–from which initial conditions and forecasts are produced for later comparative evaluation against a nature run–we implemented the WoFS-based system introduced in chapter one. Our domain of interest was centered on Winchester Regional Airport (KOKV) for all experiments, providing full coverage of the state of Maryland and extending several hundred kilometers into surrounding states located to the North, South, and West (Fig. 3.1). In Table 3.1 we specify model physical parameterization

options used in OSSE experiments. With the exception of ensemble size–which we set at 40 members to accommodate a high number of experiments–and the microphysical scheme–which was experimentally varied–selected options remain unchanged from the previous sections of this dissertation. Likewise, individual experiments were performed with a 14-h period of sequential DA, sufficient to generate 19 separate 3-h forecasts at 30-minute intervals.

Table 3.1: Selected WRF namelist options used for nature run, standard-uncertainty modeling system, and constrained-uncertainty modeling system (top) Selected DART namelist options used for all OSSE experiments (bottom)

WRF namelist option	**Nature**	**Standard-uncertainty**	**Constrained-uncertainty**
Grid spacing	1 km	3 km	3 km
Microphysics	NVD (Mansell, 2010)	Thompson (Thompson et al., 2008)	NVD
Cumulus	–		
Longwave radiative transfer	RRTM (Iacono et al., 2008)		
Shortwave radiative transfer	Dudhia (Chen and Dudhia, 2001)		
Land surface	RUC (Smirnova et al., 2016)		
Surface layer physics	Monin-Obukhov (Jimenez et al., 2012)		
Planetary boundary layer	YSU (Ghonima et al., 2017)		

DART namelist option	**Standard-uncertainty**	**Constrained-uncertainty**
Localization Radius (surface obs)	30 km	
Localization Function	Gaspari-Cohn (Gaspari and Cohn, 1999)	

3.2.3 Nature runs and generation of synthetic observations

Nature runs were produced for each event by performing single 18-h WRF integrations from corresponding initial conditions. This integration length was sufficient to encompass the period of sequential DA and resulting forecasts produced by our modeling system. As with initial ensemble members, nature run initial and lateral boundary conditions were interpolated from HRRR analysis output, but were not perturbed with spatially-correlated Gaussian noise. Given the 3-km grid spacing specified for our modeling system, we chose a finer nature run grid

spacing of 1 km to induce a source of representativeness error in synthetic observations. Physical parameterization schemes in nature runs were maintained from those used by our modeling system configurations (Table 3.1), with the exception of microphysics for which we specified the NSSL 2-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), which offers a comparatively high degree of complexity that may be useful for OSSE experiments simulating the effect of missing physical processes.

From nature run output, we generated synthetic observations corresponding to several observation types typically assimilated by operational NWP systems. Radar observations were created for Next Generation Radar (NEXRAD) radar reflectivity and radial velocity, as well as for Multi-Radar Multi-Sensor (MRMS) zero-reflectivity observations in areas not experiencing storm activity. Surface observations were created corresponding to wind and temperature observations available from MADIS. We also produced synthetic observations of temperature and wind in the lower troposphere corresponding to observations from the Aircraft Communications Addressing and Reporting System (ACARS). Synthetic conventional observations were generated from nature runs at locations corresponding to real observations available during the severe weather event of 12 to 13 August 2020 with the timestamps of these observations modified to match those from the corresponding case event. Observations generated in this manner correspond to those that would be expected from the pre-existing observing network. To facilitate experiments with new mesonet configurations, we also generated synthetic observations of 2-meter temperature and 10-meter wind at candidate network locations.

Since NWP systems typically do not assimilate radar observations below a threshold value, we used a modified process to create synthetic radar observations only at locations correspond-

ing to reflectivity at or above a value of 25 dBZ in the nature run. We first created dummy radar volumes for each radar station with sweep elevation angles typical of those from real NEXRAD observations used in part 3. We generated these volumes at time intervals that were likewise representative. These dummy volumes contained non-empty values at every possible location within the volume, subject to the same re-gridding procedure applied to real observations for our work in part 3. After interpolating truth values from the nature run, we then removed observations within these volumes that were below the 25 dBZ minimum reflectivity threshold. Both steps made use of software tools provided by the Observation Processing and Wind Synthesis (OPAWS) program (Wicker, nd; Majcen et al., 2008)

In the context of NWP, observation error can be thought of as the sum of a measurement error term and a representation term consisting of contributions from (a) unresolved scales, (b) forward operator error, and (c) pre-processing error (Janjić et al., 2018a). Although some degree of error from unresolved scales was implicitly included from the interpolation of synthetic observations between the 1 km nature run grid and the 3 km modeling system grid, we accounted for measurement error and the remainder of the representation error term by adding Gaussian noise to the values of synthetic observations in a manner similar to Errico et al. (2013) but using posterior root mean departure from observations as the tuning parameter.

3.2.4 MM Configurations

Our OSSE framework considered four candidate configurations of 75 mesonet stations to be constructed as part of the MM project. These configurations differ in aspects including geographic uniformity of station density, locus of maximum density for non-uniform configurations,

as well as consideration of surface observations provided by the pre-existing observation network. We note that 28 of these stations have fixed locations as part of the first phase of construction, leaving 47 stations with spatial locations that may vary between configurations. Candidate networks were generated using a method based on simulated annealing (Bertsimas and Tsitsiklis, 1993) that maximizes equidistance between stations subject to constraints and forcing terms that impact the final distribution of station placement. A summary of station configurations and the specific constraints used by the simulated annealing method is provided in Table 3.2. All networks were required to place stations within the state of Maryland and on land surfaces only. The west-centered (WEST) and population-centered (POP) MM configurations included forcing terms that increased station density in favored areas: near western boundaries–coastal or state administrative–upstream of climatological storm propagation for the former case, and in densely populated regions of the Washington and Baltimore metropolitan areas for the latter. Forcing terms are not provided for the EQD or EQD+ configurations, and these configurations therefore produce a more uniform spatial density of candidate stations within state boundaries. Only MM stations were considered by the simulated annealing method, with the sole exception of the modified-equidistant (EQD+) configuration which also considered pre-existing (non-MM) surface observing stations when maximizing equidistance.

Table 3.2: Configurations evaluated by OSSE experiments and parameters used in their generation by a synthetic annealing algorithm

Configuration Parameters	**Equidistant (EQD)**	**Modified-equidistant (EQD+)**	**West-centered (WEST)**	**Population-centered (POP)**
Geographic constraints	Within the state of Maryland			
Surface constraints	Land surface only			
Geographic forcing	–	–	Western boundaries	BWI metro area
Equidistance maximized for	MM	MM & non-MM surface	MM	MM

Specific geographic locations of individual surface observing stations within each network

are shown in Fig. 3.2. For the EQD+ configuration, pre-existing stations were taken as observing locations that report surface level measurements of temperature, wind, and humidity to NOAA's Meteorological Assimilation Data Ingest System at a frequency of at least one measurement every 30 minutes. In addition to the four candidate networks, our experiments also consider a default or 'no-build' observation network without new mesonet stations. The no-build configuration provides the baseline from which all MM configurations will be compared.

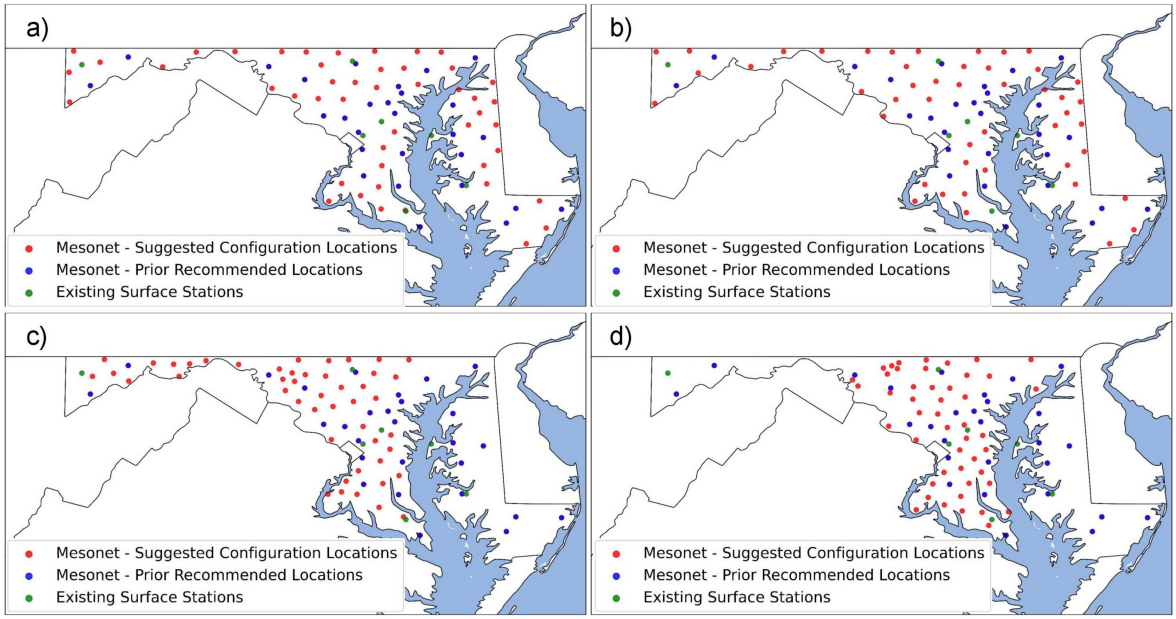

Figure 3.2: OSSE candidate MM configurations: (a) network placed to preserve equidistance between MM stations, (b) network placed to preserve equidistance between MM and pre-existing (non-MM) stations, (c) network with greater density in the Appalachian plateau region of western Maryland, (d) network with greater density in heavily populated areas of the Washington, DC and Baltimore metropolitan areas. Configuration-specific candidate locations indicated by red marks. pre-determined locations for new MM stations indicated by blue marks. Non-MM surface stations with frequent reporting intervals indicated by purple marks.

3.2.5 Simulations of mesonet impact with approximations for current and future modeling system uncertainty

To accurately represent the impact of assimilating MM observations in both contemporary and future NWP systems, we performed two separate sets of OSSE experiments using configurations that appropriately simulated modeling system uncertainty. For standard-uncertainty experiments simulating the capabilities of contemporary NWP systems, we ran our forecast model using the double-moment Thompson microphysics scheme (Thompson et al., 2008). This choice induces a model process error in the sense that our parameterization for clouds differs from the NVD scheme used by the nature run, acting in addition to model error induced by scale discrepancy between our 3-km modeling system and 1-km nature run. In an effort to realistically approximate the uncertainty that is introduced to limited-area modeling systems by parent global models via initial and boundary condition error, we added an additional error source in the form of low wave-number Gaussian noise applied to initial and lateral boundary conditions used by our modeling system. This additive noise has a wavelength roughly double that of our domain size and simulates the impact that may be expected from synoptic scale error structures found in global model output. Using forecast output for a case occurring on April 13, 2020, we tuned the amount of added noise in an iterative process and chose the amplitude that produced a 90-min fractions skill score closest to that produced by our modeling system configuration for the same case event when assimilating real observations. For tuning process results see Fig. A.1.

For 'constrained-uncertainty' experiments simulating the potential capabilities of future NWP systems, we used a modeling system configuration with the same NVD microphysical scheme as in our nature runs, removing the approximated source of physical parameterization

error. To simulate improved constraints on the synoptic scale environment associated with future global NWP systems, we did not add low wave-number Gaussian noise to model initial and boundary conditions. Given the short wavelengths of Gaussian noise added to maintain ensemble spread, boundary conditions used for constrained-uncertainty experiments may be thought of as perfect on the synoptic scale. The relatively high fidelity of initial conditions produced under these conditions further allowed us to halve the amplitude of additive noise used to maintain spread during sequential DA without significantly degrading model spread during convective initiation. Because additive inflation methods can themselves act as sources of sampling error that induce bias in state estimates (Sobash and Wicker, 2015), this modification further imitated hypothetical improvements in sampling characteristics of future NWP systems.

3.2.6 Verification of MM configuration experiments

We evaluated the performance of candidate MM configurations by generating verification scores for forecasts and analyses that may be compared against those expected from the pre-existing observing system. We first considered RMSE for ensemble-mean wind and potential temperature fields at various levels of interest. RMSE was evaluated using grid point locations within a verification region corresponding to the state of Maryland, which were re-gridded to 6-km resolution and then compared to similarly re-gridded nature-run values at corresponding locations. For our second metric, we used FSS produced for two verifying events: the occurrence of composite radar reflectivity over 25 dBZ, to correspond with areas of convection and severe storm activity, as well as the occurrence of precipitation rates more than 0.625 mm/hr. We chose a neighborhood length scale of 15 km, which provides a balanced metric informed by both the

resolution of discrete storm structures as well as more general performance characteristics such as precipitation bias. Spatially-averaging for both RMSE and FSS considered a verification region consisting of areas within the state of Maryland.

3.3 Impact of MM observations on the fidelity of analysis fields and derived quantities

3.3.1 Results for standard-uncertainty and constrained-uncertainty OSSE configurations

Accurate model analyses and derived, diagnostic quantities are crucial for forecaster decision-making alongside numerical forecast output. To gauge the utility of assimilating MM observations on the accuracy of state estimates, we examined prior and posterior quantities over the duration of sequential DA. In Fig. 3.3 we present several related plots that characterize the spatial distribution of mesonet impact for potential temperature during experiments with our standard-uncertainty modeling system. Corresponding plots for zonal-wind speed are shown in Fig. 3.4. All plots presented consider averages for all forecast initialization times across cases and are stratified column-wise by applied MM configuration. The grey lines in Fig. 3.3d and Fig. 3.4d depict the vertical structure of MM impact on initial condition error, using the first 20 members of the analysis ensemble. The impact is calculated as a percentage improvement in ensemble-mean RMSE over the no-build case and averaged longitudinally and meridionally within the boundaries of the state of Maryland. Results across MM configurations indicate maximal improvements of 6-7% for potential temperature and 3-4% for zonal-wind speed that decline with elevation. Im-

provement is mostly constrained to a vertical range extending from the surface to roughly the height of the planetary boundary layer, with impact becoming negligible above roughly 1000 m. Besides a slight advantage for near-surface zonal-wind speed shown for the modified equidistant MM configuration, there is little difference in impacts produced by candidate networks.

We extend this vertical stratification by examining the geographic distribution of MM configuration impact on analysis-mean quantities, considering only a layer roughly below 330 m-corresponding to the five lowest eta levels in our WRF configuration-that appears most strongly affected by mesonet observations (Figs. 3.3c,3.4c). We subsequently refer to impacts averaged over this layer as 'near-surface-averaged'. Resulting fields show geographically heterogeneous impact distributions for both temperature and zonal wind, with all MM configurations producing benefits that are roughly concentrated within the state boundaries of Maryland and yielding roughly equivalent areas of improvement and degradation in neighboring states. With few exceptions, zones of contiguous improvement are located near MM observation sites, well within the 30 km localization radius of MM observations. The location of maximum near-surface-averaged improvement to analysis-means over the default observation network varies with MM configuration, occurring in southern Maryland and the north-western portion of the Delmarva peninsula for the two equidistant type configurations. For the western and population-centered networks, these maxima occur in southern Maryland and in the north-central portion of the state. Near-surface-averaged improvements near these maxima are on the order of 15–20% of the posterior RMSE values associated with the default observation network for potential temperature and 10–15% for zonal wind, with potential temperature showing more consistent and geographically contiguous improvements throughout the state.

Plots of MM configuration-mediated changes to analysis-increment magnitude (Δ Inc)

(Figs. 3.3a,3.4a) reveal a notable correlation with local site density for most regions outside of central Maryland, with higher configuration site densities producing locally enhanced Δ Inc. Since site density also tends to correlate positively with analysis impact in these regions, this suggests that Kalman gains for near-surface variables are reasonably unbiased in such areas leading to positive marginal utility of additional observations. A strong example of this effect can be seen in the northwestern Delmarva peninsula, where the west and population-centered MM configurations place comparatively few stations compared to the equidistant MM configurations, and are in turn associated with significantly smaller changes to analysis increments and positive analysis error impact for both potential temperature and zonal-wind speed. Conversely, regions of central Maryland between the Washington, DC and Baltimore MD metropolitan areas, and portions of northeastern Maryland show no distinguishable correlations between site density and Δ Inc, and the correlation between site density and positive analysis error impact is greatly muted. We note that these regions correspond rather well with areas where high average densities (25+ observations/100 km) of pre-existing, non-MM surface observations are assimilated in our experiments (Fig. A.4).

We repeat the preceding analysis for our constrained-uncertainty modeling system configuration and present corresponding figures for potential temperature (Fig. 3.5) and zonal-wind speed (Fig. 3.6). Although the vertical distribution of impact on initial-condition error appears quite similar to results for the standard-uncertainty OSSE when considering potential temperature (Fig. 3.5d), the same results for zonal-wind speed show stark differences below 1000 m, with maximal improvements over the no-build configuration reaching only 2-3% (Fig. 3.6d) near the surface. While the geographical distribution of impact retains improvement maxima in regions of southern Maryland and the Eastern shore indicated for the standard-uncertainty OSSE, there

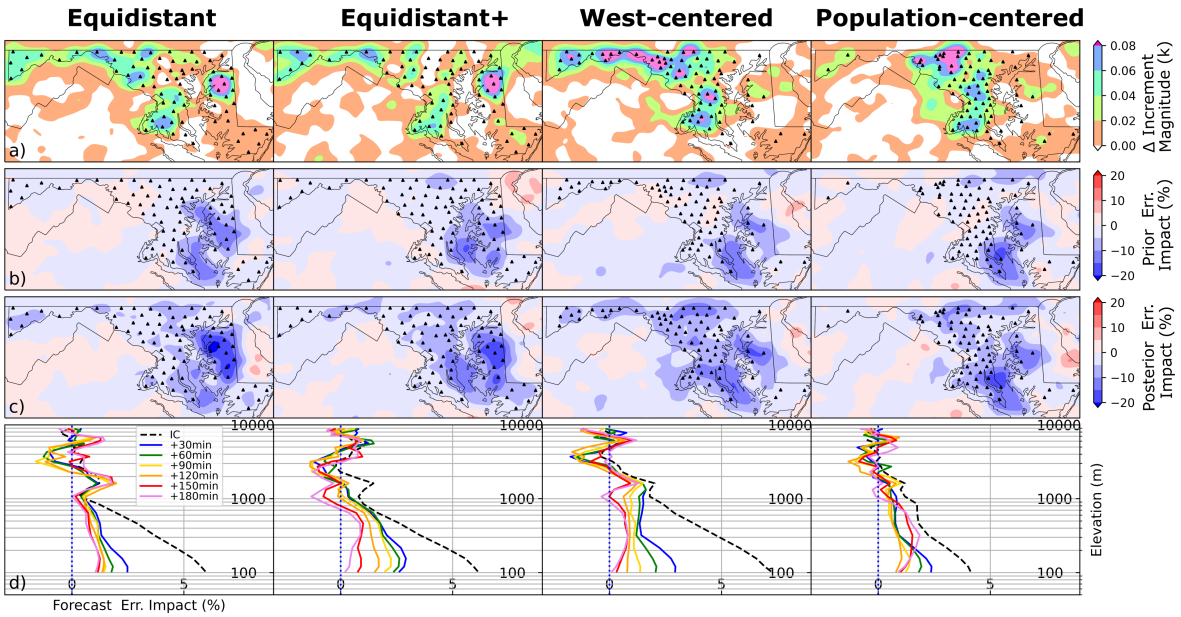

Figure 3.3: (a) Near-surface-averaged MM configuration change in analysis-increment magnitude for potential temperature, (b) near-surface-averaged MM configuration impact on prior-mean RMSE difference for potential temperature, (c) near-surface-averaged MM configuration impact on analysis-mean RMSE difference for potential temperature, and (d) Vertical profile of MM configuration impact on initial condition and forecast-mean RMSE for potential temperature. Results for standard-uncertainty OSSE

are distinct reductions in analysis-mean improvement for both variables in the central Maryland region, and for potential temperature in the southern Chesapeake Bay. For zonal-wind speed, there is a large increase in the extent of regions suffering analysis-mean degradation compared to the standard-uncertainty experiments (Figs. 3.5c,3.6c). Interestingly–despite shifts in the geographical distribution of analysis-mean impact–the corresponding distribution of MM mediated Δ Inc appears quite close to that from the standard-uncertainty OSSE (Figs. 3.5a,3.6a), which suggests that subdued analysis-mean improvement in the constrained-uncertainty experiments is mediated by other factors.

To investigate temporal and case-specific factors affecting near-surface-averaged MM impact on analysis error, we examined smaller sets of analysis means aggregated in 2-h intervals, commencing 3 h after ensemble initialization and for potential temperature only (Fig. A.2). We

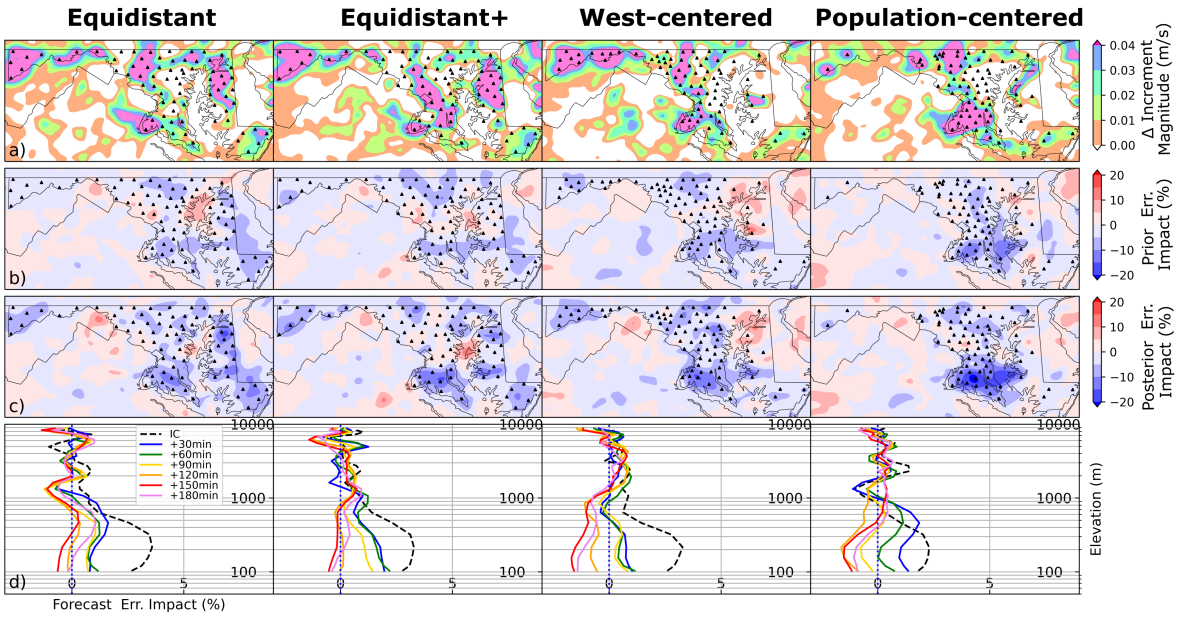

Figure 3.4: Same as Fig. 3.3 but for zonal-wind speed.

indicate the presence of composite reflectivity above 25 dBZ within these temporal intervals with cross-hatching. Results strongly suggest that the impact of MM observations achieves the greatest amplitude, positive or negative, only after a given local region has been impacted by moist convection, regardless of event set-up. It also appears that the majority of statewide reduction to analysis-mean RMSE can be attributed to particular cases and even intervals within specific events. Events centered on 12 August 2020 and 16 July 2022 benefit the most from MM assimilation, with the latter showing an especially positive impact after the passage of a quasi-linear-convective system (QLCS) across the Chesapeake Bay as shown in panels D,E, and F. Conversely, cases centered on 13 April 2020 and 3 September 2020 produce the least benefit from MM assimilation, even after convection has spread throughout the state. Both of these cases feature strong synoptic-scale convective forcing driven by frontal passages, in contrast to other cases where convection is either air-mass driven or else forced as part of a mesoscale convective system (MCS). For experiments using a constrained-uncertainty modeling system configuration

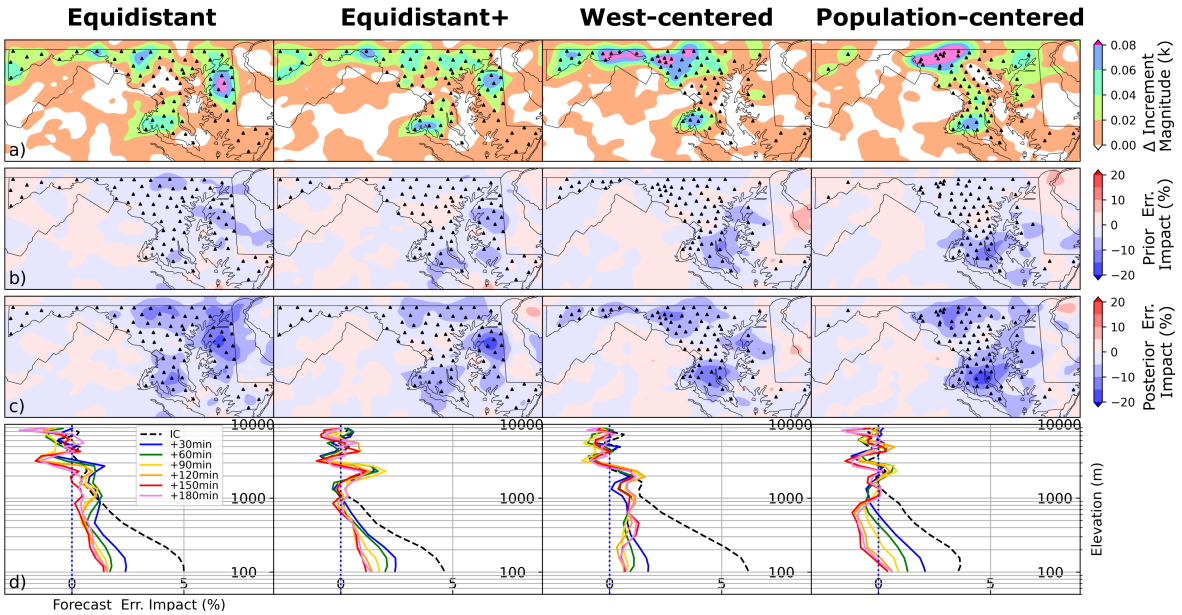

Figure 3.5: (a) Near-surface-averaged MM configuration change in analysis-increment magnitude for potential temperature, (b) near-surface-averaged MM configuration impact on prior-mean RMSE difference for potential temperature, (c) near-surface-averaged MM configuration impact on analysis-mean RMSE difference for potential temperature, and (d) Vertical profile of MM configuration impact on initial condition and forecast-mean RMSE for potential temperature. Results for constrained-uncertainty OSSE.

(Fig. A.3), the aforementioned synoptically-driven cases show even less improvement relative to other events. Although the air-mass-driven event centered on 12 August 2020 retains the large improvements demonstrated in standard-uncertainty experiments, analyses for the 16 July 2022 case no longer show especially large improvements associated with the QLCS passage occurring towards the end of this case.

Although ensemble-mean RMSE is a powerful metric for quantifying the accuracy of a given field, differences in RMSE between sets of initial conditions contain contributions from errors that are more or less important for forecast evolutions, which may be concentrated in regions containing information directly relevant to the development of localized convective storms. To directly evaluate the representation of dynamically relevant structures within our OSSE framework we calculated CSI scores–averaged across initial members of all collected forecast ensembles–

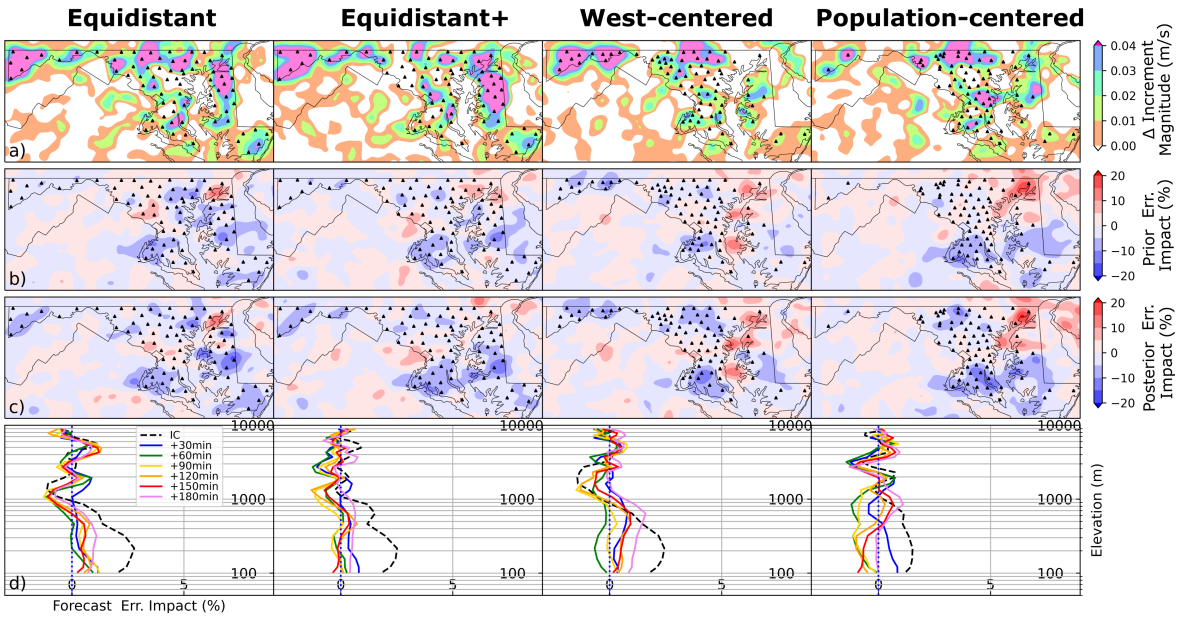

Figure 3.6: Same as Fig. 3.5 but for zonal-wind speed.

for the detection of reduced virtual potential temperature (θ_v) indicative of developing cold pools or sea breezes, and also for the detection of elevated horizontal convergence at the lowest model level indicative of colliding cold pool gust fronts and other mass boundaries that favor forced convection. The θ_v threshold for cold pool detection was dynamically set based on environmental means following Torri et al. (2015), while for horizontal convergence we chose to use the 90th percentile of convergence present in the nature run at the corresponding time stamp. We calculated these average scores across all cases and for both the standard-uncertainty and constrained-uncertainty modeling system configurations, with results presented in terms of the case-averaged difference between each candidate MM configuration and the no-build configuration (Fig. 3.7). Contrast is evident between scores for cold pool identification and the identification of elevated horizontal convergence, with the former showing robust benefit from MM assimilation for all MM configurations and the latter showing only modest median improvement with values for lower quartiles showing degraded member representation of convergence zones.

Results therefore suggest that the assimilation of MM observations provides utility in improving the resolution of near-surface thermodynamic features but is much less helpful in refining model depiction of mesoscale frontal boundaries. We note that with the possible exception of cold pool scores for equidistant configurations, there is remarkably little difference in CSI scores associated with individual MM configurations.

As a specific example that highlights how changes to small-scale structures may occur in analyses from the assimilation of MM observations, we examined density-current structures–inclusive of both convectively-generated cold pools and sea breezes but referred to here as 'cold pools' for simplicity–identifiable at a single timestamp during the 12 August 2020 case event. We again used a threshold θ_v to identify the presence of these features in the nature run, before applying the same thresholding technique to the analysis means produced by our standard-uncertainty OSSE ensemble both with the default observation network and when assimilating the equidistant MM configuration. Focusing on a region-of-interest centered on the Baltimore-Washington metropolitan area, we plot the outlines of cold pools identifiable in the nature run along with those identifiable from both observation network configurations, as well as prevailing 10-m winds from all three sources (Fig. 3.8). Assimilating MM observations produces cold pool boundaries that are a closer match to the nature run across central and northeastern Maryland when compared to boundaries placed with the default observation network. This improvement stems primarily from an increase in analysis mean θ_v at the edge of cold pool regions with the effect of reducing their overall extent. While slight improvements to 10-m wind vectors are noted from assimilating MM observations, these occur haphazardly and do appear concentrated near outflow boundaries.

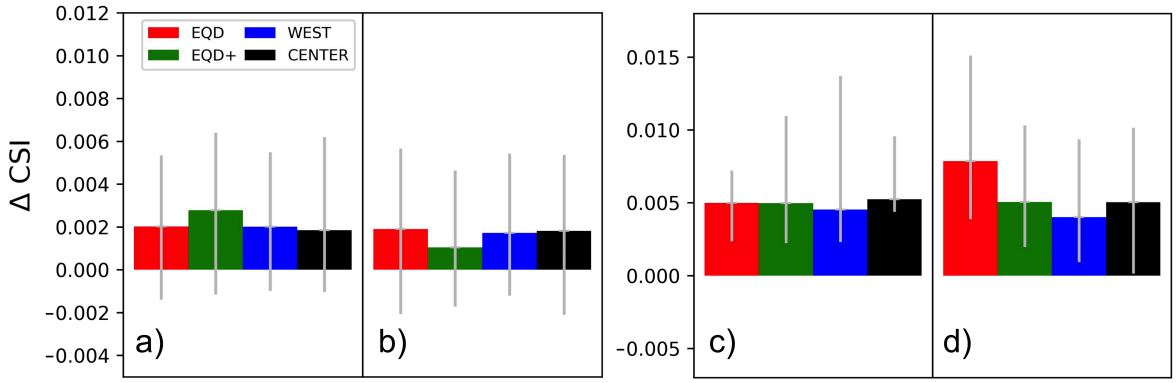

Figure 3.7: Case-averaged CSI score differences between each candidate MM configuration and the default observation network for the identification of near-surface horizontal convergence above nature run 90th percentile values for (a) standard-uncertainty modeling system (b) constrained-uncertainty modeling system. CSI score differences for the identification of θ_v below a threshold indicative of cold pool development for (c) standard-uncertainty modeling system (d) constrained-uncertainty modeling system. Error bars indicate quartile ranges among considered cases.

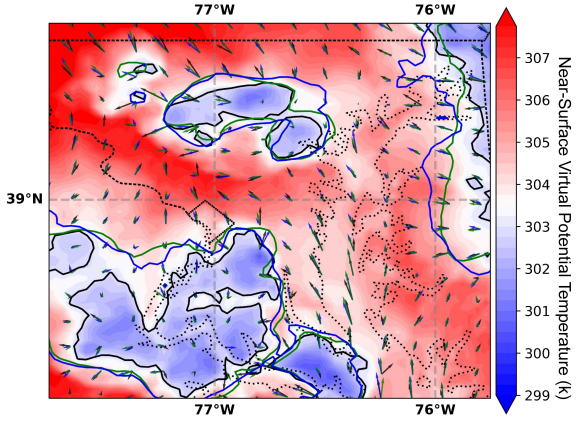

Figure 3.8: Cold pool boundaries and 10-meter wind vectors valid 2200 UTC 12 August 2020 as depicted by nature run (black), standard-uncertainty OSSE assimilating the default observation network (blue) and standard-uncertainty OSSE assimilating the EQD MM configuration (green). Shading shows base level θ_v for nature run. Dotted black lines indicate state and coastal boundaries.

3.4 Impact of MM observations on forecast verifications

3.4.1 Results for standard-uncertainty OSSE configuration

Using our standard-uncertainty modeling system configuration, we generated sets of initial conditions and 3-hr forecasts corresponding to sequential assimilation of observations from each of our four candidate MM configurations, as well as a set corresponding to assimilation with only the default 'no build' observation network. We chose to limit verification for resulting forecasts to the earlier defined Maryland verification region given the steep spatial gradient in near-surface-averaged mesonet impact to analysis fields of wind and temperature seen in Figs. 3.3-3.6. In Fig. 3.9a we present score impacts–defined as the percentage-wise difference between forecast FSS of candidate configurations and that of the no-build configuration–for the occurrence of elevated radar reflectivity and precipitation indicative of severe storm activity. Median and quartile score impacts were produced according to lead time and consider all forecasts produced across case events. This provides, for every MM configuration, a sample size of 133 discrete forecast verifications at 15-minute increments of lead time starting from initialization. Median score impacts in this context may be thought of as the effect on verification scores at a given lead time that may be expected from the assimilation of a particular MM configuration, with quartiles acting as an indicator for the variability in improvement (or degradation) at that lead time across sampled forecasts. For the prediction of elevated composite reflectivity (25+ dBZ), results are inconclusive and show a highly variable range of configuration impacts. Although the median differences between FSS using the no-build configuration and that produced using the candidate MM configurations is small, the 25th and 75th percentile score impacts from initialization to 45

minutes of lead-time indicate that a subset of forecasts experience larger benefits of up to 2% from the assimilation of MM observations, with comparatively fewer forecasts showing degradation of similar magnitude. Median results for POD and FAR (Fig. 3.9b,c) show no change in the detectability of events, but do indicate a 1–2.5% reduction in false alarm rate at short lead times roughly corresponding to the first 1.5 h after forecast initialization. For the prediction of 15-minute accumulated precipitation rates corresponding to moderate rainfall (0.625+ mm/h), we see no significant impact from the assimilation of MM observations for any of the aforementioned measures (data not shown). None of the aforementioned metrics showed differences between MM configurations that were large compared to the interquartile range.

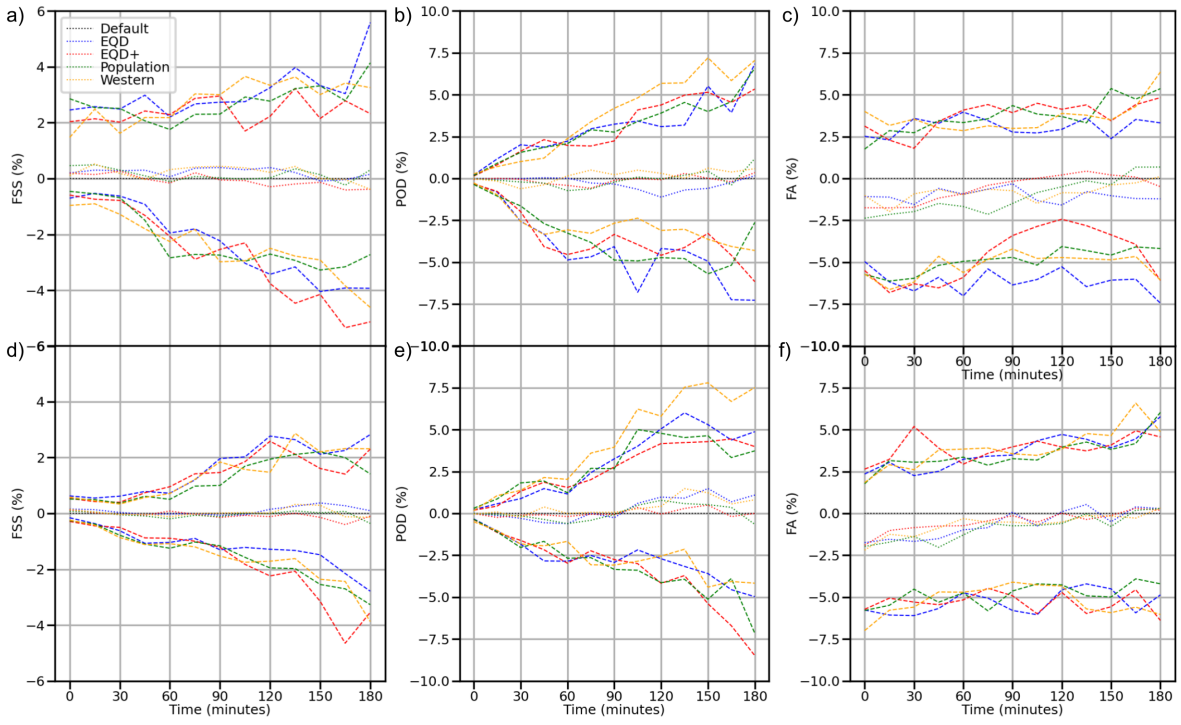

Figure 3.9: Network median and quartile percentage score difference from default observation network for (a) forecast fractions skill score, (b) forecast false alarm rate, (c) forecast probability of detection, valid for standard-uncertainty OSSE. Same for (d), (e), and (f) but valid for constrained-uncertainty OSSE

To augment event-based metrics, we examined the percentage-wise RMSE difference (RMSE

impact) between no-build and candidate configurations for forecast temperature and zonal wind fields, again averaging these scores within the Maryland verification region and across all available forecasts. We present color-coded vertical profiles of mean RMSE impacts for wind and temperature fields at lead times from 30 to 180 minutes (Figs. 3.3d,3.4d), overlaying corresponding impact profiles for forecast initial conditions introduced in the previous section. This figure depicts how forecast error relates to improvements afforded to analysis fields, both above and below the planetary boundary layer. Results for potential temperature show that–with the exception of the west-centered network–analysis improvements below 1000 m persist throughout the 180-minute forecast period, at which point they have decayed to roughly 0.5 – 1.5% from 5 – 7% present in initial conditions. For zonal-wind speed, improvements decay to near zero within the same layer by 60 – 90 minutes. Neither variable shows robust forecast improvements above 1000 m at any lead time, indicating that improvements to analysis fields at lower elevations do not translate to the free atmosphere during forecasts.

Although we do not break down the geographical distribution of forecast improvements by lead-time, in figures 3.3b and 3.4b we present plots of near-surface-averaged MM configuration impact on prior-mean error for potential temperature and zonal-wind speed. These were produced in the same way as corresponding plots for analysis means (Figs. 3.3c,3.4c), and may be seen as a proxy for short-term, 15-minute forecasts produced by our modeling system over the course of sequential DA. Comparisons with analysis-mean impacts show that for potential temperature, improvements are preferentially degraded in western and central Maryland, with stronger persistence for southeastern regions of the state surrounding the Chesapeake Bay. For zonal wind, regions of positive analysis impact are degraded somewhat more evenly, though the quickest decay still occurs for western Maryland.

3.4.2 Results for constrained-uncertainty modeling system featuring perfect physics

To examine how the utility of candidate mesonet configurations would change given advances in physical parameterization schemes and limited initial-condition uncertainty for large-scale flow–as depicted by reduced boundary condition uncertainty–we generated further sets of initial conditions and forecasts using our constrained-uncertainty modeling system configuration. As before, we examine fractions skill score curves for radar reflectivity (Fig. 3.9d) and precipitation (not shown) corresponding to aggregated forecast performance across events stratified by lead time. For verification of composite reflectivity above 25 dBZ, we find that the range of impact from assimilation of MM observations is smaller than that for the standard-uncertainty set of experiments, both in the median and for the interquartile spread of score impact. This is in part due to an overall increase in FSS at analysis times associated with reduced model error and improved boundary conditions. Unlike the standard-uncertainty case, any benefit from assimilation of MM observations is limited to the analysis time where it is again present only in a subset of forecasts contributing to 25th and 75th percentile score impacts. As before, verification of precipitation rates above 0.625 mm/hr shows no detectable benefit from MM observations, although the variance of skill-score impact is also reduced in a manner similar to that for composite reflectivity.

Vertical profiles of forecast-mean impacts for potential temperature (Fig. 3.5d) show similar lead-time progressions to that for the standard-uncertainty experiments, maintaining small improvements below 1000 m for the duration of forecasts. However, equivalent profiles for zonal-mean wind (Fig. 3.6d) show a much sharper lead-time cutoff in terms of useful impact near the

surface. For most MM configurations, there is no discernible trend in forecast mean-impact past 30 minutes. As before, no forecast improvement is seen above 1000 m for either potential temperature or zonal wind. Likewise, the geographical distribution of prior-mean impacts presented in figures 3.3b and 3.4b indicates that improvements to analysis accuracy associated with MM observations are preferentially retained in southeastern Maryland for both variables.

3.5 Discussion

3.5.1 Findings on the utility of a Maryland mesonet

Notwithstanding the comparative performance of candidate MM configurations, we find that the assimilation of MM observations as a whole leads to clear but localized improvements in analysis and forecast accuracy. Results show a distinct reduction in RMSE for near-surface, analysis-mean fields that persists within corresponding forecast fields for up to 3-h in the case of temperature and for 30-90 minutes in the case of zonal-wind speed. The vertical extent of benefit is restricted to a surface-based layer below 1000 m, roughly corresponding to the convective boundary layer, with little or no impact seen above. We note that this is consistent with previous findings in Sobash and Stensrud (2015) and Marquis et al. (2014) that showed limitations in the ability of surface observations to influence analyses in the free atmosphere. Geographically, improvement is confined to regions that are directly adjacent to MM site locations, well within the 30-km localization length scale for surface observations. This is especially apparent in analyses produced with the constrained-uncertainty configuration of our modeling system, where analysis benefits retreat from remote portions of the southern Chesapeake Bay. Temporally, within the sequential cycling period, reduced RMSE for analysis-mean fields occurs pre-

dominately after convective activity has passed through a given region, and is more distinct for MCS cases than for cases driven by synoptic-scale frontal passages. Of our considered analysis variables, potential temperature appears to benefit to a slightly greater degree than wind speed for our standard-uncertainty experiments, and this difference increases significantly for simulations with constrained uncertainty. Together, these effects suggest that the primary effect of MM observations is to improve model representation of near-surface outflow structures produced by convective storms, with limitations imposed by the ability of mesonet observations to inform spatially distant variables via cross-covariances. When and where benefit to analysis-mean RMSE is present, improvements extend to the identification of mesoscale surface-based structures such as cold pools and gust fronts in individual posterior members.

Despite the aforementioned benefits, the results of our OSSE experiments are inconclusive in demonstrating that the assimilation of additional surface observations is itself beneficial to regional forecast skill for moist convective events. We find only limited forecast improvement associated with MM observations for simulations with realistic error structures, as well as for simulations featuring reduced model error and constrained lateral boundary conditions. When considering fractions skill scores for the occurrence of elevated composite reflectivity as well as those for the exceedance of threshold accumulated precipitation values, median paired skill-score differences between the default observation network and evaluated mesonet configurations are consistently negligible across OSSE experiments. Although score impacts for false-alarm rate–especially at their 25th percentile–indicate that at least a subset of forecasts benefit from reduced over-prediction of convection, we note that such outliers may be impacted by our choice of case events and that–on average–the location, timing, and strength of forecast convective events is not improved by assimilating additional mesonet data.

We offer two possible explanations for why improvements to near-surface analysis fields and outflow features do not translate into significantly increased NWP forecast skill for simulated moist convective events. The first is an inherent limitation of geographically constrained networks like the MM, caused by the rapid turnover of information exported from the domain covered by MM observations and replaced by information from outside the local vicinity. Indeed, we see evidence of this in the relative absence of near-surface-averaged improvement to prior-mean fields for wind and potential temperature in western Maryland, where all candidate MM configurations are necessarily thin in the meridional direction, suggesting that improvements to these fields in posterior members are washed out of the region completely during the 15-minute model advance. Secondly, we find it highly likely that predictability aspects of moist convection act to constrain the ability of surface observations to correct the dynamic evolution of convective storms even when they result in more accurate model depictions of the near-surface environment.

3.5.2 Findings on the importance of station placement within MM networks

Given inconclusive results for the impact of MM observations on NWP forecast skill, our evaluation of candidate MM configurations and the site placement of individual stations focuses on the effect that these may be expected to have in improving analysis accuracy. Results for the geographic distribution of analysis-impact indicate that there is a generally positive relationship between the local density of stations placed by MM configurations and the magnitude of nearby analysis-mean improvement, although this relationship does not hold in areas of northern and eastern Maryland that coincide with high concentrations of non-MM surface observations. In line with the above discussion on limitations for MM observations, we contend that the

marginal utility of additional site density is positive in so much as uncertainty for nearby variables is both present and well-sampled in the prior ensemble. The fulfillment of these conditions depends greatly on the scale of resolved uncertainty structures, which in turn may depend on modeling system configurations, dynamic conditions, and the particular variables involved. The behavior described here would explain reductions in positive analysis impact between standard-uncertainty and constrained-uncertainty experiments since the former contains large-scale uncertainty features that are resolvable with our 40-member ensemble, whereas uncertainty structures in the latter experiments are dominated by meso-γ scale storm structures that may be incorrectly sampled during DA. Similarly, we may explain the lack of positive analysis impact in well-observed regions of central and northeastern Maryland by recognizing that a sufficient density of observations from pre-existing surface stations would significantly constrain large-scale uncertainty structures in prior ensembles and therefore cause innovations in these regions to be drawn from smaller, less resolvable scales. These considerations should favor a network configuration that maximizes useful information by maintaining rough equidistance while avoiding areas where pre-existing observations are capable of restraining the growth of error structures at scales amenable to operational DA systems. Of the MM configurations evaluated in this paper, the modified equidistant configuration best fulfills these requirements. However, this is not confirmed by results for analysis impact–which do not show a clear winner. This is most likely because we required that pre-existing surface stations report data at intervals of less than or equal to 30 minutes in order to influence site placement in our simulated-annealing algorithm, which did not account for large numbers of surface stations reporting data at less frequent–often hourly–intervals. A network of this type, generated in a manner that accounts for all pre-existing stations, may show an improved impact on analysis accuracy beyond that demonstrated for evaluated MM

configurations.

3.5.3 Implications for convective-scale DA and future observing systems

We may extrapolate the results of our OSSE experiments to provide more generalized insight into factors that drive forecast errors in moist convective regimes and how they may limit the current and potential utility of assimilating denser networks of surface observations in limited-area, convective-allowing NWP systems. From a simplified perspective governed by Bayesian inference, forecast error arises from a combination of initial condition and model errors (Privé and Errico, 2013), with the interaction of the two depending on the particular nature of the dynamical system, its discretization, and consequent structure of both true and modeled phase space trajectories (Shaw, 1981). Although not quite as rigorous as formal sensitivity analysis (Torn and Hakim, 2008), our standard-uncertainty and constrained-uncertainty OSSE configurations sampled two possible permutations of these errors from which to conclude: the former featuring elevated model error from choice of microphysics scheme and initial conditions degraded by synoptic-scale perturbations to boundary conditions, and the latter featuring no such synoptic scale forcing and only a small discretization component of model error. While both cases show reductions in near-surface initial condition errors attributable to assimilating MM observations, this improvement does not scale monotonically with station density in the sense of reductions to squared errors and is insufficient to improve forecast skill for the prediction of moist convective events even when forecasts are integrated with near perfect boundary conditions and physics. Furthermore, residual improvements to near-surface forecast fields of wind and temperature that exist for lead times below one hour for our standard-uncertainty experiments are notably reduced

in our latter experiments where error is predominantly generated in regions of convective activity. While the lack of improvement to forecast verifications at extended lead times is undoubtedly consistent with the restricted spatial extent of the Maryland mesonet and consequent advection of information out of the verification domain, that such improvements are minimal even at short lead times suggests that the ability of surface observations to improve model representations of moist convection may be limited by aspects of intrinsic predictability.

Some studies that have taken an analytical approach to quantifying intrinsic predictability limits of moist convection (Ramanathan et al., 2019; Ramanathan and Satyanarayana, 2021) suggest that storm features near the upper end of the meso-γ scale may retain a degree of predictability for up to 6 hrs, with further extended predictability possible for highly organized systems. Though this would appear to leave room for improvement on practical predictability time scales achievable by NWP systems, the damped-driven dynamics of moist convective systems may result in a high rate of 'information turnover' compared to synoptic scale processes. As an example motivated by Fig. 3.7, in situations where convection is heavily favored at particular locations it may be unnecessary to refine the model depiction of outflow boundaries beyond a certain level of accuracy to obtain a reasonable forecast of convective activity. Conversely, in weakly sheared, conditionally unstable environments where convection is disorganized and heavily dependent on outflow forcing, noticeably improving forecast accuracy may require an unachievable amount of constraint on outflow boundaries. Alternatively, improvements to near-surface analysis fields and even to resolved cold-pool structures may be diluted by poorly-resolved processes occurring above 1000 m–including entrainment, detrainment and downdraft formation–where surface observations have little impact. We note that deep moist convective events occur in a high dimensional phase space involving thermodynamic, momentum, and moisture fields over the full

height of the troposphere (Rotunno et al., 1988). In the context of our experiments, limitations specific to moist convection would be most evident in the constrained-uncertainty experiment where 'low-hanging fruit' in the form of large-scale initial condition error leading to downscale error propagation has been addressed, and where indeed we see the least evidence of persistent RMSE benefit to forecast fields. While the role of model error in this scenario is less clear, our results suggest that even drastic improvements to physical parameterizations are insufficient to significantly increase the utility of surface observations. This could be the case if dynamical attractors relevant to moist convection are already reasonably well defined by an imperfect model, making model error less important than sampling deficiency and other factors such as improper assumptions made by DA algorithms. If predictability aspects of moist convective systems do in fact reduce the amenability of such regimes to surface DA, then we would expect the marginal utility of additional surface observations to decrease and reach saturation at lower density as improvements to global models increasingly reduce the amplitude of error at synoptic-scales.

3.5.4 Experiment limitations

Errico and Privé (2018) notes that regional OSSEs of the type presented in this chapter may present issues in terms of accurate representation of boundary conditions, and further warns that ill-posed OSSE frameworks may spuriously indicate utility for candidate observing systems. We note that although we do not use a global ensemble for initial and boundary condition generation, this would not be a factor in our constrained-uncertainty OSSE configuration that removes LBC perturbations larger than the meso-γ spatial scale. Another potential limitation in our experimental framework is the non-consideration of observation error correlations that are sometimes

parameterized in highly tuned OSSEs for operational NWP systems. We suspect that given our assimilated observation types, the primary effect of considering these correlations would be a slightly reduced impact from radar observations in areas of moist convection. Although this could feasibly lead to greater benefit from assimilating MM observations in our standard-uncertainty OSSE, parameterized correlations would likely not hold for future changes to radar networks, or improved statistical techniques to reduce existing error correlations. Furthermore, we note that although humidity observations are often ingested by NWP systems, we chose to assimilate only temperature and wind measurements provided by mesonet stations due to difficulties in properly representing the assimilation of bounded quantities in an OSSE framework. Given the importance of lower-level moisture in properly simulating conditional instability, it is conceivable that these observations may mediate yet-undetected NWP benefits from additional surface observing stations. Finally, we acknowledge that our sample of warm-season moist convective case events necessarily excluded other common causes of severe weather. Further study may be warranted for examining the impact of mesonet observations for the prediction of winter-weather events, particularly those with convective elements or shear gradients in surface precipitation type that may be amenable to the high resolution possible with mesonets.

3.6 Summary and conclusions

This chapter presents results from a set of observing system simulation experiments that evaluated the impact of a new Maryland mesonet on analyses and forecasts produced using a convective-allowing model framework. Although the Maryland mesonet has wide-ranging aims, this study was focused on implications for the predictability of severe convective storms due to

their prevalence and impact on the local region. Accordingly the adopted OSSE framework simulated mesonet impact for seven moist convective case-events, with final selection influenced by event severity in densely populated areas of the state. We hypothesized that additional surface observations may reduce analysis error for quantities relevant to the development and propagation of convective storms–especially near-surface wind and temperature anomalies associated with cold pools–and result in lengthier periods of predictability for moist-convective events than if these observations were not assimilated. We further proposed that such benefits may be mediated by the spatial configuration of mesonet site locations, making it desirable to study and optimize impacts associated with specific configurations. Initial conditions and forecasts were produced using several candidate spatial configurations for mesonet site location and then verified against a 'truth-state' defined by our nature run using both continuous and categorical statistics. Simulations also considered the role of common modeling system errors, and how anticipated improvements to boundary condition constraint and model process error may affect the utility of mesonet observations in future modeling systems. To answer these questions in an OSSE framework, we performed two sets of experiments for every set of configurations: one using a 'standard-uncertainty' configuration with the simulated model process and boundary condition error with synoptic-scale structure, and a 'constrained-uncertainty' configuration with perfect physics and low-amplitude, high-frequency boundary condition errors. For analysis uncertainty, we found that assimilation of any of our mesonet configurations is associated with modest reductions in analysis error for near-surface temperature and zonal wind fields that is restricted to a layer below 1000 m, consistent with previous studies that assimilated similar observing systems. Analysis error reduction is slightly more enhanced and extensive for experiments with our standard-uncertainty modeling system compared to those performed with the constrained-

uncertainty configuration, especially for near-surface zonal wind which we attribute to the relative ease of sampling error structures induced by large-scale boundary perturbations compared to errors propagating upscale in moist-convective regimes. While increased accuracy for thermodynamic and wind fields is maintained in short-term forecasts–most visibly with our standard-uncertainty modeling system configuration–it does not translate to detectable improvement in forecast skill for the occurrence of discrete convective events. This may suggest limitations in the ability of dense surface observing networks to extend the practical predictability of moist convective regimes due to scale-dependent aspects of involved dynamical systems and the nature of their coupling to surface processes. Although our results do not show drastic differences in analysis or forecast improvement between candidate mesonet configurations, we conclude that optimal improvements to analysis accuracy may be achieved by roughly equidistant placement that avoids areas where pre-existing surface data heavily constrains error structures resolvable by operational DA systems and where remaining uncertainty pertaining to near-surface variables is mediated by meso-γ-scale structures.

Chapter 4: Exploring choices of model physics and observation pre-processing for convective-scale data assimilation

4.1 Introduction

4.1.1 Potential and remaining challenges for state estimation using a localized particle filter

The past 10 years have seen rapid improvements to regional prediction systems explicitly tuned for forecasting the near-term evolution of MCS's and other weather hazards governed by mesoscale dynamics. State-of-the-art prediction systems such as the NSSL Warn-on-Forecast-System feature self-cycled, convective-allowing ensembles, comprehensive multi-moment microphysics parameterizations, sub-hourly assimilation intervals, and direct assimilation of radar reflectivity rather than latent-heat nudging or other indirect assimilation methods. These methods allow for powerful, real-time state estimates for variables that characterize the internal components of MCS structure within advanced weather models. The accuracy of such systems exceeds that of earlier-generation regional prediction systems by estimating not only the local mass and thermodynamic fields, but also the distribution of hydrometeor species, mixing ratios, and number concentrations within discrete convective cells or groups of convective cells. However, state estimates produced by operational prediction systems still remain tied to approximations

in contemporary DA methods and to ad-hoc observing strategies primarily motivated by impacts to forecast skill. Further improving DA accuracy for applications of moist convection may require alternative state-estimation and observing strategies that systematically remove assumptions present in existing systems.

Particle filtering (PF) methods are a Bayesian approach to state estimation that avoid parametric assumptions for prior and posterior distributions. PF's have long been noted for their potential to improve state estimates for systems featuring highly non-linear dynamics and non-Gaussian distributed quantities of the kind found in moist-convective systems, but suffer from scalability issues for high dimensional applications. The recent advent of computationally affordable, localized particle filtering methods such as the iterative local particle filter (LPF)(Poterjoy, 2016; Poterjoy et al., 2019; Poterjoy, 2022b) addresses scalability problems and allows for the application of Bayesian DA with models that exhibit clear error independence across subsets of state variables on short time scales. As demonstrated in McCurry et al. (2023) and Chapter 2 of this thesis, the LPF offers advantages over Kalman filter-based methods in constraining microphysical variables associated with moist convection. In particular, LPF state estimates are closer to climatological attractors produced by model physics, while producing forecast skill that is becoming increasingly more competitive with EnKF-initialized predictions. We hypothesize that further improvement in state estimation with the LPF can be obtained by taking advantage of the algorithm's flexibility in accommodating arbitrary forms of likelihood functions, which has been an unexplored area of research thus far.

At the same time, the use of a delta function-estimated prior density in the LPF poses a separate set of challenges that continue to hinder state estimation for convective storms. In contrast to linear estimation methods that apply covariance-based regression, the LPF and other

PF methods cannot easily form posterior samples that fall outside the span of a prior ensemble. This becomes especially problematic where prior estimates are affected by significant model error and therefore repeatedly driven far from the truth. The primary expression of this deficiency for the Warn-on-Forecast system appears to be an underestimation of surface precipitation rate and cold-pool strength, mediated by persistently negative DA updates to hydrometeor mixing ratios. Although model error leading to spurious hydrometeor updates may originate from multiple sources, microphysical parameterization schemes offer the most direct explanation. To this point, McCurry et al. (2023) found that comparative forecast skill between initial conditions from the LPF and EnKF was strongly impacted by the choice of microphysics scheme; the LPF generally verified worse than the EnKF using the NVD scheme but yielded more accurate forecasts for a test case implementing Thompson microphysics.

To overcome challenges related to model error and exploit unique capabilities offered by the LPF, the current study proposes a synergistic set of assimilation and observing strategies. We begin by modifying the microphysical scheme used for sequential DA to prevent systemic underestimation of hydrometeor mixing ratios. We then increase the constraint on filter posteriors by implementing a novel observing strategy for radar reflectivity that eschews the separation of radar observations into separate types based on the strength of returns. Having minimized analysis error using the previous two approaches, we use analysis residuals to derive an empirical, state-dependent likelihood function for radar reflectivity, which we then apply for DA with the LPF.

While changes to microphysics and reflectivity observing strategies could be implemented with any type of DA algorithm, we emphasize that the modifications in this experiment are specifically intended to correct deficiencies noted in the LPF. To this point, preliminary experiments

performed with an EnKF have shown significantly less sensitivity to changes in microphysics. For these reasons, we restrict discussion in this chapter to experiments with an LPF.

4.2 Background

4.2.1 Current DA strategies for radar reflectivity

Although we do not go into great detail on assimilation strategies for radar observations, we note that the direct assimilation of reflectivity values below roughly 25 dBZ poses more of a challenge than the assimilation of higher values. Moderate returns between roughly 10-25 dBZ may be found in both precipitating and non-precipitating regions of storm activityStarzec et al. (2017), and may experience greater biases than higher returns from ground clutter signals that occasionally contaminate radar data Dixon et al. (2007). This presents the risk that analysis increments influenced by misspecified likelihoods or poor estimates of prior covariances could introduce spurious convection in forecasts when assimilating returns inside this range. 'Clear-air' returns featuring near-zero or negative values present a separate challenge since they often produce large innovations when assimilated near regions of modeled convection Gastaldo et al. (2018). These large innovations can cause observations to be discarded by the DA system or introduce spurious increments when using a typical Gaussian likelihood function. The aforementioned challenges represent trade-offs that must be compared against the potential benefits of assimilating clear-air and moderate returns, which include removing modeled convection not supported by observations, and reducing analysis error in regions dominated by ice and snow hydrometeors – which produce low reflectivity even at high mixing ratios due to the dielectric properties of ice Smith (1984).

A common compromise is to separate radar reflectivity into two separate observation types with major differences in pre-processing, as with the operational WoFS. Gridded reflectivity returns above a cut-off threshold are assimilated without further modification, while returns lower than the cut-off value are rejected and used only as an indicator of radar coverage. A "clear-air reflectivity" observation type set to zero dBZ is then created at discrete vertical levels where coverage exists and no returns within an adjacent vertical layer exceed the cut-off value. This vertical layer may be quite large, to the point of encompassing the entire atmospheric column for pre-processing configurations that apply the cut-off threshold to composite reflectivity.

The assimilation of radar reflectivity is also impacted by the specification of the likelihood function. When applying Bayes' rule a likelihood function determines how states that exist within the prior domain are re-weighted to reflect observed quantities. These likelihood functions depend primarily on the distribution used to characterize observation errors and are typically a function of "prior innovations" or observation departures from prior model states. An observation error in this case represents a sum of contributions from – usually well-known – instrument-based measurement error, measurement operator error, a "representativeness error" component that depends on internal aspects of the modeling system, as well as other sources of error such as observation pre-processing failures Janjić et al. (2018b). We reproduce from (Janjić et al., 2018b) a summary of relations contributing to observation error:

$$e^o = e^m + e^{'} + e^{''} + e^{'''} \tag{4.1}$$

$$e^{'} = h^c(w) - h^c(w^r) \tag{4.2}$$

$$e^{''} = h^c(w^r) - h(w^r) \tag{4.3}$$

Here, e^o, e^m, and $e^{'''}$ respectively signify total observation error and contributions from measurement and pre-processing error. $e^{'}$ signifies the contribution to observation error from unresolved scales and processes, referred to henceforth as representation error. As shown in Eqn. 4.2, representation error can be described as the difference between the fully correct observation operator applied to discretized model fields–$h^c(w)$–and the same operator applied to hypothetical continuous fields that fully describe the underlying physical state–$h^c(w^r)$. Likewise, we may quantify error induced by the measurement operator as the difference in output between a hypothetical perfect measurement operator and the actual, implemented measurement operator when applied to the same fully represented physical state (Eqn. 4.3).

Although not inherent to linear estimation, it is common practice for DA implementations to assume the same observational error variance for all observations of a given type. This may be suboptimal when the true observational error variance is state-dependent. Likewise, commonly used Gaussian likelihood functions may not be appropriate for observations with non-Gaussian errors. Studies performed with the Kilometer-scale Ensemble DA (KENDA) modeling system, a mesoscale weather prediction similar to the NSSL WoFs, have shown that the representational component of observation errors for radar reflectivity are significantly state-dependent Zeng. et al. (2021). Likewise, large deviations from Gaussianity are noted with background innovation statistics collected using a WRF-based modeling system in Gao et al. (2024).

4.3 Methodology

4.3.1 Modifications to the NVD microphysics scheme

From results presented in McCurry et al. (2023), we surmised that excessively negative adjustments to hydrometeor mixing ratios may best be explained by bias in parameterizations – such as hydrometeor size distribution shape parameter – that influence the relationship between mixing ratios and diagnosed radar reflectivity. We specifically considered two possible regions within convective systems that may disproportionately contribute to the interaction of this bias with reflectivity increments: areas of moderate to heavy rainfall within convective cores, and the bright-band region associated with trailing stratiform regions in squall-line producing systems.

To address noted deficiencies, we made changes to five parameters in the scheme's internal namelist (Table 4.1). The shape parameter for rain was changed from the default value of 0 to 1, which consequently decreased model-diagnosed reflectivity for fixed mixing ratios. Lower values of reflectivity should lead to more positive innovations in regions of heavy liquid precipitation, and thereby smaller reductions in hydrometeor mixing ratios. We made further changes to the snow aggregation efficiency factor, which we reduced from 1 to 0.5, and to the range of temperatures between which aggregation efficiency increases linearly. These two changes together decreased the intensity and vertical extent of the bright-band region primarily associated with trailing stratiform regions within QCLS-type events. As before, these changes are intended to increase precipitation efficiency by modifying expected innovation values for reflectivity. Finally, we increased the snow fall-speed factor from 1 to 1.25 to avoid excessive hydrometeor loading at upper levels arising from the aforementioned changes to aggregation that produce smaller, slower

falling hydrometeors. (Ted Mansell, personal communication).

Table 4.1: Parameters varied for modifications of NVD microphysics scheme, description of parameter function, and values for default and modified settings.

Varied Parameter name	Description of parameter	Value for default NVD scheme	Value for modified NVD scheme
alphar	rain droplet size distribution shape parameter	0.0	1.0
ess0	snow aggregation efficiency factor	1.0	0.5
esstem1	snow aggregation temperature minimum for aggregation	-25 K	-15 K
esstem2	snow aggregation temperature of maximum aggregation efficiency	-20 K	-10 K
snowfallfac	snow fall speed factor	1.0	1.25

4.3.2 Description of alternative observing strategy for radar reflectivity

In addition to microphysics changes, our study evaluates two strategies for generating radar reflectivity observations. We refer to the first as the "standard observing" configuration. This configuration separately assimilates two types of radar reflectivity observations using distinct pre-processing methods. Reflectivity returns above a cut-off threshold of 25 dBZ are assimilated as typical radar reflectivity observations with 3-km grid spacing, while a distinct "clear-air" observation type is created in regions without nearby returns exceeding the cut-off threshold. This method is reminiscent of the observing strategy currently employed for the WoFS and is the default observing configuration for experiments described in Chapters 2 and 3. For our experiments, the cut-off threshold was implemented for total composite reflectivity, which means that clear-air observations are only placed where no return within a column exceeds the cut-off. Clear-air observations are placed at two vertical levels at 1000 m and 6000 m, and gridded at 6-km grid spacing, compared to 3-km grid-spacing for other reflectivity observations. We note that current practice for the WoFS has removed the use of composite reflectivity in favor of an approach based on vertical intervals, and uses a lower, 15 dBZ threshold (Derek Stratman, personal communication). Our experiments therefore reflect a slightly out-of-date configuration for

observation processing.

A second, "alternative observing" configuration creates reflectivity observations in a homogeneous manner by processing all returns into a single reflectivity observation type and removing special treatment for clear-air observations. The only distinction made between return values during pre-processing is in setting gridded reflectivities that are negative to 0 dBZ, which is retained from the standard observing configuration. The resulting observations differ greatly from the first strategy in containing reflectivity values between 0 and 25 dBZ as opposed to solely zeros. The combined density of observations below 25 dBZ is also significantly higher than that of 0 dBZ observations produced by the first strategy. In Figure 4.1, we provide a simple graphical overview of the differences between standard and alternative observing configurations.

Changes introduced with the alternative observing configuration are intended to reduce analysis errors in impacted regions by a combination of increased observing density relative to that provided by clear-air observations, as well as reductions in observation error introduced by assumptions during pre-processing. The latter may be especially important for portions of convective systems featuring composite radar reflectivity slightly below the 25 dBZ cut-off for the standard observing configuration.

4.3.3 A non-parametric specification of likelihood functions

As an addendum to microphysics and observing system modifications, we take advantage of the flexibility offered by the LPF algorithm by implementing empirical likelihood functions that may better represent observation errors than current approaches, which ignore state dependence and non-Gaussian distributions. Modifying the LPF to work with empirical likelihood

functions is relatively straightforward, and accomplished by assigning state and observation-space weights with a look-up table rather than the default Gaussian function.

Creating state-dependent empirical likelihood functions requires estimating full observational error distributions that are uniquely defined for each hypothesized true state. Although several methods exist for estimating observational error variances using innovation statistics combined with linear estimation theory (Desroziers et al., 2005; Karspeck, 2016)], these methods only consider the second moment of the observation error distribution. Alternatively, it is possible to deconvolve observational and background error distributions of arbitrary form from innovation statistics by constructing a sufficient model of background error in observation space (Hu et al., 2024). Although this is a promising approach, constructing a full-fledged model for background error can be quite difficult and the deconvolution process relies on the assumption of uncorrelated background and observation errors that may not hold given sufficiently complex background-error models for observation types like radar reflectivity. Lastly, the resulting deconvolution is not guaranteed to be unique.

As an initial attempt at reducing form assumptions for likelihood, we apply a simpler method for forming empirical likelihoods which assumes that observation error is given solely by the observation departure from analyses. We implement this strategy by collecting and binning analysis residuals from sets of real sequential DA experiments performed with the LPF using a standard Gaussian likelihood function. This method does not involve deconvolution, but instead becomes equivalent to the deconvolution method of Hu et al. (2024) as analysis errors approach zero. In this case, background error is implicitly defined as the difference between background innovations and analysis residuals, and is completely removed during DA. Given the rapid DA updates and dense observing network featured in our case events, we may reasonably expect

analysis errors to be small compared to observation error variance. This approximation may be especially valid for cases where the effect of DA updates is primarily to correct persistent simulation errors. Binned analysis residuals represent an approximation of the observational error PDF contingent on the model and observing system configuration and were collected separately based on experiment configuration. For experiments considering the default configuration, likelihoods were formed using data from archived experiments without changes to microphysics or observing strategy, but using a 5 rather than 15 minute cycling frequency. To introduce a form of state dependence we sorted residuals based on their corresponding posterior-mean value before separately binning the residuals within each band. Although the method described above is applicable to any given observation type, only radar reflectivity was considered for the purposes of our experiments.

Our current approach with analysis residuals is a relatively inexpensive method that provides rough estimates for observational error distributions of any form, with the caveat that sampled PDF's will represent contributions from both observation and analysis error. Residuals collected from sequential DA experiments featuring minimal analysis error will reflect a purer estimate of the observational error distribution than residuals collected from experiments with lesser constraint. In this way, our experiments using empirical likelihood specification serve a two-fold purpose: firstly as a preliminary investigation of challenges and impacts that come with removing Gaussian assumptions for likelihood, and secondly by acting as a barometer for the accuracy of posterior estimates yielded by the two preceding modifications.

4.3.4 Experiment Configurations

Using the standard LPF, we performed a set of 10 total sequential DA experiments to evaluate the effect of modifications to microphysics settings, observing strategy for radar reflectivity, and likelihood function formulation. These experiments are performed in five pairs using distinct combinations of base modeling system configuration and independently varied settings, as expanded upon below in Table 4.2. With the exception of modifications introduced above, most aspects of methodology relevant to these experiments are unchanged from those described in the first chapter of this dissertation.

Three variables – reflectivity, radial velocity, and 2-m temperature – were chosen for root-mean-square-departure (RMSD) verification of observation-space output produced by sequential DA and forecasts. These variables tie directly to the model representation of convective processes via hydrometeor distributions and subsequent impacts on latent heating and cold-pool propagation. Because experiments using the alternative observing configuration assimilate a substantially different set of observations below the cut-off threshold for the standard configuration, we verify all experiments using observations above 25 dBZ unless otherwise stated.

All experiments described in this chapter were performed for a single event featuring a squall-line type MCS with multiple bowing segments that impacted portions of the midwestern United States from July 17th to 18th 2019. This case, in particular, was chosen because the quasi-linear mode of convective organization allows for a greatly simplified comparison of state estimates and forecasts between considered approaches, and because of its previous inclusion in the set of cases considered in McCurry et al. (2023). Snapshots of composite reflectivity showing MCS propagation across the model domain are shown in Figure 4.2.

Table 4.2: Name of broad experiment type based on the independently varied setting, the base modeling system configuration used for paired experiments and relevant settings used for individual sequential DA experiments

Experiment type (varied strategy)	**Base configuration for paired experiments**	**Settings considered for sequential DA**
Microphysics	Default	Default NVD scheme Modified NVD scheme
Observing	Modified microphysics	Standard observing configuration Alternative observing configuration
Likelihood	Default	Gaussian likelihood functi on Empirical likelihood function
Likelihood	Modified microphysics	Gaussian likelihood function Empirical likelihood function
Likelihood	Modified microphysics & observing	Gaussian likelihood function Empirical likelihood function

4.4 Experiment Results

4.4.1 Implementation of a modified NVD microphysical scheme

To demonstrate the overall impact of parameter modifications on microphysical state estimates, we show domain-averaged, column-integrated hydrometeor masses in Figure 4.3. Results for the default scheme are indicated in blue, with the modified NVD scheme indicated in red. For clarity, we plot only prior values for column mass but note that posterior values follow very similar trends. We can see a clear increase produced for both rain and snow column masses, beginning from the onset of organized convective activity at 1000 UTC and persisting over the entirety of the event. Conversely, there is little to no change to column-integrated graupel mass. Although not shown, we note that increased hydrometeor column masses with the modified NVD scheme translate to subsequent increases in surface precipitation rate throughout the duration of the squall-line event. Domain-averaged differences past 1300 UTC remain stable at roughly 0.08 mm/hr, representing a 50% increase over the standard NVD scheme.

In Figure 4.4, we present prior-mean RMSD results for experiments that compare experiments using modified microphysical parameters. Notably, the revised microphysics scheme yields a reduction in short-term forecast error for all variables considered in the verification. In the case of radar reflectivity, these improvements are small in magnitude– generally less than 1 dBZ–for the entirety of the sequential cycling period. Error reductions for radar radial velocity are also small for most of the cycling period. The most striking improvement is for surface temperature, where modified parameters lead to an advantage in error that grows continuously over the experiment to reach 0.6 k by 2200 UTC. We further extend this verification to forecasts at different lead times by computing forecast-mean errors for 0 – 180 min forecasts (Fig. 4.5). Results for default and modified microphysics schemes are once again indicated with blue and red, respectively. Trends established for prior-mean errors generally persist in forecasts. Both default and modified microphysics experiments show a similar evolution for reflectivity errors, while initial advantages for the modified physics configuration are maintained throughout the forecast period for radial velocity and 2-m temperature. Because of the close relationship between modifications made to microphysics and precipitation, we supplement the previous results with event-based verification statistics for the occurrence of hourly accumulated precipitation above 5 mm (Fig. 4.6). Fraction skill scores indicate a large advantage for the modified NVD scheme over the entirety of the 3-h forecast period. This appears to be mediated almost entirely by increases in POD, suggesting that additional precipitation present during sequential cycling is maintained for forecasts.

4.4.2 Implementation of an alternative observing configuration for radar reflectivity

To determine how DA with standard and alternative observing configurations impacts hydrometeors within our case study, we return to Figure 4.3, where we now consider results for combined microphysics and observing modifications shown in green. We focus comparisons on results shown in blue for the configuration with standard observing configuration and modified microphysics. The primary changes noted with the alternative observing strategy are a large, consistent increase in column-integrated snow mass and an initial decrease in column graupel mass. The relative decrease in graupel column mass over the standard observing configuration peaks from approximately 1200 UTC to 1500 UTC in the maturing phase of the squall line. Column rainwater mass is largely unchanged from the modified microphysics experiment using the standard observing configuration.

To complement our set of sequential DA experiments, we also ran several "identical prior" experiments to examine the immediate impact of observing configuration on hydrometeor quantities. The set of priors is valid for 1400 UTC and collected from a sequential DA experiment using the alternative observing configuration with modified microphysics and empirical likelihoods. From these experiments, we show a vertical cross-section of analysis increments for hydrometeor mixing ratios generated by both radar observing strategies (4.7). This cross-section is taken along an axis parallel to the direction of storm propagation that roughly splits the developing squall line. Increments are presented for mixing ratios of rain, snow, and graupel, along with contour lines indicating prior reflectivity. To reduce noise from small spatial scales, all values reflect grid-cell averages within a 6 km meridional distance perpendicular to the vertical

cross-section.

Comparing results for both observing configurations in Fig. 4.7, we see strong altitude-dependent differences in increments for snow mixing ratio. Above 9000 m, the standard observing configuration produces increment magnitudes that trail off to negligible values. By comparison, increment magnitudes for the alternative configuration reach their peak near 10 km and only trail off to near zero above 14 km. Increments for the alternative configuration remain large within an anvil and trailing stratiform regions of the MCS even at significant horizontal displacements from the convective core, which may indicate a vector by which observing configuration impacts MCS cloud shield extent. Between 5500 and 9000 m in the same region, near the level of clear-air observations placed at 6000 m by the standard configuration, we find increments of similar magnitude but of opposite sign. The standard configuration removes snow in these regions, while it is being added by the alternative observing configuration. Below 5500m, we find no significant disagreement between increments produced by observing configurations, even in the bright-band region of trailing stratiform precipitation.

Both observing configurations produce negative increments for rainwater mixing ratio for regions within the convective core and portions of the MCS immediately behind the convective core. In the latter region, the alternative observing configuration yields smaller increments at elevations roughly below 1000 m. We suspect that this is once again related to the standard observing configuration's placement of a clear-air observation level at 1000 m – closer to the surface than the freezing line at 3000 m. Interestingly, the impacted part of the trailing stratiform region is within the transition zone between convective and stratiform precipitation, where composite reflectivity considered for the generation of clear-air observations is depressed compared to adjacent regions.

For graupel, little difference is seen other than slightly stronger negative increments above 5000 m within the convective core. Due to the absence of clear-air observations in this region for the standard configuration, changes are most likely caused by similar factors to those that impact snow mixing ratios at high elevations.

Though not shown, the sequential cycling experiments using the alternative observing configuration show significantly extended upper-level cloud shields even in prior state estimates, which would support the idea of a lasting structural impact above-mentioned trends in upper-level hydrometeor increments. As with earlier changes to microphysical parameters, we examine prior-mean RMSD for observing configuration experiments shown in green in Fig. 4.4. These may be compared to results shown in blue for the modified microphysics experiment with standard observing configuration. The primary impact, in this case, is for reflectivity, where the alternative observing configuration produces a clear decrease in error for the majority of timestamps occurring after the advection of organized convective activity into the domain at approximately 1000 UTC. This reduction is typically on the order of 1-2 dBZ. Unlike changes made to microphysics, we do not see accompanying improvements to wind fields via radar radial velocity, or to 2-m temperature. While the error for 2-m temperature is largely unchanged, there is a slight degradation to radial velocity that may be indicative of spurious convection initiated from increments with this strategy.

Verification of forecast-mean error (Fig. 4.5) shows that radial velocity and 2-m temperature errors for the configuration featuring modified microphysics and alternative observing strategy eventually overcome an initial disadvantage compared to forecasts launched with only modified microphysics while maintaining initial advantages for reflectivity verification. The reversal shown for radial velocity and 2-m temperature is likely attributable to changes in hydrometeor

distribution shown in Fig. 4.7.

4.4.3 Removal of Gaussian likelihood constraint for selected Configurations

For the final portion of our study, we present results from experiments that removed Gaussian likelihood constraints during DA by implementing a state-dependent empirical likelihood function derived from analysis residuals. While we do not explicitly describe functions produced with this technique, we include a supplementary figure showing sampled observational-error PDFs calculated for our three experimental configurations B.1. To demonstrate the effects of state dependence we include PDFs for sampled posterior-mean reflectivity ranges of 4-5, 24-25, and 44-45 dBZ.

For each of the three base configurations mentioned in Table 2, experiments with a Gaussian likelihood function were repeated using empirical likelihoods. Since the likelihood function mediates increments produced by DA, we first examine how changes to the likelihood function translate to analysis increments for each of our configurations (Fig. 4.8). Using aggregated prior and posterior output in observation space, we present contour plots showing how the standard deviation of the prior ensemble and the observed innovation value jointly impact expected analysis increments. We produce similar plots considering the prior ensemble mean and innovation values. Although our sampled analysis increments are affected by nearby observations – and by localization within the LPF– we note that under idealized conditions and a fixed Gaussian likelihood function, the combination of prior standard deviation and innovation should fully describe the observation-space analysis increment. Prior values are included here to illuminate state dependence influenced by observing configurations and likelihood formulations.

For modeling system configurations that use the standard observing configuration, we find large differences in analysis increments produced by empirical likelihoods, especially for observations featuring elevated (10+ dBZ) standard deviations for prior reflectivity, and for observations that yield positive background innovations. These changes comprise a weakening of positive analysis increments and a strengthening of negative increments relative to the value of background innovations. We note that while a degree of state dependence is noted for increments associated with Gaussian likelihoods, this is reduced for DA with empirical likelihoods. We take this as a positive sign that empirical likelihoods are accounting for state-dependent uncertainty. Modifications made to the NVD microphysics scheme do not appear to change likelihood formulation in a way that significantly modifies the overall behavior of empirical likelihoods during DA, nor do such modifications drastically alter increments produced using Gaussian assumptions for likelihood.

Conversely, the application of our alternative observing configuration to a modeling system featuring modified microphysics yields drastic changes to increment structure from that seen in preceding experiments. These changes impact increments produced from both Gaussian and empirical likelihoods, which are weakened for positive increments relative to background innovation value, and strengthened for negative innovation values. We note that this change in Gaussian increment structure is similar to changes arising from empirical likelihood application in standard observing configuration experiments. For both Gaussian and empirical experiments, analysis increments show far less state dependence than for the standard observing configuration. Removing Gaussian assumptions does not seem to have as much impact on increment structure with the alternative observing configuration, although we still note a smaller tendency in the empirical likelihoods to the effect of weakening increments for large positive innovations, and

reducing remaining state dependency. In Fig. 4.9, we present comparative verification results for sequential DA experiments with Gaussian and empirical likelihoods. Beginning with the default modeling system configuration shown in the leftmost column, we find that empirical likelihoods cause a moderate increase in error of roughly 2 dBZ which persists from 1000 UTC through the last assimilation cycle at 2200 UTC. Only small intermittent differences are noted between likelihood formulations for radar radial velocity and 2-m temperature. For subsequent experiments using the modified NVD microphysics scheme, we find a similar increase of nearly 2 dBZ for radar reflectivity assimilated with our empirical likelihood function. Likelihood functions again have little effect on error for radial velocity, but do produce a relatively substantial increase in error for 2-m temperature. This increase is on the order of 0.1 k, beginning during a period of intensifying convection near 1300 UTC and persisting throughout the remainder of the experiment. Conversely, when likelihood experiments consider a modeling system that jointly modifies the NVD microphysics scheme and implements the alternative radar observing configuration, we find little to no increase in error for reflectivity when using empirical likelihood functions. The intermittent reduction in error for 2-m temperature between 1400 UTC and 1700 UTC is the only decrease shown with empirical likelihoods for this variable for any evaluated configuration.

4.5 Chapter Summary and Conclusions

4.5.1 Efficacy of a modified NVD microphysical scheme for DA with the LPF

We find that modifications to the raindrop shape parameter and to parameters controlling snow aggregation in the NVD scheme are mostly beneficial to state estimates produced using the LPF for our case event. These changes are effective in increasing mixing ratios during sequential

DA and at least partially counteracting the tendency of the LPF to significantly underestimate such quantities. We consider the physical justification of these changes by examining verification scores with and without modifications. Evidence in favor of physicality is provided by significant increases in skill scores for accumulated precipitation (Fig. 4.5). A realistic depiction of precipitation is crucial not only for practical forecast purposes but also for correctly resolving thermodynamic processes that drive organized convection via negative latent heating. Indeed, we show in Figs. 3 and 4 that changes to microphysical parameters lead to significant error reductions in surface temperature as well as more modest decreases in error for radial velocity. This is indicative of better resolved cold pool extents that allow for better resolution of density driven gust fronts.

Despite these advantages, we also identify two possible drawbacks associated with changes to the NVD scheme. Firstly, it is difficult to separate changes made to factors involved in the diagnosis of reflectivity from the subsequent verification of reflectivity. Since our approach intentionally reduces reflectivity by inducing smaller rain and snow hydrometeors, and by narrowing the bright-band melting region, we cannot say for sure which factor – reflectivity modifications or filter placement of hydrometeors – is influencing verification results. Secondly, while our changes are able to correct a generalized bias in the relationship between reflectivity and mixing ratio, we do not know for certain if our changes address the specific mechanisms leading to this bias rather than compensate for them. This may not be crucial for reducing known model error in state estimation, but in the sense of forecasts, two 'wrongs' do not necessarily make a 'right'. The re-appearance of model error without frequent constraint from DA would explain the rapid growth of error for 2-m temperatures for forecast timestamps past 90 minutes when run with the modified NVD scheme (Fig. 4.4), as well as the absence of any definitive advantage in forecast

FSS over the default microphysics scheme (not shown).

4.5.2 Implications of reflectivity observing configuration for resolution of MCS structure

As with changes to the microphysical scheme, modifications to the creation of radar reflectivity observations from raw measurements convey apparent benefits to state estimation. By processing all observations above 0 dBZ in the same manner and eliminating the specific treatment given to the 'clear-air' observation type in the standard observing configuration, the new method increases overall information content assimilated by the LPF, and reduces error previously introduced by assimilating moderately positive returns as 0 dBZ observations.

From identical prior experiments (Fig. 4.7), we can reasonably conclude that these differences are expressed most strongly at upper levels above 5500 m, where they significantly affect filter updates to frozen hydrometeors. Behavior noted above 9000 m seems to be mediated largely by information content in so much as the local density of reflectivity observations placed at a 3-km spacing, and at vertical elevations corresponding to that of radar swaths compares favorably to the corresponding density of 6-km spaced clear-air observations placed into two monolithic levels far below the elevations of interest at 1000 and 6000 m. Given a 3000-m vertical localization radius, the impact of observations at 6000 m may not be very large at these levels. For elevations between 5500 and 9000 m, while both configurations agree reasonably well within the convective core where the standard configuration does not place clear-air observations, this is not the case for the trailing stratiform region. Opposing increments in this region are likely the result of spuriously negative increments induced by the standard configuration's

placement of 0 dBZ observations in locations featuring composite reflectivity just slightly below the 25 dBZ threshold. For example, analysis-mean reflectivity fields produced with the standard observing configuration during sequential cycling, often produce 'holes' seemingly in the middle of glaciated portions of trailing nimbostratus that do not appear for the alternative observing configuration.

On the horizontal plane, the above-mentioned phenomena act to extend the cloud shield outward from the convective core, both in the trailing stratiform region and in the anvil. The increase in extent of glaciated stratus cloud extent appears to explain the increase in domain-averaged, column-integrated snow mass more than local changes in hydrometeor loading. Circumstantial evidence supporting the physicality of this increase in cloud-shield extent is provided by forecast verification results for 2-m temperature (Fig. 4.4), which show that the combined microphysics and observing modifications produce lower errors than microphysics modifications alone in the second half of the forecast period. Since rainwater mixing ratios and presumably near-surface precipitation flux did not increase with the alternative observing configuration (Fig. 4.3), and no increase in cold-pool extent is noted, we conclude that improvement is the result of interaction between extended cloud-shield and diurnal heating.

Based on the factors mentioned above and the reduction in prior-mean RMSD noted for reflectivity in the period of sequential DA, we conclude that the alternative observing configuration likely provides greater constraint on analysis error for reflectivity and hydrometeor variables than the standard observing configuration. However – once again – this does not necessarily translate to more accurate forecasts. The initial motivation for processing reflectivity separately based on a threshold – at least at NSSL – was to avoid spurious convection that might degrade forecasts for convective storms (Derek Stratman, personal communication), and there is at least

some evidence in our study to support this approach. Applying the alternative observing strategy increases prior-mean error for radar radial velocity during sequential assimilation. We also see a significant increase in false-alarm during the initial portion of the forecast period, although this occurs more so with a 25 dBZ event threshold that is more influenced by the extent of stratiform regions than with the 35 dBZ threshold indicative of active convection (not shown). Given our consideration of only one case event, concern for spurious convection may be sufficient to justify the standard observing approach for forecast-centric applications.

4.5.3 Evaluation of empirical likelihood functions and interactions with experimental observing strategies

Although each of the previously discussed modifications implemented in this chapter may stand on their own as strategies for improving state estimation with the LPF, the successful implementation of our final strategy for removing Gaussian assumptions on likelihood form hinged on input from modeling system configurations that minimized analysis error. Application of empirical likelihoods to three selected permutations of our modeling system did not result in any notable improvement to verification results from sequential DA or forecasts. We, therefore, conclude that even the strongest combination of approaches to minimizing analysis error in our modeling system did not produce errors small enough to isolate observation error and construct likelihood functions capable of outperforming a Gaussian function with fixed variance. Nevertheless, the combined changes to model physics and observation processing edge closer to a configuration where using empirical likelihoods during DA leads to results that are as skillful as the benchmark Gaussian likelihood specification.

Configuration-specific impacts of likelihood formulation on prior-mean error over the course of sequential DA may be useful as a proxy indicator for analysis error remaining in corresponding analysis residuals used to generate likelihoods. Between the default and modified microphysics configurations there is little difference in relative error for likelihood formulation with the exception of 2-m temperature for which the modified microphysics configuration is degraded while the default configuration experiences little change in error (Fig. 4.9c). We may offer an explanation by considering changes in aggregated increment statistics in Fig. 4.8 Here we see that the effect of empirical likelihoods over those assuming a Gaussian is to reduce the magnitude of positive increments for equivalent background innovations. Since the modifications to the NVD microphysics scheme bring about increased – and presumably more accurate – mixing ratios by lowering reflectivity at given mixing ratios and increasing innovation values, empirical likelihoods would therefore act to negate a portion of the changes induced to analyses during sequential assimilation and restore some of the surface temperature error from the default configuration. That the default and modified microphysics configurations show similar absolute differences in reflectivity error between Gaussian and empirical likelihoods (4.9b), might suggest that while the modified microphysics configuration does reduce analysis error for microphysical quantities, it is unsuccessful in doing so for reflectivity. Alternatively, the lack of a greater improvement with modified microphysics experiments may simply reflect a compensation between analysis error reductions attributable to the modified NVD scheme and those originating from the higher frequency sampling used for generating likelihoods with the default configuration.

Conversely, when comparing likelihood-induced errors from the default and modified microphysics configurations to those with a configuration modifying both microphysics and observing configuration, we see that the small impact from likelihood formulation given by the latter

compares favorably to the negative impact given by both of the former. Put in other terms, when we construct empirical likelihoods with analysis residuals that take advantage of both strategies, using these likelihoods in place of a Gaussian causes less degradation than would be caused by likelihoods from the default modeling system or a modeling system that only modified microphysics. Given the assumptions made by our method for generating likelihood functions, this is an encouraging sign that the combined microphysics and observing configuration modifications are indeed reducing analysis error and thereby increasing the accuracy of empirical distributions for observation error.

The construction of state-dependent empirical likelihoods from innovation statistics or residuals requires disentangling observation errors from other error sources, which can be very challenging for high-dimensional systems. We emphasize that our particular implementation using analysis residuals represents an initial attempt that may be improved upon with either alternative approaches to the representation of observation error, or with further improved modeling systems that build upon the technique described here. In particular, methods for the online or offline estimation of model bias may be very helpful in reducing the residual influence of analysis or background errors when sampling observational error distributions. Conditioning observation uncertainty estimates on model prognostic variables, such as hydrometeor mixing ratios, in place of observation-space priors for radar reflectivity may also prove helpful in more accurately describing state dependence.

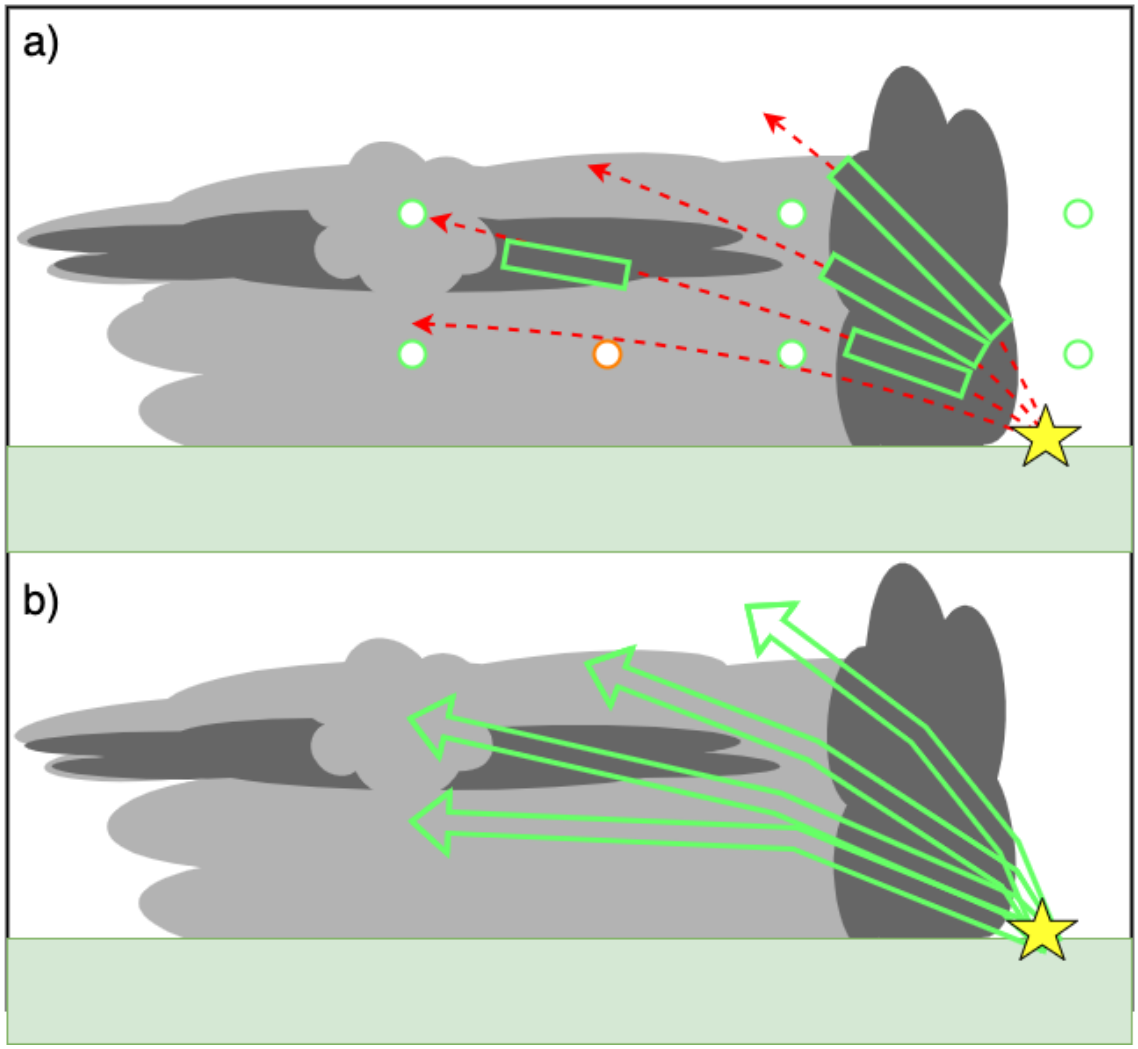

Figure 4.1: Regions highlighted in green indicate the distribution of radar observations generated using the standard observing configuration (a) and alternative observing configuration (b). Rectangular regions contain multiple observations gridded at high resolution from a volume scan pattern. Shading indicates areas of elevated (dark grey) and moderate (light grey) returns in a typical squall-line type system. Circular regions indicate individual clear-air observations derived from radar returns. The yellow star indicates the location of a hypothetical radar. Regions highlighted in orange for the top panel indicate observations generated in the observing configuration currently implemented for the NSSL WoFS, but not in the standard observing configuration used for experiments.

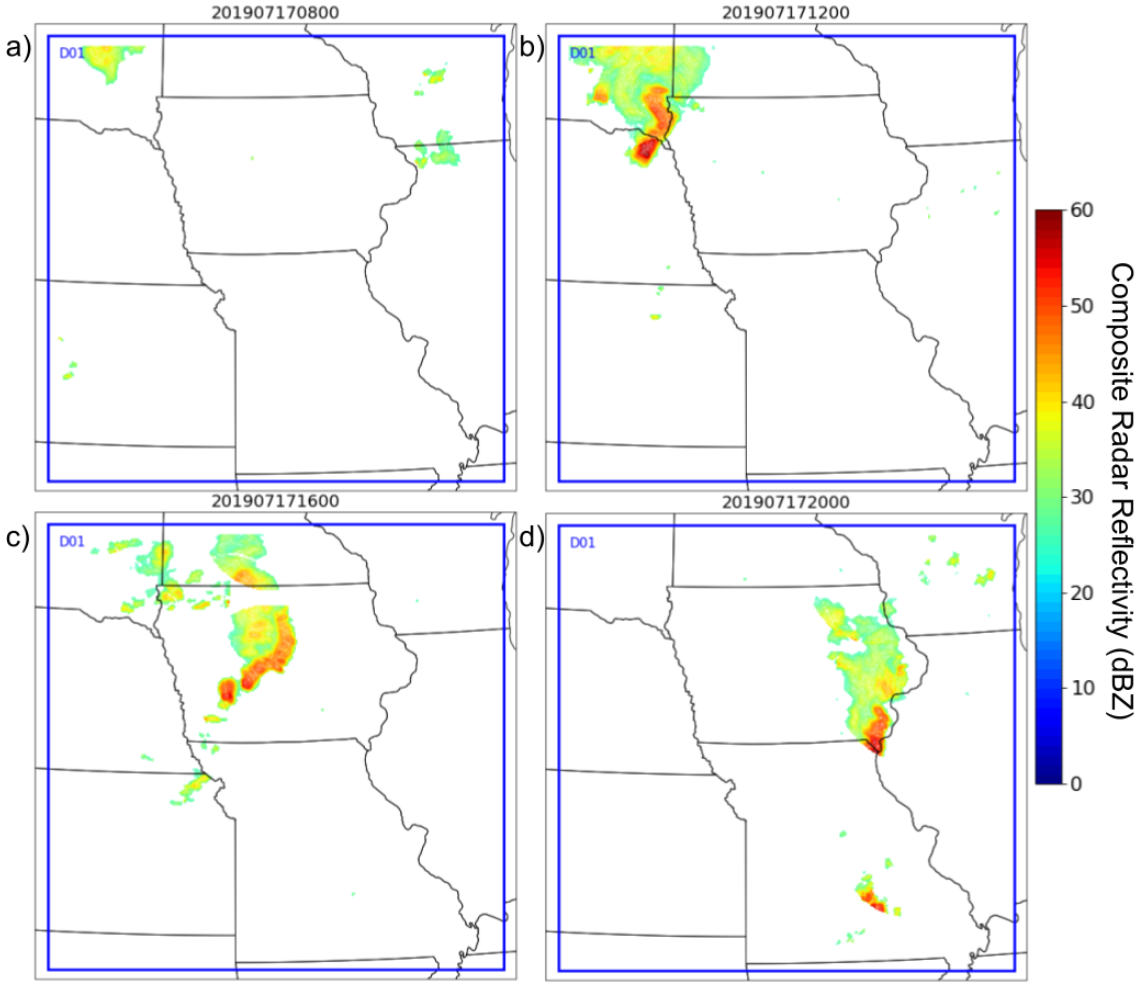

Figure 4.2: Composite reflectivity for squall-line case event occuring on July 17th 2019. Reflectivity shown for (a) 0800 UTC, (b) 1200 UTC, (c) 1600 UTC, and (d) 2000 UTC.

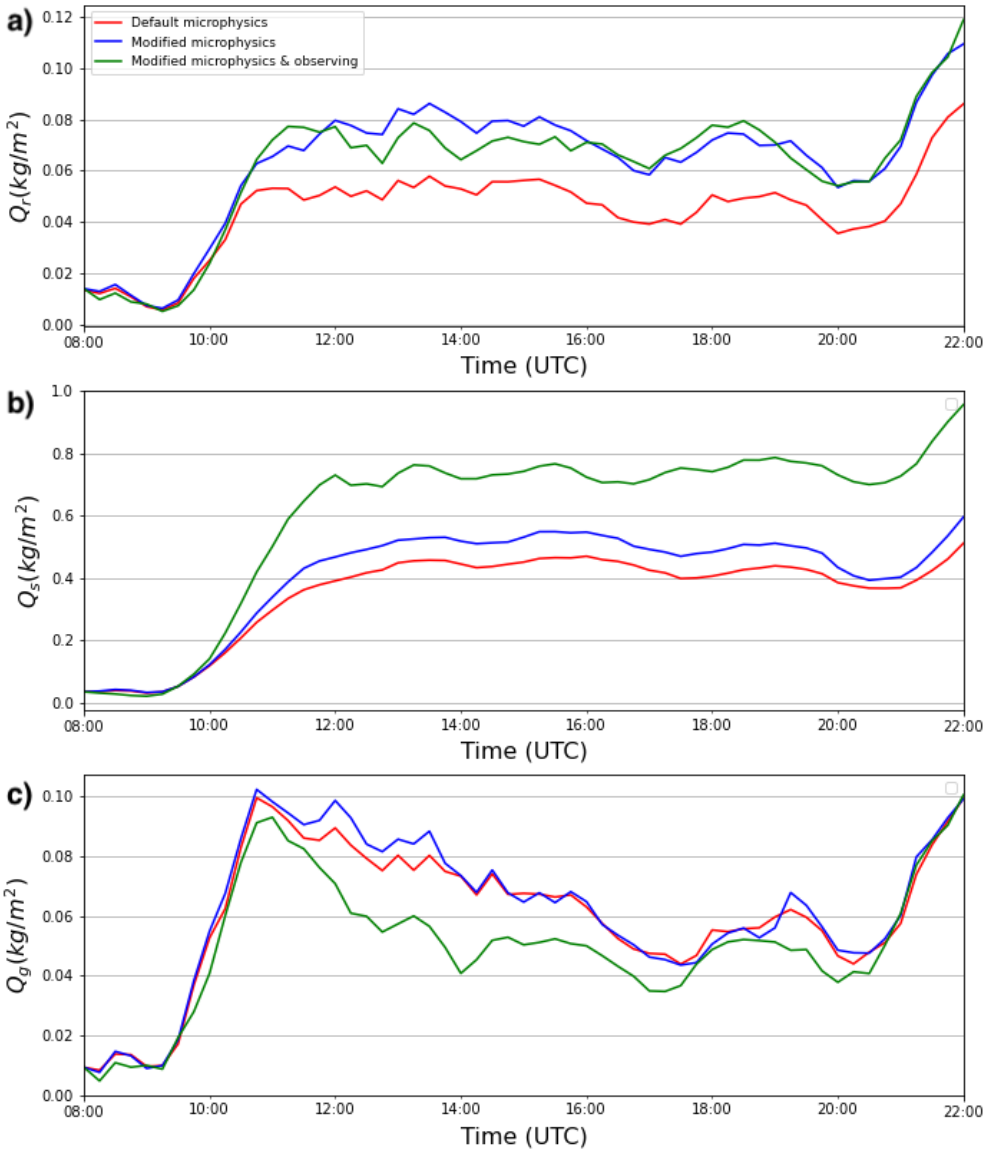

Figure 4.3: Time series plots of domain-averaged, column-integrated hydrometeor mass for (a) rain, (b) snow, and (c) graupel. Results shown for default microphysics (red), modified microphysics (blue) and modified microphysics & observing (green).

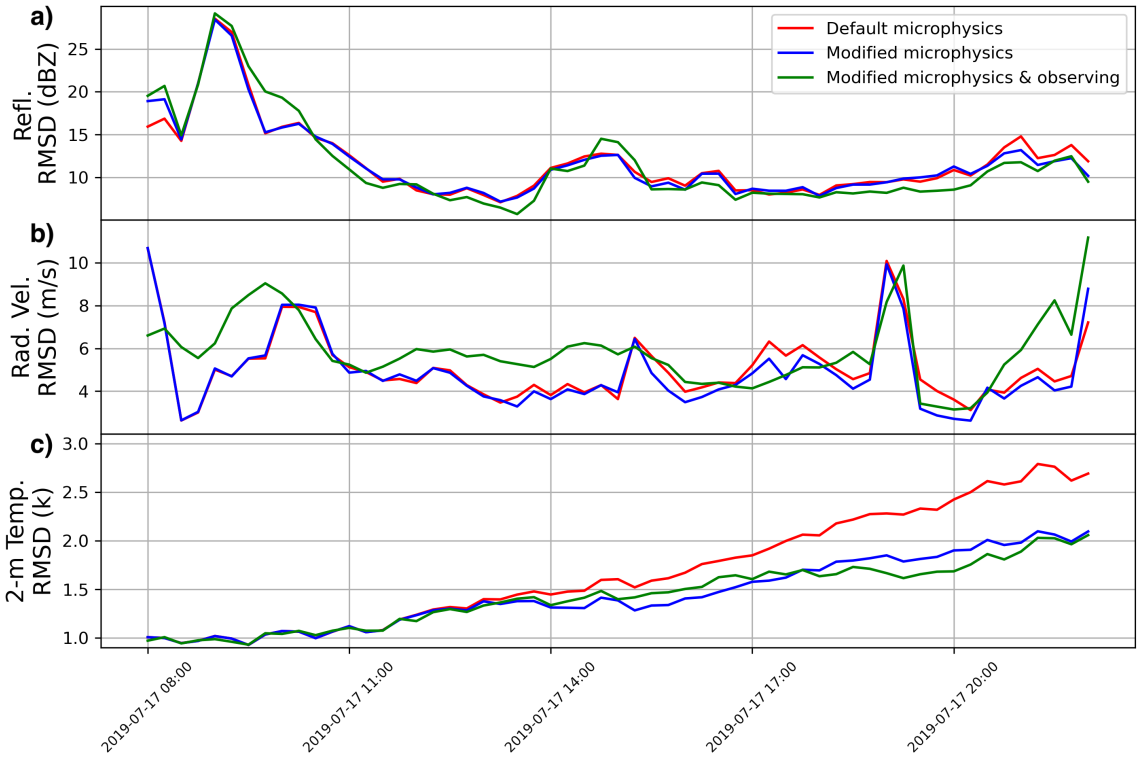

Figure 4.4: Prior-mean RMSD for radar reflectivity (a), radial velocity (b) and 2-m temperature (c) during the period of sequential DA. Results shown for default microphysics (red), modified microphysics (blue) and modified microphysics & observing (green).

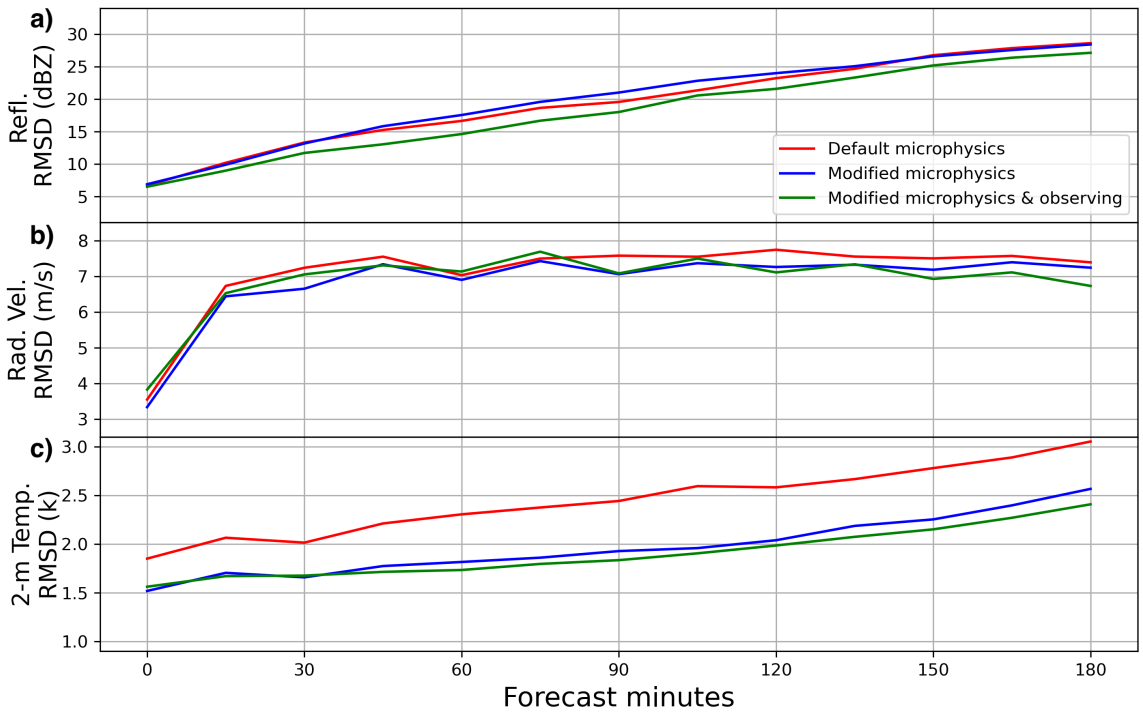

Figure 4.5: Average forecast-mean RMSD for radar reflectivity (a), radial velocity (b) and 2-m temperature (c) for forecast leadtimes from 0 to 180 minutes. Results shown for default microphysics (red), modified microphysics (blue) and modified microphysics & observing (green).

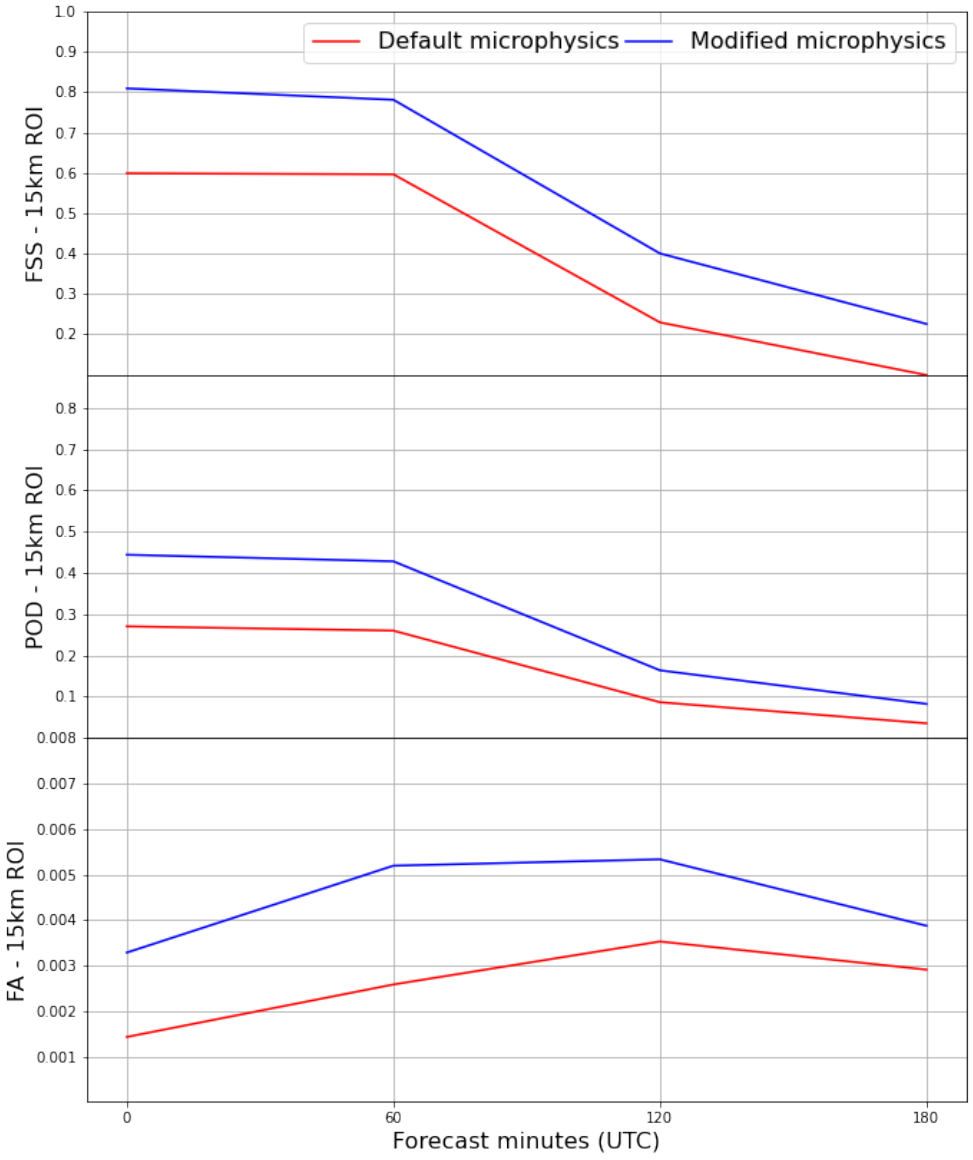

Figure 4.6: Averaged scores for FSS (top row), POD (middle row), and FAR (bottom row), considering the occurrence of hourly precipitation over 5 mm for forecast lead times from 0 to 180 minutes. Results shown for default (red) and modified (blue) microphysics configurations.

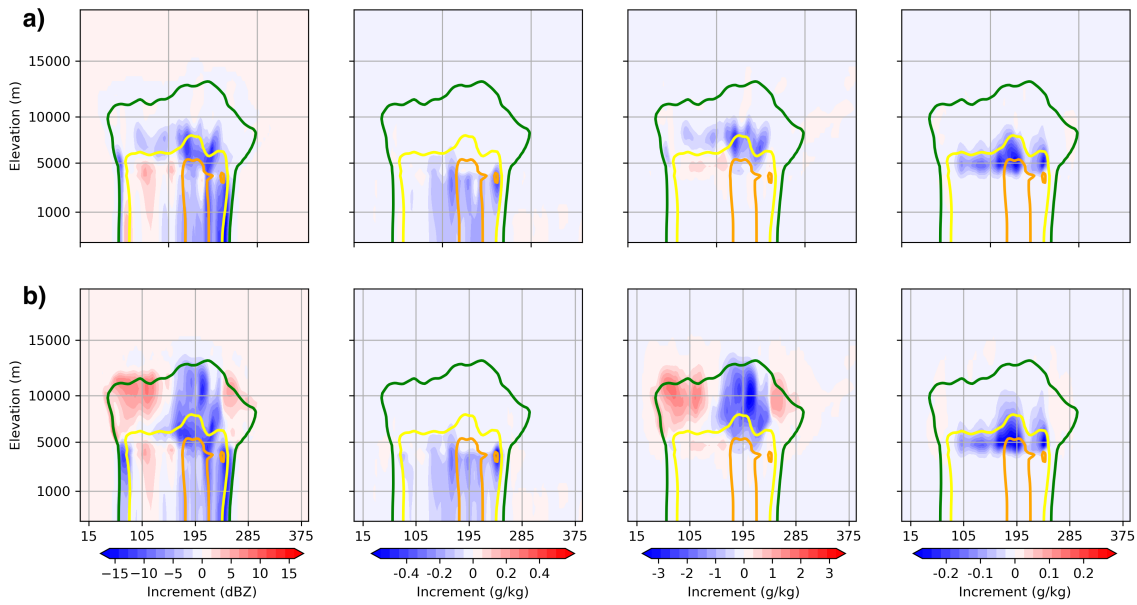

Figure 4.7: Vertical cross-sections showing analysis increments produced from identical priors with (a) standard observing configuration and (b) alternative observing configuration. Increments shown for radar reflectivity (leftmost column), rain mixing ratio (second to left column), snow mixing ratio (second to right column), and graupel mixing ratio (rightmost column). For all increment panels, contours of prior reflectivity are shown at 5 (green), 25 (yellow), and 35 (orange) dBZ.

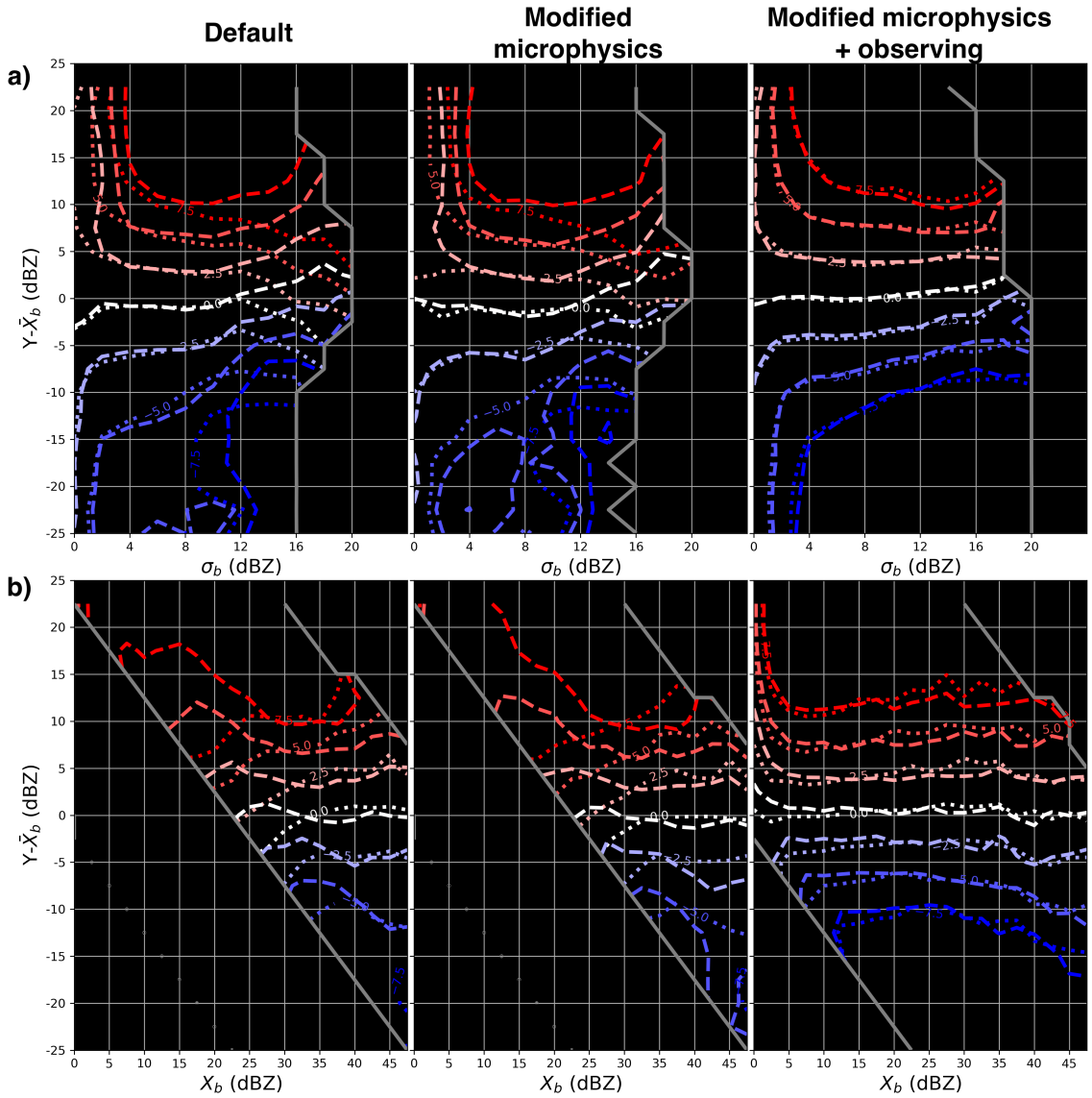

Figure 4.8: Contours of analysis increments plotted as a function of innovation and prior standard deviation (a) and as a function of innovation and observed value (b). Likelihood function formulation is indicated by line style — dotted for Gaussian, and dashed for empirical. Modeling system configuration for likelihood experiments is indicated by column title.

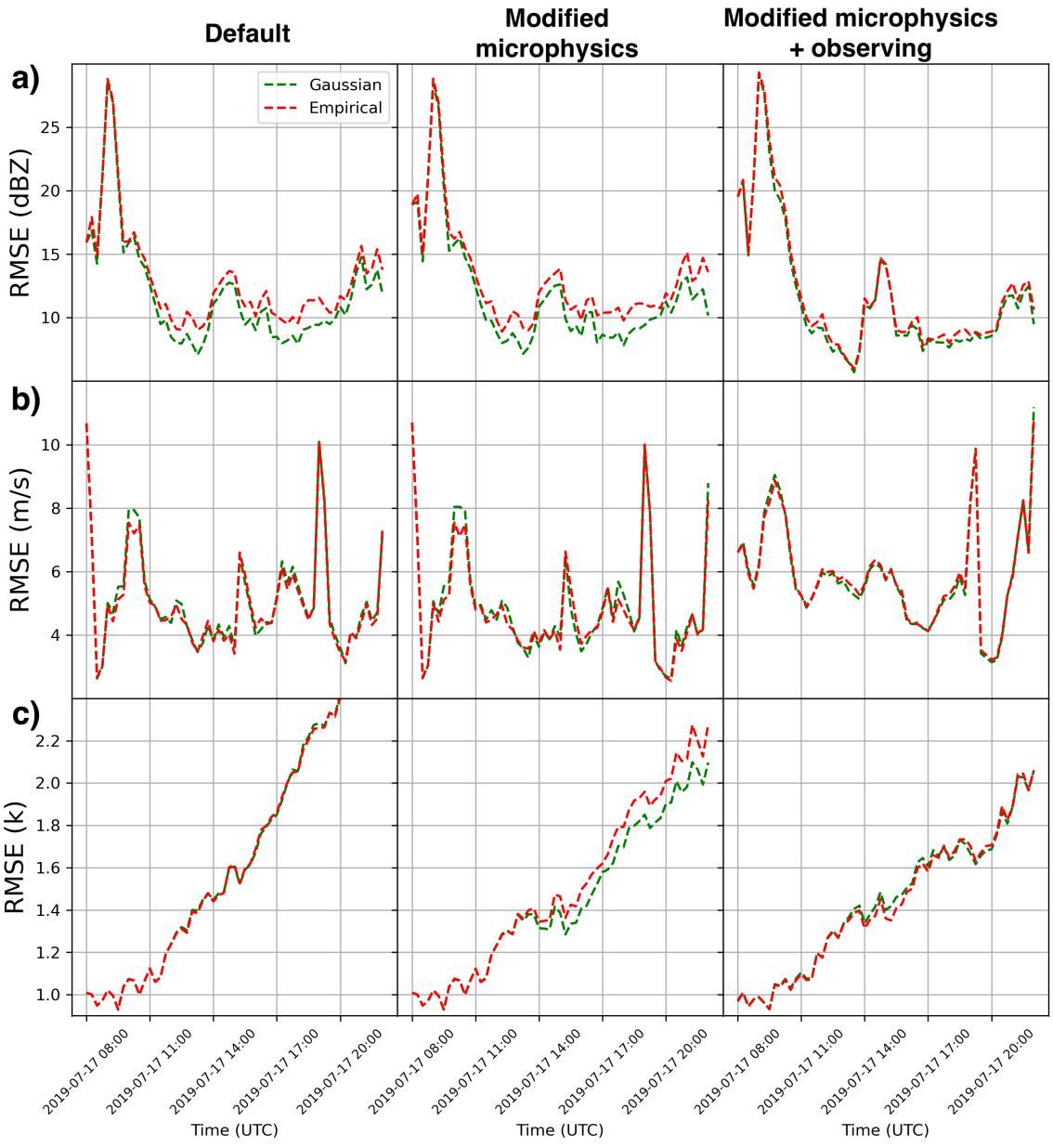

Figure 4.9: Prior-mean RMSD for radar reflectivity (a), radial velocity (b) and 2-m temperature (c) during the period of sequential DA. Results shown for Gaussian likelihood experiments (red) and experiments using empirical likelihoods (green). Modeling system configuration for likelihood experiments is indicated by column title.

Chapter 5: Conclusions and future implications

5.1 Findings from evaluated assimilation and observing strategies

The work considered in this dissertation has been aimed first and foremost at improving the accuracy of state estimates and forecasts produced by DA for moist-convective regimes. In seeking these improvements, we have built upon the already quite advanced capabilities offered by the NSSL WoFS by evaluating an array of strategies that systematically remove assumptions made by DA algorithms, implement greater observational constraint on moist-convective structures, and address the impact of model and observational errors on filter updates to hydrometeor variables. Our findings from sequential DA, forecast, and other experiments shed light on challenges hindering state estimation and prediction for moist-convective systems, and how they relate to uncertainty generated as a byproduct of highly nonlinear dynamics.

In chapter two, containing work published in McCurry et al. (2023), we provide a comprehensive inter-comparison for DA with an EnKF and two variants of the LPF applied to selected moist-convective cases. We find that assimilating with the LPF produces posterior quantities for microphysical variables that are more consistent with model climatology than comparable quantities from an EnKF, which we attribute to the removal of bias induced by Gaussian assumptions of prior form. These differences are significant enough to impact the dynamic evolution of convective systems through significant modifications to rainwater column mass and precipita-

tion rates. Lower precipitation rates with the LPF ultimately lead to weaker cold pools that are much less extensive than those produced when assimilating with an EnKF. Further experiments reveal that comparative differences in forecast verification scores between the EnKF and LPF change substantially when NVD microphysical parameterizations are replaced with an alternative double-moment scheme. These results are indicative of model bias in the NVD scheme that impacts prior estimates for microphysical quantities. Deficient priors disproportionately degrade state estimates for PF based methods that cannot rely on regression when observed values are not contained by model spread.

Chapter three deviates in some respects from the consideration of DA algorithms in chapters two and four because it uses an OSSE framework with an EnKF to evaluate various configurations of a statewide mesonet system on their ability to improve state estimates and predictions for moist-convective cases. Work from this chapter has been accepted for publication in Mccurry et al. (in press). We find that the assimilation of mesonet observations for near-surface wind and temperature produces definitive improvements to analysis fields below 1000 m that are mediated by the degree of uncertainty prescribed by the OSSE framework through perturbations to LBCs and parameterization error for microphysical processes. Experiments prescribing higher amounts of uncertainty show larger improvements to state estimates than subsequent experiments with restrained perturbations and parameterization error. For both experimental configurations, the impact of observing density on analyses appears limited by a saturation effect that caps local improvement to analyses past a minimal density of observing stations. Regardless of uncertainty, mesonet impact on forecast verification is inconclusive and strongly variable across verification metrics. Both the density-saturation effect for state-estimation error and the lack of improvement to forecasts from assimilating mesonet observations may ultimately be attributed to the effect of

moist-convective processes on resolved uncertainty in the lower boundary layer. Although our modeling system is able to capture uncertainty in the general placement of cold pools and precipitation, it is not nearly as capable in resolving smaller-scale uncertainty in the exact outline of cold pools and gust fronts. This limits the ability for surface-based observations to inform state estimates, especially when estimates have already been informed by nearby observations, or when boundary conditions are sufficiently constrained to limit displacement errors in prior estimates. Likewise, we conclude that strong non-linearities induced by parameterizations and resolved dynamics at lower levels are able to quickly remove information added by mesonet observations, limiting their ability to reduce forecast error.

In chapter four, we expand upon the results of McCurry et al. (2023) by implementing modifications to the NVD microphysical scheme intended to address deficiencies in precipitation efficiency mediated by filter increments, as well as an observing strategy for radar reflectivity that increases the availability of observational data for the trailing stratiform region by dispensing with a separate clear-air observation type. Additional experiments are performed exploiting the utility offered by the LPF algorithm for implementing likelihoods of arbitrary form. These experiments are used to double-effect in investigating the impact of removing Gaussian assumptions for likelihood function form, and in gauging the efficacy of observing and microphysical strategies used to generate empirical likelihoods. We find that changes to the NVD scheme can correct large deficiencies in precipitation rate previously found in the LPF, and lead to better state estimates and forecasts for near-surface temperature which we tie to improved latent heating and consequent generation of larger cold pools. Likewise, the homogeneous processing of radar reflectivity returns leads to reductions in prior-mean and forecast RMSD values that suggest significant improvements to the depiction of hydrometeors in state estimates, likely from

better constraint on their spatial distribution and mixing ratio at upper levels. Although experiments with empirical likelihoods did not reduce prior-mean error over the cycling period, they did produce characteristic changes to aggregate analysis increment that show a more physically justifiable state dependence. Prior-mean errors over the period of sequential DA provide circumstantial evidence that changes to the microphysical scheme and observing strategy for reflectivity reduce analysis errors enough to improve observational error PDFs generated from sampled analysis residuals. Further improvements to assimilation strategies leading to further reductions in analysis error may lead to positive impacts from relaxing Gaussian assumptions for likelihood.

5.2 Extensions to introduced strategies and future applications

While our experiments have successfully demonstrated ways in which the LPF provides qualitative advantages for moist-convective state estimation, and have provided means to correct some of its deficiencies, there remains significant work to be done in applying the LPF to a modeling system framework capable of maximizing the utility of nonparametric state estimation and producing conclusive improvement to moist-convective forecasts over existing approaches. A possible extension to strategies developed in this dissertation could involve the implementation of modeling system configurations feature large (1000+) ensembles that allow for the relaxation of localization strategies in the LPF algorithm. Localization is typically necessary where constraints such as sampling error from small ensemble size produce spurious cross-covariances between spatially separated variables. However, ignoring cross-covariances with localization can distort dynamical balances and other physical relationships in filter updates, and it has been shown that localization with EnKFs introduces imbalance to wind and mass fields when using

small localization radii Greybush et al. (2011); Kepert (2009). Increasing the ensemble size within a modeling system can be expected to reduce sampling error, and therefore allows for a corresponding increase in optimal localization radii. This presents advantages beyond those gained from sampling error reductions alone and would synergize with previous approaches by further reducing introduced biases that cause state estimates to deviate from solutions consistent with model physics. Large ensemble experiments would also provide the simultaneous opportunity to conduct observation sensitivity experiments for moist-convective cases that could identify specific observation types or locations that disproportionately benefit from relaxed localization.

Future directions may also include the continuation of work with empirical likelihoods developed in chapter four. The successful removal of Gaussian likelihood assumptions may require the use of alternative methods for generating nonparametric, state-dependent observational error PDFs, either with de-convolution methods Hu et al. (2024) or improvements to the current method using analysis residuals. The latter approach may involve progressively removing model error signals contained in PDFs sampled from analysis residuals by iteratively generating new likelihoods that are used in subsequent sequential DA experiments that are then used again in the sampling process. Such a method would be reminiscent of iterative procedures for fixing observational error variances in Desroziers et al. (2005). Alternatively, the approach with analysis residuals may be improved using methods that apply corrections to state estimates based on an online or offline estimation of model bias. As a final consideration, we note that the utility of our empirical likelihood functions may be greatly enhanced by broadening the state dependence of empirical functions from prior reflectivity only towards a wider range of microphysical variables. Since observational errors may reasonably be considered to vary jointly across a high dimensional space the current approach may be limited by sampling only the marginal distribution.

We hypothesize that incorporating factors such as mixing ratios for discrete hydrometeor species would allow empirical likelihoods to improve the accuracy of state estimates by better-informing analysis increments.

As a concluding thought, we note that the assimilation and observing strategies discussed in this dissertation may be beneficial for training generative AI models for the prediction of severe weather threats. AI models for weather prediction produce forecasts without a traditional numerical model by applying machine-learning techniques to find statistical relationships between input and output variables from training datasets. Existent AI weather-prediction models differ on their treatment–or lack thereof–for the DA process, but the overwhelming majority rely on training datasets that include reanalysis data, often from ERA5 Liu et al. (2024). The accuracy of training data, and therefore of reanalyses, can greatly affect the skill of resulting AI weather predictions Whang and Lee (2020). Although the use of AI models for convective-scale weather threats is an active area of research, there appears to be a lack of a comprehensive regional mesoscale reanalysis dataset that may fill a role similar to ERA5 for global medium-range forecasts. This role may currently be best approximated with the real-time mesoscale analysis (RTMA) and unrestricted mesoscale analysis (URTMA) products generated by NCEP, however, these products are currently generated only for two-dimensional fields and rely on products such as the HRRR and Rapid Refresh (RAP) system, that may feature greater error at small-scales than the WoFS due to less frequent DA updates and indirect methods for the assimilation of radar data Morris et al. (2020). A significant update to these products is planned in the form of a 3D-RTMA, which will rely on the new Rapid Refresh Forecast System (RRFS) to produce high-resolution three-dimensional analyses Ge et al. (2023). While this would represent a significant improvement over existing datasets, the Rapid Refresh Forecast System Data Assimilation System (RDAS)–which

models uncertainty in the RRFS–relies on similar convection-allowing ensemble techniques used in the WoFS and HRRRE Alexander and Carley (2024). We note that recent work at NSSL has foregone the use of dedicated analysis products by directly training AI models with output from WoFS Flora and Potvin (2024). Both analyses produced for a 3D-RTMA, or directly within modeling systems like the WoFS, may benefit from observing strategies specifically intended to constrain analysis error for moist-convective events and assimilation strategies that remove errors induced by Gaussian assumptions. Further experiments explicitly aimed at AI prediction may be useful in determining if state estimates produced using strategies described in this dissertation lead to increased accuracy for AI forecasts.

A.0.1 OSSE tuning and regional breakdown of mesonet impact

Fig. A.1 shows a breakdown of binary verification scores associated with OSSE tuning configurations intended to simulate the capabilities of contemporary NWP systems. Scores are calculated from forecasts initialized for the case event occurring on April 13th 2020.

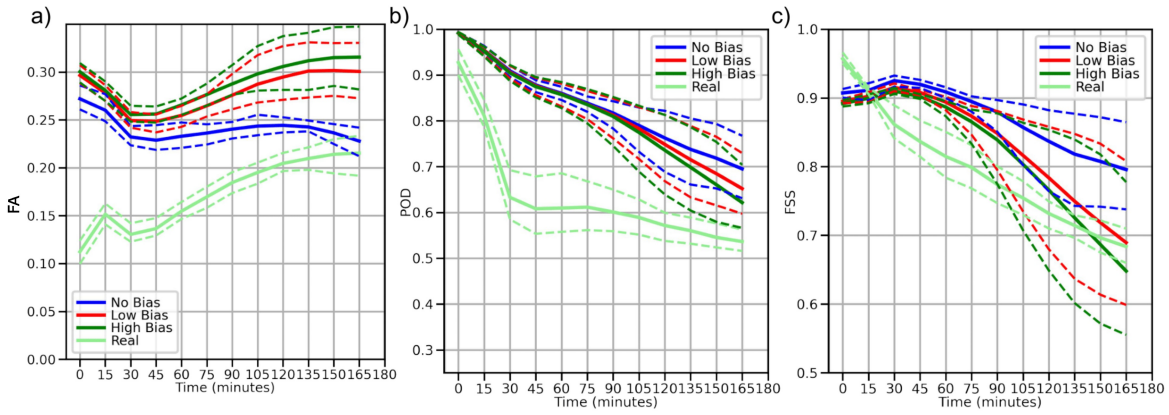

Figure A.1: (a) False alarm rate, (b probability of detection, and (c) fractions skill score for forecasts with a real test case, and for corresponding OSSE forecasts without bias, with low amplitude applied bias, and with high amplitude applied bias

Figs. A.2 and A.3 show a breakdown of near-surface-averaged mesonet impact on posterior error for potential temperature during the period of sequential cycling for standard-uncertainty and constrained-uncertainty experiments respectively. Individual panels consider data in 2-h intervals commencing 3-hrs after ensemble initialization.

Fig. A.4 shows the average density of pre-existing surface observations considered in a 15-minute assimilation cycle within out OSSE modeling system.

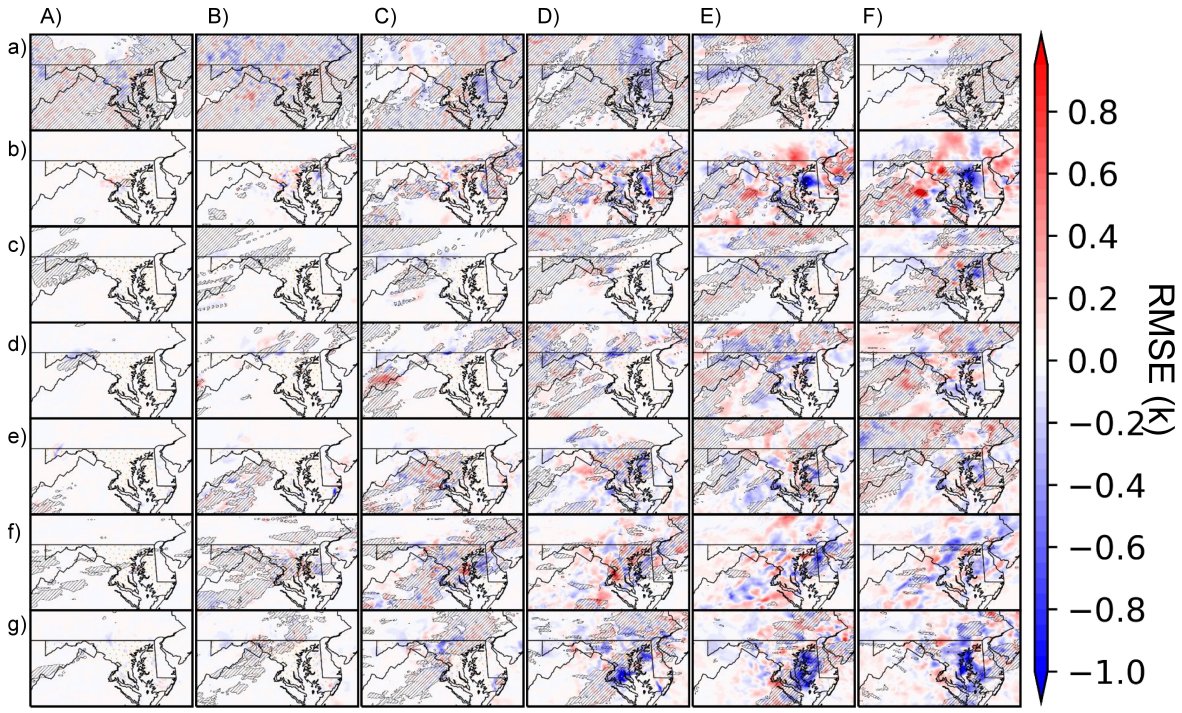

Figure A.2: Posterior RMSE for near-surface potential temperature field for case events beginning on (a) April 13th 2020 (b) August 12th 2020 (c) September 3rd 2020 (d) July 17th 2021 (e) June 8th 2022 (f) July 2nd 2022 (g) July 16th 2022, for 2-hr intervals starting (A) 3-hrs (B) 5-hrs (C) 7-hrs (D) 9-hrs (E) 11-hrs (F) 13-hrs after ensemble initialization. Occurrence of composite reflectivity above 30 dBZ during interval indicated with cross-hatching. Valid for standard-uncertainty experiments.

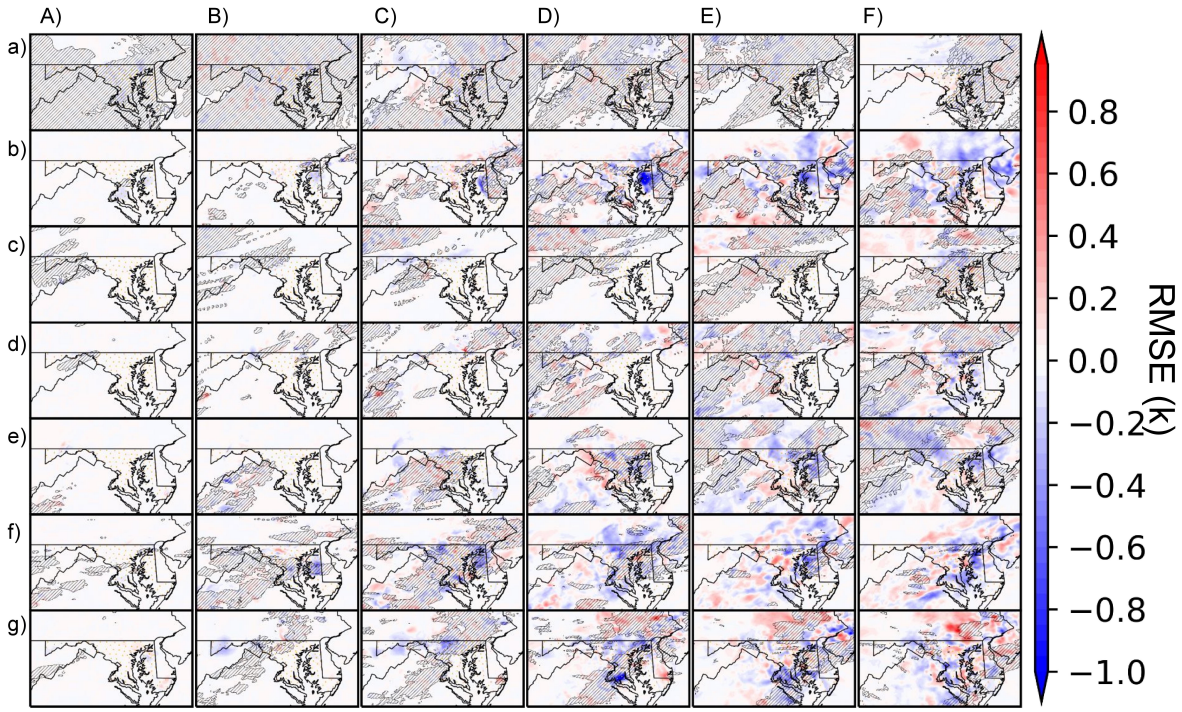

Figure A.3: Same as Fig. A.2 but for constrained-uncertainty experiments.

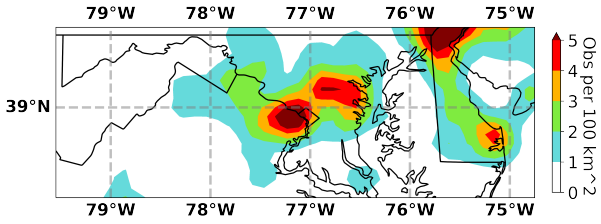

Figure A.4: Density per 15-minute assimilation window of surface-based observations publicly available from MADIS

B.0.2 Empirical likelihood functions calculated with technique using analysis residuals

Figure B.1 shows state-dependent observational-error PDFs used for the implementation of empirical likelihood functions in chapter four.

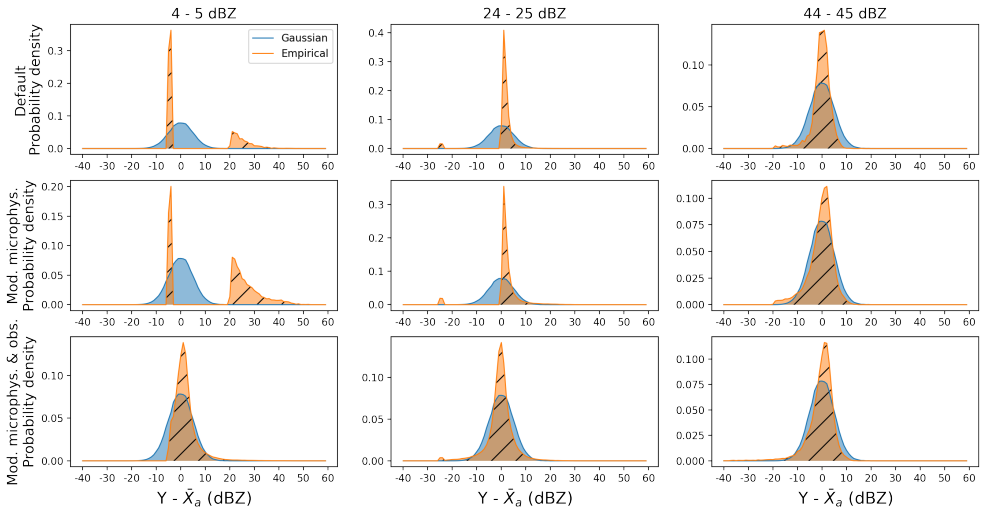

Figure B.1: Observational error PDFs calculated for the default (top), modified microphysics (middle) and modified microphysics & observing (bottom) experiment configurations. Columns reflect ranges of posterior-mean reflectivity sampled by analysis-residual technique to account for state dependence: 4-5 dBZ (left), 24-25 dBZ (center), and 44-45 dBZ (right).

Bibliography

Alexander, C. and J. Carley (2024). The rapid refresh forecast system:looking beyond the first operational version. *American Meteorological Society Annual Meeting*, 2024, Baltimore, MD, Amer. Meteorological Soc., 4B.1.

Anderson, J. and N. Collins (2007). Scaleable implementation of ensemble filter algorithms for data assimilation. *Journal of Atmospheric and Oceanic Technology 24*, 1452–1464.

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev. 129*, 2884–2903.

Bengtsson, T., 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. 2*, 316–334.

Bertsimas, D. and J. Tsitsiklis (1993). Simulated annealing. *Statist. Sci. 8*, 10–15.

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 3*, 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.

Choate, J., A. Clark, D. Imy, and P. Skinner (2018). First demonstration of the nssl experimental warn-on-forecast system as part of the 2017 spring experiment. *Eighth Conf. on Transition of Research to Operations*, Austin, TX, Amer. Meteor. Soc., 1194, https://ams.confex.com/ams/98Annual/webprogram/Paper335289.html.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *QJRMS 131*, 3385–3396.

Dixon, M., G. Meymaris, S. Ellis, and J. Hubbert (2007). Improving nexrad data - data qualty algorithm progress. *NCAR Technical Note*, NCAR/TN–475+STR.

Dowell, D., C. Alexander, E. James, S. Weygandt, S. Benjamin, G. Manikin, B. Blake, J. Brown, J. Olson, M. Hu, T. Smirnova, T. Ladwig, J. Kenyon, R. Ahmadov, D. Turner, J. Duda, and T. Alcott (2022). The high-resolution rapid refresh (hrrr): An hourly updating convection-allowing forecast model. part i: Motivation and system description. *Weather and Forecasting 37*, 1371–1395.

Dowell, D. and L. Wicker (2009). Additive Noise for Storm-Scale Ensemble Data Assimilation. *Journal of Atmospheric and Oceanic Technology 26*, 911–927.

Errico, R. M. and N. Privé (2018). Some general and fundamental requirements for designing observing system simulation experiments [white paper]. *WMO Rep. WWRP*.

Errico, R. M., R. Yang, N. C. Privé, K. Tai, R. Todling, and M. E. S. amd J. Guo (2013). Development and validation of observing-system simulation experiments at nasa's global modeling and assimilation office. *Quarterly Journal of the Royal Meteorological Society 139*, 1162–1178.

Fiebrich, C. A., K. Brinson, R. Mahmood, S. Foster, M. Schargorodski, N. Edwards, C. Redmond, J. Atkins, J. Anderson, and X. Lin (2020). Toward the standardization of mesoscale meteorological networks. *Journal of Atmospheric and Oceanic Technology 37*, 2033–2049.

Flora, M. and C. Potvin (2024). Wofscast: A machine learning model for predicting thunderstorms at watch-to-warning scales. *ESS Open Archive*.

Gallo, B. and coauthors (2017). Breaking new ground in severe weather prediction: The 2015 noaa/hazardous weather testbed spring forecasting experiment. *Weather and Forecasting 32*, 1541–1568.

Gao, Y., L. Huyan, and B. Liu (2024). Improving the gaussianity of radar reflectivity departures between observations and simulations using symmetric rain rates. *Atmos. Meas. Tech. 17*, 4675–4686.

Gaspari, G. and S. E. Cohn (1999). Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. 125*, 723–757.

Gastaldo, T., V. Poli, C. Marsigli, and P. A. T. Paccagnella (2018). Data assimilation of radar reflectivity volumes in a letkf scheme. *Nonlinear Processes in Geophysics 25*, 747–764.

Ge, G., T. Ladwig, M. Pondeca, M. Hu, E. Colon, A. Gibbs, M. Morris, G. Zhao, C. Hartsough, M. Rancic, J. Carley, S. Weygandt, and C. Alexander (2023). Towards a digital twin of the current atmosphere: a 3d/4d real-time mesoscale analysis system at noaa. *AGU Fall Meeting*, 2024, San Francisco, CA, Amer. Geophysical Union, IN51A–08.

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. Mod. Earth Sys. 9*, 2635–2653.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). Novel approach to nonlinear/non-Gaussian state estimation. *Radar and Signal Processing 140*, 107–113.

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.

Ha, S. and C. Snyder (2014). Influence of surface observations in mesoscale data assimilation using an ensemble kalman filter. *Mon. Wea. Rev. 142*, 1489–1508.

Hodyss, D. and M. Morzfeld (2023). How sampling errors in covariance estimates cause bias in the kalman gain and impact ensemble data assimilation. *Mon. Wea. Rev. 151*, 2413–2426.

Hu, C., L. Peter, and G. Alan (2024). A non-parametric way to estimate observation errors based on ensemble innovations. *QJRMS 150*, 2296–2315.

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. Geo. Research. Atmos. 113*.

James, E. P., C. R. Alexander, D. C. Dowell, S. S. Weygandt, S. G. Benjamin, G. S. Manikin, J. M. Brown, J. B. Olson, M. Hu, T. G. Smirnova, T. Ladwig, J. S. Kenyon, and D. D. Turner (2022). The high-resolution rapid refresh (hrrr): An hourly updating convection-allowing forecast model. part ii: Forecast performance, weather and forecasting,. *Weather and Forecasting 37*, 1397–1417.

Janjić, T., N. Bormann, M. Bocquet, J. A. Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, and P. Weston (2018a). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society 144*, 1257–1278.

Janjić, T., N. Bormann, M. Bocquet, J. A. Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, and P. Weston (2018b). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society 144*, 1257–1278.

Jimenez, P. A., J. Dudhia, J. F. Gonzalez-Rouco, J. Navarro, and J. P. Montavez (2012). A revised scheme for the wrf surface layer formulation. *Mon. Wea. Rev. 140*, 898–918.

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 Graduate College*.

Jones, T., P. Skinner, K. Knopfmeier, E. Mansell, P. Minnis, R. Palikonda, and W. S. Jr. (2018). Comparison of cloud microphysics schemes in a warn-on-forecast system using synthetic sattelite objects. *Weather and Forecasting 6*, 1681–1708.

Karspeck, A. (2016). An ensemble approach for the estimation of observational error illustrated for a nominal 1° global ocean model. *MWR 144*, 1713–1728.

Kepert, J. D. (2009). Covariance localisation and balance in an ensemble Kalman filter. *Q. J. R. Meteorolog. Soc. 135*, 1157–1176.

Khare, S. P. and J. L. Anderson (2006). An examination of ensemble filter based adaptive observation methodologies. *Tellus A: Dynamic Meteorology and Oceanography 58*, 179–195.

Knopfmeier, H. K. and J. D. Stensrud (2013). Influence of mesonet observations on the accuracy of surface analyses generated by an ensemble kalman filter,. *Weather and Forecasting 28*, 815–841.

Leitman and Thompson (2019). Mesoscale discussion 851. *https://www.spc.noaa.gov/products/md/2019/md0851.html*.

Li, H., Y. Huang, Y. Luo, H. Xiao, M. Xue, X. Liu, and L. Feng (2023). Does "right" simulated extreme rainfall result from the "right" representation of rain microphysics? *QJRMS 149*, 3220–3249.

Liu, C., K. Hsu, M. S. Peng, D. Chen, P. Chang, L. Hsiao, C. Fong, J. Hong, K. Lu, C. Chen, and H. Kuo (2024). Evaluation of five global ai models for predicting weather in eastern asia and western pacific. *npj Climate and Atmospheric Science 7*.

M., A., M. Gobashy, M. H. Khalil, and M. Abdrabou (2019). A complete model parameter optimization from self-potential data using hale algorithm. *Journal of Applied Geophysics 170*.

Majcen, M., P. Markowski, Y. Richardson, D. Dowell, and J. Wurman (2008). Multipass objective analyses of doppler radar data. *J. Atmos. Oceanic Technol. 25*, 1845–1858.

Mansell, E. R. (2010). On sedimentation and advection in multimoment bulk microphysics. *J. Atmos Sci. 70*, 2032–2050.

Marquis, J., Y. Richardson, P. Markowski, D. Dowell, J. Wurman, K. Kosiba, P. Robinson, and G. Romine (2014). An investigation of the goshen county, wyoming, tornadic supercell of 5 june 2009 using enkf assimilation of mobile mesonet and radar observations collected during vortex2. part i: Experiment design and verification of the enkf analyses. *Mon. Wea. Rev. 142*, 530–554.

McCurry, J. and J. Poterjoy (2024). Optimizing numerical weather prediction utility of the maryland mesonet with observing system simulation experiments. *WAF*.

McCurry, J. M., J. Poterjoy, K. Knopfmeier, and L. Wicker (2023). An Evaluation of Non-Gaussian Data Assimilation Methods in Moist Convective Regimes. *Mon. Wea. Rev.*.

Morris, M., J. Carley, E. Colon, M. Pondeca, and S. Levine (2020). A quality assessment of the real-time mesoscale analysis (rtma) for aviation. *Weather and Forecasting 35*, 977–996.

Necker, T., S. Geiss, M. Weissmann, J. Ruiz, T. Miyoshi, and G. Lien (2019). A convective-scale 1,000-member ensemble simulation and potential applications. *QJRMS 146*, 1423–1442.

Pannekoucke, O., L. Berre, and G. Desroziers (2008). Background-error correlation length-scale estimates and their sampling statistics. *QJRMS 134*, 497–508.

Posselt, D. J. (2016). A bayesian examination of deep convective squall-line sensitivity to changes in cloud microphysical paramete rs. *Journal of the Atmospheric Sciences 73*(2), 637 – 665.

Poterjoy, J. (2016). A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev. 144*, 59–76.

Poterjoy, J. (2022a). Implications of multivariate non-gaussian data assimilation for multiscale weather prediction. *Mon. Wea. Rev. 150*, 1475–1493.

Poterjoy, J. (2022b). Regularization and tempering for a moment-matching localized particle filter. *Quart. J. Roy. Meteor. Soc.*, In press.

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.

Poterjoy, J., 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.

Potvin, C., P. Skinner, K. Hoogewind, M. Coniglio, J. Gibbs, A. Clark, M. Flora, A. Reinhart, J. Carley, and E. Smith (2020). Assessing systematic impacts of pbl schemes on storm evolution in the noaa warn-on-forecast system. *Monthly Weather Review 148*, 2567–2590.

Privé, N. and R. M. Errico (2013). The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment. *Tellus A: Dynamic Meteorology and Oceanography 65*.

Ramanathan, A., A. Satyanarayana, and M. Mandal (2019). Theoretical predictability limits of spatially anisotropic multifractal processes: Implications for weather prediction. *Earth and Space Science 6*, 1067–1080.

Ramanathan, A. and A. N. V. Satyanarayana (2021). Satellite-based estimate of intrinsic predictability limits at convective scales over northeast india. *Earth and Space Science 8*.

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.

Rotunno, R., J. B. Klemp, and M. L. Weisman (1988). A theory for strong, long-lived squall lines. *JAS 45*, 463–485.

Ruckstuhl, Y. M. and T. Janjic (2018). Parameter and state estimation with ensemble kalman filter based algorithms for convective-scale applications. *Quarterly Journal of the Royal Meteorological Society 144*, 826–841.

Shaw, R. (1981). Strange attractors, chaotic behavior, and information flow. *Zeitschrift für Naturforschung A 36*, 80–112.

Siriwardene, N. R. and B. J. C. Perera (2006). Selection of genetic algorithm operators for urban drainage model parameter optimisation. *Mathematical & Computer Modelling 44*, 415–429.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, W. Wang, J. G. Powers, M. G. Duda, D. M. Barker, and X.-Y. Huang (2019). A description of the advanced research wrf version 4. *NCAR Tech. Note*.

Skinner, P., D. Wheatley, K. K. and A. Reinhart, J. Choate, T. Jones, G. C. and D. Dowell, C. Alexander, T. Ladwig, L. Wicker, and P. H. and P. Minnis and R. Palikonda (2018). Object-based verification of a prototype warn-on-forecast system. *Weather and Forecasting 33*, 1225–1250.

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.

Smith, P. (1984). Equivalent radar reflectivity factors for snow and ice particles. *Journal of Applied Meteorology and Climatology 23*, 1258–1260.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson (2008). Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev. 136*, 4629–4640.

Sobash, R. A. and J. D. Stensrud (2015). Assimilating surface mesonet observations with the enkf to improve ensemble forecasts of convection initiation on 29 may 2012. *Monthly Weather Review 143*, 3700–3725.

Sobash, R. A. and L. Wicker (2015). On the impact of additive noise in storm-scale enkf experiments. *Monthly Weather Review 143*, 3067–3086.

Starzec, M., C. Homeyer, and G. Mullendore (2017). Storm labeling in three dimensions (sl3d): A volumetric radar echo and dual-polarization updraft classification algorithm. *MWR 145*, 1127–1145.

Stensrud, D. J. and Coauthors (2013). Progress and challenges with Warn-on-Forecast. *Atmos. Res. 123*, 2–16.

Stensrud, D. J., M. Xue, L. J. Wicker, K. E. Kelleher, M. P. Foster, J. T. Schaefer, R. S. Schneider, S. G. Benjamin, S. S. Weygandt, J. T. Ferree, and J. P. Tuell (2009). Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, 1487–1499.

Strensud, D., N. Yussouf, D. Dowell, and M. Coniglio (2009). Assimilating surface data into a mesoscale model ensemble: Cold pool analyses from spring 2007. *Atmospheric Research 93*, 207–220.

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.

Torn, R. D. and G. Hakim (2008). Ensemble-based sensitivity analysis. *MWR 136*, 663–677.

Torri, G., Z. Kuang, and Y. Tian (2015). Mechanisms for convection triggering by cold pools. *Geophys. Res. Lett. 42*, 1943–1950.

Wang, Q. J. (1997). Using genetic algorithms to optimise model parameters. *Environmental Modelling & Software 12*, 27–34.

Whang, S. and J. Lee (2020). Data collection and quality challenges for deep learning. *Proc. VLDB Endow. 13*, 3429–3432.

Wicker, L. (n.d). python replacement for the observation processing analysis and wind synthesis program (opaws). *Github Accessed Aug. 2019*, https://github.com/louiswicker/pyOPAWS.

Zeng., Y., T. Janjic, Y. Feng, U. Blahak, A. Lozar, E. Bauernschubert, K. Stephan, and J. Min (2021). Interpreting estimated observation error statistics of weather radar measurements using the icon-lam-kenda system. *Atmos. Meas. Tech. 14*, 5735–5756.