Optimizing Numerical Weather Prediction Utility of the Maryland Mesonet with Observing System Simulation Experiments

JOSHUA MCCURRY[a] AND JONATHAN POTERJOY[a]

[a] *University of Maryland, College Park, College Park, Maryland*

(Manuscript received 10 June 2024, in final form 1 August 2024, accepted 12 September 2024)

ABSTRACT: The Maryland Mesonet project will construct a network of 75 surface observing stations with aims that include mitigating the statewide impact of severe convective storms and improving analyses of records. The spatial configuration of mesonet stations is expected to affect the utility newly provided observations will have via data assimilation, making it desirable to study the effects of mesonet configuration. Furthermore, the impact associated with any observing system configuration is constrained by errors inherent to the prediction systems used to generate forecasts, which may change with future advances in data assimilation methodology, physical parameterization schemes, and resource availability. To address such possibilities, we perform sets of observing system simulation experiments using a high-resolution regional modeling system to assess the expected impact of four candidate mesonet configurations. Experiments cover seven 18-h case study events featuring moist convective regimes associated with severe weather over the state of Maryland and are performed using two versions of our experimental modeling system: a "standard-uncertainty" configuration tuned to be representative of existing convective-allowing prediction systems and a "constrained-uncertainty" configuration with reduced boundary condition and model error that reflects a possible trajectory for future prediction systems. We find that the assimilation of mesonet data produces definitive improvements to analysis fields below 1000 m that are mediated by modeling system uncertainty. Conversely, mesonet impact on forecast verification is inconclusive and strongly variable across verification metrics. The impact of mesonet configuration appears limited by a saturation effect that caps local analysis improvements past a minimal density of observing stations.

SIGNIFICANCE STATEMENT: The Maryland Mesonet project will construct 75 surface observing stations to improve the analysis of records for Maryland's surface weather conditions as well as predictions for severe weather events. The spatial placement of sensors is expected to influence the utility of a mesonet, making it desirable to optimize mesonet layouts. The utility provided by a mesonet may also be impacted by errors in prediction systems used to generate analyses and forecasts, which are themselves subject to change given future advances in prediction methods and resources. This study uses observing system simulation experiments (OSSEs)—which comprehensively simulate numerical weather prediction for a known "truth state" —to characterize improvement we may expect from mesonet observations and evaluate four potential mesonet configurations.

KEYWORDS: Severe storms; In situ atmospheric observations; Sensitivity studies; Mesoscale forecasting; Regional models

1. Introduction

The forecast skill offered by numerical weather prediction (NWP) systems has improved in tandem with advances in observing strategies, model physics, and data assimilation (DA) methodology, combined with increases in computing capacity and observational coverage. Furthermore, the widespread adoption of ensemble modeling over the past two decades has yielded a coherent basis for analyzing the sources of error that impact practical predictability. Given the rate of such advances, there is a need for resources to be organized in a way that considers the future trajectories of NWP system components. Carefully structured observing system simulation experiments (OSSEs) may be used to evaluate—and potentially optimize—the NWP utility that can be expected from an observing system given existing NWP systems while providing insight on how afforded utility may evolve under future NWP systems that feature improvements such as more accurate model parameterization schemes, better constrained initial and boundary conditions, and alternative DA strategies. The present study adopts this strategy to evaluate the potential impacts of a new mesonet network to be established in the state of Maryland. The new network aims to improve analyses and numerical predictions for high-impact weather events, with a specific emphasis on hazards associated with severe convective storms.

a. *Maryland Mesonet project*

The Maryland Mesonet (MM) project will oversee the placement of 75 state-of-the-art, in situ surface observing stations, with the goal of improving climate records, producing more accurate regional analyses and forecasts, and mitigating the impact of severe weather events across the state. It will join other similarly motivated mesonet systems that exist across the country, such as the pioneering Oklahoma Mesonet established in 1994. As the definition of a mesonet has historically been contingent on local requirements (Fiebrich et al. 2020), for our study, we define a mesonet as a network of

Corresponding author: Joshua McCurry, jmccurry@umd.edu

DOI: 10.1175/WAF-D-24-0089.1

observing stations that collects surface wind velocity, temperature, humidity, etc. with a spatial and temporal frequency sufficient to capture mesoscale processes and the development of severe convective storms. Such observing systems typically provide information at a granularity beyond the resolution available from pre-existing surface observing stations. The MM will be constructed in two phases. The first 28 stations will be distributed on a roughly county-by-county basis with specific locations contingent on the input of local officials and emergency managers. The second phase of the network design will place the remaining 47 stations based on an optimization strategy that maximizes network utility. The current paper outlines the optimization strategy for the second phase of this project and discusses findings from numerical experiments that were performed for this purpose.

b. *Challenges and limitations of assimilating surface observations for NWP*

The effective use of in situ surface data from mesonets in atmospheric modeling systems is a nuanced problem that requires consideration of error sources in convective-scale NWP systems. Subgrid-scale physical parameterization schemes for components such as land surface, cloud microphysics, and planetary boundary layer in numerical models may not accurately reflect physical processes involved in moist convection. Model process error can often be highly correlated, difficult to estimate, and treated heuristically by inflating background error covariance during DA. As such, a model error can produce large biases in estimates of model states and the respective uncertainty in the state estimate. Furthermore, true error correlation length scales across point locations for state variables are nonstationary. They may, for instance, decrease as large-scale synoptic flow gives way to convective regimes and vice versa. Multiple factors, including the grid spacing of the modeling system, ensemble size, model error, and choices of heuristic DA parameters, such as localization and inflation, can determine the accuracy with which cross covariances are dynamically estimated using ensembles (Pannekoucke et al. 2008; Hodyss and Morzfeld 2023). The meso-γ- and microscale correlation structures that characterize organized moist convection could require finer grid spacing and larger ensemble sizes than presently available in operational contexts (Necker et al. 2020). Importantly, sampling error associated with unresolved variability can translate to biased analysis increments (Hodyss and Morzfeld 2023) that degrade or fail to improve state estimates and resulting forecasts. Even where uncertainty is accurately estimated with ensembles, Gaussian assumptions made by current DA methods such as ensemble Kalman filters (EnKFs) and variational methods may introduce errors in state estimates due to the nonlinearity of moist convective dynamics and resulting non-Gaussianity of prior probability distributions (McCurry et al. 2023).

The assimilation of mesonet observations has been the subject of multiple observing system sensitivity studies, although the diversity of observing system configurations, modeling system parameters, case studies, and broader experiment methodology within this body of research makes it difficult to draw firm conclusions on their practical utility for improving analyses and forecasts for severe events in the context of operational, limited-area NWP systems. Results from Stensrud et al. (2009) demonstrate the utility of mesonet observations in terms of improving the depiction of cold pools and storm wind structures in state estimates. Ha and Snyder (2014) likewise showed reduced low-level moisture and temperature biases and improved estimates for the location and structure of frontal boundaries. For the latter study, error reductions for wind and temperature were seen to extend vertically into the midtroposphere, and forecasts initialized with mesonet data outperformed verification scores from a control run on several metrics—including accumulated precipitation—even 6-h postinitialization. We note that both of the preceding studies used coarse (>10 km) grid spacing and infrequent assimilation intervals (>1 h), with no assimilation of Doppler wind observations in areas of storm activity. Studies using higher-resolution modeling systems that are perhaps more relevant for convective-scale data assimilation have shown mixed results for utility gained from assimilating mesonet observations. When performing hourly data assimilation for a convective-allowing modeling system using conventional surface and airborne observations from National Oceanic and Atmospheric Administration (NOAA) Global Systems Division Meteorological Analysis and Data Ingest System (MADIS; Miller et al. 2005), Knopfmeier and Stensrud (2013) found no significant difference between verification scores of state estimates produced using all possible mesonet observations and those from data denial experiments that assimilated only 25% of available mesonet observations. Sobash and Stensrud (2015) assimilated MADIS mesonet observations along with other real surface observation types in a modeling system with 3-km grid spacing and found that the addition of mesonet observations improved model timing of convective initiation by reducing surface biases for temperature and dewpoint, but only when observations were assimilated at a 5-min frequency. Sobash and Stensrud (2015) also noted limitations in the ability of surface observations to improve state estimates of quantities above the boundary layer, which contributed to limited mesonet utility for an event featuring a strong capping inversion. This issue also appears in Marquis et al. (2014), which evaluated a mobile mesonet system for real cases using a modeling system with 500-m grid spacing that also assimilated Doppler radar observations. The authors noted that surface observations provided by the mobile mesonet did not produce significant change to state estimates above a near-surface cold pool layer 2–3 km in depth. Although both Knopfmeier and Stensrud (2013) and Sobash and Stensrud (2015) examined the impact of mesonet observations separately from METAR and other conventional surface observations, no study so far has comparatively evaluated the impact associated with distinct spatial configurations of mesonet systems. Due to a relative lack of comprehensive OSSEs experiments involving modeling system uncertainty, it is also not well understood how factors like model process error and boundary condition error mediate information available to modeling systems from surface observing networks. It is conceivable, for instance, that reductions in

uncertainty from improved model physics parameterizations and boundary conditions that remove synoptic-scale errors may isolate error sources associated with mesoscale dynamics to the point where surface observing systems can help inform state estimates for the free atmosphere in the vicinity of severe storms. Alternatively, improved error covariance estimates for surface-based mesoscale features like cold pools and horizontal convergence zones may allow surface observations to inform model representations of such phenomena at finer scales and therefore produce more accurate forecasts even without reducing analysis error above the near-surface environment.

c. *Observing system simulation experiments*

OSSEs are a commonly used method for numerically estimating the impact that may be expected from changes applied to any component of an NWP system, including changes to the type and density of observations ingested during the DA process. All OSSEs involve the creation of a nature run—a single model integration defined as the benchmark "truth" from which errors may be unambiguously defined. "Synthetic" observations are then generated from the nature run and assimilated by a modeling system that produces initial conditions and forecasts. OSSEs provide a natural framework for investigating the impact of nonexistent observing systems such as the Maryland Mesonet and also offer flexibility in tuning modeling system error structures against a reference state. Our study makes use of the latter property of OSSEs to comparatively evaluate the utility of four candidate networks for a "standard-uncertainty" modeling system configuration representative of contemporary limited-area, convective-permitting NWP systems and a "constrained-uncertainty" modeling system configuration imitating future NWP systems with improved model physics and well-constrained boundary conditions. Although alternative methods exist for optimizing observing networks, such as the adaptive observation framework mentioned in Khare and Anderson (2006), our choice of a simple comparative evaluation of pregenerated mesonet configurations avoided assumptions for sampling and model error that may complicate such approaches.

2. Methodology

a. *Case event selection*

Our experiments simulated seven 18-h real event cases featuring warm-season moist convection within the state of Maryland. Cases were chosen from within the archival range for forecast data from the High-Resolution Rapid Refresh (HRRR) model (James et al. 2022), spanning 2018–22 at the time of our experiments. Our selection process emphasized events featuring multiple Storm Prediction Center (SPC) storm reports within state boundaries, with further filtering based on storm severity within the densely populated Baltimore–Washington metropolitan area. Severity in this context was judged by available observations of composite reflectivity and accumulated event precipitation. Limiting events to those that impacted densely populated areas allows human impact

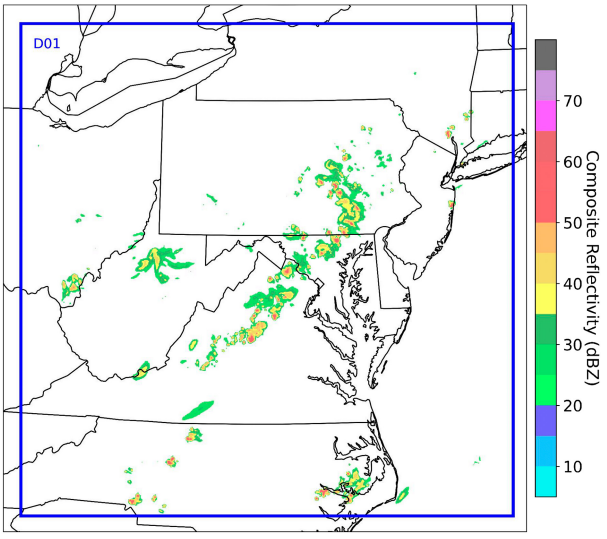

FIG. 1. Mid-Atlantic domain used for OSSE experiments. Colored shading shows the nature run composite reflectivity from 5 to 80 dBZ at 1800 UTC for one of the selected events on 16 Jul 2022.

factors to play an implicit role in the evaluation of mesonet configurations without dividing our verification region into smaller subdomains that may suffer from issues of sampling error related to the spatial distribution of impact from a limited number of test cases. We chose initial conditions for our test case events so that the peak of convective activity occurs after the beginning of forecast initiation following a 5-h spinup period of sequential DA.

b. *Modeling system and data assimilation*

For the analysis and forecasting component of our OSSE framework—from which initial conditions and forecasts are produced for later comparative evaluation against a nature run—we implemented a modeling system similar to the Warn-on-Forecast System (WoFS) developed by the NOAA National Severe Storms Laboratory (Wheatley et al. 2015; Jones et al. 2016; Lawson et al. 2018) and previously applied for experiments evaluating non-Gaussian data assimilation methods in moist convective regimes (McCurry et al. 2023). All experiments were performed with the Advanced Research version of WRF (version 4.2) (Skamarock et al. 2019) and implemented data assimilation using the DA Research Testbed (DART) software package (Anderson et al. 2009). The domain size was set at 900 × 900 km^2, with 50 vertical levels and a 3-km horizontal grid spacing, which is sufficient to explicitly resolve meso-γ-scale convective features in numerical experiments. The domain was centered on Winchester Regional Airport (KOKV) for all experiments, providing full coverage of the state of Maryland and extending several hundred kilometers into surrounding states located to the north, south, and west (Fig. 1). In Table 1, we specify model physical parameterization options used in OSSE experiments. With the exception of ensemble size—which we set at 40 members—and the microphysical scheme—which is experimentally varied—selected

TABLE 1. Selected WRF namelist options used for nature run, standard-uncertainty modeling system, and constrained-uncertainty modeling system (top) and selected DART namelist options used for all OSSE experiments (bottom).

WRF namelist option	Nature	Standard uncertainty	Constrained uncertainty
Grid spacing	1 km	3 km	3 km
Microphysics	NSSL (Mansell 2010)	Thompson (Thompson et al. 2008)	NSSL
Cumulus		—	
Longwave radiative transfer		RRTM (Iacono et al. 2008)	
Shortwave radiative transfer		Dudhia (Chen and Dudhia 2001)	
Land surface		RUC (Smirnova et al. 2016)	
Surface layer physics		Monin–Obukhov (Jiménez et al. 2012)	
Planetary boundary layer		YSU (Ghonima et al. 2017)	
DART namelist option		Standard uncertainty	Constrained uncertainty
Localization radius (surface obs)		30 km	
Localization function		Gaspari–Cohn (Gaspari and Cohn 1999)	

options were motivated by validation experiments performed at NSSL for the WoFS (Jones et al. 2018; Potvin et al. 2020).

Initial ensemble members and lateral boundary conditions for our modeling system were generated by perturbing output from HRRR analyses and forecasts with spatially correlated Gaussian noise for horizontal wind, potential temperature, and moisture fields. After a 1-h spinup period intended to develop flow-dependent ensemble covariance, we performed 14 h of sequential DA, with an assimilation frequency of 15 min. Hourly updates were applied to lateral boundary conditions for each member during DA, to reflect solutions from the posterior ensemble and HRRR forecasts. To cope with underdispersion in the ensemble, we also used additive inflation at 15-min intervals near areas of high observed reflectivity (Dowell and Wicker 2009). We initialized 180-min, 20-member ensemble forecasts at 30-min intervals starting 5 h after the first DA cycle, giving a total of 19 ensemble forecasts per experiment.

c. *Nature runs and generation of synthetic observations*

Nature runs were produced for each event by performing single 18-h WRF integrations from corresponding initial conditions. This integration length was sufficient to encompass the period of sequential DA and resulting forecasts produced by our modeling system. As with initial ensemble members, nature run initial and lateral boundary conditions were interpolated from HRRR analysis output but were not perturbed with spatially correlated Gaussian noise. Given the 3-km grid spacing specified for our modeling system, we chose a finer nature run grid spacing of 1 km to induce a source of representativeness error in synthetic observations. Physical parameterization schemes in nature runs were maintained from those used by our modeling system configurations (Table 1), with the exception of microphysics for which we specified the NSSL 2-moment variable density (NVD) scheme. The NVD scheme is notable for its fully double-moment representation of all hydrometeor classes and the inclusion of graupel density as a diagnosable parameter (Mansell 2010), which offers a comparatively high degree of complexity that may be useful for OSSE experiments simulating the effect of missing physical processes.

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

Since NWP systems typically do not assimilate radar observations below a threshold value, we used a modified process to create synthetic radar observations only at locations corresponding to reflectivity at or above a value of 25 dBZ in the nature run. We first created dummy radar volumes for each radar station with sweep elevation angles typical of those from real NEXRAD observations used in part 3. We generated these volumes at time intervals that were likewise representative. These dummy volumes contained nonempty values at every possible location within the volume, subject to the same regridding procedure applied to real observations for our work in part 3. After interpolating truth values from the nature run, we then removed observations within these volumes that were below the 25-dBZ minimum reflectivity threshold. Both steps made use of software tools provided by the Observation Processing and Wind Synthesis (OPAWS) program (Wicker 2017; Majcen et al. 2008).

In the context of NWP, observation error can be thought of as the sum of a measurement error term and a representation

TABLE 2. Configurations evaluated by OSSE experiments and parameters used in their generation by a synthetic annealing algorithm.

Configuration parameters	EQD	EQD+	WEST	POP
Geographic constraints	Within the state of Maryland			
Surface constraints	Land surface only			
Geographic forcing	—	—	Western boundaries	BWI metro area
Equidistance maximized for	MM	MM and non-MM surface	MM	MM

term consisting of contributions from 1) unresolved scales, 2) forward operator error, and 3) preprocessing error (Janjić et al. 2017). Although some degree of error from unresolved scales was implicitly included from the interpolation of synthetic observations between the 1-km nature run grid and the 3-km modeling system grid, we accounted for measurement error and the remainder of the representation error term by adding Gaussian noise to the values of synthetic observations in a manner similar to Errico et al. (2013) but using posterior root mean departure from observations as the tuning parameter.

d. MM configurations

Our OSSE framework considered four candidate configurations of 75 mesonet stations to be constructed as part of the MM project. These configurations differ in aspects including geographic uniformity of station density, locus of maximum density for nonuniform configurations, as well as consideration of surface observations provided by the pre-existing observation network. We note that 28 of these stations have fixed locations as part of the first phase of construction, leaving 47 stations with spatial locations that may vary between configurations. Candidate networks were generated using a method based on simulated annealing (Bertsimas and Tsitsiklis 1993) that maximizes equidistance between stations subject to constraints and forcing terms that impact the final distribution of station placement. A summary of station configurations and the specific constraints used by the simulated annealing method is provided in Table 2. All networks were required to place stations within the state of Maryland and on land surfaces only. The west-centered (WEST) and population-centered (POP) MM configurations included forcing terms that increased station density in favored areas: near western boundaries—coastal or state administrative—upstream of climatological storm propagation for the former case and in densely populated regions of the Washington and Baltimore metropolitan areas for the latter. Forcing terms are not provided for the equidistant (EQD) or modified equidistant (EQD+) configurations, and these configurations therefore produce a more uniform spatial density of candidate stations within state boundaries. Only MM stations were considered by the simulated annealing method, with the sole exception of the EQD+ configuration which also considered pre-existing (non-MM) surface observing stations when maximizing equidistance.

Specific geographic locations of individual surface observing stations within each network are shown in Fig. 2. For the

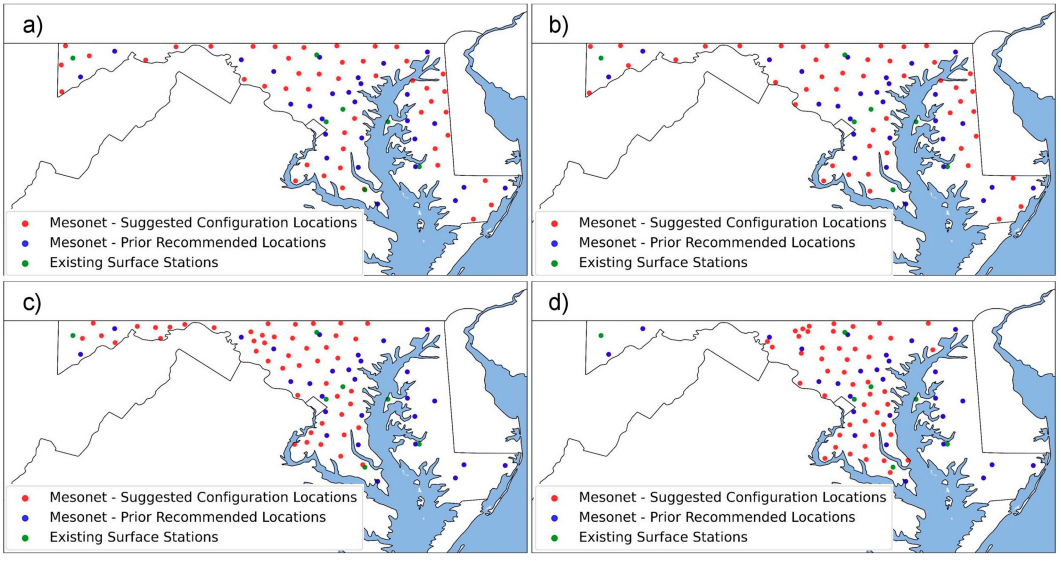

FIG. 2. OSSE candidate MM configurations: (a) network placed to preserve equidistance between MM stations, (b) network placed to preserve equidistance between MM and pre-existing (non-MM) stations, (c) network with greater density in the Appalachian Plateau region of western Maryland, (d) network with greater density in heavily populated areas of the Washington, D.C., and Baltimore metropolitan areas. Configuration-specific candidate locations are indicated by red marks. Predetermined locations for new MM stations are indicated by blue marks. Non-MM surface stations with frequent reporting intervals are indicated by purple marks.

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

e. *Simulations of mesonet impact with approximations for current and future modeling system uncertainty*

To accurately represent the impact of assimilating MM observations in both contemporary and future NWP systems, we performed two separate sets of OSSE experiments using configurations that appropriately simulated modeling system uncertainty. For standard-uncertainty experiments simulating the capabilities of contemporary NWP systems, we ran our forecast model using the double-moment Thompson microphysics scheme (Thompson et al. 2008). This choice induces a model process error in the sense that our parameterization for clouds differs from the NSSL scheme used by the nature run, acting in addition to model error induced by scale discrepancy between our 3-km modeling system and 1-km nature run. In an effort to realistically approximate the uncertainty that is introduced to limited-area modeling systems by parent global models via initial and boundary condition error, we added an additional error source in the form of low wavenumber Gaussian noise applied to initial and lateral boundary conditions used by our modeling system. This additive noise has a wavelength roughly double that of our domain size and simulates the impact that may be expected from synoptic-scale error structures found in global model output. Using forecast output for a case occurring on 13 April 2020, we tuned the amount of added noise in an iterative process and chose the amplitude that produced a 90-min fractions skill score closest to that produced by our modeling system configuration for the same case event when assimilating real observations. For tuning process results, see Fig. A1 in the appendix.

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

f. *Verification metrics for evaluation of MM configurations*

We evaluated the performance of candidate MM configurations by generating verification scores for forecasts and analyses that may be compared against those expected from the pre-existing observing system. Scores were produced with two primary metrics, the first of which was a root-mean-square error (RMSE) for ensemble-mean wind and potential temperature fields at various levels of interest. RMSE is evaluated using gridpoint locations within a verification region corresponding to the state of Maryland, which are regridded to 6-km resolution and then compared to similarly regridded nature run values at corresponding locations. For our second metric, we chose fractions skill score (FSS) (Roberts and Lean 2008), a spatial verification metric that may better represent the qualitative performance of forecasts. FSS considers the ratio of modeled and observed occurrence of events for defined "neighborhoods" of grid points. In doing so, it does not overly penalize small displacement errors in forecasts of discrete weather features, including convective cells and precipitation maxima. We chose two verifying events for these metrics: the occurrence of composite radar reflectivity over 25 dBZ, to correspond with areas of convection and severe storm activity, as well as the occurrence of precipitation rates more than 0.625 mm h^{-1}. Calculations for FSS considered only grid points within a neighborhood window centered on a single gridpoint location. Spatially averaged values were then found by centering a window at each grid point within the verification region—taken again as the state of Maryland—and normalizing the resulting sum. We chose a neighborhood length scale of 15 km, which provides a balanced metric informed by both the resolution of discrete storm structures as well as more general performance characteristics such as precipitation bias.

In addition to RMSE and FSS, we make occasional use of three other event-based metrics: probability of detection (POD), false alarm (FA) rate, and critical success index (CSI), which are given by the following relations:

$$\mathrm{FA} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}, \tag{1}$$

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

$$\mathrm{CSI} = \frac{\mathrm{TP}}{\mathrm{FN} + \mathrm{FP} + \mathrm{TP}}, \tag{3}$$

where TP and FP indicate the number of true and false positives, respectively, within the verification region and FN and TN likewise indicate the number of false negatives and true positives, respectively. These metrics provide qualitative information on forecast behavior beyond that available from our primary metrics.

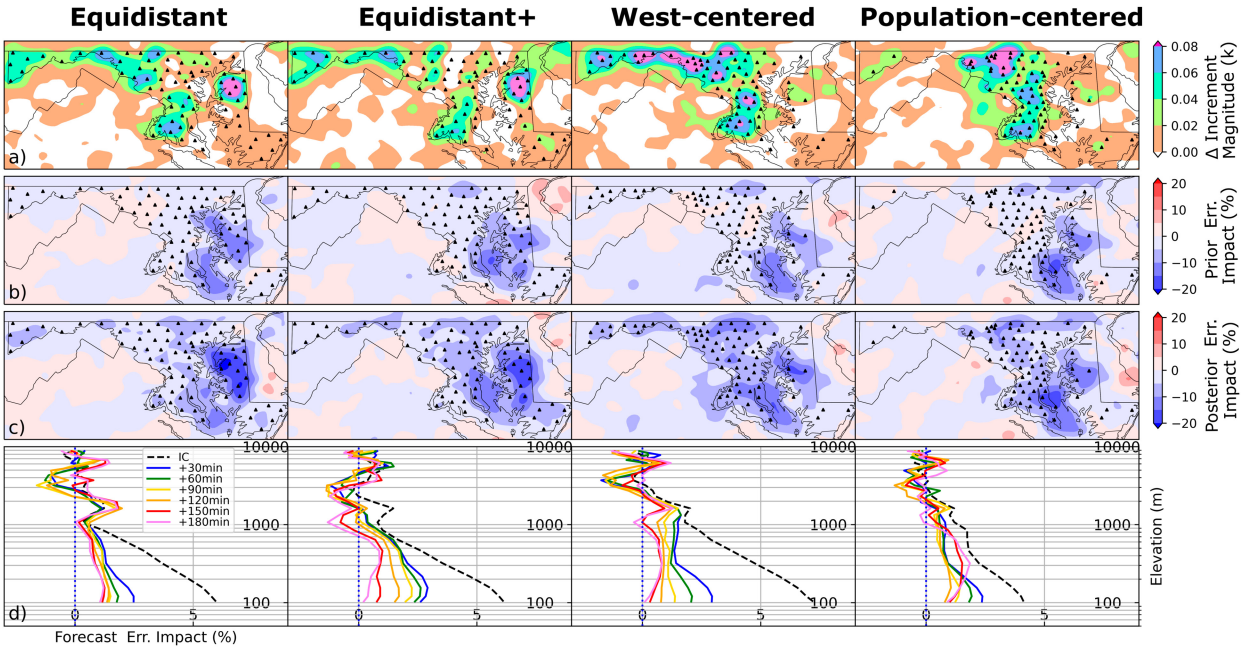

FIG. 3. (a) Near-surface-averaged MM configuration change in analysis increment magnitude for potential temperature, (b) near-surface-averaged MM configuration impact on prior-mean RMSE difference for potential temperature, (c) near-surface-averaged MM configuration impact on analysis-mean RMSE difference for potential temperature, and (d) vertical profile of MM configuration impact on the initial condition and forecast-mean RMSE for potential temperature. Results for standard-uncertainty OSSE.

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

Results for standard-uncertainty and constrained-uncertainty OSSE configurations

Accurate model analyses and derived, diagnostic quantities are crucial for forecaster decision-making alongside numerical forecast output. To gauge the utility of assimilating MM observations on the accuracy of state estimates, we examined prior and posterior quantities over the duration of sequential DA. In Fig. 3, we present several related plots that characterize the spatial distribution of mesonet impact for potential temperature during experiments with our standard-uncertainty modeling system. Corresponding plots for zonal wind speed are shown in Fig. 4. All plots presented consider averages for all forecast initialization times across cases and are stratified columnwise by applied MM configuration. The gray lines in Figs. 3d and 4d depict the vertical structure of MM impact on initial condition error, using the first 20 members of the analysis ensemble. The impact is calculated as a percentage improvement in ensemble-mean RMSE over the no-build case and averaged longitudinally and meridionally within the boundaries of the state of Maryland. Results across MM configurations indicate maximal improvements of 6%–7% for potential temperature and 3%–4% for zonal wind speed that decline with elevation. Improvement is mostly constrained to a vertical range extending from the surface to roughly the height of the planetary boundary layer, with impact becoming negligible above roughly 1000 m. Besides a slight advantage for near-surface zonal wind speed shown for the modified equidistant MM configuration, there is little difference in impacts produced by candidate networks.

We extend this vertical stratification by examining the geographic distribution of MM configuration impact on analysis-mean quantities, considering only a layer roughly below 330 m—corresponding to the five lowest eta levels in our WRF configuration—that appears most strongly affected by mesonet observations (Figs. 3c and 4c). We subsequently refer to impacts averaged over this layer as "near-surface-averaged." Resulting fields show geographically heterogeneous impact distributions for both temperature and zonal wind, with all MM configurations producing benefits that are roughly concentrated within the state boundaries of Maryland and yielding roughly equivalent areas of improvement and degradation in neighboring states. With few exceptions, zones of contiguous improvement are located near MM observation sites, well within the 30-km localization radius of MM observations. The location of maximum near-surface-averaged improvement to analysis means over the default observation network varies with MM configuration, occurring in southern Maryland and the northwestern portion of the Delmarva Peninsula for the two equidistant type configurations. For the west- and population-centered networks, these maxima occur in southern Maryland and in the north-central portion of the state. Near-surface-averaged improvements near these maxima are on the order of 15%–20% of the posterior RMSE values associated with the default observation network for potential temperature and 10%–15% for zonal wind, with potential temperature showing more consistent and geographically contiguous improvements throughout the state.

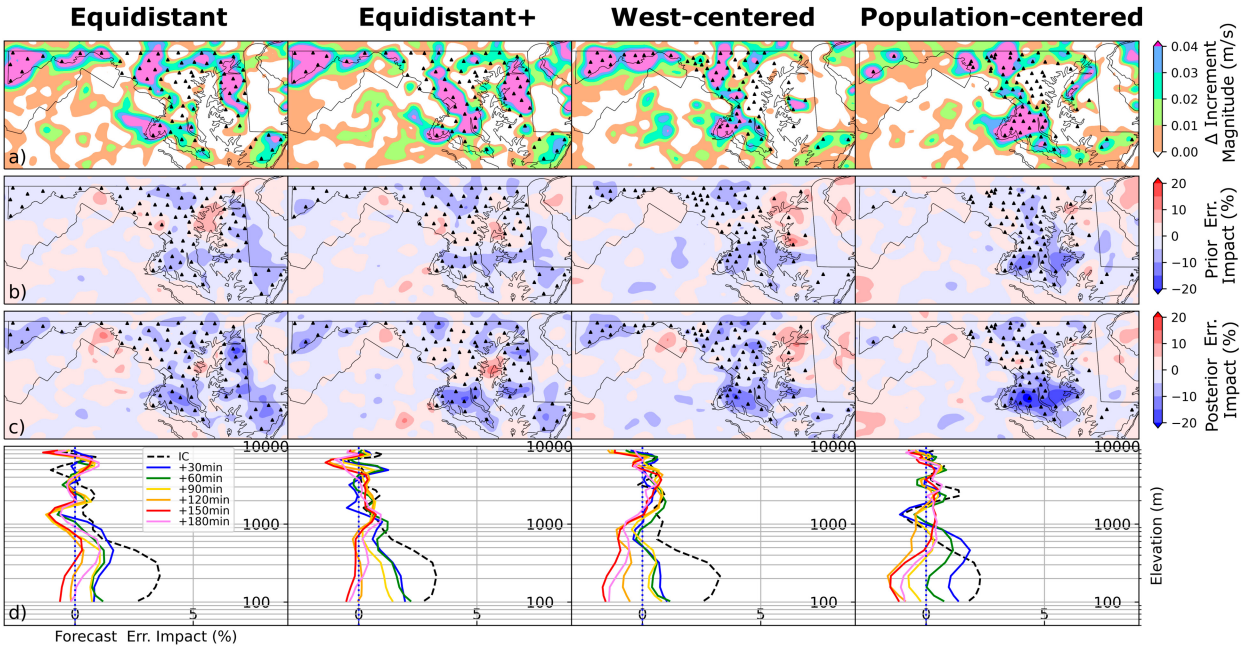

FIG. 4. As in Fig. 3, but for zonal wind speed.

Plots of MM configuration-mediated changes to analysis increment magnitude (ΔInc) (Figs. 3a and 4a) reveal a notable correlation with local site density for most regions outside of central Maryland, with higher configuration site densities producing locally enhanced ΔInc. Since site density also tends to correlate positively with analysis impact in these regions, this suggests that Kalman gains for near-surface variables are reasonably unbiased in such areas leading to positive marginal utility of additional observations. A strong example of this effect can be seen in the northwestern Delmarva Peninsula, where the west- and population-centered MM configurations place comparatively few stations compared to the equidistant MM configurations and are in turn associated with significantly smaller changes to analysis increments and positive analysis error impact for both potential temperature and zonal wind speed. Conversely, regions of central Maryland between Washington, D.C., and Baltimore, Maryland, metropolitan areas, and portions of northeastern Maryland show no distinguishable correlations between site density and ΔInc, and the correlation between site density and positive analysis error impact is greatly muted. We note that these regions correspond rather well with areas where high average densities [25+ observations $(100\ \text{km})^{-1}$] of pre-existing, non-MM surface observations are assimilated in our experiments (Fig. A4).

We repeat the preceding analysis for our constrained-uncertainty modeling system configuration and present corresponding figures for potential temperature (Fig. 5) and zonal wind speed (Fig. 6a). Although the vertical distribution of impact on initial condition error appears quite similar to results for the standard-uncertainty OSSE when considering potential temperature (Fig. 5d), the same results for zonal wind speed show stark differences below 1000 m, with maximal improvements over the no-build configuration reaching only 2%–3% (Fig. 6d) near the surface. While the geographical distribution of impact retains improvement maxima in regions of southern Maryland and the eastern shore indicated for the standard-uncertainty OSSE, there are distinct reductions in analysis-mean improvement for both variables in the central Maryland region and potential temperature in the southern Chesapeake Bay. For zonal wind speed, there is a large increase in the extent of regions suffering analysis-mean degradation compared to the standard-uncertainty experiments (Figs. 5c and 6c). Interestingly—despite shifts in the geographical distribution of analysis-mean impact—the corresponding distribution of MM mediated ΔInc appears quite close to that from the standard-uncertainty OSSE (Figs. 5a and 6a), which suggests that subdued analysis-mean improvement in the constrained-uncertainty experiments is mediated by other factors.

To investigate temporal and case-specific factors affecting near-surface-averaged MM impact on analysis error, we examined smaller sets of analysis means aggregated in 2-h intervals, commencing 3 h after ensemble initialization and for potential temperature only (Fig. A2). We indicate the presence of composite reflectivity above 25 dBZ within these temporal intervals with cross-hatching. Results strongly suggest that the impact of MM observations achieves the greatest amplitude, positive or negative, only after a given local region has been impacted by moist convection, regardless of event setup. It also appears that the majority of statewide reduction to analysis-mean RMSE can be attributed to particular cases and even intervals within specific events. Events centered on 12 August 2020 and 16 July 2022 benefit the most from MM assimilation, with the latter showing an especially positive impact after the passage of a quasilinear convective system (QLCS) across the Chesapeake Bay as shown in Figs. A2d–f. Conversely, cases

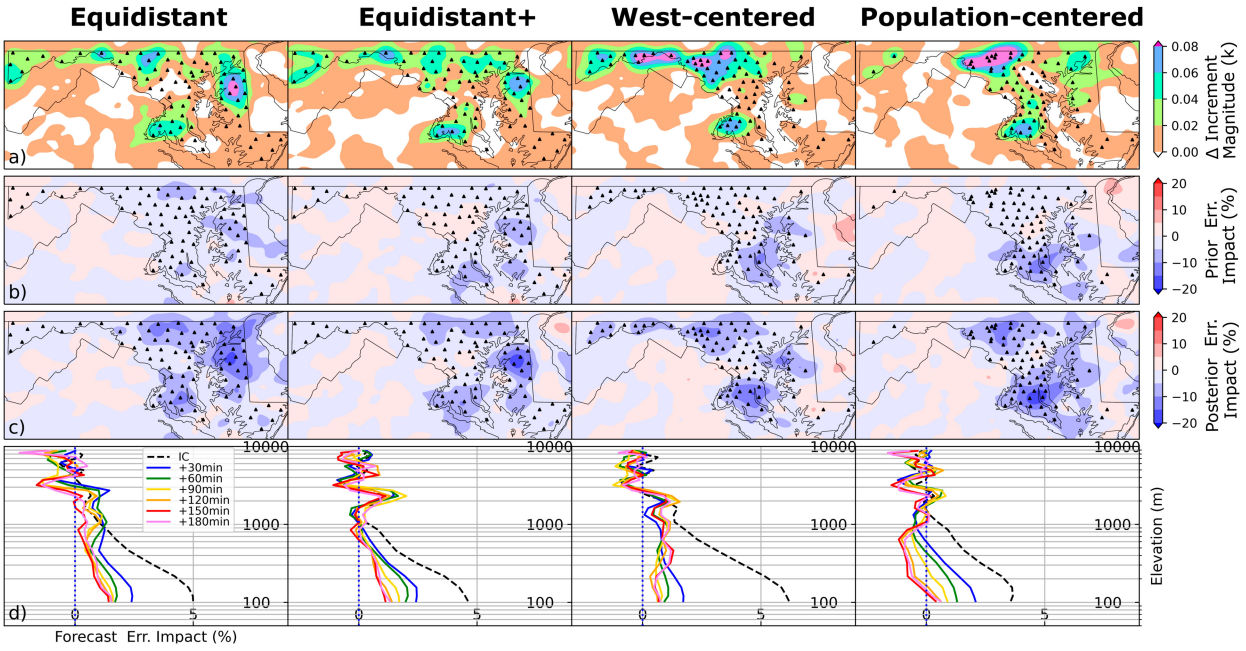

FIG. 5. (a) Near-surface-averaged MM configuration change in analysis increment magnitude for potential temperature, (b) near-surface-averaged MM configuration impact on prior-mean RMSE difference for potential temperature, (c) near-surface-averaged MM configuration impact on analysis-mean RMSE difference for potential temperature, and (d) vertical profile of MM configuration impact on the initial condition and forecast-mean RMSE for potential temperature. Results for constrained-uncertainty OSSE.

centered on 13 April 2020 and 3 September 2020 produce the least benefit from MM assimilation, even after convection has spread throughout the state. Both of these cases feature strong synoptic-scale convective forcing driven by frontal passages, in contrast to other cases where convection is either airmass-driven or else forced as part of a mesoscale convective system (MCS). For experiments using a constrained-uncertainty modeling system configuration (Fig. A3), the aforementioned synoptically driven cases show even less improvement relative to other events. Although the airmass-driven event centered on

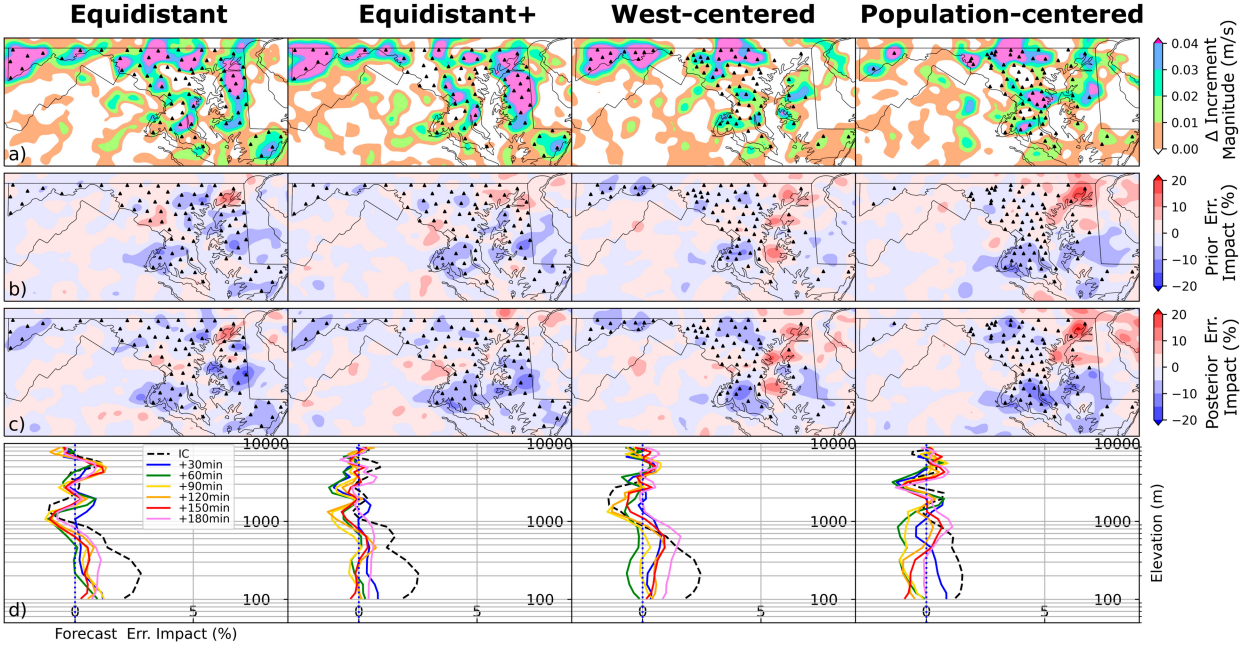

FIG. 6. As in Fig. 5, but for zonal wind speed.

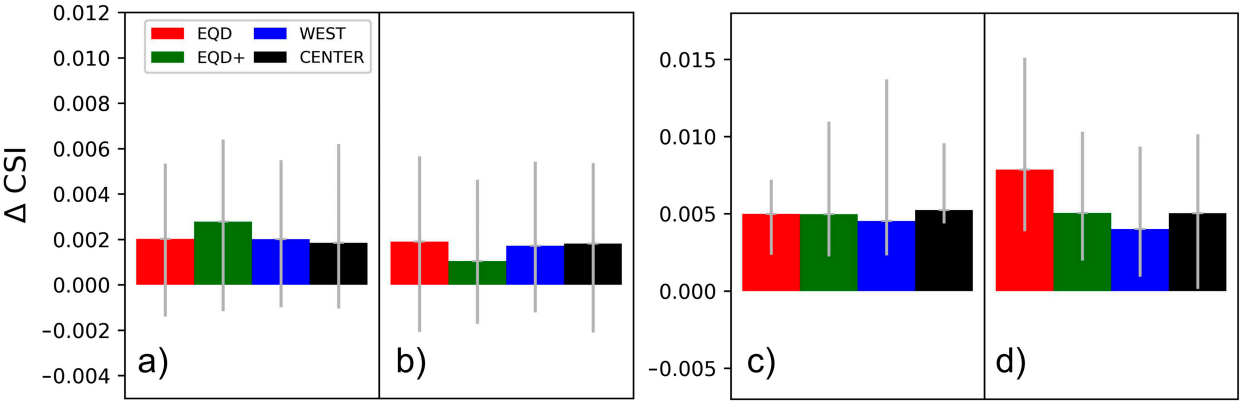

FIG. 7. Case-averaged CSI score differences between each candidate MM configuration and the default observation network for the identification of near-surface horizontal convergence above nature run 90th percentile values for (a) standard-uncertainty modeling system and (b) constrained-uncertainty modeling system. CSI score differences for the identification of θ_v below a threshold indicative of cold pool development for (c) standard-uncertainty modeling system and (d) constrained-uncertainty modeling system. Error bars indicate quartile ranges among considered cases.

12 August 2020 retains the large improvements demonstrated in standard-uncertainty experiments, analyses for the 16 July 2022 case no longer show especially large improvements associated with the QLCS passage occurring toward the end of this case.

Although ensemble-mean RMSE is a powerful metric for quantifying the accuracy of a given field, differences in RMSE between sets of initial conditions contain contributions from errors that are more or less important for forecast evolutions, which may be concentrated in regions containing information directly relevant to the development of localized convective storms. To directly evaluate the representation of dynamically relevant structures within our OSSE framework, we calculated CSI scores—averaged across initial members of all collected forecast ensembles—for the detection of reduced virtual potential temperature (θ_v) indicative of developing cold pools or sea breezes and also for the detection of elevated horizontal convergence at the lowest model level indicative of colliding cold pool gust fronts and other mass boundaries that favor forced convection. The θ_v threshold for cold pool detection was dynamically set based on environmental means following Torri et al. (2015), while for horizontal convergence, we chose to use the 90th percentile of convergence present in the nature run at the corresponding time stamp. We calculated these average scores across all cases and for both the standard-uncertainty and constrained-uncertainty modeling system configurations, with results presented in terms of the case-averaged difference between each candidate MM configuration and the no-build configuration (Fig. 7). Contrast is evident between scores for cold pool identification and the identification of elevated horizontal convergence, with the former showing robust benefit from MM assimilation for all MM configurations and the latter showing only modest median improvement with values for lower quartiles showing degraded member representation of convergence zones. Results therefore suggest that the assimilation of MM observations provides utility in improving the resolution of near-surface thermodynamic features but is much less helpful in refining model depiction of mesoscale frontal boundaries. We note that with the possible exception of cold pool scores for equidistant configurations, there is remarkably little difference in CSI scores associated with individual MM configurations.

As a specific example that highlights how changes to small-scale structures may occur in analyses from the assimilation of MM observations, we examined density-current structures—inclusive of both convectively generated cold pools and sea breezes but referred to here as "cold pools" for simplicity—identifiable at a single time stamp during the 12 August 2020 case event. We again used a threshold θ_v to identify the presence of these features in the nature run, before applying the same thresholding technique to the analysis means produced by our standard-uncertainty OSSE ensemble both with the default observation network and when assimilating the equidistant MM configuration. Focusing on a region of interest centered on the Baltimore–Washington metropolitan area, we plot the outlines of cold pools identifiable in the nature run along with those identifiable from both observation network configurations, as well as prevailing 10-m winds from all three sources (Fig. 8). Assimilating MM observations produces cold pool boundaries that are a closer match to the nature run across central and northeastern Maryland when compared to boundaries placed with the default observation network. This improvement stems primarily from an increase in analysis mean θ_v at the edge of cold pool regions with the effect of reducing their overall extent. While slight improvements to 10-m wind vectors are noted from assimilating MM observations, these occur haphazardly and do appear concentrated near outflow boundaries.

4. Impact of MM observations on forecast verifications

a. Results for standard-uncertainty OSSE configuration

Using our standard-uncertainty modeling system configuration, we generated sets of initial conditions and 3-h forecasts

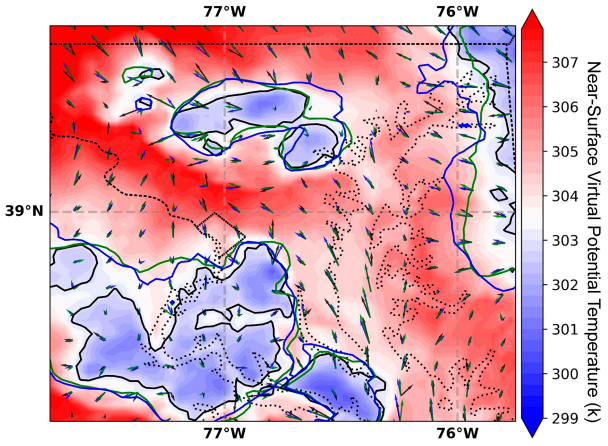

FIG. 8. Cold pool boundaries and 10-m wind vectors valid at 2200 UTC 12 Aug 2020 as depicted by nature run (black), standard-uncertainty OSSE assimilating the default observation network (blue), and standard-uncertainty OSSE assimilating the EQD MM configuration (green). Shading shows base-level θ_v for nature run. Dotted black lines indicate state and coastal boundaries.

corresponding to sequential assimilation of observations from each of our four candidate MM configurations, as well as a set corresponding to assimilation with only the default no-build observation network. We chose to limit verification for resulting forecasts to the earlier defined Maryland verification region given the steep spatial gradient in near-surface-averaged mesonet impact to analyze fields of wind and temperature seen in Figs. 3–6. In Fig. 9a, we present score impacts—defined as the percentage-wise difference between forecast FSS of candidate configurations and that of the no-build configuration—for the occurrence of elevated radar reflectivity and precipitation indicative of severe storm activity. Median and quartile score impacts were produced according to lead time and consider all forecasts produced across case events. This provides, for every MM configuration, a sample size of 133 discrete forecast verifications at 15-min increments of lead time starting from initialization. Median score impacts in this context may be thought of as the effect on verification scores at a given lead time that may be expected from the assimilation of a particular MM configuration, with quartiles acting as an indicator for the variability in improvement (or degradation) at that lead time across sampled forecasts. For the prediction of elevated composite reflectivity (25+ dBZ), results are inconclusive and show a highly variable range of configuration impacts. Although the median differences between FSS using the no-build configuration and that produced using the candidate MM configurations is small, the 25th and 75th percentile score impacts from initialization to 45 min of lead time indicate that a subset of forecasts experience larger benefits of up to 2% from the assimilation of MM observations, with comparatively fewer forecasts showing degradation of similar magnitude. Median results for POD and FA rate (Figs. 9b,c) show no change in the detectability of events, but do indicate a 1%–2.5% reduction in false alarm rate at short

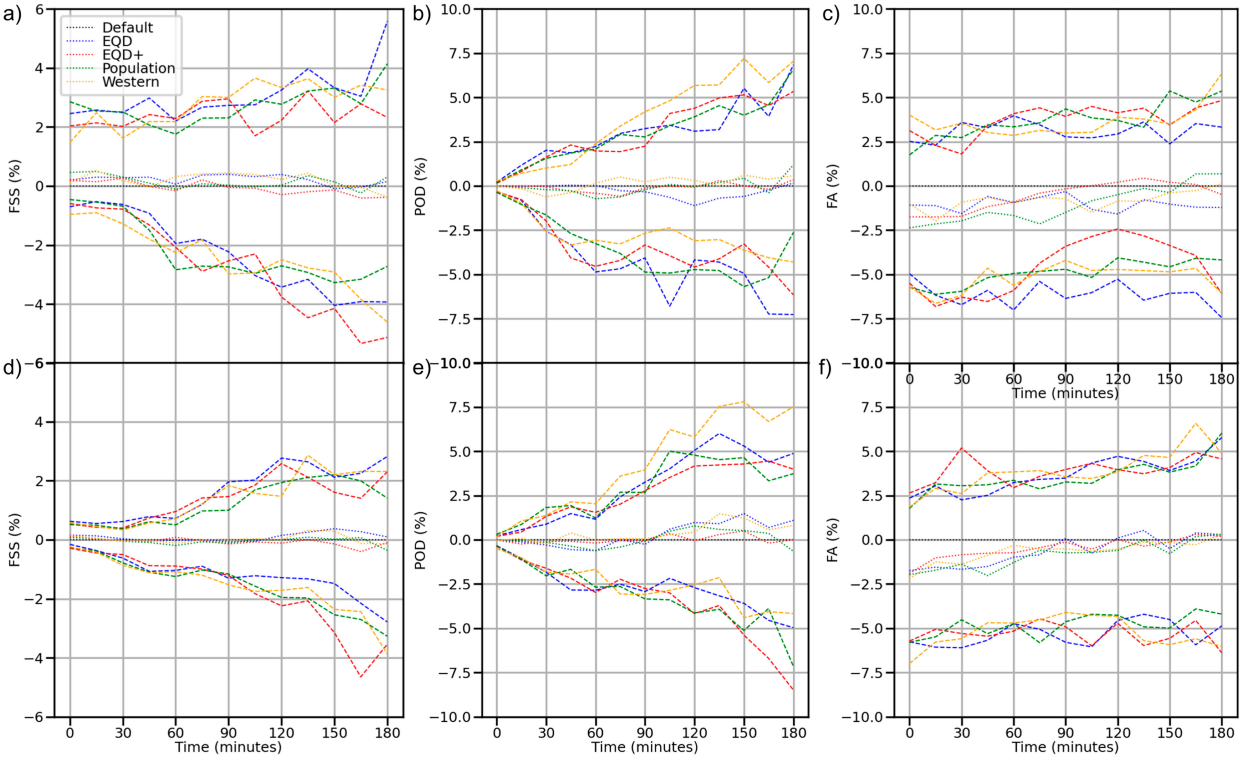

FIG. 9. Network median and quartile percentage score difference from default observation network for (a) forecast FSS, (b) forecast POD, and (c) forecast FA rate, valid for standard-uncertainty OSSE. (d)–(f) As in (a)–(c), but for constrained-uncertainty OSSE.

lead times roughly corresponding to the first 1.5 h after forecast initialization. For the prediction of 15-min accumulated precipitation rates corresponding to moderate rainfall (0.625+ mm h^{-1}), we see no significant impact from the assimilation of MM observations for any of the aforementioned measures (data not shown). None of the aforementioned metrics showed differences between MM configurations that were large compared to the interquartile range.

To augment event-based metrics, we examined the percentagewise RMSE difference (RMSE impact) between no-build and candidate configurations for forecast temperature and zonal wind fields, again averaging these scores within the Maryland verification region and across all available forecasts. We present color-coded vertical profiles of mean RMSE impacts for wind and temperature fields at lead times from 30 to 180 min (Figs. 3d and 4d), overlaying corresponding impact profiles for forecast initial conditions introduced in the previous section. This figure depicts how forecast error relates to improvements afforded to analysis fields, both above and below the planetary boundary layer. Results for potential temperature show that—with the exception of the west-centered network—analysis improvements below 1000 m persist throughout the 180-min forecast period, at which point they have decayed to roughly 0.5%–1.5% from 5% to 7% present in initial conditions. For zonal wind speed, improvements decay to near zero within the same layer by 60–90 min. Neither variable shows robust forecast improvements above 1000 m at any lead time, indicating that improvements to analysis fields at lower elevations do not translate to the free atmosphere during forecasts.

Although we do not break down the geographical distribution of forecast improvements by lead time, in Figs. 3b and 4b, we present plots of near-surface-averaged MM configuration impact on prior-mean error for potential temperature and zonal wind speed. These were produced in the same way as corresponding plots for analysis means (Figs. 3c and 4c) and may be seen as a proxy for short-term, 15-min forecasts produced by our modeling system over the course of sequential DA. Comparisons with analysis-mean impacts show that for potential temperature, improvements are preferentially degraded in western and central Maryland, with stronger persistence for southeastern regions of the state surrounding the Chesapeake Bay. For zonal wind, regions of positive analysis impact are degraded somewhat more evenly though the quickest decay still occurs for western Maryland.

b. *Results for constrained-uncertainty modeling system featuring perfect physics*

To examine how the utility of candidate mesonet configurations would change given advances in physical parameterization schemes and limited initial condition uncertainty for large-scale flow—as depicted by reduced boundary condition uncertainty—we generated further sets of initial conditions and forecasts using our constrained-uncertainty modeling system configuration. As before, we examine fractions skill score curves for radar reflectivity (Fig. 9d) and precipitation (not shown) corresponding to aggregated forecast performance across events stratified by lead time. For verification of composite reflectivity above 25 dBZ, we find that the range of impact from the assimilation of MM observations is smaller than that for the standard-uncertainty set of experiments, both in the median and for the interquartile spread of score impact. This is in part due to an overall increase in FSS at analysis times associated with reduced model error and improved boundary conditions. Unlike the standard-uncertainty case, any benefit from the assimilation of MM observations is limited to the analysis time where it is again present only in a subset of forecasts contributing to 25th and 75th percentile score impacts. As before, verification of precipitation rates above 0.625 mm h^{-1} shows no detectable benefit from MM observations although the variance of skill-score impact is also reduced in a manner similar to that for composite reflectivity.

Vertical profiles of forecast-mean impacts for potential temperature (Fig. 5d) show similar lead-time progressions to that for the standard-uncertainty experiments, maintaining small improvements below 1000 m for the duration of forecasts. However, equivalent profiles for zonal-mean wind (Fig. 6d) show a much sharper lead-time cutoff in terms of useful impact near the surface, with little improvement remaining past 30 min for most MM configurations. As before, no forecast improvement is seen above 1000 m for either potential temperature or zonal wind. Likewise, the geographical distribution of prior-mean impacts presented in Figs. 3b and 4b indicates that improvements to analysis accuracy associated with MM observations are preferentially retained in southeastern Maryland for both variables.

5. Discussion

a. *Findings on the utility of a Maryland Mesonet*

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

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

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

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

b. *Findings on the importance of station placement within MM networks*

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

c. *Implications for convective-scale DA and future observing systems*

We may extrapolate the results of our OSSE experiments to provide more generalized insight into factors that drive forecast errors in moist convective regimes and how they may limit the current and potential utility of assimilating denser networks of surface observations in limited-area, convective-allowing NWP systems. From a simplified perspective governed by Bayesian inference, forecast error arises from a combination of initial condition and model errors (Privé and Errico 2013), with

the interaction of the two depending on the particular nature of the dynamical system, its discretization, and consequent structure of both true and modeled phase space trajectories (Shaw 1981). Although not quite as rigorous as formal sensitivity analysis (Torn and Hakim 2008), our standard-uncertainty and constrained-uncertainty OSSE configurations sampled two possible permutations of these errors from which to conclude the former featuring elevated model error from choice of microphysics scheme and initial conditions degraded by synoptic-scale perturbations to boundary conditions and the latter featuring no such synoptic-scale forcing and only a small discretization component of model error. While both cases show reductions in near-surface initial condition errors attributable to assimilating MM observations, this improvement does not scale monotonically with station density in the sense of reductions to squared errors and is insufficient to improve forecast skill for the prediction of moist convective events even when forecasts are integrated with near perfect boundary conditions and physics. Furthermore, residual improvements to near-surface forecast fields of wind and temperature that exist for lead times below 1 h for our standard-uncertainty experiments are notably reduced in our latter experiments where error is predominantly generated in regions of convective activity. While the lack of improvement to forecast verifications at extended lead times is undoubtedly consistent with the restricted spatial extent of the Maryland Mesonet and consequent advection of information out of the verification domain, that such improvements are minimal even at short lead times suggests that the ability of surface observations to improve model representations of moist convection may be limited by aspects of intrinsic predictability.

Some studies that have taken an analytical approach to quantifying intrinsic predictability limits of moist convection (Ramanathan et al. 2019; Ramanathan and Satyanarayana 2021) suggest that storm features near the upper end of the meso-γ scale may retain a degree of predictability for up to 6 h, with further extended predictability possible for highly organized systems. Though this would appear to leave room for improvement on practical predictability time scales achievable by NWP systems, the damped-driven dynamics of moist convective systems may result in a high rate of "information turnover" compared to synoptic-scale processes. As an example motivated by Fig. 7, in situations where convection is heavily favored at particular locations, it may be unnecessary to refine the model depiction of outflow boundaries beyond a certain level of accuracy to obtain a reasonable forecast of convective activity. Conversely, in weakly sheared, conditionally unstable environments where convection is disorganized and heavily dependent on outflow forcing, noticeably improving forecast accuracy may require an unachievable amount of constraint on outflow boundaries. Alternatively, improvements to near-surface analysis fields and even to resolved cold pool structures may be diluted by poorly resolved processes occurring above 1000 m—including entrainment, detrainment, and downdraft formation—where surface observations have little impact. We note that deep moist convective events occur in a high-dimensional phase space involving thermodynamic, momentum, and moisture fields over the full height of the troposphere (Rotunno et al. 1988). In the context of our experiments, limitations specific to moist convection would be most evident in the constrained-uncertainty experiment where "low-hanging fruit" in the form of large-scale initial condition error leading to downscale error propagation has been addressed and where indeed we see the least evidence of persistent RMSE benefit to forecast fields. While the role of model error in this scenario is less clear, our results suggest that even drastic improvements to physical parameterizations are insufficient to significantly increase the utility of surface observations. This could be the case if dynamical attractors relevant to moist convection are already reasonably well defined by an imperfect model, making model error less important than sampling deficiency and other factors such as improper assumptions made by data assimilation algorithms. If predictability aspects of moist convective systems do in fact reduce the amenability of such regimes to surface data assimilation, then we would expect the marginal utility of additional surface observations to decrease and reach saturation at lower density as improvements to global models increasingly reduce the amplitude of error at synoptic scales.

d. *Experiment limitations*

Errico and Privé (2018) note that regional OSSEs of the type used for this study may present issues in terms of accurate representation of boundary conditions and further warn that ill-posed OSSE frameworks may spuriously indicate utility for candidate observing systems. We note that although we do not use a global ensemble for initial and boundary condition generation, this would not be a factor in our constrained-uncertainty OSSE configuration that removes lateral boundary condition (LBC) perturbations larger than the meso-γ spatial scale. Another potential limitation in our experimental framework is the nonconsideration of observation error correlations that are sometimes parameterized in highly tuned OSSEs for operational NWP systems. We suspect that given our assimilated observation types, the primary effect of considering these correlations would be a slightly reduced impact from radar observations in areas of moist convection. Although this could feasibly lead to greater benefit from assimilating MM observations in our standard-uncertainty OSSE, parameterized correlations would likely not hold for future changes to radar networks or improved statistical techniques to reduce existing error correlations. Furthermore, we note that although humidity observations are often ingested by NWP systems, we chose to assimilate only temperature and wind measurements provided by mesonet stations due to difficulties in properly representing the assimilation of bounded quantities in an OSSE framework. Given the importance of lower-level moisture in properly simulating conditional instability, it is conceivable that these observations may mediate yet-undetected NWP benefits from additional surface observing stations. Finally, we acknowledge that our sample of warm-season moist convective case events necessarily excluded other common causes of severe weather. Further study may be warranted for examining the impact of mesonet observations for the prediction of winter weather events, particularly those with convective elements or shear gradients in surface

precipitation type that may be amenable to the high resolution possible with mesonets.

6. Summary and conclusions

The current study presents results from a set of observing system simulation experiments that evaluated the impact of a new Maryland Mesonet on analyses and forecasts produced using a convective-allowing model framework. Although the Maryland Mesonet has wide-ranging aims, this study focused on implications for the predictability of severe convective storms due to their prevalence and impact on the local region. Accordingly, the adopted OSSE framework simulated mesonet impact for seven moist convective case events, with the final selection influenced by event severity in densely populated areas of the state. We hypothesized that additional surface observations may reduce analysis error for quantities relevant to the development and propagation of convective storms—especially near-surface wind and temperature anomalies associated with cold pools—and result in lengthier periods of predictability for moist convective events than if these observations were not assimilated. We further proposed that such benefits may be mediated by the spatial configuration of mesonet site locations, making it desirable to study and optimize impacts associated with specific configurations. Initial conditions and forecasts were produced using several candidate spatial configurations for mesonet site location and then verified against a "truth state" defined by our nature run using both continuous and categorical statistics. Simulations also considered the role of common modeling system errors and how anticipated improvements to boundary condition constraints and model process error may affect the utility of mesonet observations in future modeling systems. To answer these questions in an OSSE framework, we performed two sets of experiments for every set of configurations: one using a "standard uncertainty" configuration with the simulated model process and boundary condition error with synoptic-scale structure and a "constrained uncertainty" configuration with perfect physics and low-amplitude, high-frequency boundary condition errors. For analysis uncertainty, we found that assimilation of any of our mesonet configurations is associated with modest reductions in analysis error for near-surface temperature and zonal wind fields that is restricted to a layer below 1000 m, consistent with previous studies that assimilated similar observing systems. Analysis error reduction is slightly more enhanced and extensive for experiments with our standard-uncertainty modeling system compared to those performed with the constrained-uncertainty configuration, especially for near-surface zonal wind which we attribute to the relative ease of sampling error structures induced by large-scale boundary perturbations compared to errors propagating upscale in moist convective regimes. While increased accuracy for thermodynamic and wind fields is maintained in short-term forecasts—most visibly with our standard-uncertainty modeling system configuration—it does not translate to detectable improvement in forecast skill for the occurrence of discrete convective events. This may suggest limitations in the ability of dense surface observing networks to extend the practical predictability of moist convective regimes due to scale-dependent aspects of involved dynamical systems and the nature of their coupling to surface processes. Although our results do not show drastic differences in analysis or forecast improvement between candidate mesonet configurations, we conclude that optimal improvements to analysis accuracy may be achieved by roughly equidistant placement that avoids areas where pre-existing surface data heavily constrain error structures resolvable by operational DA systems and where remaining uncertainty pertaining to near-surface variables is mediated by meso-γ-scale structures.

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

Data availability statement. HRRR data are publicly available from NOAA at archives hosted by Amazon Web Services (https://registry.opendata.aws/noaa-hrrr-pds/) and Google Cloud Platform (https://console.cloud.google.com/marketplace/product/noaa-public/hrrr?project=python-232920&pli=1). Code used to generate ensemble analyses and forecasts, along with synthetic observations used in OSSEs, is available on a GitHub repository (https://github.com/jmccurry95/Maryland-Mesonet-OSSE).

APPENDIX

OSSE Tuning and Regional Breakdown of Mesonet Impact

Figure A1 shows a breakdown of binary verification scores associated with OSSE tuning configurations intended to simulate the capabilities of contemporary NWP systems. Scores are calculated from forecasts initialized for the case event occurring on 13 April 2020.

Figures A2 and A3 show a breakdown of near-surface-averaged mesonet impact on posterior error for potential temperature during the period of sequential cycling for standard-uncertainty and constrained-uncertainty experiments, respectively. Individual panels consider data in 2-h intervals commencing 3 h after ensemble initialization.

Figure A4 shows the average density of pre-existing surface observations considered in a 15-min assimilation cycle within our OSSE modeling system.

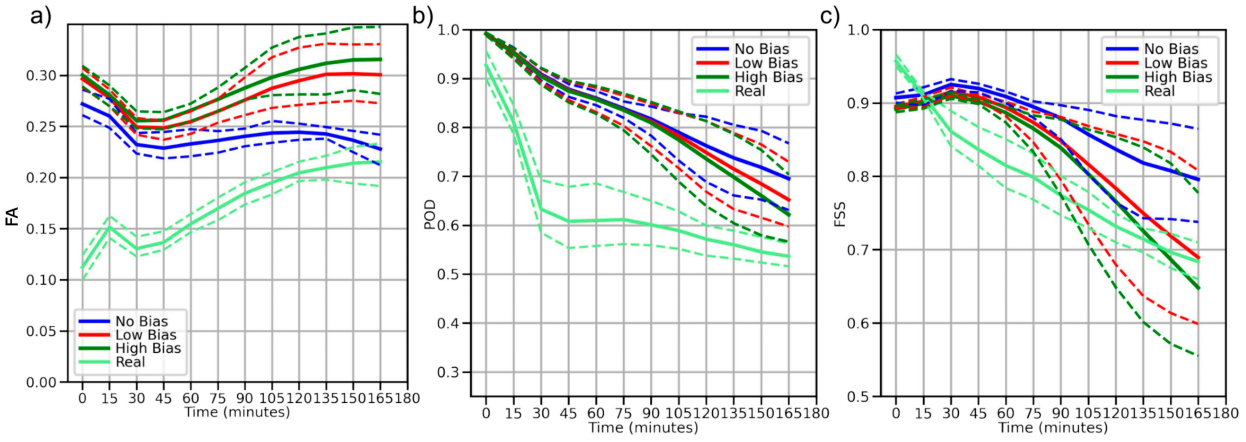

FIG. A1. (a) FA rate, (b) POD, and (c) FSS for forecasts with a real test case and for corresponding OSSE forecasts without bias, with low-amplitude applied bias, and with high-amplitude applied bias.

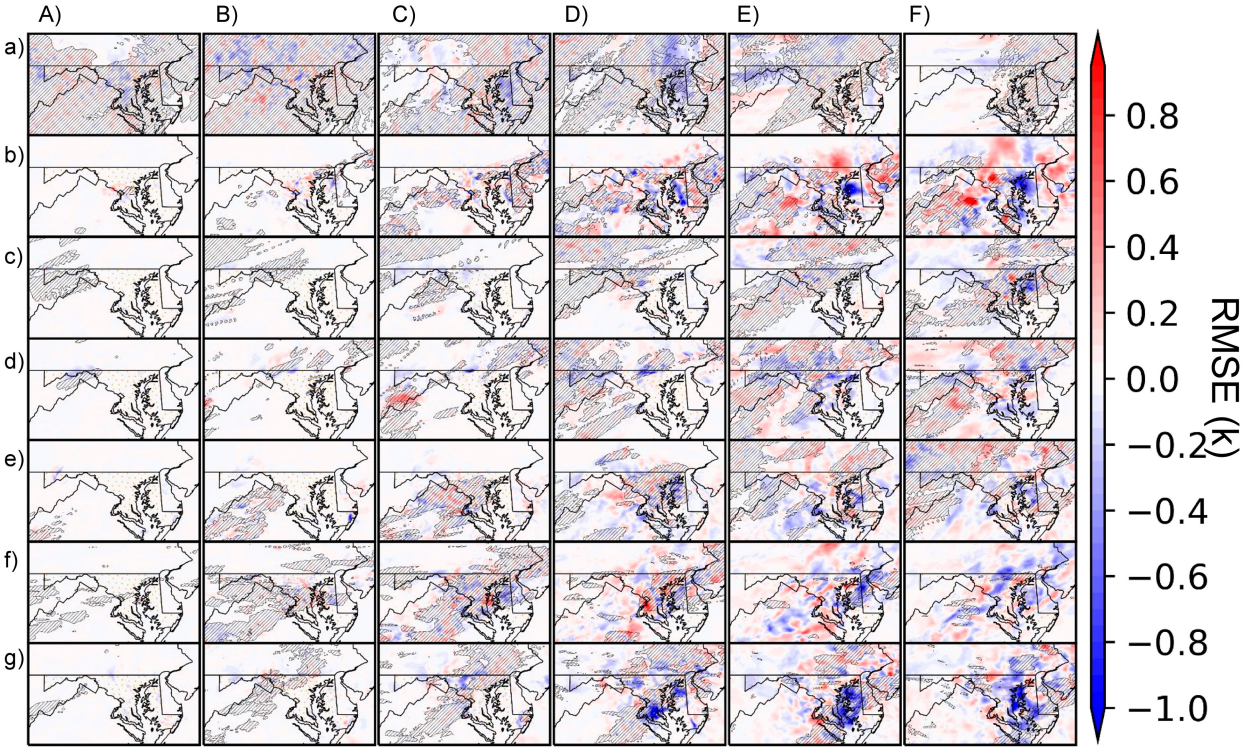

FIG. A2. Posterior RMSE for near-surface potential temperature field for case events beginning on (a) 13 Apr 2020, (b) 12 Aug 2020, (c) 3 Sep 2020, (d) 17 Jul 2021, (e) 8 Jun 2022, (f) 2 Jul 2022, and (g) 16 Jul 2022 for 2-h intervals starting (A) 3 h, (B) 5 h, (C) 7 h, (D) 9 h, (E) 11 h, and (F) 13 h after ensemble initialization. Occurrence of composite reflectivity above 30 dBZ during interval indicated with cross-hatching. Valid for standard-uncertainty experiments.

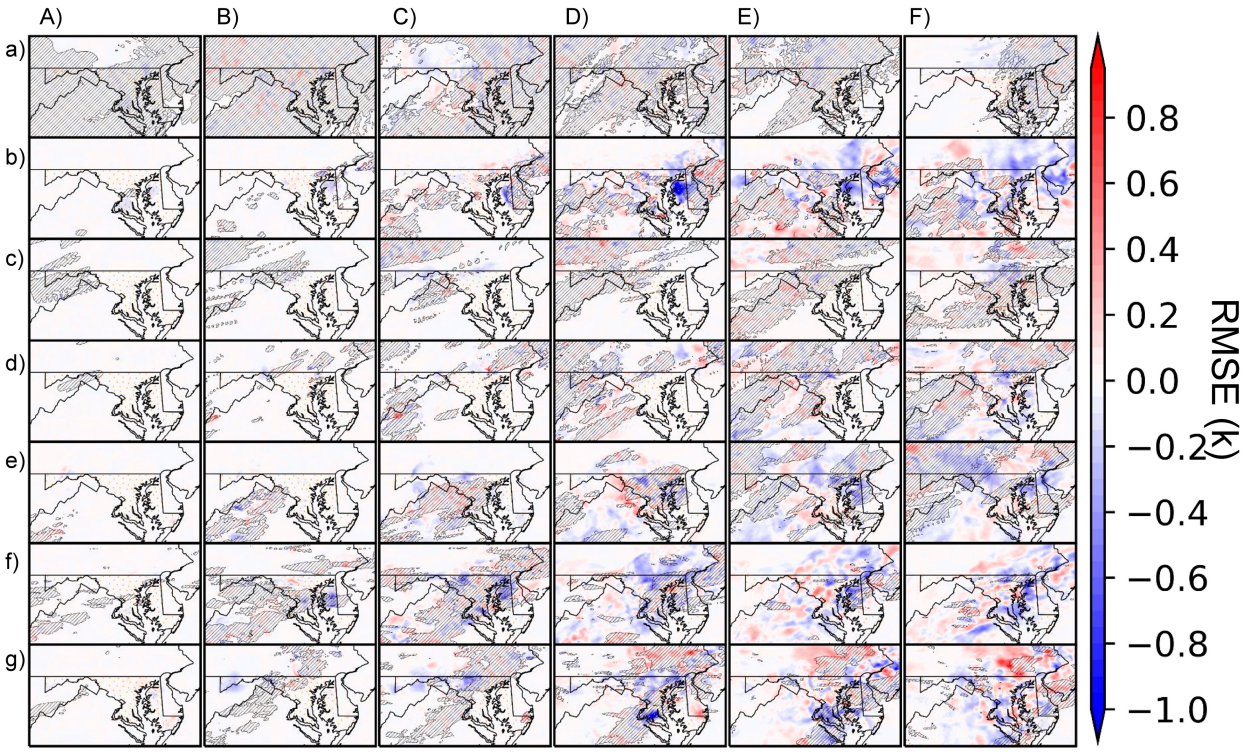

FIG. A3. As in Fig. A2, but for constrained-uncertainty experiments.

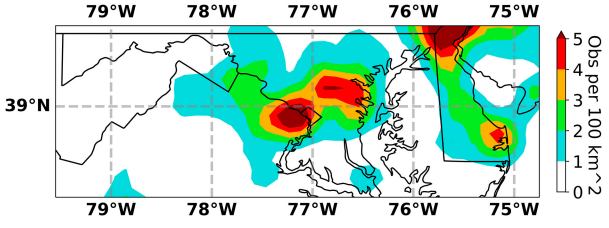

FIG. A4. Density per 15-min assimilation window of surface-based observations publicly available from MADIS.

REFERENCES

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, https://doi.org/10.1175/2009BAMS2618.1.

Bertsimas, D., and J. Tsitsiklis, 1993: Simulated annealing. *Stat. Sci.*, **8**, 10–15, https://doi.org/10.1214/ss/1177011077.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927, https://doi.org/10.1175/2008JTECHA1156.1.

Errico, R. M., and N. C. Privé, 2018: Some general and fundamental requirements for designing Observing System Simulation Experiments (OSSEa). WMO Rep. WWRP, 24 pp., https://ntrs.nasa.gov/api/citations/20190025338/downloads/20190025338.pdf.

——, R. Yang, N. C. Privé, K.-S. Tai, R. Todling, M. E. Sienkiewicz, and J. Guo, 2013: Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office. *Quart. J. Roy. Meteor. Soc.*, **139**, 1162–1178, https://doi.org/10.1002/qj.2027.

Fiebrich, C. A., and Coauthors, 2020: Toward the standardization of mesoscale meteorological networks. *J. Atmos. Oceanic Technol.*, **37**, 2033–2049, https://doi.org/10.1175/JTECH-D-20-0078.1.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

Ghonima, M. S., H. Yang, C. K. Kim, T. Heus, and J. Kleissl, 2017: Evaluation of WRF SCM simulations of stratocumulus-topped marine and coastal boundary layers and improvements to turbulence and entrainment parameterizations. *J. Adv. Model. Earth Syst.*, **9**, 2635–2653, https://doi.org/10.1002/2017MS001092.

Ha, S.-Y., and C. Snyder, 2014: Influence of surface observations in mesoscale data assimilation using an ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 1489–1508, https://doi.org/10.1175/MWR-D-13-00108.1.

Hodyss, D., and M. Morzfeld, 2023: How sampling errors in covariance estimates cause bias in the Kalman gain and impact ensemble data assimilation. *Mon. Wea. Rev.*, **151**, 2413–2426, https://doi.org/10.1175/MWR-D-23-0029.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shepard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

James, E. P., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, https://doi.org/10.1175/WAF-D-21-0130.1.

Janjić, T., and Coauthors, 2017: On the representation error in data assimilation. *Quart. J. Roy. Meteor. Soc.*, **144**, 1257–1278, https://doi.org/10.1002/qj.3130.

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012: A revised scheme for the WRF surface layer formulation. *Mon. Wea. Rev.*, **140**, 898–918, https://doi.org/10.1175/MWR-D-11-00056.1.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, https://doi.org/10.1175/WAF-D-15-0107.1.

——, P. Skinner, K. Knopfmeier, E. Mansell, P. Minnis, R. Palikonda, and W. Smith Jr., 2018: Comparison of cloud microphysics schemes in a Warn-on-Forecast system using synthetic satellite objects. *Wea. Forecasting*, **33**, 1681–1708, https://doi.org/10.1175/WAF-D-18-0112.1.

Khare, S. P., and J. L. Anderson, 2006: An examination of ensemble filter based adaptive observation methodologies. *Tellus*, **58A**, 179–195, https://doi.org/10.1111/j.1600-0870.2006.00163.x.

Knopfmeier, K. H., and D. J. Stensrud, 2013: Influence of mesonet observations on the accuracy of surface analyses generated by an ensemble Kalman filter. *Wea. Forecasting*, **28**, 815–841, https://doi.org/10.1175/WAF-D-12-00078.1.

Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, https://doi.org/10.1175/WAF-D-17-0145.1.

Majcen, M., P. Markowski, Y. Richardson, D. Dowell, and J. Wurman, 2008: Multipass objective analyses of Doppler radar data. *J. Atmos. Oceanic Technol.*, **25**, 1845–1858, https://doi.org/10.1175/2008JTECHA1089.1.

Mansell, E. R., 2010: On sedimentation and advection in multimoment bulk microphysics. *J. Atmos. Sci.*, **70**, 3084–3094, https://doi.org/10.1175/2010JAS3341.1.

Marquis, J., Y. Richardson, P. Markowski, D. Dowell, J. Wurman, K. Kosiba, P. Robinson, and G. Romine, 2014: An investigation of the Goshen County, Wyoming, tornadic supercell of 5 June 2009 using EnKF assimilation of mobile mesonet and radar observations collected during VORTEX2. Part I: Experiment design and verification of the EnKF analyses. *Mon. Wea. Rev.*, **142**, 530–554, https://doi.org/10.1175/MWR-D-13-00007.1.

McCurry, J., J. Poterjoy, K. Knopfmeier, and L. Wicker, 2023: An evaluation of non-Gaussian data assimilation methods in moist convective regimes. *Mon. Wea. Rev.*, **151**, 1609–1629, https://doi.org/10.1175/MWR-D-22-0260.1.

Miller, P. A., M. F. Barth, and L. A. Benjamin, 2005: An update on MADIS observation ingest, integration, quality control, and distribution capabilities. *21st Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, San Diego, CA, Amer. Meteor. Soc., J7.12, https://ams.confex.com/ams/Annual2005/techprogram/paper_86703.htm.

Necker, T., S. Geiss, M. Weissmann, J. Ruiz, T. Miyoshi, and G.-Y. Lien, 2020: A convective-scale 1,000-member ensemble simulation and potential applications. *Quart. J. Roy. Meteor. Soc.*, **146**, 1423–1442, https://doi.org/10.1002/qj.3744.

Pannekoucke, O., L. Berre, and G. Desroziers, 2008: Background-error correlation length-scale estimates and their sampling statistics. *Quart. J. Roy. Meteor. Soc.*, **134**, 497–508, https://doi.org/10.1002/qj.212.

Potvin, C. K., and Coauthors, 2020: Assessing systematic impacts of PBL schemes on storm evolution in the NOAA Warn-on-Forecast System. *Mon. Wea. Rev.*, **148**, 2567–2590, https://doi.org/10.1175/MWR-D-19-0389.1.

Privé, N. C., and R. M. Errico, 2013: The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment. *Tellus*, **65A**, 21740, https://doi.org/10.3402/tellusa.v65i0.21740.

Ramanathan, A., and A. N. V. Satyanarayana, 2021: Satellite-based estimate of intrinsic predictability limits at convective scales over northeast India. *Earth Space Sci.*, **8**, e2019EA000797, https://doi.org/10.1029/2019EA000797.

——, ——, and M. Mandal, 2019: Theoretical predictability limits of spatially anisotropic multifractal processes: Implications for weather prediction. *Earth Space Sci.*, **6**, 1067–1080, https://doi.org/10.1029/2018EA000528.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A theory for strong, long-lived squall lines. *J. Atmos. Sci.*, **45**, 463–485, https://doi.org/10.1175/1520-0469(1988)045%3C0463:ATFSLL%3E2.0.CO;2.

Shaw, R., 1981: Strange attractors, chaotic behavior, and information flow. *Z. Naturforsch. A*, **36**, 80–112, https://doi.org/10.1515/zna-1981-0115.

Skamarock, W. C., and Coauthors, 2019: A description of the Advanced Research WRF Model version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp., https://doi.org/10.5065/1dfh-6p97.

Smirnova, T. G., J. M. Brown, S. G. Benjamin, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle land surface model (RUC LSM) available in the Weather Research and Forecasting (WRF) model. *Mon. Wea. Rev.*, **144**, 1851–1865, https://doi.org/10.1175/MWR-D-15-0198.1.

Sobash, R. A., and D. J. Stensrud, 2015: Assimilating surface mesonet observations with the EnKF to improve ensemble forecasts of convection initiation on 29 May 2012. *Mon. Wea. Rev.*, **143**, 3700–3725, https://doi.org/10.1175/MWR-D-14-00126.1.

——, and L. J. Wicker, 2015: On the impact of additive noise in storm-scale EnKF experiments. *Mon. Wea. Rev.*, **143**, 3067–3086, https://doi.org/10.1175/MWR-D-14-00323.1.

Stensrud, D. J., N. Yussouf, D. C. Dowell, and M. C. Coniglio, 2009: Assimilating surface data into a mesoscale model ensemble: Cold pool analyses from spring 2007. *Atmos. Res.*, **93**, 207–220, https://doi.org/10.1016/j.atmosres.2008.10.009.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Torn, R. D., and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677, https://doi.org/10.1175/2007MWR2132.1.

Torri, G., Z. Kuang, and Y. Tian, 2015: Mechanisms for convection triggering by cold pools. *Geophys. Res. Lett.*, **42**, 1943–1950, https://doi.org/10.1002/2015GL063227.

Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

Wicker, L., 2017: python replacement for the Observation Processing Analysis and Wind Synthesis Program (OPAWS). Github, accessed 15 August 2019, https://github.com/louiswicker/pyOPAWS.