UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

EXPLORING THE OPTIMAL DESIGN OF A MESOSCALE BOUNDARY LAYER PROFILING NETWORK FOR MODEL FORECASTING APPLICATIONS IN CASES OF SIGNIFICANT SEVERE WEATHER

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
MASTER OF SCIENCE

By

MATTHEW BRADLEY AMMON
Norman, Oklahoma
2025

EXPLORING THE OPTIMAL DESIGN OF A MESOSCALE BOUNDARY LAYER PROFILING NETWORK FOR MODEL FORECASTING APPLICATIONS IN CASES OF SIGNIFICANT SEVERE WEATHER

A THESIS APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Elizabeth Smith, Chair

Dr. Stacey Hitchcock

Dr. Aaron Hill

Acknowledgments

First and foremost, I want to express my utmost gratitude to my advisors; Dr. Elizabeth Smith, Dr. Josh Gebauer, and Dr. Tyler Bell for providing me with the opportunity to participate in this incredible research project. I cannot express how grateful I am for their support and faith in my work, as well as their persistence through unconventional circumstances to make this opportunity available to me in particular. Their continued support and flexibility has been invaluable in making me the researcher, scholar, and person that I am today.

I would also like to thank the University of Oklahoma School of Meteorology and CIWRO for supporting me through my research and academic journey. The community associated with these institutions creates a unique environment that cultivates an experience unlike anything that can be found anywhere else in the world, and I will forever be honored to be a part of it. I am also thankful for the support and expertise provided by my committee members, Dr. Aaron Hill and Dr. Stacey Hitchcock. Their insights and feedback helped to make my research better and gave me ideas for future research projects. The guidance and expertise of Dr. Lou Wicker, Dr. Jeremy Gibbs, and Kent Knopfmeier was also instrumental in making this project a success. Our weekly meetings never had a dull moment and I always left having learned something that I did not expect.

Finally, I would like to thank my parents and my fiance Sarah for their unconditional love and encouragement that made me the person that I am today. I would never have been able to even start the journey that ultimately has me writing these words without their many sacrifices and their endless support.

Table of Contents

List Of Tables

List Of Figures

Abstract

Rapid advancements to computational capabilities, and the resulting proliferation of high-resolution convective allowing models (CAMs), have recently revolutionized severe weather forecasting. However, CAMs are still far from perfect, and are limited by the availability of observational data. Operational observations directly above Earth's surface are relatively sparse in space and time, leaving much of the three-dimensional structure of the lower troposphere poorly resolved. This under-sampling is likely a major limiting factor to the performance of CAM forecasts, which rely on observations to inform accurate initial conditions of the atmosphere. This is especially relevant to weather phenomena that occur in the atmospheric boundary layer (ABL), where small-scale heterogeneities driven by turbulence are merely parameterized, even in the most sophisticated operational models. Numerous features and processes, some of which play an important role in the evolution of convective storms, also exist in the ABL. A mesoscale network of vertical profiling systems could provide high-resolution observations of these features in addition to the general kinematic and thermodynamic structure of the ABL. Such observations would help to fill some of the data gaps that are likely detrimental to model forecasts of severe convective storms and their hazards. This study uses observing system simulation experiments (OSSEs) to explore how assimilating simulated observations from such networks into CAMs may impact forecast performance for high-impact severe weather hazards such as tornadoes. Additionally, various components of the design and operation of a profiling network are evaluated with the intent of gaining insight into how the network should be optimally designed to provide value while minimizing cost.

Chapter 1

Introduction

Weather prediction is perhaps the most enduring scientific endeavor of the human race. The unrelenting force of mother nature has eluded centuries of innovation, and continues to impact the daily lives of individuals around the globe. Since we cannot control the weather, understanding the atmosphere and striving to better forecast its behavior is the only way to prepare for the hazards it poses and mitigate the inevitable damages that result. Both our understanding of how the atmosphere works and our ability to predict it have improved remarkably, particularly over the past 60-70 years. These improvements were accelerated by the adoption of operational numerical weather prediction (NWP) models, which were brought to light in the midst of continuously evolving computational capabilities (Benjamin et al. 2019). As illustrated in Fig. 1.1 the skillfulness of synoptic scale NWP forecasts has increased immensely over the last few decades. Today, week-out forecasts have a similar degree of accuracy to 3-day-out forecasts from just four decades ago (Bauer et al. 2015). More recently, artificial intelligence (AI) has begun to emerge as a new tool in the relentless pursuit to improve weather forecasting. When used alongside NWP models, AI has helped to streamline data assimilation, model integration, and other computationally expensive components of the modeling process (Jian et al. 2021). AI methods have also shown promise toward improving the prediction and identification of high-impact and severe weather hazards (McGovern et al. 2017). While AI continues to gain leverage as a prospect for the future of weather prediction, it still has many limitations that need to be addressed, such as limited observational data to train on, before it can be trusted for operational use (Bonavita 2024). Nevertheless, the emergence of AI in the field of meteorology highlights the continued pursuit for more accurate and efficient weather forecasting.

Although the skill of NWP forecasts have improved appreciably in general, the predictability of small-scale weather remains relatively low in comparison to the synoptic scale. Severe convective storms and their hazards remain particularly difficult to accurately forecast. The predictability of tornadoes, for instance, has improved marginally over the past few decades. In fact, the lead time and false alarm ratio of tornado warnings have shown no statistically significant improvements since the early 1990s (Brooks and Correia Jr 2018). This suggests that large-scale atmospheric features are only part of the story. Although a favorable large-scale dynamic environment is required to support severe thunderstorms, mesoscale and storm-scale features are often the primary driver for the initiation and evolution of individual convective storms. The precise timing and location of surface-based convective initiation is highly dependent upon local variations in thermodynamic structure of the lower troposphere (Weckwerth 2000). Numerical simulations indicate that variability in thermodynamic structure on scales as small as hundreds of meters can substantially influence the evolution of deep moist convection (Bryan et al. 2003; Kirshbaum 2022). This scale is an order of magnitude smaller than the grid spacing of today's highest resolution operational NWP models. As such, many features in the atmosphere that play an important role in the evolution of convective storms are unresolved and parameterized in forecast models. This problem punctuates the need for more thorough observations of the atmosphere on smaller scales.

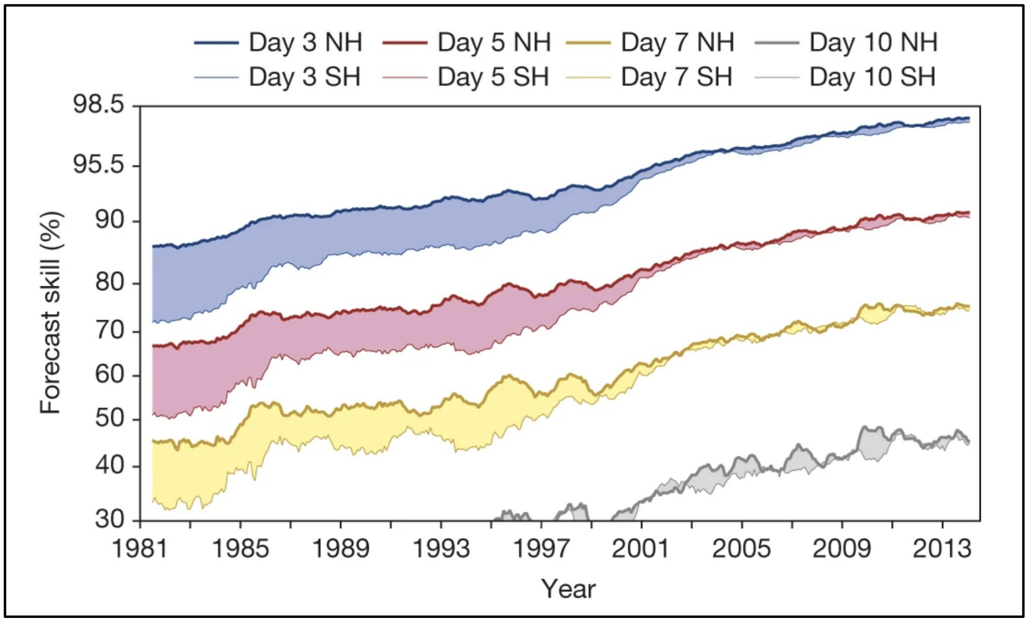

Figure 1.1: Diagram from Bauer et al. (2015) illustrating the improvements in forecast skill of global NWP models over time at three (blue), five (red), seven (yellow), and ten (grey) day ranges for the northern (NH) and southern (SH) hemispheres.

1.1 The Atmospheric Boundary Layer

Thermodynamic quantities change quickly and tumultuously, especially in the atmospheric boundary layer (ABL), where turbulence dictates the transport of heat and moisture. The ABL is characterized by the relatively thin layer of the troposphere that is directly influenced by the underlying surface of Earth (Stull 2012). The depth of the ABL is highly variable in time and can range anywhere from tens of meters to multiple kilometers, depending on a wide range of factors such as surface terrain structure, time of year, the diurnal cycle, and cloud cover (Garratt 1994). These factors modulate surface heating from solar radiation, and thus, the thermodynamic stability of the lower atmosphere. When surface insolation is reduced, such as at night and in overcast conditions, surface heat flux is dominated by radiative cooling which causes the ABL to remain stably stratified. These conditions suppress vertical momentum transport which keeps the boundary layer relatively shallow. During the daylight hours and in the absence of cloud cover, radiative heating destabilizes the surface which promotes turbulent vertical momentum transport that contributes to a deep, convective boundary layer (CBL).

The continuous fluctuations in boundary layer depth and stability that coincide with the diurnal cycle drive inertial oscillations in horizontal flow patterns and thermodynamic structure. The turbulent mixing in CBLs, for example, destratifies the lower levels of the atmosphere leading to a well-mixed boundary layer with enhanced drag imparted on low-level flow by viscous eddies. After sunset, convective turbulence ceases promptly which promotes re-stratification of the ABL and a reduction of surface drag. This causes flow above the surface to accelerate, often to supergeostrophic levels, which prompts the development of a nocturnal low-level jet (LLJ) and an associated enhancement of low-level vertical wind shear (Blackadar 1957; Shapiro and Fedorovich 2010).

The formation of LLJs can also be caused by horizontal variations in diurnal cooling across sloped terrain (Holton 1967). Shapiro et al. (2016) showed that these two mechanisms work synergistically to provide the forcing for the well-known Great Plains LLJ (GP-LLJ) that forms regularly across the Southern and Central Great Plains of the United States. This feature, which is particularly prevalent during the spring, plays a fundamental role in severe weather across the region.

The strong southerly flow of the GP-LLJ inherently increases the magnitude of low-level vertical wind shear and augments northward moisture transport which fosters the proliferation of a favorable regional environment for severe thunderstorms (Wu and Raman 1998; Weaver et al. 2012; Berg et al. 2015). Additionally, heterogeneities in LLJ structure have been linked to bands of enhanced convergence and associated convective initiation (Gebauer et al. 2018; Smith et al. 2019). This can allow for severe convection to form despite the thermodynamically stable nocturnal boundary layer. With its substantial effects on the formation of convection, as well as its contribution to the development of environments favorable for severe thunderstorms, the GP-LLJ is a critical feature to understand with respect to predicting severe weather. However a proper representation of its structure and evolution is highly dependent upon the ABL.

The GP-LLJ is just one of many features in the atmosphere that impacts severe weather for which the ABL contributes substantially to its formation and morphology. On the mesoscale, features such as sea-breezes and slope flows can arise from spatial variations in surface insolation (Garratt 1994). Both of these features, and other related features, are often responsible for initiation of convection and local modifications to the three-dimensional (3D) wind field (McNider and Pielke 1984; Segal et al. 1988; Miller et al. 2003).

Organized flow alterations in the ABL can also exist on the microscale in the form of coherent turbulent structures. These features are more cohesive and longer lasting than individual turbulent eddies. Horizontal convective rolls (HCRs), for example, can span multiple kilometers in depth and width, and can persist for hours (Weckwerth et al. 1999). HCRs are frequently observed in unstably stratified turbulent boundary layers where they facilitate vertical transport of momentum from the surface to the top of the ABL (Etling and Brown 1993). Weckwerth et al. (1996) found that the ascending portion of HCRs is associated with water vapor mixing ratios (WVMR) as much as 1.5 - 2.5 g kg^{-1} higher than the surrounding environment throughout the depth of ABL. This substantial increase in moisture content directly contributes to the formation of “street” of clouds that are often visible on satellite. Additionally, bands of enhanced near-surface convergence along the interface of two individual rolls can induce sufficient dynamic ascent throughout the ABL to trigger the initiation of deep moist convection (Xue and Martin 2006; Banghoff et al. 2020).

1.2 The Boundary Layer Data Gap

The ABL is home to a wide variety of turbulent features, structures, and processes that can impact how convection forms and evolves. In fact, according to Bryan et al. (2003), convection can be sensitive to turbulent features on scales as small as 100 m. These features can exist on scales ranging from sub-microscale to synoptic scale, and can persist on timescales ranging from seconds to hours. The complex underlying surface of Earth plays a major role in the proliferation of these ABL features and processes. Spatial and temporal heterogeneity in surface structure makes the interaction between the ABL and the surface highly variable and nearly impossible to quantify or predict in a meaningful way. In addition, the chaotic nature of the ABL lends to minimal spatial covariance, which makes the value of available observations deteriorate quickly with increasing displacement from the observation site. The accuracy of localized weather forecasts are therefore likely highly dependent upon the availability of boundary layer observations in the area of interest. As such, the current lack of operational ABL observations may be a major source of error in convection forecasts. ABL observations above 10 m are primarily provided by radiosondes, which are launched operationally twice a day (at 0000 UTC and 1200 UTC) at just 1500 sites across the globe, and only about 100 in the United States (Durre et al. 2006), leaving expansive data gaps. Considering the small spatial and temporal scales of potentially important ABL features and processes, it is clear that the ABL is sorely under-sampled. Some above-surface data gaps are filled by observations from commercial aircraft and satellites. However, these observations are mostly confined to the upper levels (i.e. cruising altitude) and the vicinity of major airports, therefore largely neglecting the ABL.

Unlike upper-air observations, operational surface observing stations are quite dense in many parts of the world, especially across the continental United States (CONUS). The majority of the CONUS is blanketed by a surface network made up of stations that are separated by just hundreds, in some cases as little as tens of kilometers. The Oklahoma Mesonet (OK-Mesonet, Brock et al. 1995), for example, provides continuous observations of numerous parameters up to 10 m above the surface at 110 sites across the state of Oklahoma with an average spacing of just 35 km between sites. The country-wide surface station network, coupled with smaller supplemental networks such as the OK-Mesonet, provide relatively thorough observational coverage of surface features and quantities at the mesoscale and higher. Surface networks, however, usually only observe the lowest 10 m of the troposphere which accounts for less than a tenth of a percent of its total depth, and just one percent of the depth of the ABL (assuming a typical ABL depth of 1 km). Therefore, while the networks of operational surface stations in the CONUS, as well as some other regions of the globe, are quite robust, they still fail to capture the vast majority of the ABL.

1.3 The 3D Mesonet Concept

In 2009, the National Research Council (NRC) Board on Atmospheric Sciences and Climate called for a coordinated nation-wide effort to expand upon meteorological observations above the surface, noting that the current network of operational upper-air observations provided by radiosondes only addresses the synoptic scale, and is much too dispersed in space and time to adequately capture lower-tropospheric variability (Council et al. 2009). To address this issue, the NRC suggested implementing a unified nation-wide "network of networks" (NoN) that would provide lower-tropospheric (i.e. ABL) observations at mesoscale resolution across the country. They emphasized that such a network would have a wide range of applications that would be beneficial for weather and climate prediction, air quality monitoring, agriculture, transportation, and more.

In pursuit of bringing the NoN to fruition, a variety of remote and in-situ profiling systems have been developed to observe kinematic and thermodynamic properties of the lower-troposphere with high spatial and temporal resolution (Bell 2021). Distributing numerous of these profiling systems across a mesoscale domain collocated with pre-existing surface station network is the fundamental basis of the 3D-Mesonet concept, a possible component of the NoN proposed by Chilson et al. (2019). The ABL observations that a 3D-Mesonet would provide could assist various facets of weather prediction, including real-time forecasting, reanalysis, and numerical modeling. A crucial step in the development of an operational 3D-Mesonet is identifying which profiling instruments, or combinations thereof, are best suited for large-scale deployment. Additionally, it is important to consider how a 3D-Mesonet should be organized and operated to optimize its utility while keeping costs to a minimum. This study will address these topics primarily in the context of NWP. More specifically, the utility of a 3D-Mesonet, and its dependence on instrument type, organization, and operation, will be assessed in terms of how the provided observations influence the performance of convective-allowing model (CAM) forecasts.

1.4 Assimilating Boundary Layer Observations in NWP

Much of the existing literature on the impact of ABL observations on CAMS is based on real observations collected from profiling facilities deployed during various field campaigns. In general, the spatial coverage and amount of profiling sites during these projects were quite limited. As such, the resulting prototype networks were more localized and coarse in resolution than the more robust mesoscale network proposed by the 3D-Mesonet concept. However, the results of assimilating observations from just a few sites give initial insight into the potential impact of high resolution ABL observations on model forecasts. Much of these results are based on datasets from he Plains Elevated Convection at Night (PECAN) field campaign, which sought to improve the understanding and prediction of nocturnal convective storms in the Great Plain of the United States (Geerts et al. 2017). PECAN featured an unprecedented network of 11 profiling systems spread across Oklahoma, Kansas, and Nebraska, five of which were mobile and deployed in variable locations that were strategically chosen based upon the area of interest for a given period. The resulting network, while not quite the resolution of a 3D-Mesonet, still provided observations of the lower troposphere at much better spatial and temporal resolution than the current operational radiosonde network.

A case study done by Degelia et al. (2019), for example, assessed the impact of assimilating observations from the PECAN profiling network on an ensemble forecast of a nocturnal mesoscale convective system (MCS) from the Advanced Research version of the Weather Research and Forecasting (WRF) Model. They found that assimilating thermodynamic observations from four atmospheric emitted radiance interferometers (AERIs) enhanced an elevated layer of moisture in the area of CI when compared to a baseline ensemble. They also found that when kinematic observations from four Doppler Lidar (lidar) sites were assimilated, outflow from pre-existing convection was stronger and the ascent associated with it was enhanced. Overall, the study showed that model representation of this particular MCS was improved in terms of its timing, location, and intensity when compared to a baseline experiment in which profiler observations were not assimilated. A similar study by Chipilski et al. (2020) assessed the impact of the PECAN network on model forecasts of nocturnal bore-driven convection. They found that AERI observations increased the neighborhood ensemble probability (NEP) of convection with maximum reflectivity of 30 dBZ or greater by as much as 50%. This was traced back to substantially greater boundary layer moisture ahead of the bore which translated to increased elevated convective available potential energy (CAPE) by as much as 500 J kg^{-1}. As a result, the modeled pre-bore environment became more supportive of sustained strong convection when boundary layer observations were assimilated. Notable improvements to model forecasts of CI were also found by Hu et al. (2019) in the case of a discrete supercell thunderstorm that produced a significant tornado. Assimilating observations from the AERI and lidar sites in the PECAN network resulted in a more moist boundary layer with enhanced low-level convergence in the near-storm environment when compared to a control experiment. These changes reflected a boundary layer more supportive of surface-based convection, and therefore contributed to CI that did not occur in the control experiment.

The more recent Propagation, Evolution, and Rotation in Linear Storms (PERiLS; Kosiba et al. 2024) field research campaign in 2023, which focused on using observations to study QLCS tornadoes in the southeast U.S., also involved coordinated deployment of boundary layer profiling systems. Specifically three weather-sensing uncrewed aerial systems (WxUAS) were deployed in a triangular configuration over a portion of the southern Mississippi Delta region. Tweedie (2024) assimilated real observations collected from these sites into an ensemble of the National Severe Storm Laboratory's (NSSL) Warn-on-Forecast System (WoFS; Stensrud et al. 2009; Heinselman et al. 2024) to assess the impact on forecasts of the devastating, high-end EF-4 tornado that impacted Rolling Fork, Mississippi and surrounding areas on March 24, 2023. In these experiments, they found enhanced backing of modeled near-surface winds and more moist, unstable air in the storm's inflow than what was modeled when the ABL observations were omitted from the data assimilation. Furthermore, the simulated location, intensity, and shape of the supercell that produced the tornado was more accurate to reality in the experiments that included the WxUAS observations.

Assimilating observations from WxUAS platforms have also been shown to influence model forecasts of ABL structure in the vicinity of complex terrain. Jensen et al. (2021) found significant improvements to WRF model profiles of temperature, humidity, and horizontal winds associated with terrain-driven flow within Colorado's San Luis Valley when WxUAS observations were assimilated. The improvements in model representation of the boundary layer structure fostered more accurate forecasts of the diurnal transition between katabatic and anabatic flow in the valley compared to baseline. These results emphasize the utility of profiling systems like the WxUAS for proper modeling of features and processes that exist on relatively brief time scales and small spatial scales, something that current NWP struggles to do.

It has become increasingly evident in recent literature that the accuracy of CAMs can greatly benefit from assimilation of ABL observations that are not available in current operational observing networks. These promising results provide evidence for the value of the 3D-Mesonet concept as a potential solution to the boundary layer data gap and a component of the NoN. However, the optimal design and operation of a 3D-Mesonet for use in NWP is far from obvious, especially when trying to optimize the balance between utility and cost. Factors such as the profiling instruments used, the horizontal spacing between sites, and the number of sites in the network can have implications on the potential impact that such a network could have on model forecasts. These factors will be the primary focus of this particular work. However, other factors such as localization radius, latency of data assimilation, filter methodology, vertical resolution of assimilated observations, and more cannot be ignored and should be addressed in future work.

1.5 Research Objectives

The high monetary cost of vertical profiling instruments makes finding an optimal balance between the potential value that a 3D-Mesonet could provide to forecasting and its total cost crucial to justify its implementation on an operational scale. In search of this balance, various factors of network design and functionally must be thoroughly considered. These factors can be assessed for various applications and situations, each of which may lead to different interpretations for how a profiling network should be designed. For the purpose of this particular analysis, aspects of network design will be considered in terms of how they relate to changes is ensemble-based model forecast performance for high-impact severe weather.

Specifically, this research will execute numerous OSSEs to evaluate how forecast performance is impacted by: a) The types of observations collected at each site and the instruments that provide them, and b) The number of individual sites that make up the network. In addition, the resulting performance differences will be related back to cost differences associated with these factors. This analysis will give insight into how a 3D-Network can be most efficiently implemented. Finally, the concept of small targeted networks will be discussed. This will involve exploring how mobile profiling systems can be leveraged to create a dynamic network of only a few sites strategically deployed in a focused area that is relevant to the event of focus. This concept presents a cheaper alternative to a larger, fixed network of profilers. The impact of these types of networks on convective model forecasts will be examined and compared to that of the more robust networks. The ultimate objective is to gain insight into how a mesoscale network of ABL profiling systems can be designed in a way that provides value to convective-scale model forecasts while mitigating cost.

Chapter 2

Overview of Cases

2.1 26 February 2023 (Case-1)

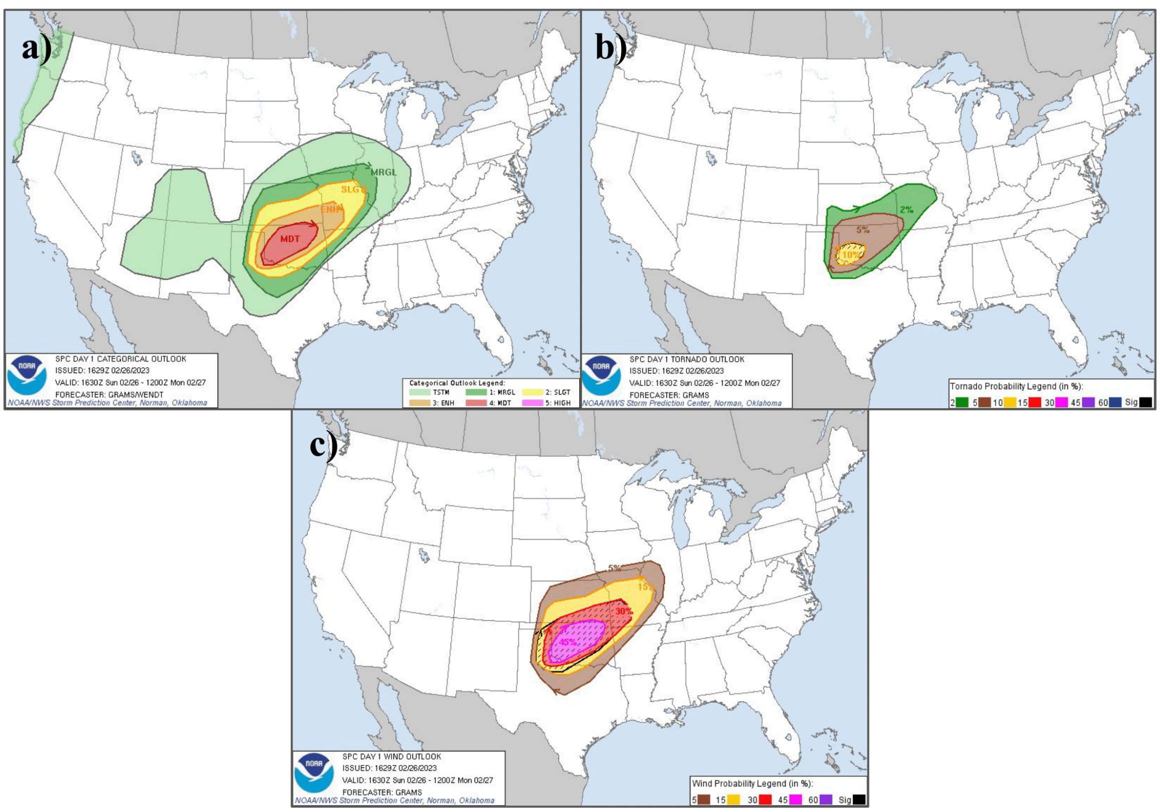

Figure 2.1: 26 February 2023, 1630 UTC SPC Convective Outlook: a) categorical risk, b) probabilistic tornado risk, and c) probabilistic damaging wind risk.

The two cases that serve as the basis of this exploration involve significant severe weather events in central Oklahoma. Both events are examples of situations with complex environments in which high-resolution ABL observations from a 3D-Mesonet may have been beneficial to the performance of model forecasts. Additionally, both events featured a relatively high degree of uncertainty regarding some aspect of the environment that a profiling network could have helped to make more clear. Events such as these provide the best opportunity to demonstrate the potential utility of a 3D-Mesonet.

The first case, hereon referred to as Case-1, is February 26, 2023. This day featured a robust, fast-moving quasi-linear convective system (QLCS) in a high-shear, low CAPE (HSLC) environment that spawned numerous tornadoes across OK. In response to the expectation for a high-end kinematic parameter space that would evolve through the evening, the SPC issued a moderate risk of severe thunderstorms across much of OK in their 1630 UTC day-1 convective outlook, driven by a 45% hatched risk of damaging winds (Fig. 2.1a,c). In addition to numerous tornadoes, the QLCS produced a large swath of severe straight-line winds from the central TX panhandle, all the way into southwestern MO, with some areas experiencing gusts in excess of 100 mph. The line of convection was supported synoptically by a high-amplitude, negatively-tilted mid-level trough that overspread western OK by early evening (Fig. 2.2a). This fast-moving trough was preceded by robust west-southwesterly mid-level flow in excess of 100 kts and powerful LLJ with southerly 850-mb flow as high as 80 kts (Fig. 2.2b). This vigorous kinematic environment blanketed an unseasonably warm and moist warm-sector that featured surface temperatures and dewpoints in the low-to-mid 60s (Fig. 2.2c,d). While a diffuse warm front pushed north over central KS, a sharp cold front extended south from the surface low and pushed eastward, driving the rapid eastward advance of the QLCS. Robust synoptic and mesoscale forcing, along with exceptionally strong low-level wind shear prompted NOAA's Storm Prediction Center (SPC) to issue a large 5% risk area for tornadoes and a smaller 10% hatched region in southwestern OK where the addition of unseasonably strong instability was expected to support the possibility for a couple of significant (EF-2+) tornadoes (Fig. 2.1c).

Despite the potent kinematic environment, some uncertainty existed regarding the degree of destabilization in the eastern portions of the risk area. Most model guidance forecast surface-based CAPE to remain below 500 J/kg across the region after sunset in the midst of a quickly stabilizing nocturnal boundary layer. The confidence in the tornado threat therefore diminished with eastward extent and time. Nevertheless, two of the three significant tornadoes of the evening occurred to the east of the 10%-hatched risk area in central OK, where minimal instability and relatively strong inhibition was forecast. Observations from a 3D-Mesonet could have supported more accurate model initial conditions in this region which, in turn, may have improved the accuracy of the modeled environment and increased confidence in the tornado forecast.

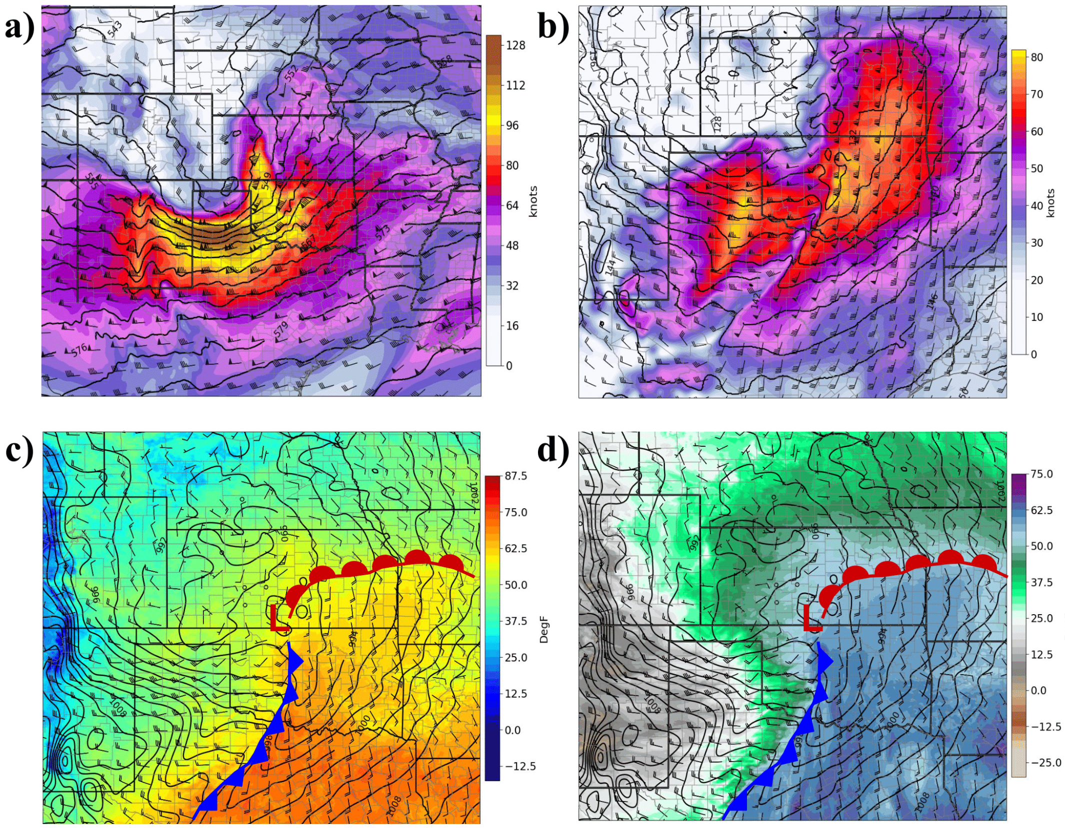

Figure 2.2: 27 February 2023, 0300 UTC HRRR analysis of a) 500 mb heights (dam; contours) and wind (knots; shading and barbs), b) 850 mb heights and wind, c) surface MSLP (hPa; contours), 10 m winds (barbs), 2 m temperature (˚F; shading), d) surface MSLP, 10 m winds, 2 m dewpoint. Fronts and surface lows are objectively analyzed in (c) and (d) for reference.

2.2 19 April 2023 (Case-2)

The second case in this analysis is April 19, 2023. This event featured a significant localized outbreak of more than a dozen tornadoes across a relatively small region of central OK. Most of the tornadoes, including two that were rated EF-3, were produced by a pair of discrete cyclic supercells that persisted for numerous hours. This outbreak occurred despite a high-degree of uncertainty regarding the coverage of convection due to a persistent capping inversion and nebulous synoptic forcing for ascent. Until

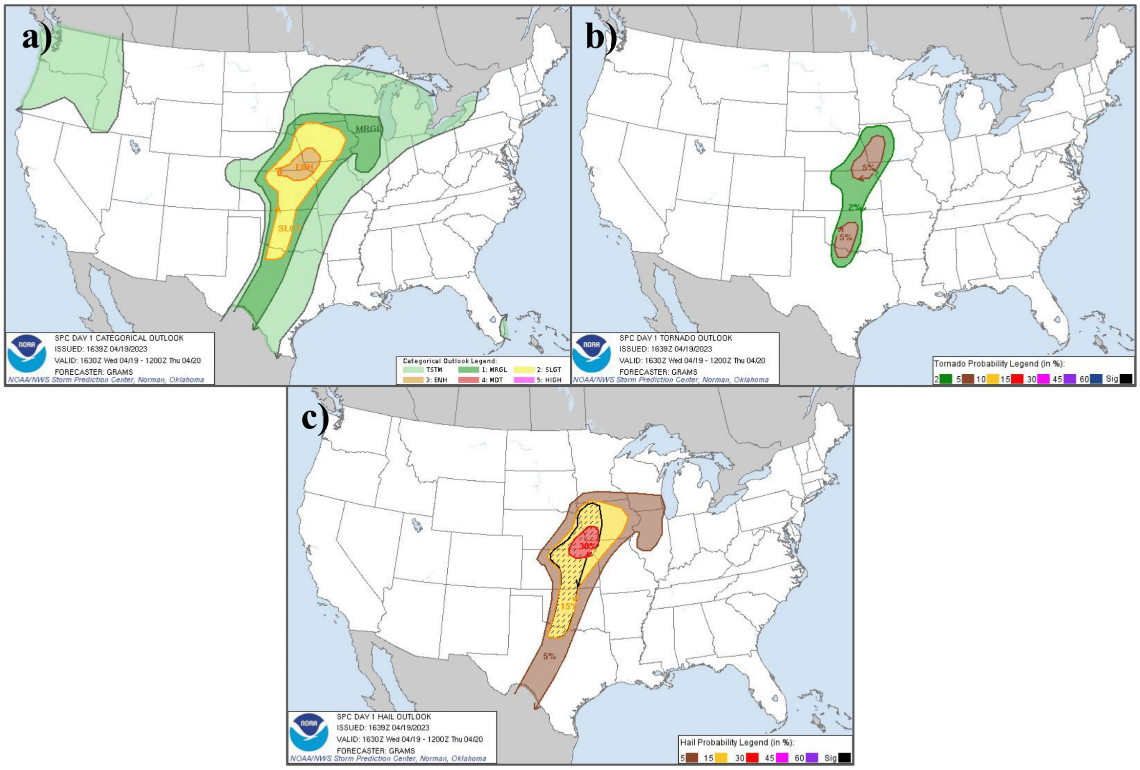

Figure 2.3: 19 April 2023, 1300 UTC SPC Convective Outlook: a) categorical risk, b) probabilistic tornado risk, and c) probabilistic severe hail risk.

just before the onset of the event, the majority of CAMs suggested that the capping inversion would prevent robust convection in central OK. The SPC noted in their 1300 UTC Day 1 convective outlook, however, that a conditional threat for significant hail and tornadoes existed if convection was able to initiate alone the dryline given strong instability and the expectation for diurnal strengthening of the LLJ as the evening progressed. As such, they delineated a small 5% tornado risk area that coincided with a 15% hatched risk for significant hail (Fig. 2.3).

Despite low confidence in storm initiation, the environment was generally favorable for severe thunderstorms. At 2100 UTC, about the time of convection initiation, a broad swath of enhanced west-southwesterly mid-level flow delineated the leading edge of a long-wave trough oriented across a quasi-meridional axis approximately co-linear with the Rocky Mountains (Fig. 2.4a). A subtle, negatively-tilted shortwave trough and associated corridor of locally enhanced mid-level flow embedded within this broader regime stretched from eastern CO into the TX and OK panhandles. A relatively modest but strengthening southerly LLJ with 850 mb flow upwards of 40 kts overspread a warm, moist surface layer across central and eastern KS and OK. Throughout most of the region temperatures reached well into the 80s (°F) and dewpoints were in the mid-to-upper 60s (Fig. 2.4 b-d). The western periphery of the warm sector was bounded by a sharp dryline that served as the impetus for the initiation of discrete convection. Once convection began, strong instability with surface-based CAPE in excess of 3,000 J/kg, sufficient deep-layer vertical wind shear upwards of 40 kts, and a strengthening LLJ combined to create a favorable environment for significant severe weather hazards.

Aside from the couple of dominant supercells, numerous other attempts at convective initiation occurred at various locations along the dryline, but were unable to fully mature and were stifled by the strong inhibition. This indicates that the localized environment surrounding the dominant supercells may have been anomalously favorable for severe convection in comparison to the broader environment forecast by CAMs. The lack of above-surface observations likely prevented the atmospheric properties necessary to properly forecast such a localized environment from being accurately represented in model initial conditions. Observations from a 3D-Mesonet may have helped to resolve this issue by improving the accuracy and resolution of the initial conditions, thereby potentially improving model forecast performance.

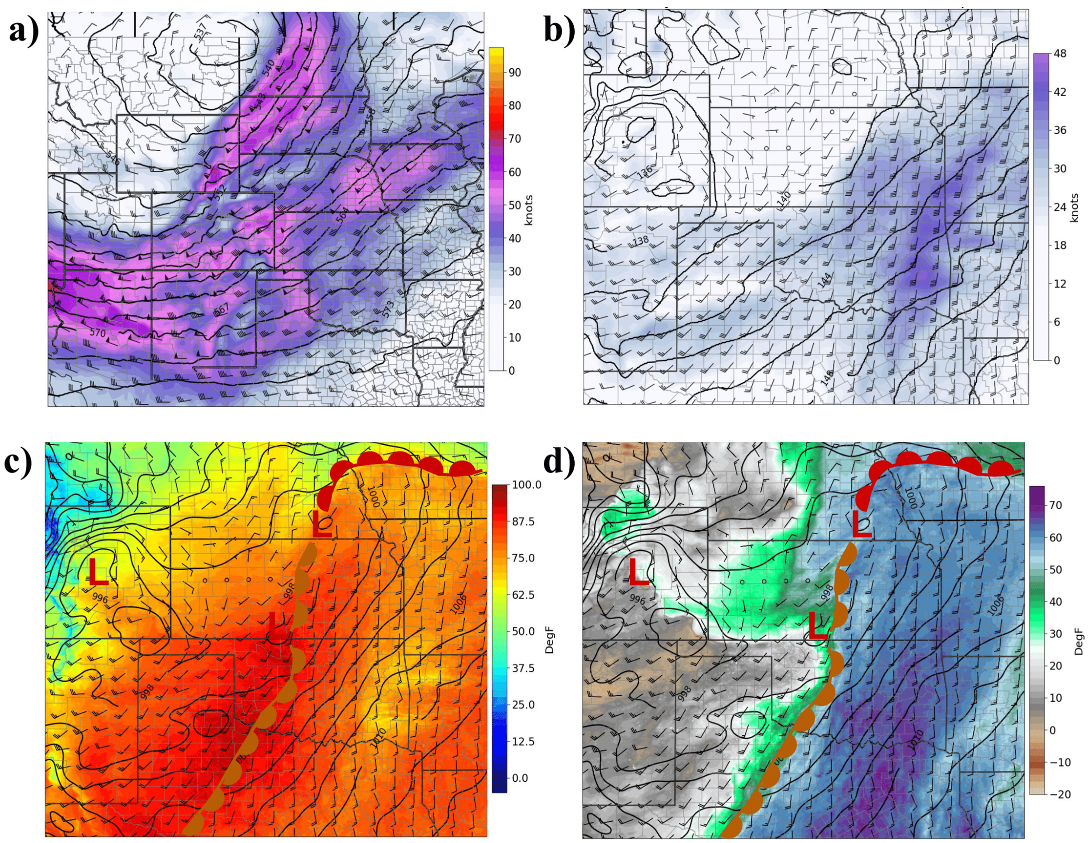

Figure 2.4: 19 April 2023, 2100 UTC HRRR analysis of a) 500 mb heights (dam; contours) and wind (knots; shading and barbs), b) 850 mb heights and wind, c) surface MSLP (hPa; contours), 10 m winds (barbs), 2 m temperature (°F; shading), d) surface MSLP, 10 m winds, 2 m dewpoint. Fronts, dryline, and surface lows are objectively analyzed in (c) and (d) for reference.

Chapter 3

Experiment Methodology

3.1 Profiling Instrumentation and Simulated Observations

The need for improved observation of the atmospheric boundary layer necessitates the development of new, innovative observing systems. As emphasized by the National Research Council (2009), the current operational radiosonde network is "in decline" and already suffering from depleting resources. Thus, expanding it is not feasible without implementing novel, more sustainable systems. A wide variety of both in-situ and remote profiling systems have been developed to address this need. For the purpose of this research, three proven remote profiling systems, including the Doppler Lidar, AERI, and microwave radiometer (MWR) will be assessed as candidates for a potential profiling network. An in-situ system, the University of Oklahoma's (OU) CopterSonde 3D (CS-3D) WxUAS (Segales et al. 2020), is also considered as a potentially more versatile alternative to the remote profiling instruments. While the CS-3D is a newer profiling system that currently requires manual operation, its portability, coupled with its ability to collect both thermodynamic and kinematic observations, make it an attractive option for a future operational network. However, the CS-3D does have some drawbacks that are important to consider such as: the requirement for preemptive approval from the Federal Aviation Administration (FAA) to operate, potentially long flight duration for a single vertical profile, and environmental limitations such as strong winds that can disrupt proper operation. Since these disadvantages can be mostly avoided by using remote profiling instruments, the CS-3D is not far and away the most suitable candidate for a 3D-Mesonet so the remote instrument options should also be considered. Various experiments were run to evaluate the impact of each of the profiling instrument candidates, and different combinations thereof, on NWP forecasts for the two severe weather cases described in the previous section. Since none of these instruments were actually deployed during either of the two cases, simulated observations based on the instrument specifications and associated errors were created. The details of how these simulated observations were generated, along with background information about aspects of design and functionality, are described below for each of the profiling instruments.

3.1.1 Doppler Lidar

The Doppler lidar is a ground-based active remote sensing instrument that is used to collect high-resolution kinematic observations. Lidars actively transmit radiation that is scattered by aerosols moving within the ambient flow, causing a Doppler shift in the transmitted radiation that can be translated into radial velocity. Simulated lidar radial velocity observations were created by applying range weighting functions associated with the Halo Photonics Streamline Doppler lidar (Pearson et al. 2009) to the u, v, and w wind components from background model profile at a desired location. The quality of lidar radial velocity observations is highly sensitive to the signal-to-noise ratio (SNR), which is the ratio of the signal power to the noise power across the full bandwidth of the lidar beam (O'Connor et al. 2010). The SNR is largely dependent upon the concentration of scattering aerosols, and therefore decreases with height, particularly in poorly-mixed boundary layers. To account for this, errors are applied to simulated lidar radial velocity observations based on a SNR climatology derived from DOE ARM site Doppler Lidar observation statistics (see Newsom et al. 2019).

Simulated vertical profiles of u, v, and w winds are then derived from the radial velocity observations using the velocity-azimuth-display (VAD) technique (Browning and Wexler 1968; Newsom et al. 2017). This technique considers the observed radial velocity from numerous azimuth angles at a fixed elevation angle (70° in this case) to derive 3D wind components using the trigonometric relationship:

$$V_r(\theta) = u\mathrm{sin}\theta\mathrm{cos}\phi + v\mathrm{cos}\theta\mathrm{cos}\phi + w\mathrm{sin}\phi \tag{3.1}$$

where $\mathrm{V}_r(\theta)$ is the radial velocity at a given azimuth angle (θ) and ϕ is the fixed elevation angle of 70°. The error in the resulting simulated 3D wind components is calculated using the trial-1 method in Newsom et al. (2017), which is what is done operationally by the US Department of Energy's Atmospheric Radiation Measurement (ARM) program (Mather and Voyles 2013).

3.1.2 AERI

The AERI is a passive ground-based infrared spectrometer that observes a wide spectrum of downwelling radiation spanning across a wave number range of 550-3,500 cm^{-1} with a temporal resolution of 20 s (Knuteson et al. 2004). The high spectral resolution of less than 1 cm^{-1} allows the AERI to resolve a vast assortment of spectral bands that can be strategically isolated in retrieval algorithms to extract quantitative information about atmospheric composition and state (Wagner et al. 2019). This is particularly useful for obtaining vertical profiles of thermodynamic quantities such as temperature and water vapor mixing ratio (q), as well as cloud microphysical properties. To extract these variables from the raw spectral data, the Tropospheric Remotely Observed Profiling via Optimal Estimation (TROPoe; formerly AERIoe) retrieval algorithm is used (Turner and Löhnert 2014; Turner et al. 2023). TROPoe utilizes an iterative Line-by-Line Radiative Transfer Model (LBLRTM) and a climatological prior dataset of radiance observations from the ARM southern Great Plains (ARM-SGP) site to translate the observed spectra to retrieved thermodynamic profiles. An added benefit of the AERI is that is self-calibrates by referencing two on-site blackbodies, one of which is heated to a constant temperature while the other assumes the temperature of the ambient environment. This method allows for the AERI to maintain a radiometric accuracy of higher than 1%. As such, the AERI is a good candidate for providing thermodynamic observations in a profiling network. It is important to consider, however, that the effective resolution of AERI observations decreases with height, and observations cannot be made above cloud-base height (Knuteson et al. 2004). Additionally, there are relatively few water vapor channels in the observed spectra, making the water vapor information content low compared to the temperature information content.

To simulate AERI observations, the LBLRTM is applied to a modeled profile of temperature, water vapor mixing ratio, and cloud properties. This results in a simulated infrared spectrum across the spectral range that is targeted by the AERI with the appropriate spectral resolution for the instrument's capabilities. Errors are then added to the resulting simulated spectrum based on the climatological ARM-SGP prior. This process results in a simulated radiance spectrum that is representative of a modeled atmospheric profile with the appropriate resolution and uncertainty characteristics of real AERI observations. Simulated profiles of temperature and water vapor are then derived from this simulated radiance spectrum using TROPoe.

3.1.3 MWR

The final remote profiling instrument examined in this work is the MWR, a passive, ground-based instrument that observes downwelling atmospheric radiation in the microwave spectrum. The MWR observes 14 microwave channels at 1-second temporal resolution with seven channels in the K-band (22.2 - 31.4 GHz) and seven in the V-band (51.8 - 58.8 GHz) (Rose et al. 2005). The K-band encompasses the water-vapor absorption lines, while the V-band targets oxygen absorption, allowing the two channels to diagnose vertical profiles of q and temperature, respectively (Löhnert et al. 2008). In order to ensure accurate observations, MWRs must be well calibrated. While an internal blackbody periodically facilitates autonomous relative calibration, manual absolute calibration is also necessary at regular intervals of several months (see Rose et al. 2005), and requires the use of liquid nitrogen. This intensive calibration process may be a factor that impedes large-scale deployment of MWRs in an operational network.

While both the accuracy and information content of the MWR is generally lower than that of the AERI (Blumberg et al. 2015), using observations from both instruments tends to result in more information content over a deeper layer than using either instrument individually (Turner 2007). As such, thermodynamic retrievals that consider observations from both instruments in tandem may be more useful and representative of atmospheric structure. This is important to consider in the design of a 3D-Mesonet as is has major implications on the monetary cost of such a network. Thus, comparing the utility of each individual instrument to that of using both simultaneously is a major focus of this work.

Simulated MWR observations were created from modeled vertical profiles of temperature, water vapor, and cloud properties by using a monochromatic radiative transfer model (monoRTM; Payne et al. 2011) to simulate brightness temperatures for the 14 MWR channels. Errors derived empirically from past comparisons between collocated

MWR and radiosonde observations (see Turner and Gero 2011) are then applied to make the simulated observations as realistic as possible.

3.1.4 CopterSonde-3D

The CS-3D is a state of the art WxUAS that is designed, maintained, and operated in house by OU's Cooperative Institute for High-Impact and Severe Weather Research and Operations (CIWRO). The CS-3D employs a unique wind vane flight mode that orients the drone into the wind at all times which allows for wind speed and direction to be derived from changes in aircraft inclination angle and yaw angle, respectively (Segales 2022). A scoop at the front of the aircraft exposes temperature and moisture sensors within the body of the drone to the surrounding air while the ambient wind flowing into the scoop helps to aspirate the sensors which maintains quality readings. Inter-comparison studies between the CS-3D and radiosondes found that wind speed and direction estimations are accurate to ± 0.65 m s^{-1} and $\pm 4°$, respectively (Bell et al. 2020). Additionally, the temperature sensor is generally ± 0.3°C while the humidity sensor is accurate to $\pm 0.1\%$.

To create simulated CS-3D observations, vertical profiles of air temperature, water vapor maxing ratio, and both components of the horizontal wind vector are extracted from a background model field at the desired location. To account for the gradual ascent (typically about 5 ms^{-1}) of the drone, errors are applied to the observation time series using a simulated red-noise process such that the applied error E at height z can be expressed as:

$$E(z) = 0.5E(z-1) + \epsilon \tag{3.2}$$

where ϵ is a constant base error of 0.5 K for temperature, 3% for relative humidity, and 0.6 m s^{-1} for wind speed (Bell et al. 2020).

3.2 Observing System Simulation Experiments

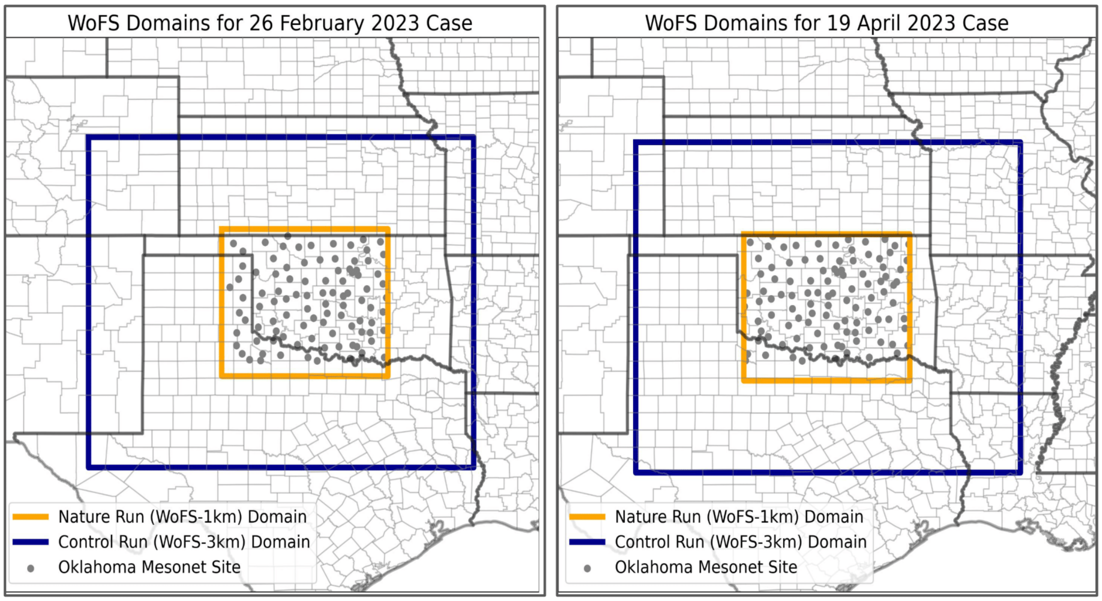

Figure 3.1: WoFS-3km (CR) and WoFS-1km (NR) domains used in the OSSE experiments for the 26 February 2023 case (left) and the 19 April 2023 case (right). Grey dots indicate Oklahoma Mesonet sites and various sites in Texas, from which simulated profiler observations were generated.

3.2.1 OSSE Framework

Assessing the utility of novel observation networks presents many challenges since a thorough analysis would require the network to be fully installed and operational. The high monetary cost of a network of instruments, especially those that would be required for vertical profiling, makes it difficult to justify large-scale deployment without quantitative evidence that it would provide value to forecasting. Simulated observations can be used to address this predicament by conducting observing system simulation experiments (OSSEs; Arnold Jr and Dey 1986). OSSEs are a valuable tool for simulating a novel observing system to gain insight into the value that it could potentially provide without the requirement of deployment and real data collection. An OSSE requires two critical components: a nature run, which serves as a model surrogate for the "real" atmosphere, and a forecast run that simulated observations are assimilated into to gauge their impact (Atlas 1997).

The nature run is a strategically chosen high-resolution model simulation of the event in question that closely resembles the true state of the atmosphere or outcome of the event. Defining the nature run can be quite ambiguous, and the methodology behind it can vary based on the context of a given OSSE. For this study, the nature run is defined by a single member of a model ensemble that most closely resembles the observed outcome of the relevant event. The modeled environment associated with this member is used to represent the true state of the atmosphere as closely as possible. Since this "true" state is only defined by a single ensemble member, model error is not accounted for in the OSSEs which requires the perfect-model assumption to be made. While having the benefit of removing a source of error in the experiment process, it is important to keep the implications of this assumption in mind while interpreting results. This perfect model nature run (hereon referred to as just the nature run) is used to generate the simulated observations used in the OSSEs.

The second main component of an OSSE, the forecast run, is a separate model simulation of the same event that is generally a less accurate representation of the observed outcome of the event. This component of an OSSE is typically chosen in a way that leaves room for the forecast performance to improve, i.e. become more closely aligned with the true outcome of the event, following the assimilation of simulated observations from the observing system being tested. For this study, the ensemble member that most poorly represents the observed outcome of the relevant event was chosen as the forecast run.

When conducting an OSSE, it is crucial to consider the "identical-twin" problem. This issue, which occurs when the same model is used for both the nature run and forecast run, can lead to artificially positive results (Arnold Jr and Dey 1986) due to an under-representation of model error and the possibility of biases associated with the model's dynamical core and parameterization schema. To avoid this, it is generally encouraged to use two different model frameworks for the two OSSE components. Williamson and Kasahara (1971) adopted an alternative solution, known as a "fraternal-twin" experiment, in which the same model is used for both OSSE components, but with different grid resolutions. Fraternal-twin experiments can adequately mitigate the negative effects of using an identical model for both components by introducing errors associated with different grid resolutions. This strategy can somewhat compensate for the lack of model error that comes with the identical-twin problem while also helping to avoid added computational expenses associated with running two entirely different model frameworks (Privé et al. 2023).

For the OSSEs conducted in this study, a fraternal-twin approach is taken using two different resolution WoFS model simulations to create the nature run and forecast run. The nature run is defined as the best-performing member of a 1-km resolution WoFS (WoFS-1km) ensemble forecast with a 402 x 402 km^2 domain. Lateral boundary conditions for the nature run domain are provided by a larger 900 x 900 km^2 WoFS domain with a more coarse 3-km resolution (WoFS-3km). The WoFS-3km domain also serves as the domain for the forecast run. The lateral and boundary conditions for the WoFS-3km domain are provided by the High-Resolution Rapid Refresh Data Assimilation System (HRRR-DAS) and large scale perturbations from the Global Ensemble Forecast System (GEFS). This configuration aligns with the typical WoFS setup outlined in Heinselman et al. (2024). The spatial extent and bounds of the two domains are illustrated for both the February 26, 2023 case (Case 1) and the April 19, 2023 case (Case 2) in Fig. 3.1.

The nature run and forecast run were chosen by subjectively identifying the best and worst performing members of an 18-member WoFS-1km ensemble forecast, respectively, based on simulated storm characteristics. This was done separately for each of the two cases. It is worth noting that this subjective methodology could potentially be a point of scrutiny for this study. However, objective assessment of forecast performance for this purpose would be an open-ended process with a wide variety of potential approaches. Subjective assessment of storm characteristics shown in simulated composite reflectivity inherently takes many factors relevant to the forecast into account including the location, evolution, and structure of the modeled convection. In future work, it would be worth exploring different methods for identifying the nature and control runs to see if the choice of which ensemble member to use significantly impacts the results.

Using the perfect-model OSSE framework, experiments were designed to evaluate numerous possible designs for a 3D-Mesonet across Oklahoma. These analyses were completed using the simulated observations that were generated for the two cases introduced in section 2.1. The experiments focus on assessing how observations from variations in the instrument types, spacing, and organization of the network impact the performance of WoFS ensemble forecasts for the two cases. For each OSSE, an ensemble forecast was run on the WoFS-3km grid associated with the respective case. After a 1-hour spin-up, the simulated observations were assimilated into the 36-member forecast run every 15 minutes for a 90 minute period, adding up to seven DA cycles. Starting from the final DA cycle, an 18-member ensemble forecast then ran freely for 4.5 hours. Assimilation of the simulated observations was done in the Data Assimilation Research Testbed (DART; Anderson et al. 2009) using the Ensemble Adjustment Kalman Filter

(EAKF) method described in Anderson (2001). The timing of each step in the OSSE experiment process is denoted for each of the two cases in Table 3.2.

Table 3.1: The period of time encompassed by each step in the OSSE experiment process for each case. All times are in UTC.

Experiment Timing		
Case (date)	Case-1 (20230226)	Case-2 (20230419)
Spin-up	2300 – 0000	2000 – 2100
NR Start time	0000	2100
DA Cycling	0000 – 0130	2100 – 2230
Free Forecast	0130 – 0600	2230 – 0300

3.2.2 Mesonet Sites

For each case, the simulated profiler observations were derived from the nature run at each OK-Mesonet site (Fig. 3.1). The OK-Mesonet network consists of 120 well-maintained sites across the state that provide numerous meteorological observations up to 10 meters above the surface every 5 minutes indefinitely (McPherson et al. 2007). OK-Mesonet sites were chosen for this study since they provide the added convenience of pre-existing observations at the surface, which some vertical profiling instruments lack the ability to reliably collect. In addition, the sites are strategically placed in accessible locations that are clear of obstructions and that are permitted by landowners. These sites would likely be suitable locations for the deployment of profiling instruments. In addition to the OK-Mesonet sites, some locations in Texas were hand-chosen to effectively expand the network while maintaining a similar spatial resolution. Although the aforementioned advantages of using the existing OK-Mesonet sites do not necessarily apply to the Texas sites, these locations were added to provide coverage across the sections of the nature run domains that extend beyond the Oklahoma border.

3.2.3 Experiment Process and Network Design

Table 3.2: Number of sites and types of simulated observations assimilated for each OSSE experiment that was conducted. Each listed experiment was conducted for both case-1 and case-2. The variable n is a placeholder variable for any positive integer that describes the number of sites in the simulated network associated with a given experiment. Note that “ALL” indicates that all OK-Mesonet sites within the case-relevant nature run domain are included, which amounts to 96 sites and 103 sites for case-1 and case-2, respectively.

OSSE Experiments							
Name	Number of Sites	Surface	Radar	Lidar	AERI	MWR	WxUAS
SfcRad	ALL	X	X				
SfcRad_ALL_LAM	ALL	X	X	X	X	X	
SfcRad_ALL_LA	ALL	X	X	X	X		
SfcRad_ALL_LM	ALL	X	X	X		X	
SfcRad_ALL_AM	ALL	X	X		X	X	
SfcRad_ALL_UAS	ALL	X	X				X
SfcRad_n_LAM	n	X	X	X	X	X	
SfcRad_n_LA	n	X	X	X	X		
SfcRad_n_LM	n	X	X	X		X	
SfcRad_n_AM	n	X	X		X	X	
SfcRad_n_UAS	n	X	X				X

Numerous aspects of the design and functionality of a 3D-Mesonet could be adjusted, including the type of instruments used, the horizontal spacing of the instruments, the temporal frequency of observations, the localization radius of the observations, and much more. Each of these nodes of variability add an additional level of complexity to the design process of a such a network. As such, finding a combination of these factors that would optimize the balance between cost and utility is a daunting task. To gain a sense of how a 3D-Mesonet could be organized in pursuit of this balance, numerous perfect-model OSSE experiments were conducted for case-1 and case-2, each of which involved assimilating simulated observations from unique network configurations with variations in the number of sites that make up the network, and the types of instruments deployed at each site. To assess these experiments in the context of expected levels of change to forecast performance, a baseline OSSE experiment was also ran for each case in which no profiler observations were assimilated. These baseline experiments, referred to as the surface-radar (SfcRad) or control experiments, assimilated simulated surface observations from each OK-Mesonet site within the nature run domain, along with simulated radar reflectivity and radial velocity observations based on the specifications of NSSL's Multi-Radar Multi-Sensor (MRMS) system (Smith et al. 2016). All other experiments also include these simulated surface and radar observations in addition to the relevant configuration of simulated profiling network observations. Comparing experiments to the baseline set by the SfcRad experiment allow for the utility of the network observations to be evaluated in the context of supplementing other observational datasets that are already available operationally. In this case, differences in forecast performance between any given experiment and the SfcRad experiment can be assumed to be an effect of assimilating the simulated 3D-Mesonet observations. It is worth noting that some currently operational observations, such as those provided by satellite and Aircraft Meteorological Data Relay

(AMDAR), were not included in the experiments. These types of observations were left out since their coverage, availability, and resolution can be inconsistent, particularly in the lower-troposphere. However, these observations likely still have some impact on model forecasts so future OSSE studies that include them could be worth investigation.

To gain a sense of the utility that the most comprehensive feasible network (which for the purpose of this study is defined as a network with the density of the most robust operational surface station networks, such as the OK-Mesonet) could potentially provide, OSSE experiments were ran for a network which included a lidar, an AERI, and an MWR at every site shown in Fig. 3.2 for Case-1 (left panel) and Case-2 (right-panel). These experiments will be hereon referred to as the SfcRad_ALL_LAM experiments where "SfcRad" indicates the inclusion of simulated surface and MRMS radar observations, "ALL" indicates that observations are assimilated from all OK-Mesonet sites within the case-relevant nature run domain, and "LAM" describes the instruments included at each site, in this case lidar (L), AERI (A), and MWR (M). The SfcRad_ALL_LAM experiments are used to approximate the upper-end potential improvement to forecast performance that a mesoscale profiling network could provide, assuming that there is a positive correlation between the amount of improvement to the forecast and the number of sites in the network. While this correlation may not be perfect, having more sites increases the resolution of observations which, in turn, generally improves the accuracy of model initial conditions. Research has shown that improved initial conditions usually translate to enhanced model forecast performance (e.g. Simmons and Hollingsworth (2002)).

In subsequent OSSEs, the number of sites that make up the simulated network was gradually decreased, thus reducing the effective horizontal resolution of the assimilated observations. For these "reduced-site" experiments, the sites that make up each network were hand-picked such that they were subjectively evenly distributed across the respective nature run domain. Comparing the forecast performance of these OSSEs to that of the SfcRad and SfcRad_ALL_LAM experiments should give insight into how network size influences the extent of performance improvement. A similar process was again executed, but with a focus on removing instruments from the network sites to evaluate which instruments have the greatest impact on forecast performance improvement. For each of these "reduced-instrument" experiments, the number of sites in the network (ALL) was held constant while one of the three remote profiling instruments was removed, leaving the other two. Using this strategy, the amount of decrease in performance improvement from the SfcRad_ALL_LAM experiment is analogous to the amount of value added to the forecast by assimilating observations from the instrument that was removed. These experiments were designed to evaluate which instruments have the greatest influence on forecast performance. Some experiments include simulated observations from the CopterSonde WxUAS in place of the remote profilers. Such experiments will include "UAS" in the name, rather than the single letter indication(s) associated the remote profiling instruments. All OSSEs that were conducted for both cases are described in terms of the number of sites in the simulated network (n) and the instruments included at each site in table 3.2.

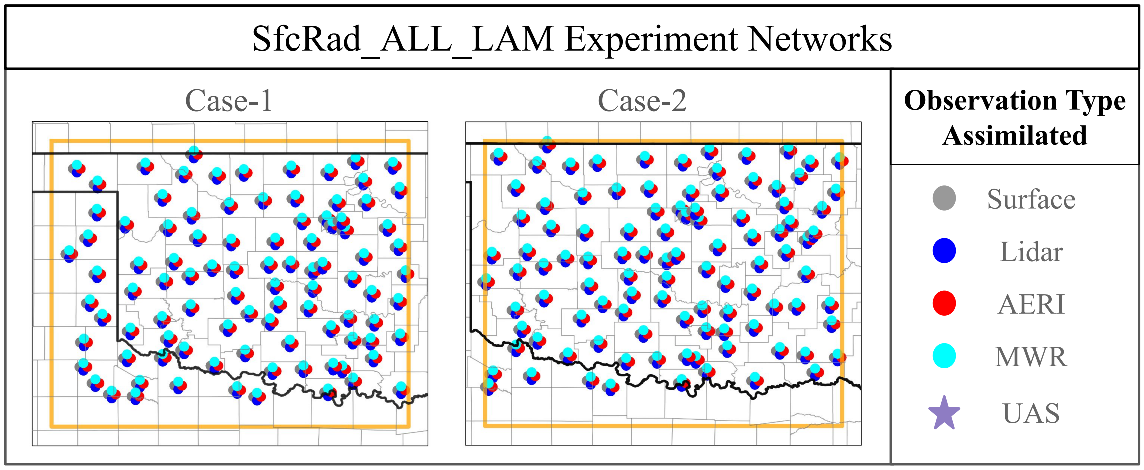

Figure 3.2: Network visualizations for the SfcRad_ALL_LAM experiments.

3.3 Defining Forecast Performance

3.3.1 Updraft Helicity

There is a wide range of possible methods that can be used to quantitatively assess model forecast performance. For any method, a crucial step is to identify a variable or parameter on which to base the assessment that is relevant to the nature of the event being modeled. For the events focused on in this study, forecast performance should be assessed based on a variable or parameter that is relevant to tornadoes. Updraft Helicity (UH), a diagnostic parameter that serves as a proxy for mesocyclone intensity, is the variable of choice for this study. UH is derived by vertically integrating the product of vertical vorticity and updraft velocity through some layer (Kain et al. 2008). The layer of vertical integration is most commonly defined as either 0–2-km (UH02) or 2–5-km (UH25), which generally serve as good approximations for the heights and depths of low-level mesocyclones, and mid-level mesocyclones, respectively (Sobash et al. 2016).

A useful tactic for visualizing the model forecast evolution of mesocyclones is to consider the hourly maximum UH at each model grid point (Kain et al. 2010). Accumulating these maxima throughout the duration of the forecast period results in continuous UH swaths that can effectively serve as surrogates for the tracks of modeled mesocyclones. Hourly maximum UH tracks have shown skill in forecasting tornado path length, location, and intensity (Kain et al. 2010; Clark et al. 2013; Sobash et al. 2019). As such, modeled UH swaths can be a useful tool for assessing model performance from the perspective of forecasting tornadoes and their associated mesocyclones. However, UH is sensitive to grid resolution which makes comparing NWP output across different model configurations challenging. This sensitivity is related to the relationship between horizontal grid spacing and simulated vertical vorticity magnitude. Adlerman and Droegemeier (2002) found that the magnitude of vertical vorticity systematically increases with increasing grid resolution. As a result, UH is intrinsically larger on more fine grid scales.

For the experiments conducted herein, UH swaths from the WoFS-3km experiment forecast output cannot be directly compared to the the WoFS-1km NR output due to this effect. For equitable comparison with the experiment forecasts, the NR forecast UH swaths were scaled down by a constant factor of 6.18 and 5.04 for UH02 and UH25, respectively. These scaling values were derived by Sobash et al. (2016) using simulated UH forecasts on both 3-km and 1-km grids for 497 cases with severe thunderstorms. For each of these 497 cases, surrogate severe weather reports (SSRs) were defined where gridpoint values exceeded a threshold value. This was done for UH01, UH03, and UH25, as well as relative vertical vorticity assessed at 500 m and 1 km AGL. The ratios of the 1-km resolution optimized threshold values to those at 3-km resolution were found to be 6.96, 5.39, and 5.04 for UH01, UH03, and UH25, respectively. Since

Sobash et al. (2016) did not directly find the optimal scaling ratio for UH02, the average of the scaling values of UH01 and UH03, which is approximately 6.18, was used.

Another problem with comparing field variables between two forecast outputs on different resolution grids is the lack of spatial alignment of the gridpoints. To address this, the WoFS-3km forecast output was interpolated to the resolution of the WoFS-1km grid using piecewise linear 2D-interpolation (see Virtanen et al. 2020).

3.3.2 Neighborhood Ensemble Probability

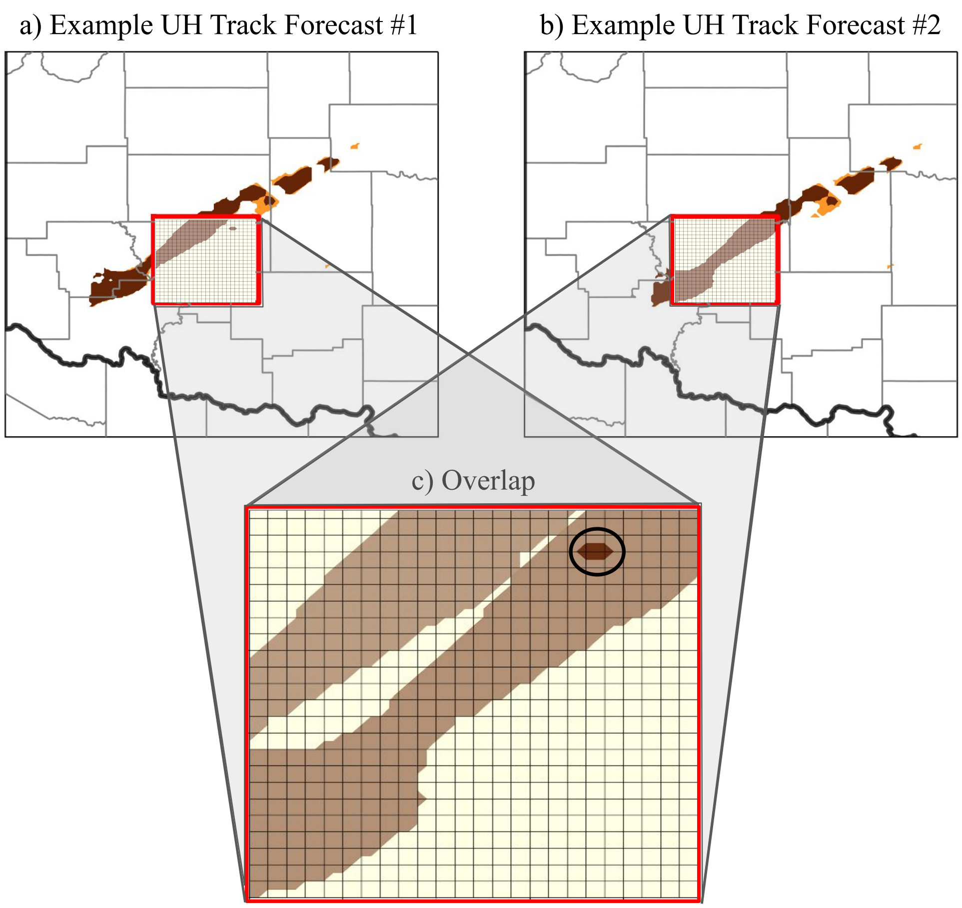

Figure 3.3: An example of how a UH track, which is similarly forecast in two different models (a and b), only commonly share a couple of grid points (c).

To assess the forecast performance of the various OSSEs, the modeled UH tracks can be compared to those of the nature run. However, the spatial overlapping of modeled positive vertical vorticity and vertical motion is generally very localized on high-resolution grids, thus creating sharp horizontal gradients in the resulting UH swaths. Furthermore, UH tracks are generally thin. Therefore, when comparing modeled UH tracks between two experiments, a very slight difference in the location of a track, that would be unlikely to impact the interpretation of the forecast as a whole, could share little or no common grid points. This predicament is illustrated in Fig. 3.3 which depicts two different example model forecasts of a UH track that share only a couple of common grid points. If the performance of forecast a (Fig. 3.3 a) were to be assessed in regard to how well it matches up with forecast b (Fig. 3.3 b), it would score very poorly if compared on a gridpoint-to-gridpoint basis. However, the difference between the two models is effectively inconsequential in the context of meaningful forecasting applications. Therefore, gridpoint-to-gridpoint performance assessment of UH tracks would be misleading in cases of slight spatial disparities such as this.

Neighborhood probability (NP; Schwartz et al. 2010) is a way to address this issue. This technique translates a spatial dataset of explicit values into a probabilistic dataset that makes each grid point somewhat affected by all other gridpoints within a surrounding "neighborhood" (Ebert 2008). A neighborhood is defined as all gridpoints within a specified radius of the individual gridpoint at which the NP is calculated. The specified neighborhood radius can be adjusted upward or downward to increase or decrease the tolerance for spatial differences, respectively. The NP technique is a useful tool for better aligning objective verification of high resolution datasets with how they are best interpreted in the context of meaningful forecast applications.

To derive NP, each gridpoint is first assigned a binary probability (BP) value based on whether or not the gridpoint meets or exceeds a defined threshold value (q) using:

$$BP(q)_i = \begin{cases} 1 & \text{if } f_i \geq q \\ 0 & \text{if } f_i \leq q \end{cases} \tag{3.3}$$

where f is the original, discrete gridpoint value for the ith gridpoint within the defined neighborhood. The NP at the ith gridpoint can subsequently be expressed as:

$$NP(q)_i = \frac{1}{N_b} \sum_{k=1}^{N_b} BP_k \tag{3.4}$$

where N_b is the total number of gridpoints in the neighborhood and BP_k is the binary probability as calculated in Eq. 3.3 for the k-th gridpoint in the neighborhood. NP is effectively the amount of gridpoints within the neighborhood of gridpoint i that exceed the specified value q. The effect is a smoothing of the field variable in a way that statistically accounts for the information content of the full-resolution data.

The NP can be averaged across all N members of an ensemble to yield the neighborhood ensemble probability NEP, expressed as:

$$NEP(q)_i = \frac{1}{N} \sum_{j=1}^{N} NP_{ij} \tag{3.5}$$

for the i-th gridpoint of the j-th ensemble member. NEPs are often used in assessing the performance of convection-allowing model ensemble (CAM-E) forecasts of field variables such as simulated composite reflectivity and accumulated precipitation (Johnson and Wang 2012; Romine et al. 2014; Ben Bouallègue and Theis 2014; Scheufele et al. 2014; Hitchcock et al. 2016; Hill et al. 2021). How NEP changes in response to varying the type, amount, or frequency of DA, or even the assimilation methodology, is particularly valuable for identifying the most useful strategies for improving CAM performance. Other studies have also applied NEP to UH swaths to assess CAM-E performance in forecasting the evolution of rotating updrafts and tornadoes (Yussouf et al. 2013; Wheatley et al. 2015; Johnson et al. 2020).

For this work, the NEP of accumulated hourly-maximum UH (i.e. UH swaths) from the entire forecast period exceeding various percentile-based thresholds of the nature run is evaluated across the experiment domains using a neighborhood radius of 9-km (3 gridpoints). This neighborhood size is often used for NEP calculations related to UH and associated probabilistic visualizations, such as in the WoFS cloud-based forecast viewer. Specifically, the NEP of UH02 and UH25 values exceeding the 90th, 95th, and 99th percentiles ($p90$, $p95$, $p99$) of the nature run is assessed for both cases. The percentiles are derived from the domain-wide accumulated hourly-maximum UH swaths of the nature run forecast. These higher percentiles place focus on the areas where intense low and mid-level mesocyclones that are capable of supporting tornadoes are most likely to occur based on the nature run.

Forecast performance for each experiment is assessed based on the experiment NEPs within the areas where the $p90$, $p95$, or $p99$ are expected to occur based on the nature run forecast. Using this technique, improvements to forecast performance from the baseline SfcRad experiment can be identified by higher NEP of UH exceeding a percentile value in areas where that value is observed or exceeded in the NR forecast. Similarly, a performance reduction from baseline would entail lower NEP values in these areas. Outside of these percentile areas, the opposite differences in NEP values would be expected (i.e. lower NEP than the SfcRad experiment would indicate relative improvement in forecast performance).

Table 3.3: 2x2 Contingency Table from Brooks 2004

	Observed			
		Yes	No	Sum
Forecast	Yes	a	b	a+b
	No	c	d	c+d
	Sum	a+c	b+d	1

3.3.3 Performance Metrics

Assessment of the performance of severe weather forecasts can be traced as far back as Finley (1884). In general, the metrics by which forecasts are assessed have not substantially changed from Finley's method, but have been expanded upon and streamlined. Murphy and Winkler (1987) introduced the 2x2 contingency table, which is used to classify a joint distribution ($p(f, x)$) of forecasts (f) and observations (x) for a dichotomous variable of interest. For each individual event in a given distribution, a binary value is assigned to both f and x, with a value of 0 representing a "no" and a value of 1 indicating a "yes". This joint distribution of f and x thus has four possible outcomes (a, b, c, and d) that are depicted in Table. 3.3. These four classifications can then be used to derive various performance metrics that are used to assess forecast "goodness" (see Murphy 1993). For the purpose of this thesis, the metrics that will be used to quantitatively assess forecast performance are the probability of detection (POD), success ratio (SR), and critical success index (CSI), which are traditionally expressed as:

$$POD = \frac{a}{a+c} \tag{3.6}$$

$$SR = \frac{a}{a+b} \tag{3.7}$$

$$CSI = \frac{a}{a+b+c} \tag{3.8}$$

where a, b, c, and d are the classifications of the joint distribution $p(f, x)$ (Table. 3.3). These traditional expressions, which are based on binary classifications, become more difficult to implement for performance assessment from an ensemble perspective since ensemble-based forecasts are generally meant to be interpreted in terms of a probability distribution, such as NEP, rather than in terms of discrete binary values. To address this challenge, an alternative contingency matrix ($P(f_{\text{NEP}}, x)$) is used for which the fractional NEP values at each gridpoint are used for the classifications instead of a binary value. In other words, the "yes" forecast classification is the NEP value of the gridpoint, and the "no" forecast classification is one minus the NEP value. For a gridpoint with an NEP value of 0.67 (67%), for example, the forecast classification would be 0.67 for "yes" and 0.33 for "no".

The observed classification (i.e. x) is defined differently since the nature run is a single, deterministic perfect-model forecast rather than an ensemble. This, along with other differences between the nature run and experiment forecasts such as grid resolution, made objectively defining x particularly difficult. Perhaps the biggest issue arises from the localized nature of high UH values associated with calculations being made on a 1-km grid. Since the high percentile values are usually isolated to a couple of gridpoints, the number of gridpoints within a neighborhood that exceed the percentile values are minimal which causes low neighborhood probability (NP). On the other hand, the experiment forecast UH calculations are made on the WoFS-3km grid, then interpolated to 1-km resolution which smooths out the high-percentile values across more gridpoints leading to relatively higher NP values. Many rounds of trial-and-error were involved in tuning the way that x was defined to account for this issue and to make it subjectively representative of the areas in which the $p90$, $p95$, and $p99$ were observed in the nature run forecast. Ultimately, a gridpoint was assigned a discrete value of 1 (observed) for x if at least one other gridpoint within a 9-km radius of the gridpoint in question had a greater than 50% NP of the nature run forecast exceeding the relevant percentile value. If this criteria was not met, a value of 0 (not observed) was assigned to the respective gridpoint. This criteria subjectively made it so that relative size and continuity of the areas that encompassed x values of 1 were generally analogous to that of the swaths of enhanced NEP in the experiment forecast output.

The subjective nature of the process of defining x is an important limitation to consider in the interpretation of this analysis. Additionally, these issues illustrate a potential disadvantage of the fraternal-twin OSSE framework, especially when assessing performance in terms of variables whose magnitudes are highly sensitive to grid resolution. Nevertheless, since this framework is applied consistently across the performance analysis of all conducted OSSE experiment forecasts in this study, the way in which x is defined likely has little impact on the inter-comparison of different experiments. With that said, a more sophisticated way of defining x may be worth exploring for similar future analyses.

With these definitions of f_{NEP} and x, the alternative contingency classifications (i.e. a, b, c, and d) can be expressed as:

$$P(f_{\text{NEP}}, x_{\text{NEP}}) \begin{cases} a = \sum_{i=1}^{N_{p+}} \text{NEP}(p)_i \\ b = \sum_{i=1}^{N_{p-}} \text{NEP}(p)_i \\ c = \sum_{i=1}^{N_{p+}} (1 - \text{NEP}(p)_i) \\ d = \sum_{i=1}^{N_{p-}} (1 - \text{NEP}(p)_i) \end{cases} \tag{3.9}$$

where $\text{NEP}(p)_i$ is the NEP of the i-th gridpoint of the relevant experiment forecast of accumulated maximum UH exceeding the p-th percentile of the nature run forecast. For a and c, the NEPs are summed across all gridpoints where at least one other gridpoint in the surrounding 3-km radius neighborhood has a nature run NP greater than 50% to exceed the relevant percentile value (i.e. N_{p+}; this is the criteria that must be met for x to be assigned a value of 1 (observed) at the relevant gridpoint). For b and d, NEPs are summed across all gridpoints where the this criteria is not met (i.e. N_{p-}) and the value of x is 0 (not observed).

Despite adjusting how the components of the contingency table are defined, the performance metrics (i.e. POD, SR, and CSI) were still calculated using the expressions given by Eqs. 3.6, 3.7, 3.8, but with the newly defined a, b, c, and d classifications outlined in Eq. 3.9. These metrics are visualized on performance diagrams (see Manning and Schutze (1999); Roebber (2009)), which conveniently express POD as a function of SR while also translating the relationship between the two metrics to a discrete CSI value. In the context of how it is defined in this analysis, the POD is a quantitative measure of the average predictability of accumulated maximum UH exceeding given percentile values for a specified experiment. Higher values of POD correspond with higher predictability and therefore represent a better performing forecast. The SR effectively scores how well the forecast correctly predicts that accumulated maximum UH will not exceed the percentile values in areas where it is not observed in the nature run forecast. Higher SR values correspond with a lower likelihood of false alarms which translates to better forecast performance. Performance diagrams were created for UH02 and UH25 for both cases, and include points for the 90th, 95th, and 99th percentile values ($p90$, $p95$, and $p99$) of the nature run forecast. The following sections analyze these performance diagrams, alongside spatial analysis plots for each case across several OSSE experiments, which provides initial insights into the impact of assimilating simulated observations from a profiling network on forecast performance, and how that may vary for different network designs.

Chapter 4

Results

4.1 Dense Network Experiments

Before exploring how changing the number of sites or types of instruments that make up a simulated profiling network influences forecast performance, the impact of a dense network with a spatial resolution that is similar to the most comprehensive operational surface station networks (i.e. statewide mesoscale networks such as the OK-Mesonet) was assessed by analyzing the performance of the SfcRad_ALL_LAM experiment forecasts and comparing them to the SfcRad experiments. This was done to gain a sense of the upper-end potential benefit that could be provided by a 3D-Mesonet in the case of minimal monetary limitations. Forecast performance improvements associated with networks simulated in other experiments can also be related back to the SfcRad_ALL_LAM experiment to gain insight into how much value is lost upon reducing the number of sites or cutting back on instrumentation.

In the context of the significant tornado events of February 26, 2023 (Case-1) and April 19, 2023 (Case-2), it is intuitive to assess forecast performance in terms of how simulated UH tracks compare to the nature run. The following comparisons between the SfcRad_ALL_LAM experiment and the SfcRad experiment forecasts focus on the most intense UH tracks where strong low-level rotating updrafts that are often associated with tornadoes are most probable. Specifically, the NEP of the SfcRad_ALL_LAM experiment forecast UH02 and UH25 being equal to or greater than the 90th, 95th, and 99th percentiles of the nature run forecast ($p90$, $p95$, and $p99$, respectively) were computed and compared to the SfcRad experiment forecast for each of the two cases.

4.1.1 Case-1

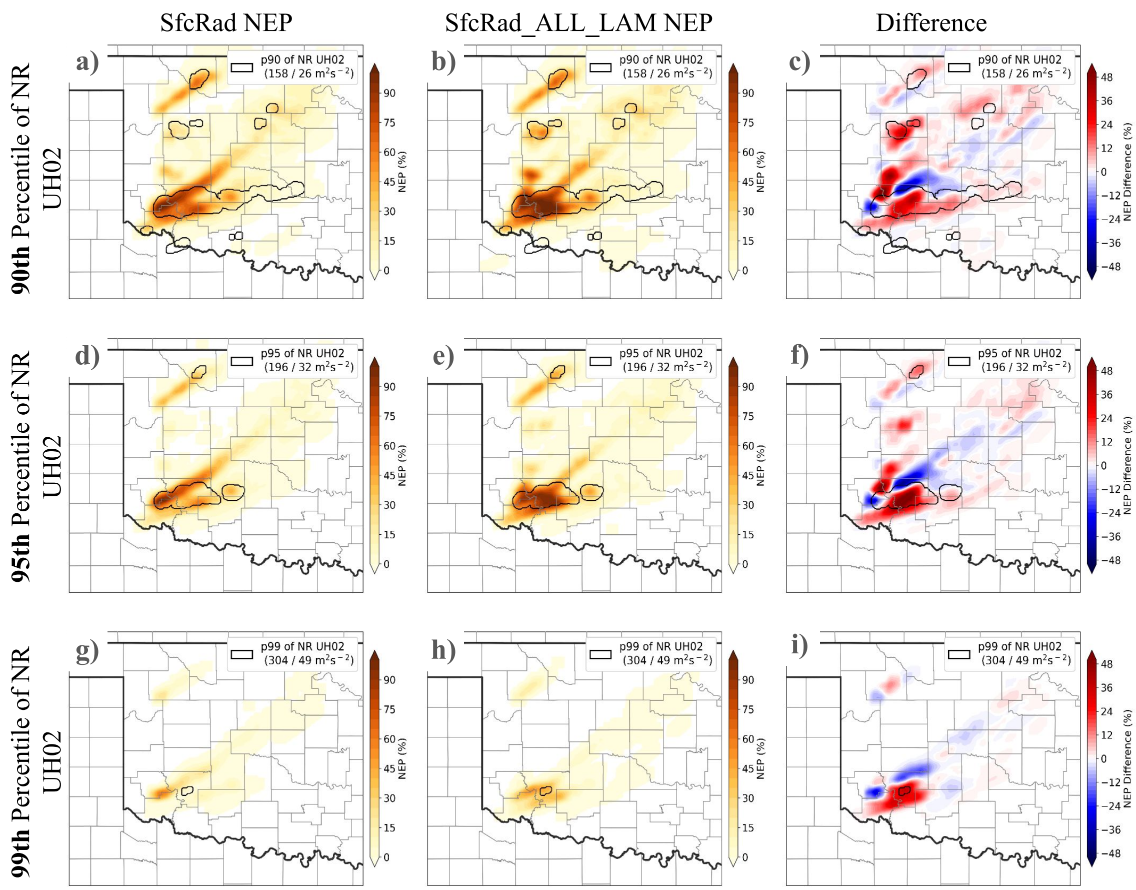

Figure 4.1: For Case-1: NEP of simulated UH02 exceeding the 90th, 95th, and 99th percentiles ($p90$, $p95$, and $p99$) of the nature run during the free-forecast period of the SfcRad experiment (a,d,g) and SfcRad_ALL_LAM experiment (b,e,h). The NEP differences between the SfcRad_ALL_LAM and SfcRad experiments are depicted in c,f, and i for $p90$, $p95$, and $p99$, respectively. Solid black contours indicate where the respective percentile values are observed or exceeded in the nature run.

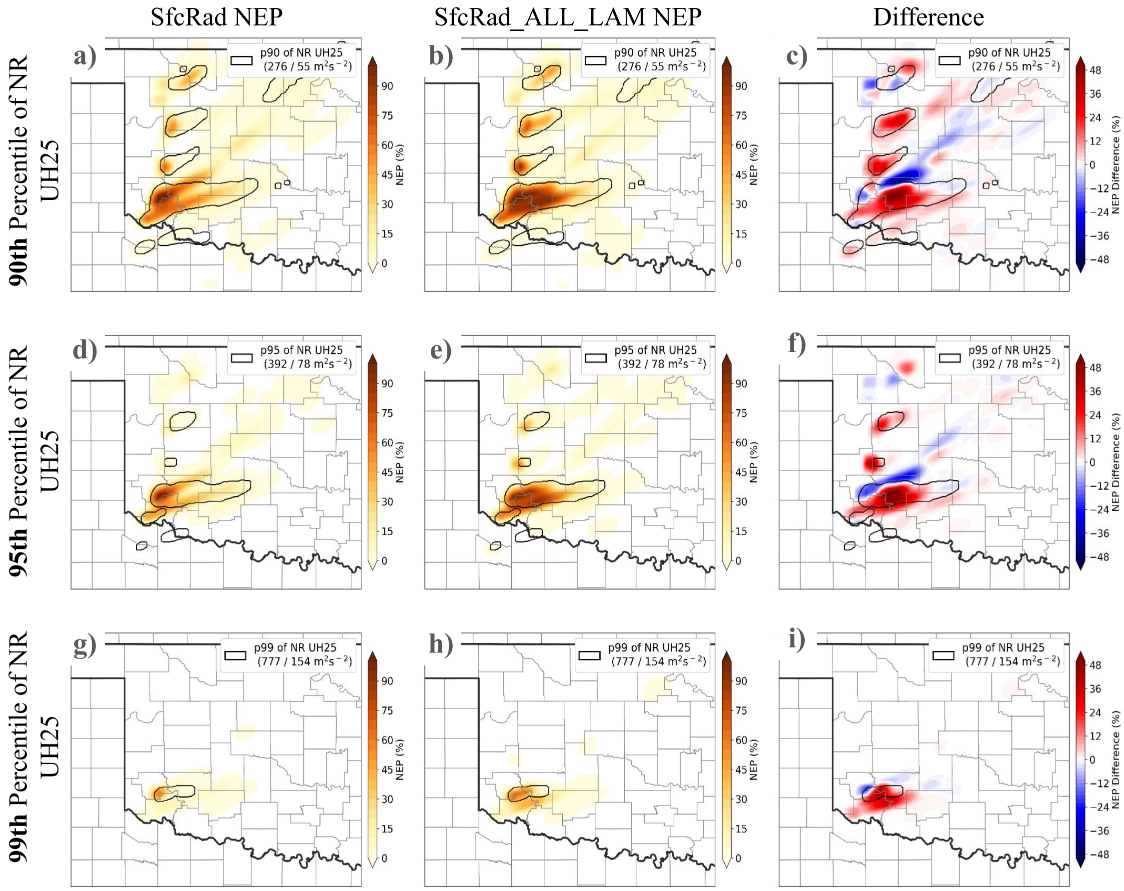

Figure 4.2: Same as Fig. 4.1 but for UH25.

To assist in comparing the two experiment forecasts and to visualize forecast performance, the NEPs of UH02 (Figs. 4.1) and UH25 (Fig. 4.2) exceeding $p90$ (top row), $p95$ (middle row), and $p99$ (bottom row) were plotted for the SfcRad (left column) and SfcRad_ALL_LAM (middle column) experiments. Darker shading indicates areas with a higher NEP of simulated UH exceeding the specified percentile values. The panels in the right column of the figures show the difference in NEP between the two experiments with positive values (red shading) indicating that the NEP is greater in the SfcRad_ALL_LAM experiment than the SfcRad experiment, and negative values (blue shading) indicating the opposite. The areas where the percentile values are observed in the nature run simulation are indicated by the black contours.

Performance diagrams were also created to quantify the differences in forecast performance between the two experiments in terms of the POD, SR, and CSI. (Fig. 4.3). In the context of how POD and SR are defined in this thesis (see Section 3.3.3), a quantitatively perfect forecast would entail both a POD and SR of 1, which would indicate a NEP of 100% at all gridpoints where the respective percentile value is observed in the nature run forecast (inside of black contoured areas), and an NEP of 0% for all other gridpoints (outside of black contoured areas). With this, darker shading (i.e. higher NEP) inside of the black percentile contour areas in Figs. 4.1 and 4.2 indicate higher POD values. Alternatively, higher SR values correspond with lighter shading (i.e. lower NEP) outside of the areas encompassed by the black percentile contours in Figs. 4.1 and 4.2. Thus, overall improvement to forecast performance from the baseline SfcRad experiment would entail darker shading inside of the black contoured areas and lighter shading elsewhere.

To evaluate the statistical significance of performance differences between the experiments, bootstrap sampling was done using 1,000 resamples of the forecast UH02 and UH25 NEP fields. The POD, SR, and CSI scores, as well as the differences in these scores between the two experiments, were computed for each resampled distribution. Finally, the mean difference across all 1,000 distributions was computed for each of the three performance indices and displayed in Table 4.1. Difference values shown in bold are statistically significant at the 95% confidence level.

For all three percentile values, the POD is higher for the SfcRad_ALL_LAM experiment (red) than the SfcRad experiment (black), especially at $p99$ where it is more than four times greater for UH02 (Fig. 4.3a) and nearly double for UH25 (Fig. 4.3b).

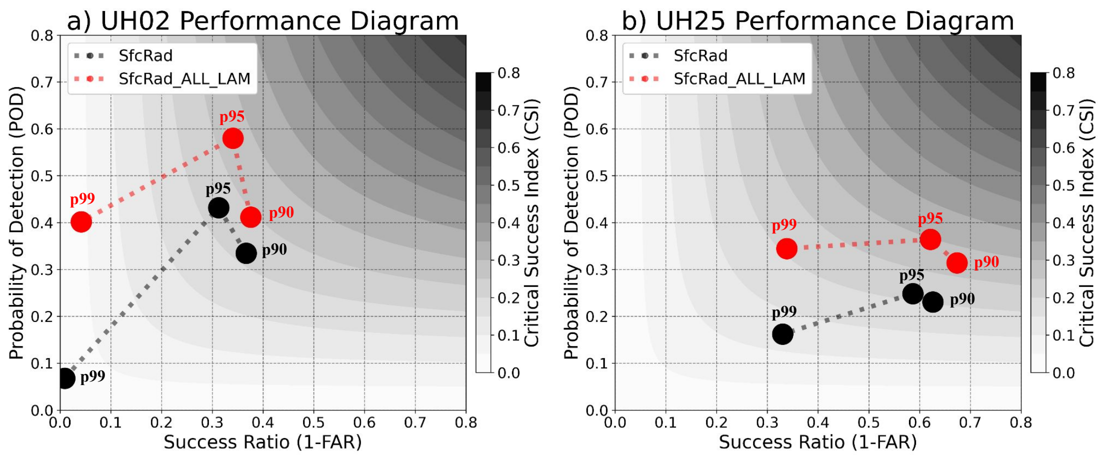

Figure 4.3: Performance Diagrams for Case-1 of a) UH02, and b) UH25 for SfcRad (black) and SfcRad_ALL_LAM (red) at $p90$, $p95$, and $p99$.

This confirms what is visually evident in Figs. 4.1 g-i and 4.2 g-i, where it is clear that the NEP of the forecast UH02 and UH25 exceeding $p99$ is much higher for the SfcRad_ALL_LAM experiment than the SfcRad experiment, especially across western portions of the domain where NEP differences are as high as 50% in and around the $p99$ contoured areas. While not as substantial of a difference is evident at $p90$ and $p95$, the POD of the SfcRad_ALL_LAM experiment is still appreciably higher than SfcRad for both UH02 and UH25. This difference is reflected by higher NEP values in the contoured areas of the respective percentiles for UH02 (Figs. 4.1 a-f) and UH25 (Figs. 4.2 a-f). As shown in Table 4.1, the difference in POD between the two experiments is statistically significant for all percentiles of both variables.

The difference in the SR between the SfcRad_ALL_LAM and SfcRad experiments is less pronounced than the POD difference, but still consistently higher for both UH02 and UH25 at all three percentiles (Fig. 4.3). However, the SR difference is only statistically significant at $p99$ for UH02, and not statistically significant at any of the three percentiles for UH25 (Table 4.1). The SR differences are visually apparent in Figs. 4.1 and 4.2, which depict some areas of lower NEP values outside of the percentile contours for the SfcRad_ALL_LAM experiment than for the SfcRad experiment. The most amplified NEP decreases are generally present outside of but in relatively close proximity to the contoured areas, effectively resulting in dipoles of NEP differences and corresponding sharp difference gradients. This suggests that some of the NEP change could be a result of corrections to the locations of the higher UH tracks in the SfcRad_ALL_LAM experiment, lending to better spatial agreement with the nature run than the SfcRad experiment.

With generally higher POD and SR values contributing to significantly greater CSI scores, the SfcRad_ALL_LAM experiment exhibits better overall performance than the baseline SfcRad in terms of forecasting the 90th, 95th, and 99th percentiles of accumulated maximum UH tracks for Case-1. This improved accuracy indicates that observations from the dense profiling network of the Case-1 SfcRad_ALL_LAM experiment likely contribute to enhancing the predictability of intense low and mid-level mesocyclones. This result, interpreted in the context of the February 26, 2023 case, indicates that more comprehensive observations of the ABL are potentially important for informing accurate model forecasts of QLCS tornado events in highly-forced, robust synoptic regimes.

It is worth noting that despite the relatively prominent NEP differences in the western half of the domain, which is where the convection is present early in the forecast period, the magnitude of these differences deteriorates with eastern extent. The NEPs for all three percentile values of both UH02 and UH25 are notably low (mostly $> 15\%$) for both the SfcRad experiment and the SfcRad_ALL_LAM experiment further east where the convection is present later in the forecast period. As such, the NEP differences between the two experiments are also minimal in these areas, largely remaining in the single digits. Since the trajectory of the simulated convection was from west to east, this suggests that the utility of the assimilated observations likely decreases with time. This is important to consider in the design of a 3D-Mesonet and is investigated further in later sections of this thesis.

Table 4.1: Average differences in the performance metrics: POD, SR, and CSI between the SfcRad_ALL_LAM experiment and the SfcRad experiment from a distribution of 1,000 bootstrap resamples with replacement. Positive values (red shaded cells) indicate forecast performance improvement from the baseline SfcRad experiment while negative values (blue shaded cells) indicate worse performance than SfcRad. The darkness of the shading is proportional to the magnitude of the difference to facilitate comparison. Values in bold are statistically significant differences to the 95% confidence level.

Performance Improvements: SfcRad_ALL_LAM						
	Case-1					
Variable:	UH02			UH25		
	p90	p95	p99	p90	p95	p99
POD	**0.085**	**0.169**	**0.325**	**0.091**	**0.132**	**0.171**
SR	0.004	0.005	**0.084**	0.008	0.003	0.000
CSI	**0.079**	**0.162**	**0.323**	**0.089**	**0.130**	**0.170**
	Case-2					
Variable:	UH02			UH25		
	p90	p95	p99	p90	p95	p99
POD	**0.029**	**0.050**	**0.023**	0.004	**0.039**	**0.054**
SR	**0.059**	**0.048**	0.018	**0.094**	**0.060**	0.018
CSI	**0.034**	**0.053**	**0.022**	0.011	**0.043**	**0.053**

4.1.2 Case-2

As with Case-1, the NEPs of UH02 and UH25 exceeding $p90$, $p95$, and $p99$ of the nature run were calculated and plotted for Case-2 (Figs. 4.4, 4.5). For this case, the nature run percentile values were calculated from a smaller, more focused area of the domain

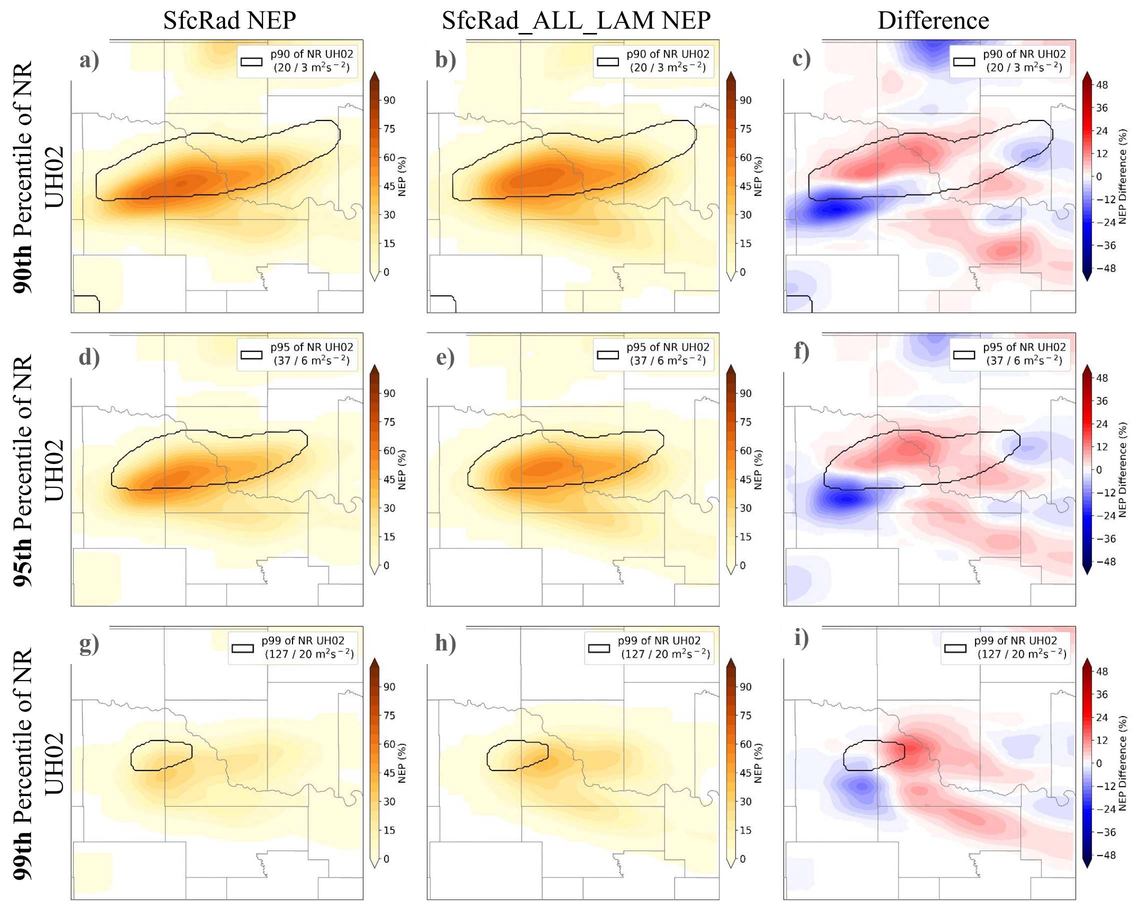

Figure 4.4: Same as Fig. 4.1, but for Case-2.

to isolate the area of interest. The area shown in each panel of Figs. 4.4 and 4.5 is the extent of this designated region. Since the primary hazards associated with this event mostly occurred within this relatively small area, including the remainder of the domain in the calculation of the percentile values and the performance metrics may cause the results to be misinterpreted and heavily influenced by zeroes. Additionally, the nature run forecast simulates two other robust convective features within the domain that did not actually occur. Since the nature run is supposed to depict the true state of the atmosphere, excluding the areas of the domain where this erroneous convection was simulated prevents the analysis of forecast performance from being skewed by data that does not accurately represent reality.

Similarly to the results from Case-1, the NEPs of both UH02 and UH25 exceeding each of the three percentiles are generally higher in the SfcRad_ALL_LAM experiment than the SfcRad experiment within the contoured areas for both UH02 and UH25 (Figs. 4.4 and 4.5). However, the magnitude of these differences between the two experiments is generally smaller than Case-1 (Figs. 4.1, 4.2). Additionally, the areas of enhanced NEP are more localized, more zonally compact, and less well aligned with the percentile contours, indicating that the profiling network observations have less positive influence on the forecast UH swaths in this case. These apparent differences between the two cases suggests that the value of assimilating ABL observations may not be consistent for all scenarios or environments. Conducting OSSEs for other cases would be necessary to better understand this variability.

The less apparent NEP differences, and the generally lower NEP values associated with Case-2 overall could also be partially due to greater ensemble spread in the forecast UH swaths. With how NEP is calculated, more ensemble spread would inherently cause lower magnitudes of NEP. As shown in Fig. 4.6, the standard deviation of UH02 across the ensemble members of the SfcRad_ALL_LAM forecast is considerably higher in the area of interest for Case-2 than for Case-1, indicating greater spread between the ensemble members. The high ensemble spread in Case-2 is potentially due to the relatively subtle forcing and conditional nature of the convective environment associated with the presence of the capping inversion that most operational CAM forecasts indicated would suppress convective initiation. Visual analysis of simulated composite reflectivity from the various ensemble members of the forecast run (not shown) reveals inconsistency regarding when, where, and if convection would form, as well as its intensity and spatial coverage. While relative comparison between different experiments should not be substantially impacted by the large ensemble spread since they all originate from the same forecast run, directly comparing experiment forecast performance metrics between Case-1 and Case-2 is not necessarily representative of the value of a profiling network and how it may change for different events. This is important to keep in mind throughout the interpretation of the results for this case.

Performance diagrams for the Case-2 SfcRad and SfcRad_ALL_LAM experiments are shown for both UH02 and UH25 in Fig. 4.7. While the POD differences between the two experiments are less pronounced for Case-2 than Case-1, the SfcRad_ALL_LAM experiment still performs better for all assessed percentile values of UH02 and UH25 than the SfcRad experiment, as evidenced by higher CSI values. The CSI differences between the two experiments are statistically significant at all percentiles for both variables with the exception of $p90$ of UH25 (Table 4.1). Additionally, the differences in the SR values are more apparent in Case-2, with statistically significant improvements from the baseline SfcRad forecast at $p90$ and $p95$. This indicates that the simulated observations contributed to a reduction in erroneous UH tracks.

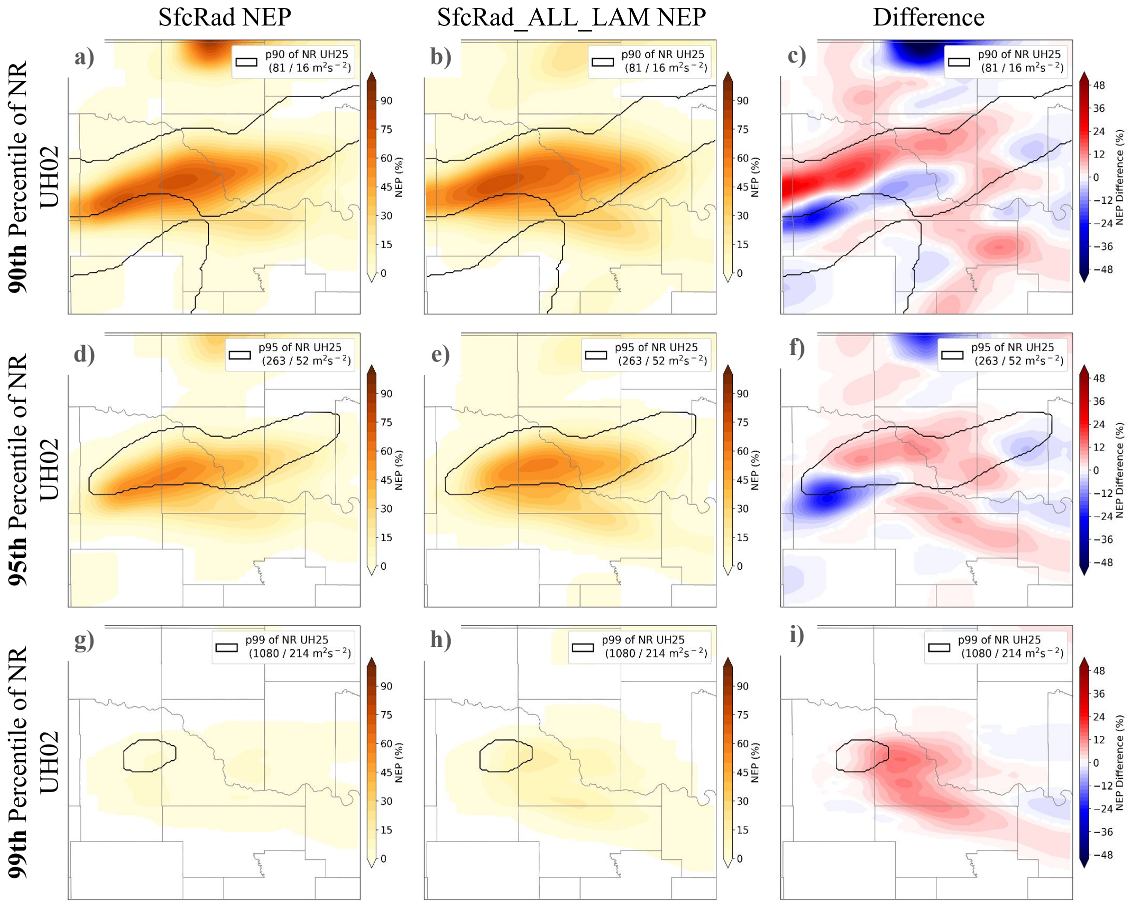

Figure 4.5: Same as Fig. 4.2, but for Case-2.

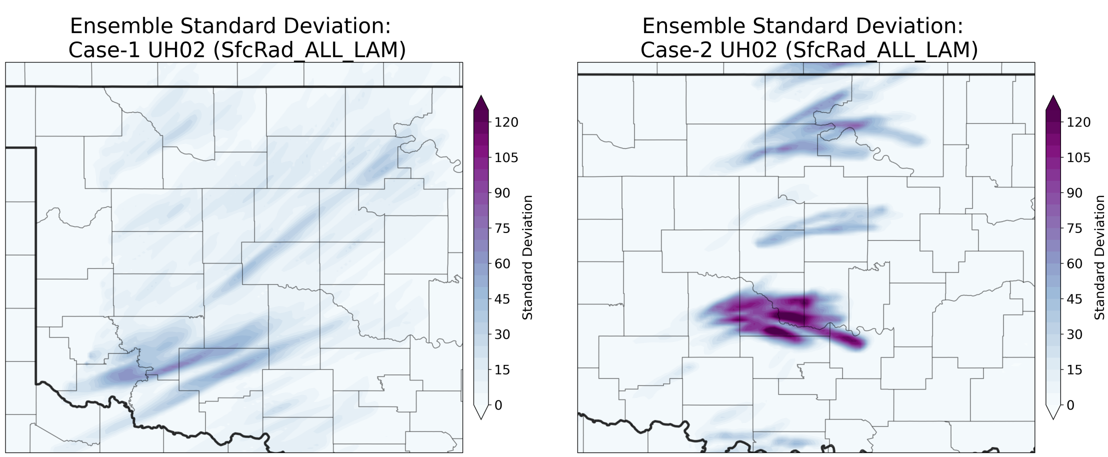

Figure 4.6: Standard deviation of forecast accumulated hourly maximum UH02 across all ensemble members for a) Case-1 and b) Case-2.

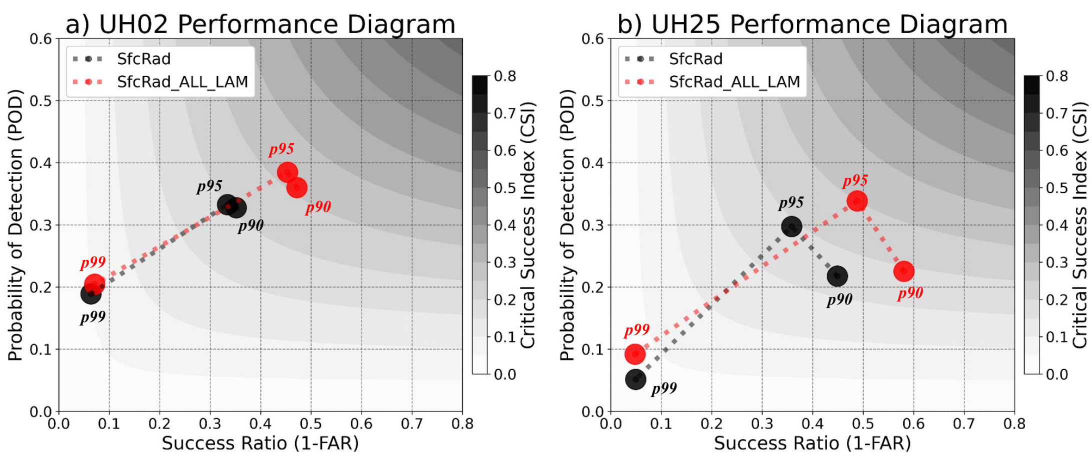

Figure 4.7: Same as Fig. 4.3, but for case-2.

4.2 Reducing Network Size

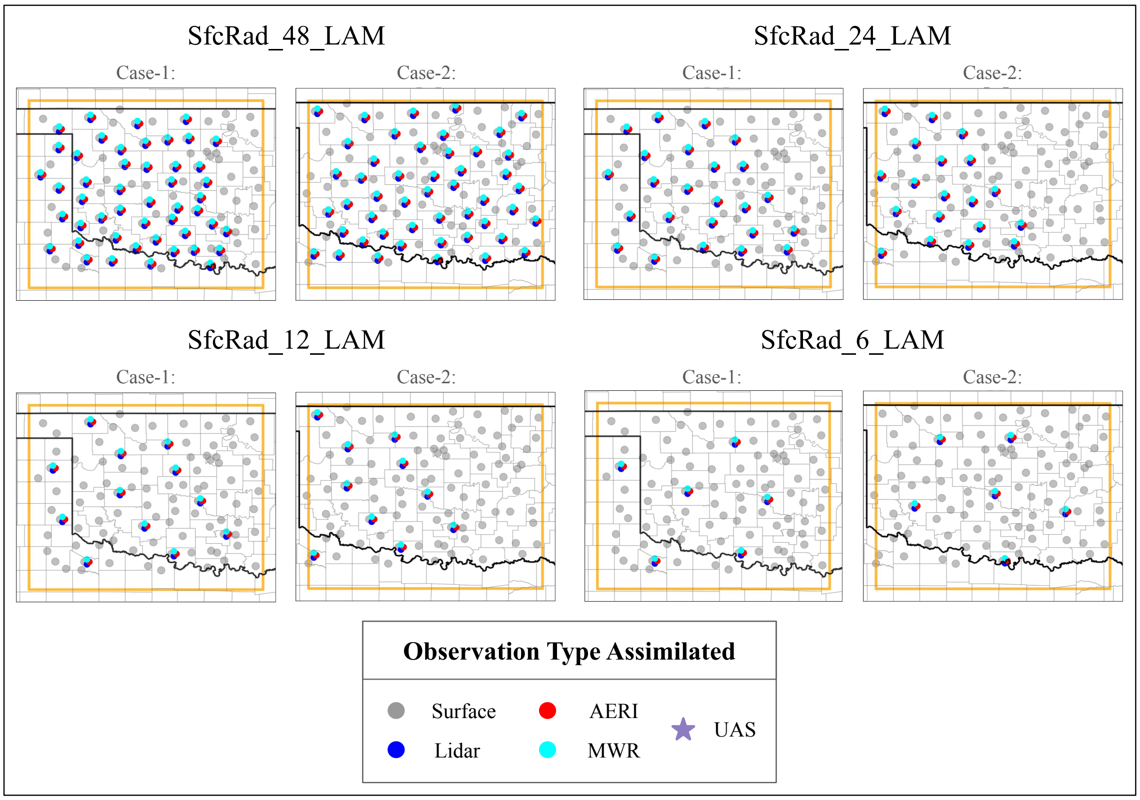

Figure 4.8: Network visualizations for the reduced-site experiments.

4.2.1 Case-1

The previous section established that assimilating simulated observations from three different remote profiling systems (lidar, MWR, and AERI) at every mesonet site within the case-respective nature run domain (i.e. the SfcRad_ALL_LAM experiments) generally contributed to statistically significant forecast performance improvements in terms of the location and intensity of accumulated hourly maximum UH tracks (i.e. UH02,

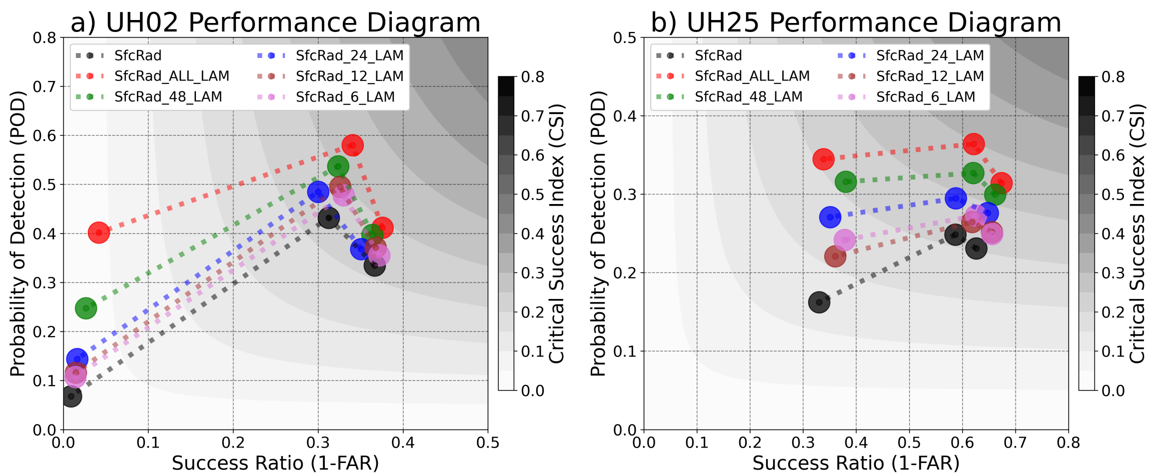

Figure 4.9: Performance Diagrams for case-1 of a) UH02 POD as a function of SR for SfcRad (black), SfcRad_ALL_LAM (red), SfcRad_48_LAM (green), SfcRad_24_LAM (blue), SfcRad_12_LAM (brown), and SfcRad_6_LAM (pink) for the $p90$ (rightmost dot), $p95$ (center dot), and $p99$ (leftmost dot) values. b) Same as (a), but for UH25.

UH25). While this result shows that a comprehensive 3D-Mesonet could be quite valuable for model forecasts of severe convection, a robust network such as this would be expensive and difficult to maintain. While the cost of these remote profiling instruments is variable, each one typically costs on the order of hundreds of thousands of dollars, not including additional costs associated with routine operation and maintenance. For the purpose of a rough approximation of the total cost for such a network, a reasonable estimate for the cost of each site is around six-hundred thousand U.S. dollars (USD). The mesoscale network of the SfcRad_ALL_LAM experiments would therefore cost upwards of sixty-million USD to implement across just one state. Profiling networks of this density are likely beyond what is currently feasible considering financial limitations. With this in mind, the SfcRad_ALL_LAM experiments primarily serve as approximations for the improvements to forecast performance that could result from a network that mirrors the spatial resolution of existing mesoscale surface observing networks.

To gauge the utility of more practical network designs, experiments were run with major reductions to the amount of sites making up the network. Specifically, four "reduced-site" experiments were conducted with 48 sites, 24 sites, 12 sites, and 6 sites (Fig. 4.8). For each of these experiments, simulated observations from all three remote profiling instruments (Lidar, AERI, and MWR) were assimilated at each site. Performance diagrams for UH02 and UH25 were created, each of which contain plots of POD with respect to SR at $p90$, $p95$, $p99$ for each of the four reduced-site experiments, as well as the SfcRad and SfcRad_ALL_LAM experiments for reference (Fig. 4.9). All reduced-site experiment forecasts clearly perform better than the SfcRad experiment for Case-1 with statistically significant improvements to POD and CSI scores at all percentiles for both UH02 and UH25 (Table 4.2). Even a network with just six sites (i.e. the SfcRad_6_LAM experiment) promotes a significantly enhanced ability to correctly forecast intense UH tracks, indicating that even relatively sparse profiling networks can still be valuable to convective model forecasts.

In general, the forecast performance scores decrease with decreasing number of sites in the simulated network. However, the rate at which the performance improvements decrease between experiments also decreases. The performance improvements associated with the SfcRad_6_LAM experiment, for instance, are comparable to the SfcRad_12_LAM experiment (Table 4.2) which has double the amount of sites and therefore would be twice as expensive to implement. This suggests that for more sparse, practical networks, adding additional sites provides relatively little added benefit to the forecast. Further research focusing on the relationship between the number of sites and forecast performance will be critical for identifying an ideal balance between the cost and utility of an operational profiling network.

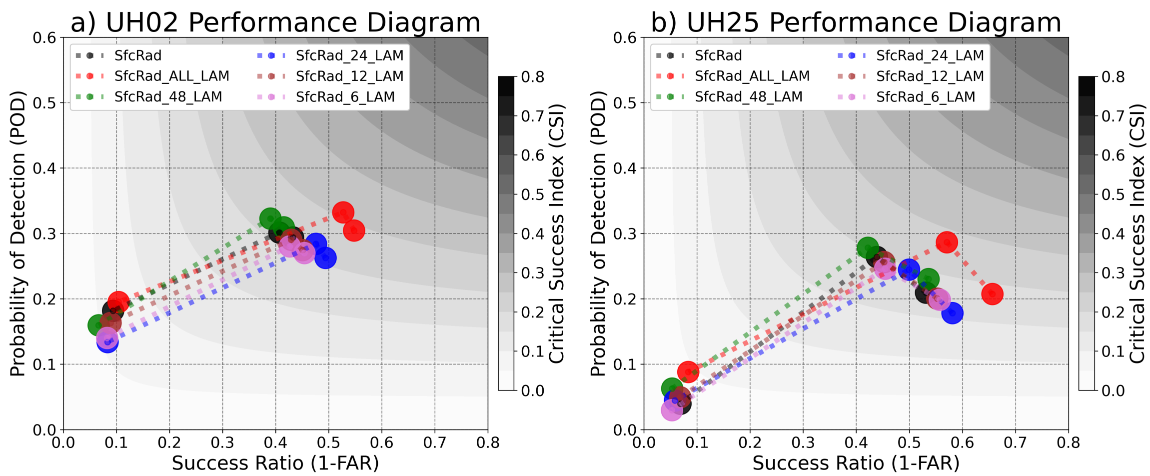

Figure 4.10: Same as Fig. 4.9 but for Case-2

4.2.2 Case-2

Reduced site experiments with networks consisting of 48 sites, 24 sites, 12 sites, and 6 sites were also conducted for Case-2 (Fig. 4.8). As depicted in Fig. 4.10 the differences in forecast performance metrics between these experiments are generally smaller in magnitude than for Case-1 (Fig. 4.7). However, a positive correlation between the number of sites in the network and the forecast performance improvement is still generally apparent at all percentiles. The most notable result of the Case-2 reduced-site experiments is the general lack of improvement from the baseline SfcRad forecast. The only experiment with statistically significant CSI improvements is the SfcRad_48_LAM experiment (Table 4.2). In fact, the other three reduced-site experiments perform worse than the SfcRad forecast, with statistically significant CSI decreases at all percentiles for both UH02 and UH25. While more cases would be required to fully understand the implications of these results, a possible interpretation is that small scale features or hyper-local gradients in the structure of the lower-atmosphere played a critical role in the outcome of this particular event. Circulations along the dryline, for example, may have been the primary impetus for the initiation of convection. Extremely high-resolution ABL observations would have likely been required to meaningfully capture the features and processes present along the dryline that may have substantially impacted how the convective storms formed and evolved. If this is the case, even the most densely spaced networks, such as those associated with the SfcRad_ALL_LAM and SfcRad_48_LAM experiments, were likely of insufficient spatial resolution. More conditional, subtly-forced convective environments such as this may benefit more from a dense, localized profiling network deployed in the vicinity of features, in this case the dryline, that may be important to the evolution of convection.

The general lack of improvement to forecast performance in the Case-2 experiments may also be related to limitations associated with this particular OSSE methodology. For example, current convective-scale models inherently lack the ability to properly resolve the small scale features that may have been crucial to the evolution of this event. Even if a profiling network was able to provide observations of such features, it is possible that some observations would be rejected by the filtering step of the data assimilation process. The likelihood of observation rejection would increase for more sparse networks that have less influence on the overall ensemble distribution. While outside the scope of this thesis, exploring the rate of observation rejection in the data assimilation and how it changes in different network designs would be a worthwhile topic of focus in future exploration into the optimal design of a profiling network.

Table 4.2: Same as Table 4.1, except for all reduced-site experiments.

Performance Improvements: Reduced-Site Experiments												
SfcRad_48_LAM												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.067**	**0.118**	**0.164**	**0.074**	**0.089**	**0.138**	**0.016**	**0.022**	**-0.024**	0.021	0.015	**0.022**
SR	0.000	0.003	**0.069**	0.006	0.003	0.001	-0.010	-0.009	**-0.052**	-0.003	-0.010	-0.043
CSI	**0.062**	**0.113**	**0.162**	**0.072**	**0.087**	**0.137**	**0.013**	**0.019**	**-0.024**	**0.017**	**0.012**	**0.022**
SfcRad_24_LAM												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.037**	**0.062**	**0.072**	**0.047**	**0.053**	**0.101**	**-0.032**	-0.017	**0.049**	**-0.031**	**-0.019**	**0.004**
SR	-0.004	-0.002	0.047	0.003	0.001	0.001	**0.033**	**0.030**	-0.017	**0.041**	**0.031**	-0.029
CSI	**0.033**	**0.058**	**0.071**	**0.045**	**0.052**	**0.100**	**-0.025**	-0.013	**-0.047**	**-0.024**	**-0.014**	**0.004**
SfcRad_12_LAM												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.040**	**0.066**	**0.045**	**0.021**	**0.018**	**0.052**	**-0.019**	**0.011**	**-0.019**	**-0.009**	**-0.007**	**0.010**
SR	0.001	0.003	0.045	0.006	0.003	0.001	0.011	**0.011**	-0.007	**0.017**	0.007	0.001
CSI	**0.037**	**0.064**	**0.045**	**0.021**	**0.018**	**0.052**	**-0.016**	**-0.009**	**-0.018**	**-0.006**	**-0.006**	**0.009**
SfcRad_6_LAM												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.025**	**0.049**	**0.036**	**0.021**	**0.027**	**0.071**	**-0.024**	**-0.021**	**-0.042**	**-0.011**	**-0.018**	**-0.010**
SR	0.003	0.004	0.04	0.007	0.004	0.002	**0.012**	0.009	**-0.018**	**0.022**	**0.009**	**-0.044**
CSI	**0.023**	**0.048**	**0.036**	**0.020**	**0.027**	**0.071**	**-0.020**	**-0.019**	**-0.040**	**-0.007**	**-0.015**	**-0.010**

4.3 Instrument-Exclusion Experiments

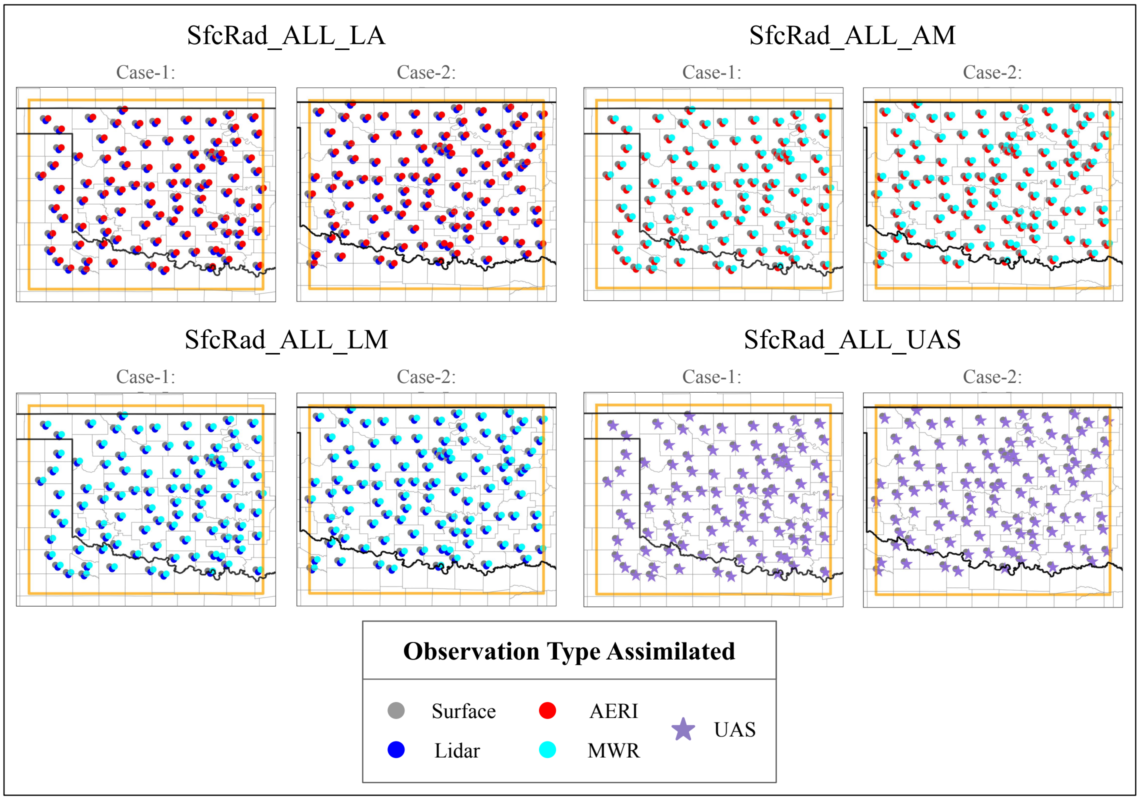

Figure 4.11: Network visualizations for the different instrument-exclusion experiments.

4.3.1 Case-1

The SfcRad_ALL_LAM experiment, and the reduced-site variations thereof, assimilated simulated observations from three remote profiling instruments at every site: a lidar for kinematic observations, an AERI for thermodynamic observations, and an MWR, also for thermodynamic observations. This combination of instruments is considered to be an optimal combination for collecting quality high-resolution information

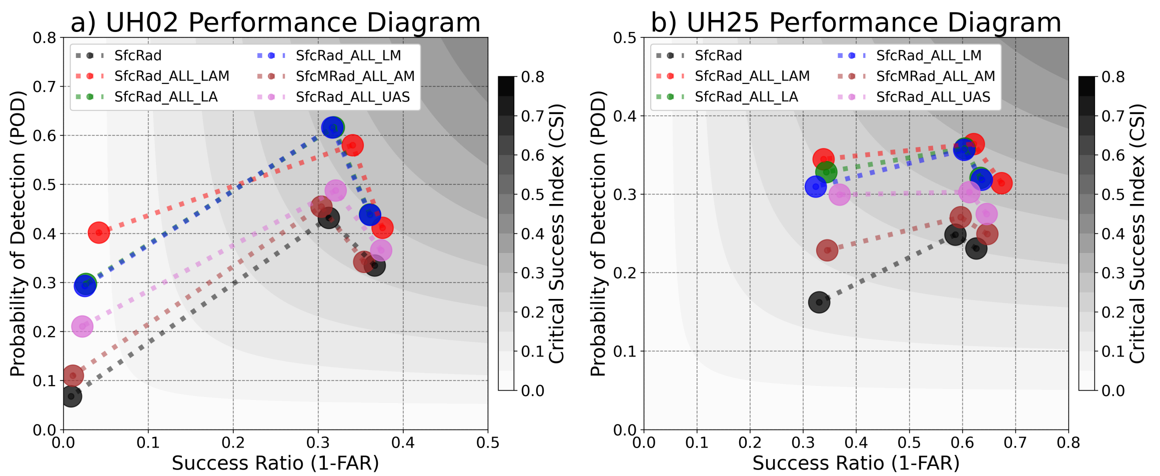

Figure 4.12: Performance Diagrams for Case-1 of a) UH02 POD as a function of SR for SfcRad (black), SfcRad_ALL_LAM (red), SfcRad_ALL_LA (green), SfcRad_ALL_LM (blue), SfcRad_ALL_AM (brown), and SfcRad_ALL_UAS (pink) for the $p90$ (rightmost dot), $p95$ (center dot), and $p99$ (leftmost dot) values. b) Same as (a), but for UH25.

about the dynamic evolution of the lower troposphere (e.g. Otkin et al. (2011)). However, including three different remote profiling systems at every site in a mesoscale network would be expensive, even for networks with fewer sites. Reducing the number of instruments at each site would help to alleviate the financial burden of a 3D-Mesonet and make operational deployment more feasible.

To evaluate the most efficient way to cut back on instrumentation, three new experiments were conducted, each of which excludes observations from a different one of the three remote profiling instruments while retaining observations from the other two (Fig. 4.11). In each of these "instrument-exclusion" experiments, observations are assimilated at all mesonet sites in the nature run domain. Any reduction in forecast performance from the SfcRad_ALL_LAM experiment is therefore analogous to the value provided by the instrument that was excluded in a given experiment. One other experiment was conducted alongside these instrument-exclusion experiments in which simulated observations at each site were only assimilated from the Coptersonde WxUAS.

Performance diagrams were created for UH02 and UH25 at $p90$, $p95$, and $p99$ (Fig. 4.12) to compare the performance of the instrument-exclusion and Coptersonde Wx-UAS experiments with the original SfcRad and SfcRad_ALL_LAM experiments. The SfcRad_ALL_LA and SfcRad_ALL_LM experiments, which exclude observations from the MWR and AERI, respectively, have nearly equivalent performance scores. This suggests that observations from the AERI and MWR are of approximately equal value to the forecast in this case. In addition, aside from at $p99$, the performance metrics of both of these experiments are comparable to the SfcRad_ALL_LAM experiment, indicating that the added value of having two thermodynamic profilers at each site rather than just one is unlikely to be worth the extra monetary investment, at least in the context of convective model DA.

The SfcRad_ALL_AM experiment, which excludes the kinematic observations from the lidar, performs the worst of all instrument exclusion experiments (Fig. 4.12). In fact, the SfcRad_ALL_AM experiment forecast exhibits minimal performance improvement over the baseline SfcRad experiment, with statistically significant CSI increases at only $p99$ (Table 4.3). This suggests that the kinematic observations provided by the lidar provide more value to the forecast than the thermodynamic observations provided by the other two remote profiling instruments, at least for this case. This could be a result of the intense low-level wind field associated with this event causing advection to dominate thermodynamic changes in the ABL. In this case, the thermodynamic structure of the ABL would be highly correlated with the kinematic structure such that direct thermodynamic observations would provide little additional information content. Analysis of OSSE experiments conducted for other cases with relatively quiescent kinematic environments would be necessary to evaluate if the impact of thermodynamic observations on forecast performance increases when the influence of advection is reduced. Nevertheless, the results of the Case-1 instrument-exclusion experiments provide some initial indication that thermodynamic profiling instruments may be unnecessary for DA applications in cases with intense low-level flow regimes.

4.3.2 Case-2

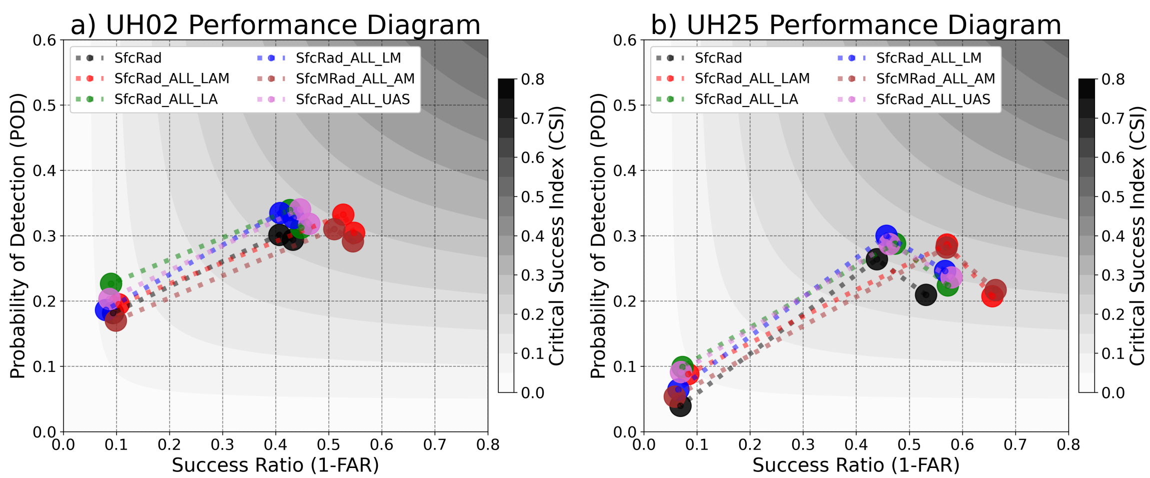

Figure 4.13: Same as Fig. 4.12 but for Case-2

The same instrument-exclusion experiments, as well as the SfcRad_ALL_UAS experiment, were conducted for Case-2 (Fig. 4.11). As evident in Fig. 4.13, the instrument-exclusion experiments for Case-2 differ less from each other in terms of forecast performance than the Case-1 experiments. Nevertheless, with the exception of the SfcRad_ALL_AM experiment, statistically significant CSI improvements from the SfcRad experiment are apparent for all instrument-exclusion experiments at all percentiles of both variables (Table 4.3). While less apparent than for Case-1, these results also emphasize that kinematic observations may have a greater impact on forecast performance than thermodynamic observations, specifically in regard to accurately predicting intense UH tracks.

Table 4.3: Same as Table 4.1, but for all instrument-exclusion experiments.

Performance Improvements: Instrument-Exclusion Experiments												
SfcRad_ALL_LA												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.077**	**0.151**	**0.220**	**0.065**	**0.083**	**0.113**	**0.019**	**0.039**	**0.044**	**0.014**	**0.024**	**0.059**
SR	0.000	0.0024	**0.073**	0.0014	0.0023	0.000	0.008	0.009	-0.006	0.034	0.018	0.008
CSI	**0.097**	**0.180**	**0.217**	**0.089**	**0.118**	**0.144**	**0.018**	**0.037**	**0.041**	**0.014**	**0.024**	**0.057**
SfcRad_ALL_LM												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.104**	**0.189**	**0.219**	**0.090**	**0.116**	**0.127**	**0.034**	**0.034**	**0.004**	**0.036**	**0.036**	**0.025**
SR	0.000	0.002	**0.066**	0.002	0.001	0.000	0.001	0.001	**-0.021**	0.029	0.009	-0.009
CSI	**0.096**	**0.179**	**0.216**	**0.087**	**0.114**	**0.127**	**0.030**	**0.031**	0.003	**0.033**	**0.034**	**0.024**
SfcRad_ALL_AM												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	0.011	**0.026**	**0.030**	0.019	**0.024**	**0.060**	-0.001	0.011	-0.012	0.009	0.020	**0.013**
SR	-0.003	-0.001	0.014	0.037	0.002	0.000	**0.056**	**0.042**	0.006	**0.104**	**0.060**	-0.029
CSI	0.009	0.025	**0.030**	0.018	0.024	**0.060**	0.005	0.014	-0.011	0.015	**0.024**	0.013
SfcRad_ALL_UAS												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.038**	**0.062**	**0.124**	**0.045**	**0.059**	**0.133**	**0.025**	**0.039**	**0.021**	**0.027**	**0.024**	**0.051**
SR	0.005	0.002	**0.062**	0.002	0.002	0.000	0.018	0.017	-0.010	0.038	0.011	0.001
CSI	**0.036**	**0.059**	**0.123**	**0.044**	**0.059**	**0.133**	**0.024**	**0.038**	**0.019**	**0.026**	**0.022**	**0.050**

4.4 Small Targeted Networks

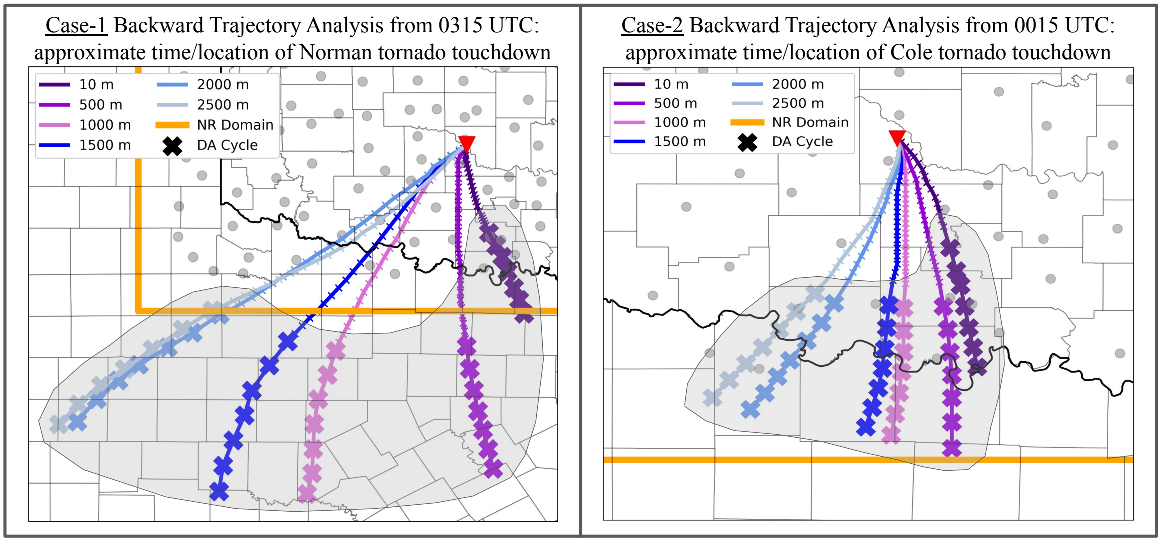

Figure 4.14: Backward trajectory analyses of parcels at various AGL levels (every 500 m from 10 m to 2500 m) originating at the approximate location and time of the start of the Norman, OK and Cole, OK tornadoes (red triangles) for Case-1 (left panel) and Case-2 (right panel), respectively. Each tick mark on the trajectory lines represents the approximate parcel location at 5-minute intervals. Each solid "X" is the approximate parcel location at each DA cycle (15-minute intervals). Light gray shaded regions are the subjectively estimated regions of origin of ABL parcels that would theoretically be observed during the DA period. Gray dots indicate site locations from the full networks associated with the SfcRad_ALL_LAM experiments.

4.4.1 Advection of Observation Impacts

For all of the network design possibilities that have been explored so far, the individual sites have been dispersed across the domain, regardless of the number of sites. This type of network distribution would accommodate the largest area and provide the opportunity for the data to be utilized in as many situations as possible. As such, this type of network configuration would probably be the most versatile option for operational implementation. There are also some downsides to this approach, especially as the number of sites in the network decreases and horizontal resolution is lost as a result. As seen with the Case-1 SfcRad_6_LAM experiment, a network of just six widely spread sites can still contribute to statistically significant improvements to model forecast performance (Table 4.2). Contrarily, however, the Case-2 SfcRad_6_LAM experiment shows that a network such as this can also provide no value in some cases, and even be detrimental to forecast performance. In cases such as this, a widely spaced network is unlikely to provide observations of small-scale features that may be a primary catalyst for the weather hazards that occur. A more dense, localized profiling network would be more capable of observing such features and would therefore potentially be more valuable for improving model forecasts of events like Case-2. With these points in mind, the optimal design of a mesoscale network of ABL profiling systems is likely case-dependent.

Another possible disadvantage of a static, spread out network becomes apparent when considering the impacts of parcel advection in the background flow. Air parcels, and the observations associated with them, inevitably translate horizontally with the wind and therefore lose relevance to the location at which they were observed with time. This is especially problematic for cases with strong low-level flow regimes, such as Case-1, where southerly flow was in excess of 80 kts just a kilometer above the surface (about 850 hpa; Fig. 2.4). In this environment, observations collected from the northern portions of the domain effectively "advect" out of the domain, thus rendering them increasingly unlikely to have any impact on the forecast, at least for within the area of interest. This effect could be a potential explanation for the decreased NEP differences between the SfcRad_ALL_LAM experiments and SfcRad experiments with eastward extend and a corresponding progression in time. As time advances, the displacement of the assimilated observations from their origin increases which degrades their relevance. This could also possibly be partly why the simulated kinematic observations from the lidar seem to generally provide the greatest value to forecast performance, especially for Case-1, as suggested by the results of the instrument-exclusion experiments. Considering this possibility, it seems valid to conjecture that assimilating observations from upstream of the primary area of interest should prolong their positive impact on the forecast performance. If this is the case, tracing air parcels back in time from the area in which the most impactful weather occurred to where they would have been during the DA period, may reveal where observations should have been assimilated from to maximize their impact on the forecast at the time and location that is most relevant to the event.

To gain insight into the origin of boundary layer parcels that were present over the primary areas of interest for each case, backward trajectory analyses were computed for each case from origins corresponding with the approximate timing and location of where the most significant observed tornadoes associated with each case occurred (Fig. 4.14). The trajectories were computed by integrating the horizontal wind components from the nature run forecast backward in time in 5-minute intervals until the last DA cycle with each time step corresponding with a forecast output file. From the last DA time, the backward trajectory was continued until the beginning of the DA period but at 15-minute intervals. A trajectory was computed every 500-m spanning from the surface (represented by 10-m AGL) to 2,500-m AGL which is the maximum height for which simulated UAS observations could be assimilated. The horizontal wind components used to calculate the trajectories were assessed at constant pressure levels that were defined based on which height level they corresponded with at the parcel origin (red triangles is Fig. 4.14). The approximate parcel location in terms of zonal (x) and meridional (y) displacement from the previous time step can be expressed as:

$$(x(t), y(t)) = \int_{t}^{t-\Delta t} \vec{u}_z(t)\, dt\imath \ + \vec{v}_z(t)\, dt\jmath$$

(4.1)

where u_z and v_z are the horizontal wind components at height z at the t-th timestep from the origin, and Δt is 5 minutes for the forecast period and 15 minutes for the DA period.

This process was done to visualize the general trajectories of low-level air parcels that were spatially and temporally relevant to the outcome of the events. The trajectories reveal where these relevant parcels were approximately located during the DA cycling period. From the left panel of Fig. 4.14, it is evident that ABL parcels in the vicinity of the tornado that was observed in Norman for Case-1 were largely outside of the nature run domain during the assimilation period. The only exception is near-surface air parcels, as represented by the 10-m AGL parcel trajectory, which are generally observable by the existing surface instrumentation at OK-Mesonet sites. The assimilated profiling network observations do not have a direct influence in the region surrounding the tornado at the time it occurs. I hypothesize that the advection of observation impacts beyond the area of interest is a possible reason for the negligible NEP differences between the SfcRad and SfcRad_ALL_LAM experiments in central and eastern Oklahoma (Figs. 4.1 and 4.2). Either increasing the size of the nature run domain or shifting it south to compensate for advection may have facilitated more substantial forecast improvements, particularly with eastward extent.

The background low-level flow was weaker across the nature run domain for Case-2 than it was for Case-1. The low-level parcel trajectories are therefore shorter which partially mitigated the negative impact of advection on simulated observations (Fig. 4.14b). As a result, the majority of the boundary layer in proximity to the event's most significant observed tornado originated and remained inside of the nature run domain during the DA period. The simulated observations should therefore have more impact on the modeled structure of the atmosphere within the area of interest in central Oklahoma where the localized tornado outbreak occurred. For similar reasons, as in Case-1, I hypothesize the reduced advection of observation impacts lead to more pronounced differences in NEP of high UH between the SfcRad and SfcRad_ALL_LAM experiments (Figs. 4.4 and 4.5). While the areas of greatest NEP differences are much more amplified in Case-1, these areas are isolated to the western portion of the domain where the convection was present early in the forecast period while the parcels that informed the simulated observations had not yet advected outside of the domain and were still directly influencing the model representation of the atmospheric structure.

4.4.2 Targeted 6-UAS Networks

Results from the previous section suggest that observations from potentially large portions of a dense profiling network may have limited utility for improving forecast performance in neighboring regions due to advection. If so, the concept of strategically targeted observations becomes attractive. Some research has already shown that targeted observations can be used to efficiently improve forecast performance (e.g. Majumdar 2016; Keclik et al. 2017; Hill et al. 2020; Arseneau and Ancell 2023). This concept, which would require more mobile profiling systems such as the WxUAS, could involve strategically choosing deployment locations for the network sites the day of, or even just hours before the targeted event. This could make observations from profiling systems more efficient and effective for forecasting, especially for events that occur with highly dynamic environments.

As depicted in the left panel of Fig. 4.14, low-level air parcels in and around the time and area that was most heavily impacted by the tornadoes in Case-1 were well

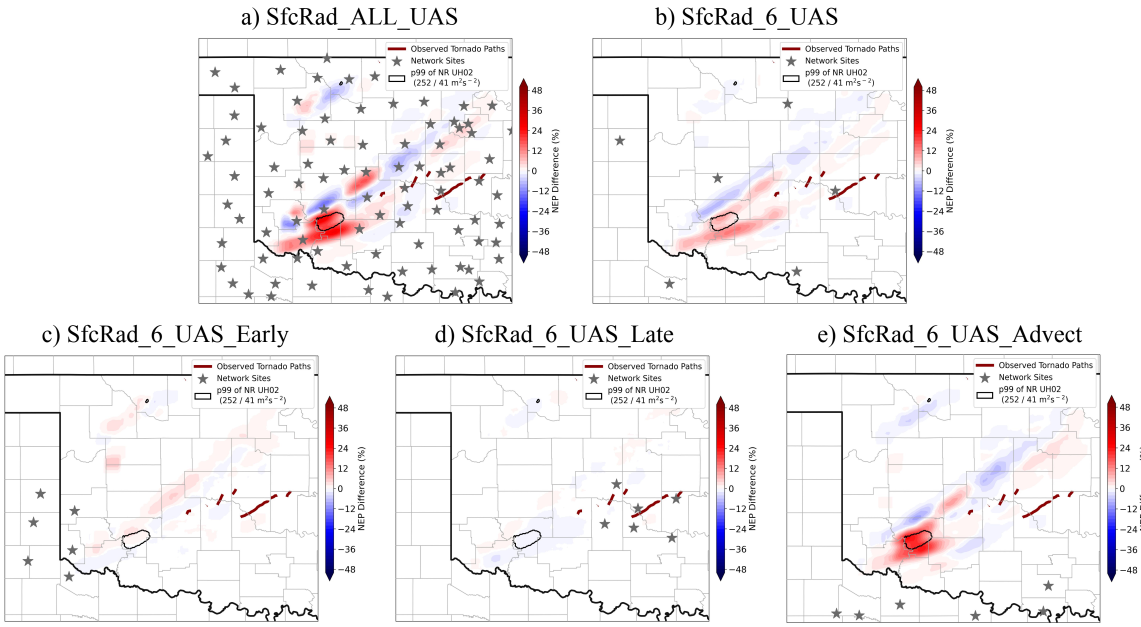

Figure 4.15: The difference in the NEP of UH02 exceeding $p99$ between the: a) SfcRad_ALL_UAS, b) SfcRad_6_UAS, c) SfcRad_6_UAS_Early, d) SfcRad_6_UAS_Late, and e) SfcRad_6_UAS_Advect experiments and the SfcRad experiment (shading) for Case-1. The black contour indicates the nature run $p99$ area. Gray stars denote the locations of the UAS sites that make up the network assessed in the respective experiment. Maroon lines depict observed tornado paths.

outside of the nature run domain throughout the DA period. Therefore, the majority of the simulated observations assimilated likely had little-to-no impact on the forecast for the primary area of focus by the time the convection arrived. Due to the intense low-level flow regime, the assimilated observations were associated with air parcels that likely resided in the northern portion of the domain or beyond when the observed tornadoes occurred in central OK, well away from their area of impact. However, a few of the sites in the southernmost portion of the SfcRad_ALL_LAM experiment network were just within, or very close to, the area where near-surface air parcels that were over central OK at the time of the tornado existed during the DA period (light gray shaded region in Fig. 4.14). As such, observations from these few sites may have influenced the model representation of the low-level atmosphere within the target area.

Additionally, observations from the later DA cycles at the south-westernmost sites in the domain may have "advected" right into the path of the convection, potentially impacting the modeled environment ahead of the developing storms. In section 3.5.1, I hypothesized that the most pronounced NEP differences of high UH values in the SfcRad_ALL_LAM experiment are early in the forecast period (Figs. 4.1, 4.2), since this is when the influence of assimilated observations is pertinent to the area that is relevant to the early evolution of the convection.

To evaluate my hypothesis, a targeted simulation was ran which assimilated simulated observations from six profiling sites strategically chosen based upon the backward trajectory analysis in Fig. 4.14. The six chosen sites, the locations of which are indicated by the grey stars in Fig. 4.15e, are intended to maximize the duration that the observations are directly relevant to the atmosphere in proximity to the targeted convection. In this experiment, which is hereon referred to as the SfcRad_6_UAS_Advect experiment, simulated Coptersonde WxUAS observations are assimilated at each site. UAS profiling systems have the advantage of increased versatility and portability over their remote-profiling counterparts which makes them the profiling instrument most adept to the application of targeted networks, which would be deployed in different locations for each unique event. Additionally, the need for deploying two different profiling systems at each site can be avoided since WxUASs collect both thermodynamic and kinematic observations simultaneously. Furthermore, since WxUAS observations are in-situ, they generally involve less frequent and less precise calibration which makes them even more suitable for rapid, flexible deployment.

As in the other assimilation experiments, the NEP of UH02 and UH25 exceeding the 90th, 95th, and 99th percentiles of the nature run forecast ($p90$, $p95$, $p99$) was computed across the domain. To visualize the relative impact of the SfcRad_6_UAS_Advect experiment network on the Case-1 simulated UH tracks, the NEP difference from the baseline SfcRad experiment for UH02 at $p99$ is plotted in Fig. 4.15e, a process identical to what was used to assess the SfcRad_ALL_LAM experiment at $p99$ in Fig. 4.1i. Within and around the $p99$ contour area, bright red shading indicates NEP values approximately 15-25% higher in the SfcRad_6_UAS_Advect experiment than the SfcRad experiment, despite observations only being assimilated from six sites.

To contextualize this result, two additional targeted WxUAS experiments were conducted for Case-1. These experiments are the SfcRad_6_UAS_Early experiment, for which the six sites encompassed the region where the convection of focus first developed (Fig. 4.15c), and the SfcRad_6_UAS_Late experiment, in which the six sites were located in the area that was most heavily impacted by the convective hazards (i.e. the tornadoes) late in the period (Fig. 4.15d). In a realistic scenario, it would make sense to chose one of these two areas for deployment of a targeted profiling network. Better observations from where the convection initiated could potentially improve how the early stages of the convective evolution were forecast due to a more accurate representation of the proximal boundary layer environment (Romine et al. 2016; Wade et al. 2018; Laser et al. 2022). It would be reasonable to presume that improving how the early stages of convection were modeled would improve the remainder of the forecast. Deploying a targeted network where the most impactful hazards are anticipated also seems like a justifiable decision, as the observations provided by the network would promote a more accurate representation of the atmosphere where it is expected to matter most. However, in this case, assimilating observations from the SfcRad_6_UAS_Early network resulted in minimal NEP changes not only within and near the $p99$ contour, but also throughout the majority of the domain (Fig. 4.15c). The SfcRad_6_UAS_Late experiment also appears to have minimal impacts on NEP values for UH02 (Fig. 4.15d). The lack of changes in NEP from SfcRad for both the SfcRad_6_UAS_Early and SfcRad_6_UAS_Late experiment is likely once again due to the intense low-level flow associated with this particular case. As before, the observations are unlikely to provide much information about the target area due to advection. For other scenarios with weaker flow regimes, perhaps even Case-2, the SfcRad_6_UAS_Early and SfcRad_6_UAS_Late networks may be more impactful due to reduced advection of the observations.

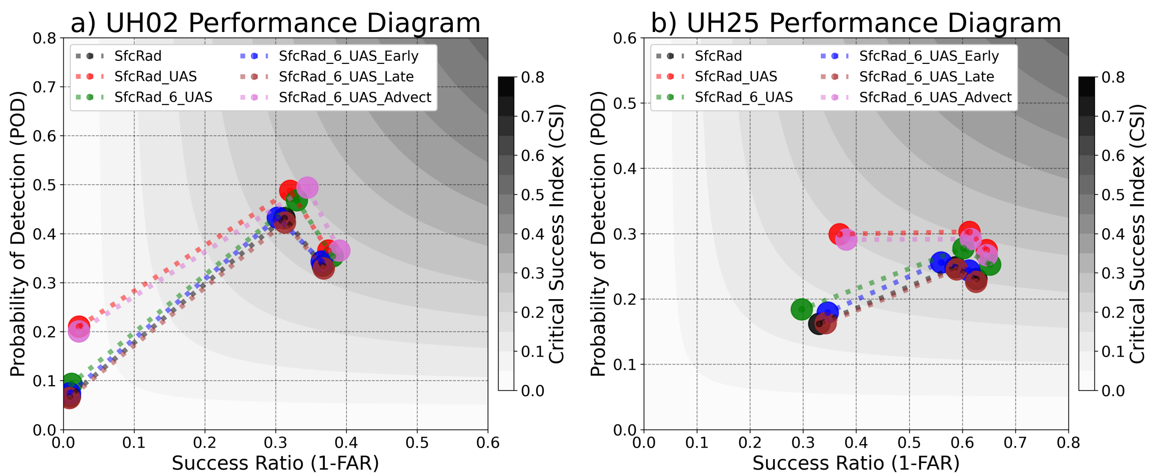

Figure 4.16: Performance Diagrams for case-1 of a) UH02 POD as a function of SR for SfcRad (black), SfcRad_ALL_UAS (red), SfcRad_6_UAS (green), SfcRad_6_UAS_Early (blue), SfcRad_6_UAS_Late (brown), and SfcRad_6_UAS_Advect (orange) for for the $p90$ (rightmost dot), $p95$ (center dot), and $p99$ (leftmost dot) threshold values. b) Same as (a), but for UH25.

The SfcRad_6_UAS_Advect experiment shows more prominent NEP changes from SfcRad than either the SfcRad_6_UAS_Early or SfcRad_6_UAS_Late experiments, and this experiment's changes are greater than in the SfcRad_6_UAS experiment with its six sites spread throughout the nature run domain (Fig. 4.15). Even though the SfcRad_6_UAS_Advect network covers only a small portion of the domain, strategic deployment of a targeted network upwind in the low-level flow appears to provide a superior capability to detect UH02 values exceeding $p99$. Furthermore, the NEP increases within the $p99$ UH02 contour for the SfcRad_6_UAS_Advect experiment (Fig. 4.15e) are comparable to the full SfcRad_ALL_UAS network (Fig. 4.15a) with more than 20 times the number of sites. The impressive performance of the SfcRad_6_UAS_Advect experiment forecast is quantitatively evident in Fig. 4.16, which shows that its POD, SR, and CSI scores are nearly equivalent to the SfcRad_ALL_UAS experiment, and generally greater than the other 6-UAS experiments for both UH02 and UH25. Additionally, the SfcRad_6_UAS_Advect experiment is the only one of the three targeted experiments for Case-1 with statistically significant CSI increases over the baseline SfcRad at all percentiles of both variables (Table 4.4).

These results, while primarily relevant for events with intense low-level flow regimes, show that notable improvements to forecast performance are possible with a relatively small, inexpensive network of WxUAS sites deployed upwind of an area of interest. Conducting reverse trajectory analysis based on the modeled low-level wind field for a particular event can help to identify deployment areas for which observations may be most useful to the forecast. It is worth considering, however, that this method is likely not the most effective way of targeting observations in all cases, particularly those with relatively weak low-level flow. Other approaches, such as ensemble sensitivity analysis (ESA), have been shown to be useful for targeted observation applications in the context of convective-scale model forecasts (see Torn et al. 2017; Hill et al. 2020; Faletti et al. 2024, etc.). Using techniques such as ESA instead of trajectory analysis would be worth exploring in future OSSEs to evaluate the impact of targeted WxUAS networks in the context of cases with less intense kinematic environments. Understanding how the impact of small targeted networks may change for different scenarios is important for informing effective operational implementation.

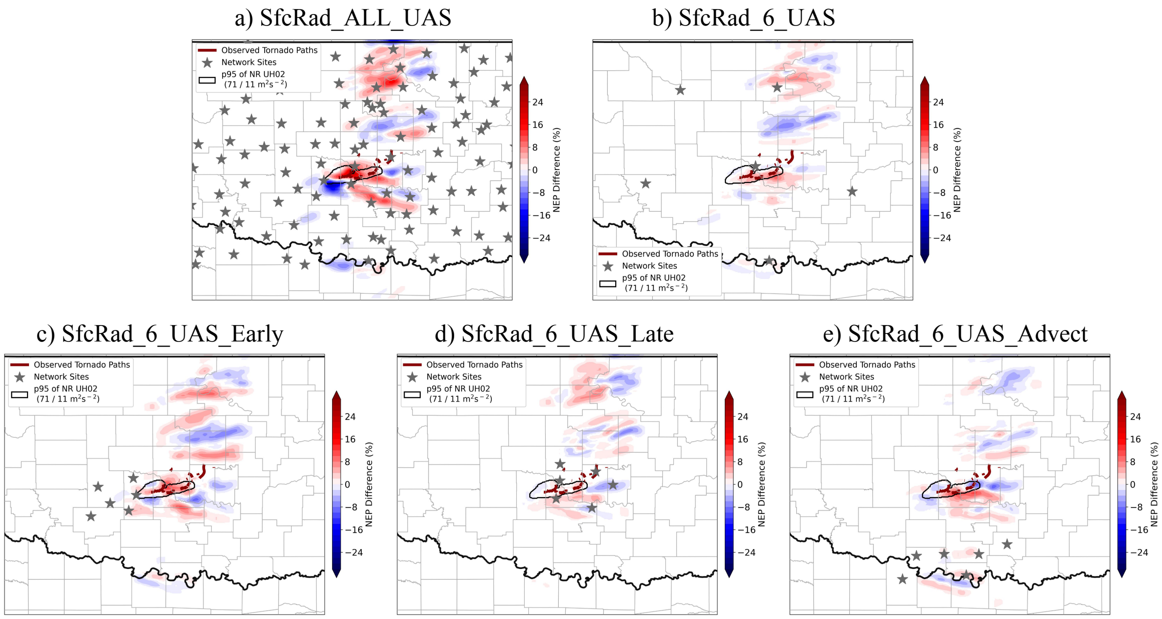

Figure 4.17: Same as Fig. 4.15, but for Case-2 and $p95$

The same three targeted 6-UAS network experiments, but with different sites, were also done for Case-2 and compared to the full SfcRad_ALL_UAS and spread out 6-UAS (SfcRad_6_UAS) experiments in Fig. 4.17. The NEP changes and percentile threshold contours that are displayed for these experiments are for $p95$ rather than $p99$. This was done simply to aid in visualization since the $p99$ contour area for UH02 is so small (see Fig. 4.4 g-i) and a larger area is easier to discern. The NEP differences between the experiments tell a much different story for Case-2 than for Case-1. Most notably, the differences in the NEP from the SfcRad experiment is much less pronounced in and around the percentile threshold contour for the SfcRad_6_UAS_Advect experiment than it was for Case-1 (see Fig. 4.15e). In fact, the SfcRad_6_UAS_Advect experiment generally performs the worst of all of the 6-UAS experiments, with the exception of for $p99$ of UH25 (Fig. 4.18).

It is worth noting that a closer examination of Fig. 4.17e reveals that this low performance is likely more due to spatial displacement of the enhanced NEP differences than low difference magnitudes. As such, the poor performance may be misleading from the perspective of broader operational forecasting applications which are not necessarily so dependent on spatial precision. Nevertheless, even if spatial displacements were ignored, the SfcRad_6_UAS_Early experiment is the best option among the targeted experiments for this case. The NEP differences throughout the domain are more pronounced in general, indicating an overall greater influence of the assimilated observations on the forecast UH tracks. Additionally, the SfcRad_6_UAS_Early experiment is the only one of the three targeted 6_UAS experiments for Case-2 with statistically significant performance improvements over the SfcRad experiment (Table 4.4).

The magnitude of the overall performance improvements, indicated by the CSI scores, are also greater than the SfcRad_6_UAS experiment at all percentiles of both variables with the single exception of $p99$ of UH02. This suggests that a targeted network, in this case focused on the area in which the convection initiated, provides more value to the forecast UH tracks than an equal-size network that is more spread out. The superiority of the SfcRad_6_UAS_Early experiment over the other 6-UAS experiments makes sense in the context of the convective storms associated with Case-2, which were likely driven primarily by the dryline, a small-scale feature whose impacts on the structure of the ABL would only be detectable by more spatially dense observations in its vicinity. Furthermore, the low-level flow, while still relatively strong compared to quiescent environments, was much less intense for Case-2 than Case-1. This would make the observations less susceptible to advection, thereby prolonging their relevancy to the region in which they were collected.

While these results differ somewhat from the findings associated with the Case-1 experiments, there is once again evidence that small, strategically placed profiling networks have the potential to yield improvements to forecast performance that are superior to more widely-spaced networks with the same amount of sites, and comparable to networks with a much greater amount of sites spread out over a large, non-targeted region. This finding has promising implications for the potential future operational application of profiling networks.

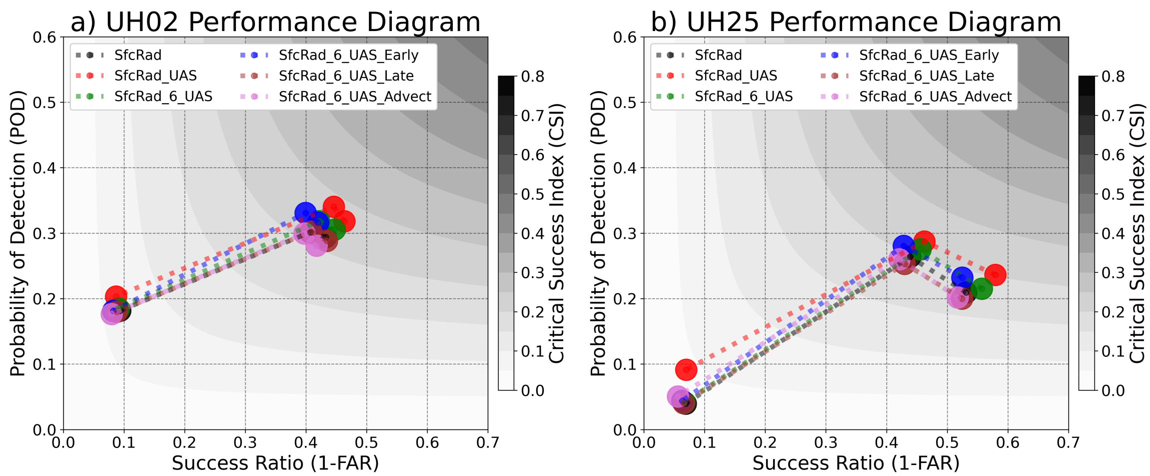

Figure 4.18: Same as Fig. 4.16 but for Case-2

Table 4.4: Same as Table 4.1, except for all 6-UAS experiments.

Performance Improvements: 6-UAS Experiments												
SfcRad_6_UAS												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.020**	**0.037**	**0.026**	**0.022**	**0.030**	**0.022**	**0.012**	**0.017**	**0.003**	**0.006**	**0.012**	**0.002**
SR	0.005	0.003	0.015	0.006	0.002	0.000	0.008	0.007	-0.003	0.021	0.009	-0.005
CSI	**0.020**	**0.036**	**0.026**	**0.022**	**0.029**	**0.022**	**0.011**	**0.016**	**0.003**	**0.007**	**0.011**	**0.002**
SfcRad_6_UAS_Early												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	0.008	0.001	0.006	**0.013**	**0.007**	**0.017**	**0.023**	**0.030**	-0.001	**0.023**	**0.017**	**0.004**
SR	-0.001	-0.002	0.001	-0.003	-0.003	0.000	-0.007	-0.005	**-0.017**	-0.007	-0.007	-0.015
CSI	0.0007	0.000	0.006	**0.013**	0.007	**0.015**	**0.019**	**0.027**	-0.002	**0.018**	**0.015**	**0.004**
SfcRad_6_UAS_Late												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	-0.006	**-0.009**	-0.004	**-0.004**	-0.003	0.001	**-0.006**	0.001	**-0.002**	**-0.009**	**-0.010**	0.000
SR	0.000	0.000	-0.003	0.000	0.000	0.000	0.012	0.001	-0.008	-0.006	-0.005	0.000
CSI	-0.005	**-0.009**	-0.004	-0.004	-0.003	0.001	**-0.005**	0.001	**-0.002**	**-0.008**	**-0.010**	0.000
SfcRad_6_UAS_Advect												
	Case-1						Case-2					
Variable:	UH02			UH25			UH02			UH25		
	p90	p95	p99	p90	p95	p99	p90	p95	p99	p90	p95	p99
POD	**0.031**	**0.064**	**0.135**	**0.037**	**0.043**	**0.131**	**-0.013**	-0.002	**-0.006**	-0.008	-0.004	**0.011**
SR	0.008	0.006	**0.046**	0.002	0.003	0.001	-0.009	-0.005	**-0.023**	-0.012	-0.011	**-0.034**
CSI	**0.031**	**0.063**	**0.134**	**0.036**	**0.043**	**0.130**	**-0.013**	-0.002	**-0.006**	-0.007	-0.004	**0.010**

Chapter 5

Summary and Conclusions

Current operational observations of the atmosphere are relatively coarse in space and time, particularly above the surface. The resulting gaps in upper-air data leave much of the atmosphere unobserved which may be a substantial source of error in numerical weather prediction (NWP) model forecasting. The predictability of the structure, behavior, and composition of atmospheric boundary layer (ABL) is particularly impacted by the shortage of observations above the surface due to its high spatial and temporal variability that arises from its direct interaction with the surface. A "network-of-networks" (NoN) was proposed by the National Research Council (NRC) in 2009 to address this issue and work towards filling observational data gaps (Council et al. 2009). The 3D-Mesonet concept, a proposed component of the NoN, would place a particular focus on improving the spatiotemporal resolution of ABL observations through implementing a mesoscale network of vertical profiling systems Chilson et al. (2019). Such a network would have a wide range of applications that would benefit a variety of industries.

Weather forecasting, which requires observations of atmospheric structure and behavior, would be a major beneficiary of an operational 3D-Mesonet. The observations that the network could provide would be valuable for real-time forecasting applications, especially in regards to localized high-impact weather such as severe thunderstorms and their hazards. The formation and evolution of severe convective storms can be substantially impacted by small features and processes in the ABL that current operational observing networks are much too coarse to adequately capture (Weckwerth 2000; Bryan et al. 2003; Kirshbaum 2022). A 3D-Mesonet could potentially resolve some of these localized phenomena and directly observe how they influence the atmosphere which could assist in improving the short-term predictability of severe weather hazards.

High-resolution convective allowing NWP models could also potentially benefit from the observations that a 3D-Mesonet would provide. Numerous studies (e.g. Hu et al. 2019; Chipilski et al. 2020; Tweedie 2024) have shown that assimilating ABL observations from just a few vertical profiling systems can positively impact the performance of model forecasts of severe thunderstorms and tornadoes. The results of these studies provide early insight into the potential benefits that assimilating ABL observations can have on NWP forecast performance. If observations from just a few sites can noticeably improve convective forecasts, a full mesoscale network of profiling systems may be worth the cost. However, the design and operation of such a network should be thoroughly investigated prior to its implementation in order to find the proper balance between its cost and the value that it could provide.

To gain initial insight into how a 3D-Mesonet could be optimally designed to improve the performance of CAM forecasts of high-impact severe weather, numerous observing system simulations experiments (OSSEs) were completed for two significant tornado events in central Oklahoma. These experiments were done using a "fraternal-twin" OSSE framework (Williamson and Kasahara 1971) in which two WoFS configurations with different grid resolutions were used to create a nature run, which served as a perfect-model forecast representation of the "true" atmosphere, and an ensemble forecast run, which was used to test various network configurations. Using realistic observation simulators, simulated lidar, AERI, MWR, and UAS observations were generated from the nature run and assimilated into the forecast run ensemble. The impact of the simulated observations on forecast performance was evaluated based on how well the experiment forecast aligned with the nature run compared to a baseline experiment that did not assimilate any profiler observations. This was done for numerous experiments, each of which involved variations in the amount of sites that made up the simulated network or the types of observations assimilated at each site. Analysis and inter-comparison of these various experiments revealed the following insights about the design of a 3D-Mesonet:

1. Assimilating simulated ABL observations from a comprehensive mesoscale profiling network that mirrors the coverage and resolution of the OK-Mesonet results in notable improvements to model forecast performance in terms of accumulated maximum UH tracks. This indicates that the increased spatial and temporal resolution of operational lower-troposphere observations that would be provided by a 3D-Mesonet may directly improve the model predictability of intense mid and low-level mesocyclones that are often associated with tornadoes.

2. In general, the value that a profiling network adds to forecast performance increases as the number of sites in the network also increases. However, a wide range exists regarding the number of sites for which adding more sites provides little additional benefit to the forecast. This suggests that the ideal number of sites to include in a 3D-Mesonet is likely a wide range rather than an explicit amount, and variable on a case-to-case basis. The lower-end of this range may be a good starting point for early adoption of an operational profiling network that provides value to convective-scale forecasts at a reasonable price.

3. Kinematic observations from a lidar generally had a more significant impact on forecast performance than thermodynamic observations from an AERI or MWR for the cases assessed, particularly Case-1. This finding prompted the hypothesis

that in cases with intense low-level flow regimes, local changes to thermodynamic variables in the ABL are dominated by advection, and are therefore highly correlated with kinematic observations. Since environments conducive to severe weather and tornadoes are often associated with strong low-level flow, it is possible that a network of only lidars, or other instruments that provide kinematic observations (e.g. the WxUAS), may be sufficient for improving model forecasts of high-impact severe weather events.

4. The influence of assimilating observations from a profiling network on forecast performance seems to diminish with time. This is likely due to observations being "advected" by the background flow away from their origin and eventually outside of the domain or area of interest. To combat this, small, strategically targeted networks of WxUASs, which are the most versatile and mobile profiling system, could be deployed upstream of the focused forecast area where they can observe parcels that will ultimately advect toward the area of interest. This could potentially maximize the duration for which the observations provide value to the forecast while placing focus on the area where improvements to forecast performance are most desired. OSSE experiments revealed that this strategy resulted in forecast performance improvements comparable to full-resolution mesoscale networks with far more sites. Therefore, small targeted WxUAS networks may be an optimal way of implementing a profiling network in a way that maximizes its value while mitigating cost.

The results of this study provide initial insight into how the number of sites in the network, the locations of the sites, and types of observations collected should be considered to optimize the design of an operational 3D-Mesonet. However, there are still numerous aspects of the design and operation of a mesoscale profiling network that should be thoroughly considered in future analyses. A non-exhaustive list of these factors includes: the temporal frequency of observations, the vertical spacing and extent of observations, the horizontal localization radius associated with the assimilation of the observations, and more. Future OSSE experiments could be conducted that place specific focus on evaluating these factors and how they should be adjusted. In addition, more OSSEs, or possibly even OSEs, should be conducted to focus on the same design factors discussed herein, but in the context of other cases or real observations to inform a more complete understanding of how the results of this thesis may or may not change in different circumstances.

Reference List

Adlerman, E. J., and K. K. Droegemeier, 2002: The sensitivity of numerically simulated cyclic mesocyclogenesis to variations in model physical and computational parameters. *Monthly weather review*, **130 (11)**, 2671–2691.

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, **90 (9)**, 1283–1296.

Anderson, J. L., 2001: An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, **129 (12)**, 2884–2903.

Arnold Jr, C. P., and C. H. Dey, 1986: Observing-systems simulation experiments: Past, present, and future. *Bulletin of the American Meteorological Society*, **67 (6)**, 687–695.

Arseneau, I., and B. Ancell, 2023: Exploring the value of a high-precision targeted observation strategy for mobile radiosonde deployment. *Monthly Weather Review*, **151 (9)**, 2233–2244.

Atlas, R., 1997: Atmospheric observations and experiments to assess their usefulness in data assimilation (gtspecial issueltdata assimilation in meteology and oceanography: theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, **75 (1B)**, 111–130.

Banghoff, J. R., J. D. Sorber, D. J. Stensrud, G. S. Young, and M. R. Kumjian, 2020: A 10-year warm-season climatology of horizontal convective rolls and cellular convection in central oklahoma. *Monthly Weather Review*, **148 (1)**, 21–42.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525 (7567)**, 47–55.

Bell, T., 2021: Examining novel profiling systems and their synergy for advancing boundary-layer research.

Bell, T. M., B. R. Greene, P. M. Klein, M. Carney, and P. B. Chilson, 2020: Confronting the boundary layer data gap: Evaluating new and existing methodologies of probing the lower atmosphere. *Atmospheric Measurement Techniques*, **13 (7)**, 3855–3872.

Ben Bouallègue, Z., and S. E. Theis, 2014: Spatial techniques applied to precipitation ensemble forecasts: From verification results to probabilistic products. *Meteorological Applications*, **21 (4)**, 922–929.

Benjamin, S. G., J. M. Brown, G. Brunet, P. Lynch, K. Saito, and T. W. Schlatter, 2019: 100 years of progress in forecasting and nwp applications. *Meteorological Monographs*, **59**, 13–1.

Berg, L. K., L. D. Riihimaki, Y. Qian, H. Yan, and M. Huang, 2015: The low-level jet over the southern great plains determined from observations and reanalyses and its impact on moisture transport. *Journal of Climate*, **28 (17)**, 6682–6706.

Blackadar, A. K., 1957: Boundary layer wind maxima and their significance for the growth of nocturnal inversions. *Bulletin of the American Meteorological Society*, **38 (5)**, 283–290.

Blumberg, W., D. Turner, U. Löhnert, and S. Castleberry, 2015: Ground-based temperature and humidity profiling using spectral infrared and microwave observations. part ii: Actual retrieval performance in clear-sky and cloudy conditions. *Journal of Applied Meteorology and Climatology*, **54 (11)**, 2305–2319.

Bonavita, M., 2024: On some limitations of current machine learning weather prediction models. *Geophysical Research Letters*, **51 (12)**, e2023GL107 377.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The oklahoma mesonet: a technical overview. *Journal of Atmospheric and Oceanic Technology*, **12 (1)**, 5–19.

Brooks, H. E., 2004: Tornado-warning performance in the past and future: A perspective from signal detection theory. *Bulletin of the American Meteorological Society*, **85 (6)**, 837–844.

Brooks, H. E., and J. Correia Jr, 2018: Long-term performance metrics for national weather service tornado warnings. *Weather and Forecasting*, **33 (6)**, 1501–1511.

Browning, K., and R. Wexler, 1968: The determination of kinematic properties of a wind field using doppler radar. *Journal of Applied meteorology and climatology*, **7 (1)**, 105–113.

Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review*, **131 (10)**, 2394–2416.

Chilson, P. B., and Coauthors, 2019: Moving towards a network of autonomous uas atmospheric profiling stations for observations in the earth's lower atmosphere: The 3d mesonet concept. *Sensors*, **19 (12)**, 2720.

Chipilski, H. G., X. Wang, and D. B. Parsons, 2020: Impact of assimilating pecan profilers on the prediction of bore-driven nocturnal convection: A multiscale forecast evaluation for the 6 july 2015 case study. *Monthly Weather Review*, **148 (3)**, 1147–1175.

Clark, A. J., J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia Jr, M. Xue, and F. Kong, 2013: Tornado pathlength forecasts from 2010 to 2011 using ensemble updraft helicity. *Weather and Forecasting*, **28 (2)**, 387–407.

Council, N. R., D. on Earth, L. Studies, B. on Atmospheric Sciences, and C. on Developing Mesoscale Meteorological Observational Capabilities to Meet Multiple National Needs, 2009: *Observing weather and climate from the ground up: A nationwide network of networks*. National Academies Press.

Degelia, S. K., X. Wang, and D. J. Stensrud, 2019: An evaluation of the impact of assimilating aeri retrievals, kinematic profilers, rawinsondes, and surface observations on a forecast of a nocturnal convection initiation event during the pecan field campaign. *Monthly Weather Review*, **147 (8)**, 2739–2764.

Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the integrated global radiosonde archive. *Journal of Climate*, **19 (1)**, 53–68.

Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, **15 (1)**, 51–64.

Etling, D., and R. A. Brown, 1993: Roll vortices in the planetary boundary layer: A review. *Boundary-Layer Meteorology*, **65**, 215–248.

Faletti, W. L., C. C. Weiss, and P. Skinner, 2024: The utility of an ensemble sensitivity-based subsetting technique to improve mesocyclone intensity forecasts in the noaa warn-on-forecast system. *104th Annual AMS Meeting 2024*, Vol. 104, 437447.

Finley, J. P., 1884: Tornado predictions. *American Meteorological Journal. A Monthly Review of Meteorology and Allied Branches of Study (1884-1896)*, **1 (3)**, 85.

Garratt, J. R., 1994: The atmospheric boundary layer. *Earth-Science Reviews*, **37 (1-2)**, 89–134.

Gebauer, J. G., A. Shapiro, E. Fedorovich, and P. Klein, 2018: Convection initiation caused by heterogeneous low-level jets over the great plains. *Monthly Weather Review*, **146 (8)**, 2615–2637.

Geerts, B., and Coauthors, 2017: The 2015 plains elevated convection at night field project. *Bulletin of the American Meteorological Society*, **98 (4)**, 767–786.

Heinselman, P. L., and Coauthors, 2024: Warn-on-forecast system: From vision to reality. *Weather and Forecasting*, **39 (1)**, 75–95.

Hill, A. J., C. C. Weiss, and B. C. Ancell, 2020: Factors influencing ensemble sensitivity–based targeted observing predictions at convection-allowing resolutions. *Monthly Weather Review*, **148 (11)**, 4497–4517.

Hill, A. J., C. C. Weiss, and D. C. Dowell, 2021: Influence of a portable near-surface observing network on experimental ensemble forecasts of deep convection hazards during vortex-se. *Weather and Forecasting*, **36 (4)**, 1141–1167.

Hitchcock, S. M., M. C. Coniglio, and K. H. Knopfmeier, 2016: Impact of mpex upsonde observations on ensemble analyses and forecasts of the 31 may 2013 convective event over oklahoma. *Monthly Weather Review*, **144 (8)**, 2889–2913.

Holton, J. R., 1967: The diurnal boundary layer wind oscillation above sloping terrain. *Tellus*, **19 (2)**, 200–205.

Hu, J., N. Yussouf, D. D. Turner, T. A. Jones, and X. Wang, 2019: Impact of ground-based remote sensing boundary layer observations on short-term probabilistic forecasts of a tornadic supercell event. *Weather and Forecasting*, **34 (5)**, 1453–1476.

Jensen, A. A., and Coauthors, 2021: Assimilation of a coordinated fleet of uncrewed aircraft system observations in complex terrain: Enkf system design and preliminary assessment. *Monthly Weather Review*, **149 (5)**, 1459–1480.

Jian, S., C. Zhuo, L. Heng, Q. Simeng, W. Xin, Y. Limin, and X. Wei, 2021: Application of artificial intelligence technology to numerical weather prediction. *Journal of Applied Meteorological Science*, **32 (1)**, 1–11.

Johnson, A., and X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble. *Monthly weather review*, **140 (9)**, 3054–3077.

Johnson, A., X. Wang, Y. Wang, A. Reinhart, A. J. Clark, and I. L. Jirak, 2020: Neighborhood-and object-based probabilistic verification of the ou map ensemble forecasts during 2017 and 2018 hazardous weather testbeds. *Weather and Forecasting*, **35 (1)**, 169–191.

Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Weather and forecasting*, **25 (5)**, 1536–1542.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing nwp. *Weather and Forecasting*, **23 (5)**, 931–952.

Keclik, A. M., C. Evans, P. J. Roebber, and G. S. Romine, 2017: The influence of assimilated upstream, preconvective dropsonde observations on ensemble forecasts of convection initiation during the mesoscale predictability experiment. *Monthly Weather Review*, **145 (12)**, 4747–4770.

Kirshbaum, D. J., 2022: Large-eddy simulations of convection initiation over heterogeneous, low terrain. *Journal of the Atmospheric Sciences*, **79 (4)**, 973–987.

Knuteson, R., and Coauthors, 2004: Atmospheric emitted radiance interferometer. part i: Instrument design. *Journal of Atmospheric and Oceanic Technology*, **21 (12)**, 1763–1776.

Kosiba, K. A., and Coauthors, 2024: The propagation, evolution, and rotation in linear storms (perils) project. *Bulletin of the American Meteorological Society*, **105 (10)**, E1768–E1799.

Laser, J. J., M. C. Coniglio, P. S. Skinner, and E. N. Smith, 2022: Doppler lidar and mobile radiosonde observation-based evaluation of warn-on-forecast system predicted near-supercell environments during torus 2019. *Weather and Forecasting*, **37 (10)**, 1783–1804.

Löhnert, U., S. Crewell, O. Krasnov, E. O'Connor, and H. Russchenberg, 2008: Advances in continuously profiling the thermodynamic state of the boundary layer: Integration of measurements and methods. *Journal of Atmospheric and Oceanic Technology*, **25 (8)**, 1251–1266.

Majumdar, S. J., 2016: A review of targeted observations. *Bulletin of the American Meteorological Society*, **97 (12)**, 2287–2303.

Manning, C., and H. Schutze, 1999: *Foundations of statistical natural language processing*. MIT press.

Mather, J. H., and J. W. Voyles, 2013: The arm climate research facility: A review of structure and capabilities. *Bulletin of the American Meteorological Society*, **94 (3)**, 377–392.

McGovern, A., K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bulletin of the American Meteorological Society*, **98 (10)**, 2073–2090.

McNider, R. T., and R. A. Pielke, 1984: Numerical simulation of slope and mountain flows. *Journal of Applied Meteorology and Climatology*, **23 (10)**, 1441–1453.

McPherson, R. A., and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the oklahoma mesonet. *Journal of Atmospheric and Oceanic Technology*, **24 (3)**, 301–321.

Miller, S., B. Keim, R. Talbot, and H. Mao, 2003: Sea breeze: Structure, forecasting, and impacts. *Reviews of geophysics*, **41 (3)**.

Murphy, A. H., 1993: What is a good forecast? an essay on the nature of goodness in weather forecasting. *Weather and forecasting*, **8 (2)**, 281–293.

Murphy, A. H., and R. L. Winkler, 1987: A general framework for forecast verification. *Monthly weather review*, **115 (7)**, 1330–1338.

Newsom, R., C. Sivaraman, T. Shippert, and L. Riihimaki, 2019: Doppler lidar vertical velocity statistics value-added product. Tech. rep., DOE ARM Climate Research Facility, Washington, DC (United States).

Newsom, R. K., W. A. Brewer, J. M. Wilczak, D. E. Wolfe, S. P. Oncley, and J. K. Lundquist, 2017: Validating precision estimates in horizontal wind measurements from a doppler lidar. *Atmospheric Measurement Techniques*, **10 (3)**, 1229–1240.

Otkin, J. A., D. C. Hartung, D. D. Turner, R. A. Petersen, W. F. Feltz, and E. Janzon, 2011: Assimilation of surface-based boundary layer profiler observations during a cool-season weather event using an observing system simulation experiment. part i: Analysis impact. *Monthly weather review*, **139 (8)**, 2309–2326.

O'Connor, E. J., A. J. Illingworth, I. M. Brooks, C. D. Westbrook, R. J. Hogan, F. Davies, and B. J. Brooks, 2010: A method for estimating the turbulent kinetic energy dissipation rate from a vertically pointing doppler lidar, and independent evaluation from balloon-borne in situ measurements. *Journal of atmospheric and oceanic technology*, **27 (10)**, 1652–1664.

Payne, V. H., E. J. Mlawer, K. E. Cady-Pereira, and J.-L. Moncet, 2011: Water vapor continuum absorption in the microwave. *IEEE transactions on geoscience and remote sensing*, **49 (6)**, 2194–2208.

Pearson, G., F. Davies, and C. Collier, 2009: An analysis of the performance of the ufam pulsed doppler lidar for observing the boundary layer. *Journal of Atmospheric and Oceanic Technology*, **26 (2)**, 240–250.

Privé, N. C., E. L. McGrath-Spangler, D. Carvalho, B. M. Karpowicz, and I. Moradi, 2023: Robustness of observing system simulation experiments. *Tellus A: Dynamic Meteorology and Oceanography*, **75 (1)**.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Weather and Forecasting*, **24 (2)**, 601–608.

Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Monthly Weather Review*, **142 (12)**, 4519–4541.

Romine, G. S., C. S. Schwartz, R. D. Torn, and M. L. Weisman, 2016: Impact of assimilating dropsonde observations from mpex on ensemble forecasts of severe weather events. *Monthly Weather Review*, **144 (10)**, 3799–3823.

Rose, T., S. Crewell, U. Löhnert, and C. Simmer, 2005: A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere. *Atmospheric research*, **75 (3)**, 183–200.

Scheufele, K., K. Kober, G. C. Craig, and C. Keil, 2014: Combining probabilistic precipitation forecasts from a nowcasting technique with a time-lagged ensemble. *Meteorological Applications*, **21 (2)**, 230–240.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather and Forecasting*, **25 (1)**, 263–280.

Segal, M., R. Avissar, M. McCumber, and R. A. Pielke, 1988: Evaluation of vegetation effects on the generation and modification of mesoscale circulations. *Journal of the Atmospheric Sciences*, **45 (16)**, 2268–2293.

Segales, A., 2022: Design and implementation of a novel multicopter unmanned aircraft system for quantitative studies of the atmosphere. Ph.D. thesis, Ph. D. thesis, University of Oklahoma, 2022. a, b.

Segales, A. R., B. R. Greene, T. M. Bell, W. Doyle, J. J. Martin, E. A. Pillar-Little, and P. B. Chilson, 2020: The coptersonde: an insight into the development of a smart unmanned aircraft system for atmospheric boundary layer research. *Atmospheric Measurement Techniques*, **13 (5)**, 2833–2848.

Shapiro, A., and E. Fedorovich, 2010: Analytical description of a nocturnal low-level jet. *Quarterly Journal of the Royal Meteorological Society*, **136 (650)**, 1255–1262.

Shapiro, A., E. Fedorovich, and S. Rahimi, 2016: A unified theory for the great plains nocturnal low-level jet. *Journal of the Atmospheric Sciences*, **73 (8)**, 3037–3057.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **128 (580)**, 647–677.

Smith, E. N., J. G. Gebauer, P. M. Klein, E. Fedorovich, and J. A. Gibbs, 2019: The great plains low-level jet during pecan: Observed and simulated characteristics. *Monthly Weather Review*, **147 (6)**, 1845–1869.

Smith, T. M., and Coauthors, 2016: Multi-radar multi-sensor (mrms) severe weather and aviation products: Initial operating capabilities. *Bulletin of the American Meteorological Society*, **97 (9)**, 1617–1630.

Sobash, R. A., G. S. Romine, C. S. Schwartz, D. J. Gagne, and M. L. Weisman, 2016: Explicit forecasts of low-level rotation from convection-allowing models for next-day tornado prediction. *Weather and Forecasting*, **31 (5)**, 1591–1614.

Sobash, R. A., C. S. Schwartz, G. S. Romine, and M. L. Weisman, 2019: Next-day prediction of tornadoes using convection-allowing models with 1-km horizontal grid spacing. *Weather and Forecasting*, **34 (4)**, 1117–1135.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bulletin of the American Meteorological Society*, **90 (10)**, 1487–1500.

Stull, R. B., 2012: *An introduction to boundary layer meteorology*, Vol. 13. Springer Science & Business Media.

Torn, R. D., G. S. Romine, and T. J. Galarneau Jr, 2017: Sensitivity of dryline convection forecasts to upstream forecast errors for two weakly forced mpex cases. *Monthly Weather Review*, **145 (5)**, 1831–1852.

Turner, D., 2007: Improved ground-based liquid water path retrievals using a combined infrared and microwave approach. *Journal of Geophysical Research: Atmospheres*, **112 (D15)**.

Turner, D., and P. Gero, 2011: Downwelling 10 μm radiance temperature climatology for the atmospheric radiation measurement southern great plains site. *Journal of Geophysical Research: Atmospheres*, **116 (D8)**.

Turner, D., and U. Löhnert, 2014: Information content and uncertainties in thermodynamic profiles and liquid cloud properties retrieved from the ground-based atmospheric emitted radiance interferometer (aeri). *Journal of Applied Meteorology and Climatology*, **53 (3)**, 752–771.

Turner, D. D., J. Gebauer, T. Bell, B. Adler, and U. Löhnert, 2023: Tropoe: A containerized thermodynamic profile retrieval system for ground-based remote sensors. *AGU Fall Meeting Abstracts*, Vol. 2023, A31M–2566.

Tweedie, J., 2024: Impact of assimilating lower atmospheric boundary layer observations from uncrewed aerial systems on short-term probabilistic predictions of a tornadic supercell event.

Virtanen, P., and Coauthors, 2020: Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, **17 (3)**, 261–272.

Wade, A. R., M. C. Coniglio, and C. L. Ziegler, 2018: Comparison of near-and far-field supercell inflow environments using radiosonde observations. *Monthly Weather Review*, **146 (8)**, 2403–2415.

Wagner, T. J., P. M. Klein, and D. D. Turner, 2019: A new generation of ground-based mobile platforms for active and passive profiling of the boundary layer. *Bulletin of the American Meteorological Society*, **100 (1)**, 137–153.

Weaver, S. J., S. Baxter, and A. Kumar, 2012: Climatic role of north american low-level jets on us regional tornado activity. *Journal of climate*, **25 (19)**, 6666–6683.

Weckwerth, T. M., 2000: The effect of small-scale moisture variability on thunderstorm initiation. *Monthly Weather Review*, **128 (12)**, 4017–4030.

Weckwerth, T. M., T. W. Horst, and J. W. Wilson, 1999: An observational study of the evolution of horizontal convective rolls. *Monthly weather review*, **127 (9)**, 2160–2179.

Weckwerth, T. M., J. W. Wilson, and R. M. Wakimoto, 1996: Thermodynamic variability within the convective boundary layer due to horizontal convective rolls. *Monthly Weather Review*, **124 (5)**, 769–784.

Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the nssl experimental warn-on-forecast system. part i: Radar data experiments. *Weather and Forecasting*, **30 (6)**, 1795–1817.

Williamson, D., and A. Kasahara, 1971: Adaptation of meteorological variables forced by updating. *Journal of Atmospheric Sciences*, **28 (8)**, 1313–1324.

Wu, Y., and S. Raman, 1998: The summertime great plains low level jet and the effect of its origin on moisture transport. *Boundary-Layer Meteorology*, **88**, 445–466.

Xue, M., and W. J. Martin, 2006: A high-resolution modeling study of the 24 may 2002 dryline case during ihop. part ii: Horizontal convective rolls and convective initiation. *Monthly Weather Review*, **134 (1)**, 172–191.

Yussouf, N., J. Gao, D. J. Stensrud, and G. Ge, 2013: The impact of mesoscale environmental uncertainty on the prediction of a tornadic supercell storm using ensemble data assimilation approach. *Advances in Meteorology*, **2013 (1)**, 731 647.