SPECIAL PECAN COLLECTION

Impacts of Methods for Estimating the Observation Error Variance for the Frequent Assimilation of Thermodynamic Profilers on Convective-Scale Forecasts

SAMUEL K. DEGELIA[a] AND XUGUANG WANG[a]

[a] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 15 March 2021, in final form 16 November 2022)

ABSTRACT: The observation error covariance partially controls the weight assigned to an observation during data assimilation (DA). True observation error statistics are rarely known and likely vary depending on the meteorological state. However, operational DA systems often apply static methods that assign constant observation errors across a dataset. Previous studies show that these methods can degrade forecast quality when assimilating ground-based remote sensing datasets. To improve the impact of assimilating such observations, we propose two novel methods for estimating the observation error variance for high-frequency thermodynamic profilers. These methods include an adaptive observation error inflation technique and the Desroziers method that directly estimates the observation error variances using paired innovation and analysis residuals. Each method is compared for a nocturnal mesoscale convective system (MCS) observed during the Plains Elevated Convection at Night (PECAN) experiment. In general, we find that these novel methods better represent the large variability of observation error statistics for high-frequency profiles collected by Atmospheric Emitted Radiance Interferometers (AERIs). When assimilating AERIs by statically inflating retrieval error variances, the trailing stratiform region of the MCS is degraded compared to a baseline simulation with no AERI data assimilated. Assimilating the AERIs using the adaptive inflation or Desroziers method results in better maintenance of the trailing stratiform region and additional suppression of spurious convection. The forecast improvements from these novel methods are primarily linked to increased error variances for some moisture retrievals. These results indicate the importance of accurately estimating observation error statistics for convective-scale DA and suggest that accounting for flow dependence can improve the impacts from assimilating remote sensing datasets.

KEYWORDS: Remote sensing; Data assimilation; Numerical weather prediction/forecasting

1. Introduction

From a Bayesian perspective, data assimilation (DA) involves computing the posterior probability density function of a model state given prior probabilities and observation likelihoods (Kalnay 2003). Much previous DA research focuses on improving the structure of the former through ensemble- or hybrid-based methods that can diagnose flow-dependent background error correlations (Whitaker and Hamill 2002; Wang et al. 2007, 2008a,b, 2009, 2013; Buehner et al. 2010; Johnson et al. 2015). Despite the background and observation error statistics playing an equally important role, relatively fewer studies explore potential improvements when estimating the observation error covariance **R**. Much of this research gap is due to the prior assumption that conventional observation sets feature observation error statistics that change little in time or space and thus can be represented by a Gaussian-distributed, constant error variance (Fowler and Van Leeuwen 2013).

Observation errors primarily consist of three components (Fowler and Van Leeuwen 2013): systematic errors (biases), instrument errors, and representation errors. Systematic errors, if nonnegligible, should be treated before DA through bias correction schemes. Instrument errors represent random observation noise that is also typically small. Finally, representation errors occur due to mismatches between the observed and modeled variable (Janjić et al. 2018). These can include errors associated with the observation operator (e.g., errors in a radiative transfer model), or due to the observation sampling a different scale than the model state. The latter can occur either when the observation samples finer-scale features than can be resolved in the coarser model grid box, or when the model has a higher resolution than the observation (Janjić et al. 2018). Given their relation to the atmospheric state, representation errors are sometimes referred to as "flow-dependent" observation errors (Minamide and Zhang 2017). Remote sensing instruments such as satellite or ground-based profilers are known to feature large flow-dependent components to their observation error statistics (Turner and Löhnert 2014; Fielding and Janiskova 2018).

Recently, novel remote sensing instruments have become increasingly common for assimilation in operational systems across the globe (e.g., Geer et al. 2018). While such instruments can be highly beneficial in improving numerical weather

Corresponding author: Samuel K. Degelia, sdegelia@ou.edu

DOI: 10.1175/MWR-D-21-0049.1

forecasts (Hu et al. 2019; Degelia et al. 2019; Chipilski et al. 2020), their observation error statistics are known to differ significantly from conventional datasets. For example, Degelia et al. (2020, hereafter DWS20) find that assimilating ground-based remote sensing data collected by the atmospheric emitted radiance interferometer (AERI; Feltz et al. 2003) can improve forecasts of nocturnal convection initiation (CI) through better analyses of midlevel moisture advection. However, DWS20 also show that using a static error inflation method across the entire network of AERIs can lead to some profiling sites featuring underestimated error variances and other sites featuring overestimated error variances. Assimilating data from the sites with underestimated error variances are then shown to weaken midlevel moisture advection and degrade subsequent forecasts of nocturnal convection initiation (CI). Turner and Löhnert (2014) also show that error variances for AERIs nearly double in cloudy scenarios. Despite the strong variation in these error statistics, most operational DA systems assume a constant (static) **R** that does not vary in time or location (Bormann et al. 2016). As such, it has become clear that novel methods are needed to estimate **R** for remote sensing datasets (Minamide and Zhang 2017; Fielding and Janiskova 2018; Fielding and Stiller 2019).

Ideally, methods for estimating **R** should account for both instrument and representation errors assuming that systematic errors are removed prior to DA. Most commonly, this is achieved by inflating known instrument errors in scenarios where representation errors are large. For example, Geer and Bauer (2011) adaptively inflate observation error variances for cloud-top observations based on an empirical function of brightness temperature. Minamide and Zhang (2017) apply a similar method for convective-scale applications, but only apply the error inflation when the difference between the observed and simulated brightness temperature (i.e., the innovation) is large. This method is shown to reduce analysis errors by limiting the use of large observation increments near clouds where representation errors are large. However, these methods are unique to satellite observations and have not been evaluated for other remote sensing datasets such as ground-based profilers.

In addition to adaptively inflating observation error variances to account for representation errors, direct estimation of **R** is possible using a method originally derived in Desroziers et al. (2005, hereafter D05). The D05 method uses observation-space diagnostics, including the innovation and analysis residuals, to compute an estimate of **R** that includes contributions from instrument and representation errors (Hodyss and Nichols 2015; Hodyss and Satterfield 2017). The D05 method has an advantage of being applicable to any observation type and has recently been shown to be beneficial even at the convective scale (e.g., Lange and Janjić 2016; Waller et al. 2016b,c; Cordoba et al. 2017; Waller et al. 2019). However, the D05 formulation features an expectation operator that requires a large sample of observations to produce an accurate estimate of **R**. As such, previous applications typically collect many observations to compute statistically robust but temporally constant observation errors (e.g., Weston et al. 2014; Bormann et al. 2016; Campbell et al. 2017; Waller et al. 2019). While this compromise removes the ability to diagnose time-dependent observation errors, the D05 method can still be individually calculated at each observation site to estimate location-dependent observation error statistics. Like the inflation methods described previously, the D05 method has also yet to be evaluated for ground-based remote sensing profilers.

To determine how to optimally assign observation errors for ground-based remote sensing profilers during convective-scale DA, we propose and evaluate three methods for estimating the observation error variance (diagonal of **R**). These methods are applied for the AERI given that this instrument is known to feature variable performance statistics depending on the meteorological condition (Turner and Löhnert 2014; Turner and Blumberg 2019). Two methods are based on inflating instrument errors to account for representation errors following Degelia et al. (2019). The final method applies the D05 diagnostic to directly estimate the full observation error variance **R**. Finally, we also compare each method when assimilating AERI observations for a nocturnal MCS observed during the Plains Elevated Convection at Night (PECAN; Geerts et al. 2017) field campaign. Given that this study is among the first work to evaluate the impact of estimating observation error variances for ground-based remote sensing profilers, a case study allows us to perform detailed analyses that trace modifications to the error statistics through the DA and forecast periods and better understand the impacts to the individual convective ingredients.

The outline of this paper is as follows: section 2 presents a general overview of the methods used to estimate the observation error variances. System configuration, description of our specific implementation of each method, and details of the 15 July nocturnal MCS are discussed in section 3. Attributes of the observation error statistics diagnosed by each method are presented in section 4. The convective-scale impacts of assimilating observations using these error profiles are detailed in section 5, and the impacts on the DA cycling are discussed in section 6. Finally, a discussion of results is found in section 7.

2. Overview of methods for estimating observation error statistics

This study compares the impact of assimilating AERI data using novel methods to estimate observation error statistics on convective-scale forecasts. These remote sensing instruments retrieve simultaneous profiles of temperature and water vapor mixing ratio up to 3 km above ground level (AGL) every 5–15 min. Given that error statistics for AERI retrievals vary in different atmospheric parameters including clouds and precipitable water vapor (Turner and Löhnert 2014; Turner and Blumberg 2019), observation error statistics for these data likely cannot be represented by a constant error variance or static inflation methods. The following is a general overview of the methods evaluated in this study. A description of our specific implementation for each method is then included in section 3.

a. Inflation of retrieval error variances from AERIoe

Degelia et al. (2019, hereafter DWS19) introduce a method for assimilating AERI observations by statically inflating retrieval error variances generated by the AERI optimal estimation retrieval algorithm (AERIoe; Turner and Löhnert 2014).

TABLE 1. List of experiments, a description of the method used to estimate their observation error variances, and values for the tunable parameters in each equation.

Experiment	Description of method	Tunable parameter values
NOAERI	— (no AERI observations assimilated)	—
AERIOE	Initial observation error variances obtained from AERIoe retrieval and statically inflated following Eq. (1)	α_T linearly increases from 0 at the surface to 4 at 3 km AGL; α_Q linearly increases from 0 at the surface to 1 at 3 km AGL
AERIOE_VRES	Initial observation error variances obtained from AERIoe retrieval and adaptively inflated using retrieval effective resolution in Eq. (2)	$\beta_T = 1.3\ \text{km}^{-1}$, $\beta_Q = 0.2\ \text{km}^{-1}$
DESROZIERS	Observation error variances computed a priori using observation-space diagnostics in Eq. (8). Unique error variances computed for each site. Residuals include all DA cycles simulated in DWS20.	$C = 1.25$

AERIoe outputs unique retrieval error statistics for each observing time that contain both instrument and forward operator errors but that need to be further inflated to account for residual errors. DWS19 account for these residual errors using the difference between the full error variance profiles used for assimilating rawinsondes in the Gridpoint Statistical Interpolation software (GSI; σ^2_{Sf}) and rawinsonde instrument error variances (σ^2_{Si}; provided by Vaisala 2017), following:

$$\sigma^2_{\text{Pf}} = \sigma^2_{\text{Pi}} + \alpha(\sigma^2_{\text{Sf}} - \sigma^2_{\text{Si}}), \tag{1}$$

where α is a unitless, tunable parameter that controls the amount of inflation, σ^2_{Pi} is the retrieval error variance profile and σ^2_{Pf} is the final observation error variance profile assigned during DA. Given that the error correlations for the AERIoe retrievals increase with height and feature different structures for temperature and moisture (Turner and Löhnert 2014), DWS19 select separate, vertically increasing values of α (Table 1).

Though the initial observation error variances σ^2_{Pi} are unique for each observing time, the DWS19 additive inflation [second term on RHS of Eq. (1)] tends to be larger than σ^2_{Pi} such that Eq. (1) produces primarily static error statistics with little variance between observing times. By comparing profiles of σ_{Pi} to RMS differences from collocated rawinsondes (a proxy of the true observation error statistics), we find that this static inflation method produces a reasonable shape and magnitude for the mean uncertainties but greatly underestimates their variability (Fig. 1). Upper-level moisture observations often differ from rawinsonde observations by 0.5–4 g kg^{-1}, while the observation error variances diagnosed using Eq. (1) only range from 1.8 to 2.2 g kg^{-1} (Fig. 1b). As such, the DWS19 inflation method can sometimes cause upper-level observations to be underweighted (overestimated error variances) or overweighted (underestimated error variances), leading to suboptimal observation increments. DWS20 further illustrate that the underestimated observation error variances for AERI retrievals can degrade convective forecasts when these observations are assimilated.

To better estimate the true observation error statistics and prevent underestimated error variances, we modify the AERIoe inflation technique using the effective vertical resolution profile for the retrieval (γ). AERIoe outputs a unique profile of γ for each observing time and site. Contrary to traditional uses of the term "resolution," γ represents the vertical smoothing applied to the retrieval such that larger values indicate smoother profiles. We refer to Turner and Löhnert (2014) for the derivation of within AERIoe. Following this modification, the error inflation becomes

$$\sigma^2_{\text{Pf}} = \sigma^2_{\text{Pi}} + \beta\gamma(\sigma^2_{\text{Sf}} - \sigma^2_{\text{Si}}), \tag{2}$$

where β represents a similar tunable parameter to α but now has units of per kilometer (km^{-1}). The AERIoe vertical

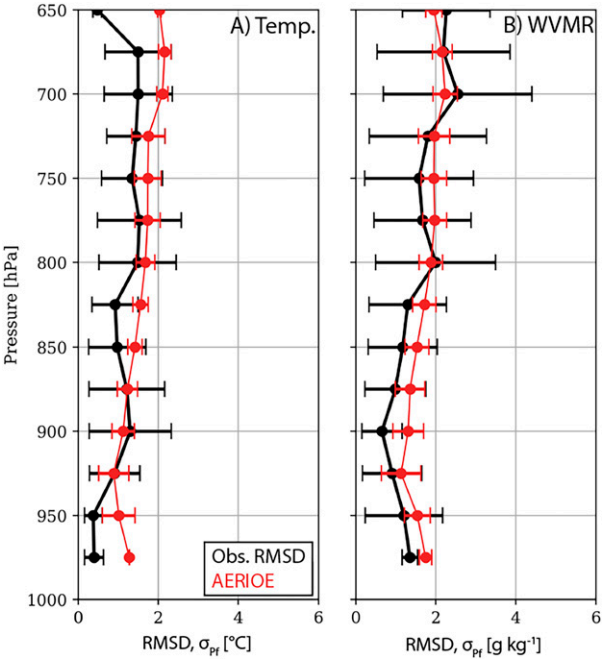

FIG. 1. RMS differences between AERI retrievals and collocated rawinsonde launches for all retrievals assimilated in DWS20 (black). The error bars in black represent one standard deviation of the absolute differences between the retrievals and rawinsondes. The RMS error for rawinsondes has also been subtracted from the black curve. Profiles are only included in the computation if a rawinsonde and AERI retrieval are available at the same site within a ±15-min window. Also shown in red are the mean and standard deviation of σ_{Pf} [Eq. (1)] assigned for the same retrievals using the AERIoe method described in section 2. The mean differences are computed by interpolating each AERI retrieval and rawinsonde onto a standard vertical grid with 25-hPa spacing.

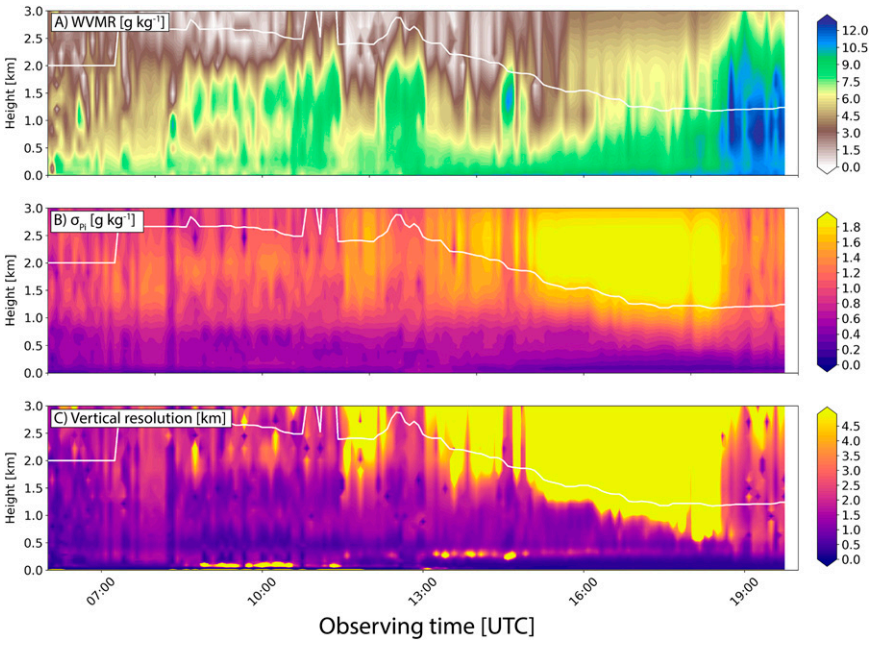

FIG. 2. Example of AERIoe moisture retrievals collected at FP4 between 0600 and 2000 UTC 23 Jun. Included are the (a) moisture observations (g kg^{-1}), (b) observation error standard deviations produced by AERIoe [σ_{Pi} in Eqs. (1) and (2); g kg^{-1}], and (c) retrieval effective resolution profiles (km). Also overlaid on each panel is the cloud base height (km) indicated by a collocated lidar or ceilometer (white).

smoothing typically increases with height, resulting in a similar effect to the vertically increasing inflation in Eq. (1). As such, we do not require β to increase with height as with α in Eq. (1).

Whereas the static method in Eq. (1) equally inflates the observation error variances, Eq. (2) *adaptively* inflates the errors given that γ varies for each site and observing time. Specifically, the modified inflation method estimates larger observation errors when the AERIoe retrieval smooths through certain features. For example, this method often inflates the error variances more in the shallow layer below cloud base, where Turner and Blumberg (2019) note that the AERI effective resolution becomes large due to problems separating cloud emission from atmospheric emission. An example of this issue is shown for a PECAN case in Fig. 2 where the AERI vertical resolution rapidly grows through a layer beginning ~500 m below cloud base after 1100 UTC. While σ_{Pi} increases by a factor of 5 relative to the surface (Fig. 2b), γ instead increases by a factor of 20 throughout the same layer (Fig. 2c). This indicates considerable smoothing in the observations that is not fully accounted for by the retrieval error statistics. Thus, applying the modified error inflation in Eq. (2) can adaptively inflate σ^2_{Pi} to account for such problems.

b. *Desroziers et al. (2005) diagnostic*

In addition to inflating the error statistics retrieved by AERIoe, we also evaluate the impact of directly estimating observation error statistics using observation-space diagnostics (Tandeo et al. 2020). The most common method for this follows from D05 who derive an estimate of $\mathbf{R}$ as a relationship between the innovation $[\mathbf{d}^o_b = \mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)]$ and analysis residual $[\mathbf{d}^o_a = \mathbf{y}^o - \mathbf{H}(\mathbf{x}^a)]$. Here, $\mathbf{y}^o$ is the observation vector, and $\mathbf{H}(\mathbf{x}^b)$ and $\mathbf{H}(\mathbf{x}^a)$ are the background and analysis vectors in observation space. Assuming that observation and background errors are uncorrelated and that the innovation and analysis residuals are unbiased, D05 show that

$$\tilde{\mathbf{R}} = E[\mathbf{d}^o_a(\mathbf{d}^o_a)^{\mathrm{T}}], \tag{3}$$

where $\tilde{\mathbf{R}}$ approximates the true $\mathbf{R}$. Because $\tilde{\mathbf{R}}$ contains information about the model departure from the observation, it should ideally contain most components of the observation error statistics including representation errors (Hodyss and Nichols 2015; Hodyss and Satterfield 2017). The D05 diagnostic is often referred to as a posterior "consistency check" on the assigned observation uncertainties given its requirement for an existing analysis vector. Many studies highlight limitations of the D05 diagnostic (e.g., Ménard et al. 2009; Ménard 2016; Waller et al. 2016a, 2017) including that a large sample of observations are required to apply the expectation operator in Eq. (3). Additionally, various studies note that $\tilde{\mathbf{R}}$ is only correct if the assigned background and observation error covariances used during DA are correct. While we have no way to verify $\tilde{\mathbf{R}}$ without a method to estimate representation errors, we note that Ménard (2016) find that the diagnostic can still provide useful information even if the assumed statistics are not exact.

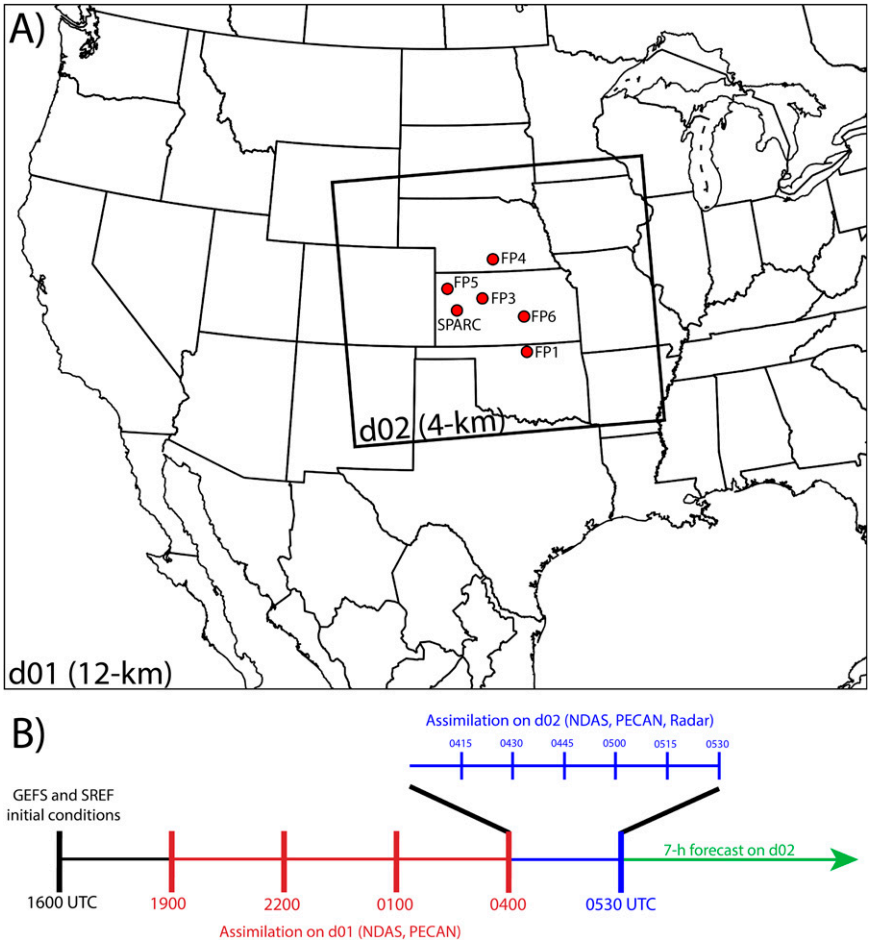

FIG. 3. Overview of experimental design including (a) domain configuration for the outer (d01, 12 km) and inner (d02, 4 km) assimilation domains. The inner assimilation domain is also used as the forecast domain. The locations of each AERI platform assimilated here are also overlaid. (b) Flowchart for the cycled DA, including the four 3-h assimilation cycles on d01, the six 15-min assimilation cycles on d02, and the 7-h forecast period.

3. Experiment design

a. *Description of cycled DA and forecast system*

All experiments described herein use the same multiscale DA and forecast system described in DWS20 (Fig. 3). The cycled system consists of two components: a GSI-based EnKF (Johnson et al. 2015; Wang and Wang 2017) and version 3.7.1 of WRF-ARW (Skamarock et al. 2008). During DA, the GSI-based EnKF uses observations ($\mathbf{y}^o$) to update a prior estimate of the model state ($\mathbf{x}^b$) into an analysis ($\mathbf{x}^a$). For this study, we apply the ensemble square root filter (EnSRF; Whitaker and Hamill 2002) version of the EnKF wherein observations are serially assimilated to update the mean state following:

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^b} + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\overline{\mathbf{x}^b}), \tag{4}$$

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}. \tag{5}$$

Here, $\mathbf{K}$ represents the Kalman gain, $\mathbf{H}$ is the linearized observation operator performed by GSI, and $\mathbf{P}^b$ is the background error covariance. Unlike static DA methods such as 3DVar, the EnKF samples a flow-dependent $\mathbf{P}^b$ from ensemble forecasts. To update the ensemble member perturbations (given by prime symbols), EnSRF computes a modified Kalman gain ($\tilde{\mathbf{K}}$) following:

$$\mathbf{x}'^a = \mathbf{x}'^b - \tilde{\mathbf{K}}\mathbf{H}\mathbf{x}'^b, \tag{6}$$

$$\tilde{\mathbf{K}} = \left[1 + \sqrt{\left(\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}}\right)}\right]^{-1}\mathbf{K}. \tag{7}$$

We note that the EnSRF is a serial filter in that each observation is assimilated sequentially. As such, $\mathbf{R}$ and $\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}}$ reduce to scalars in Eq. (7). This prevents the EnSRF from accounting for observation error correlations. As such, the error methods described in the next section are designed to only compute observation error variances instead of the full $\mathbf{R}$ matrix. Additional DA configuration settings, along with the WRF parameterization schemes, are shown in Table 2 following DWS20. We note

TABLE 2. List of WRF-ARW and EnKF settings used for all simulations. We note that a different microphysical parameterization scheme is used for the DA and forecast periods and that no cumulus parameterization is utilized on the 4-km assimilation domain (d02 in Fig. 3). The two values shown for the localization radii represent the parameters used for the outer (12 km, d01) and intermediate (4 km, d02) assimilation domains, respectively.

Parameterization or DA setting	Scheme name or value	
Microphysical parameterization (DA)	WSM6 (Hong and Lim 2006)	
Microphysical parameterization (forecast)	Lin et al. (1983)	
PBL parameterization	MYNN (Nakanishi and Niino 2006)	
Longwave radiation parameterization	RRTMG (Iacono et al. 2008)	
Shortwave radiation parameterization	Goddard (Tao et al. 2003)	
LSM	Noah (Ek et al. 2003)	
Cumulus parameterization	Grell and Freitas (2013)	
Localization radii (conventional observations, PECAN rawinsondes, AERIs)	Horizontal (km): 700, 200	Vertical [$\ln(P/P_{\text{ref}})$]: 1.1, 0.55
Localization radii (Doppler lidars, radar wind profilers)	Horizontal (km): 700, 200	Vertical [$\ln(P/P_{\text{ref}})$]: 0.20, 0.20
Localization radii (radar observations)	Horizontal (km): —, 20	Vertical [$\ln(P/P_{\text{ref}})$]: —, 0.55
Inflation factor (prior)	1.15, decreasing to 1.03 at model top	
Inflation factor (posterior)	0.95 of prior spread	

that while covariance localization can limit the quality of $\tilde{\mathbf{R}}$ estimated by Eq. (3), Waller et al. (2017) show that the observation error variances, which are the only terms modified throughout this study, will not be affected by this problem.

We now summarize the cycling configuration for each set of simulations. The cycled DA begins by generating initial and lateral boundary conditions for 40 ensemble members at 1600 UTC 14 July consisting of 20 members from the GEFS (Wei et al. 2008) and 20 members from the Short-Range Ensemble Forecast System (SREF; Du et al. 2014). Next, conventional observations and AERI retrievals are assimilated on a 12-km outer domain (d01 in Fig. 3) at 3-h intervals until 0400 UTC 15 July. The North American Mesoscale Data Assimilation System (NDAS) supplies the conventional observations containing quality controlled rawinsonde, surface, ship, and buoy data. We also preprocess AERI retrievals using the methods described at length in DWS19 and DWS20 but do not apply a bias correction scheme (see discussion section). After DA on d01, we then downscale to an inner, 4-km domain (d02 in Fig. 3) during which conventional observations, AERI retrievals, and radar data are assimilated at 15-min intervals until 0530 UTC 15 July. The radar data, including radar reflectivity and radial velocity collected from WSR-88D sites, are preprocessed using WDSS-II (Lakshmanan et al. 2007). Finally, we initialize 7-h forecasts from the first 10 ensemble members following DWS20. Unlike DWS20 who analyze small-scale nocturnal CI events on a 1-km grid, this study focuses on a larger-scale MCS that spans from southern Nebraska to northern Oklahoma. As such, we only perform forecasts on the 4-km domain to limit computational costs.

b. *Specific implementation of methods for computing observation error variances*

1) INFLATION OF RETRIEVAL ERROR VARIANCES FROM AERIOE

Equations (1) and (2) are directly computed for each DA cycle using the retrieval error variances and vertical resolution data produced by the AERIoe retrieval algorithm. Both the static and adaptive inflation methods feature a tunable parameter (α, β) to control the magnitude of the error inflation (summarized in Table 1). To be consistent with previous studies, we use the values of α derived in DWS19. These values are selected by varying α in intervals of 0.25 for a nocturnal CI event on 25 June 2015. DWS19 then select the α, which results in the highest fractions skill score (FSS; Roberts and Lean 2008) for a 6-h forecast of composite reflectivity. For the parameter β used in the adaptive inflation method [Eq. (2)], we select values that result in the lowest mean absolute difference between observation error variances estimated by the static and adaptive methods for the 25 June case. As such, this modification increases the profile-to-profile variation of the uncertainties in the adaptive method while only slightly changing the mean values (primarily at the top of the profile).

2) DESROZIERS ET AL. (2005) DIAGNOSTIC

We also compute observation error variances for AERI retrievals using the Desroziers et al. (2005) diagnostic in Eq. (3). For our implementation of this method, we first separate the residuals $\mathbf{d}_a^o$ and $\mathbf{d}_b^o$ to compute a unique solution for each AERI observing site. Next, we bin each residual pair into a 25-hPa vertical profile. Since the EnKF used here cannot account for the full observation error covariance, we only compute the observation error variances. We apply the following from D05 that corresponds to Eq. (3) but only for the mean diagonal elements of $\tilde{\mathbf{R}}$:

$$\sigma_{\text{Pf}}^2 = C\sum_{i=1}^{n}\frac{(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^a)(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^b)}{n}. \tag{8}$$

Here, n represents the number of residual pairs per bin, and $(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^b)$ and $(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^a)$ represent background and analysis residuals for individual observations in the bin, respectively. The above equation is calculated in each bin to build a vertical profile of σ_{Pf}^2. Although ignoring the error correlations is necessary for our DA system, previous studies show this assumption can result in underestimated error statistics and

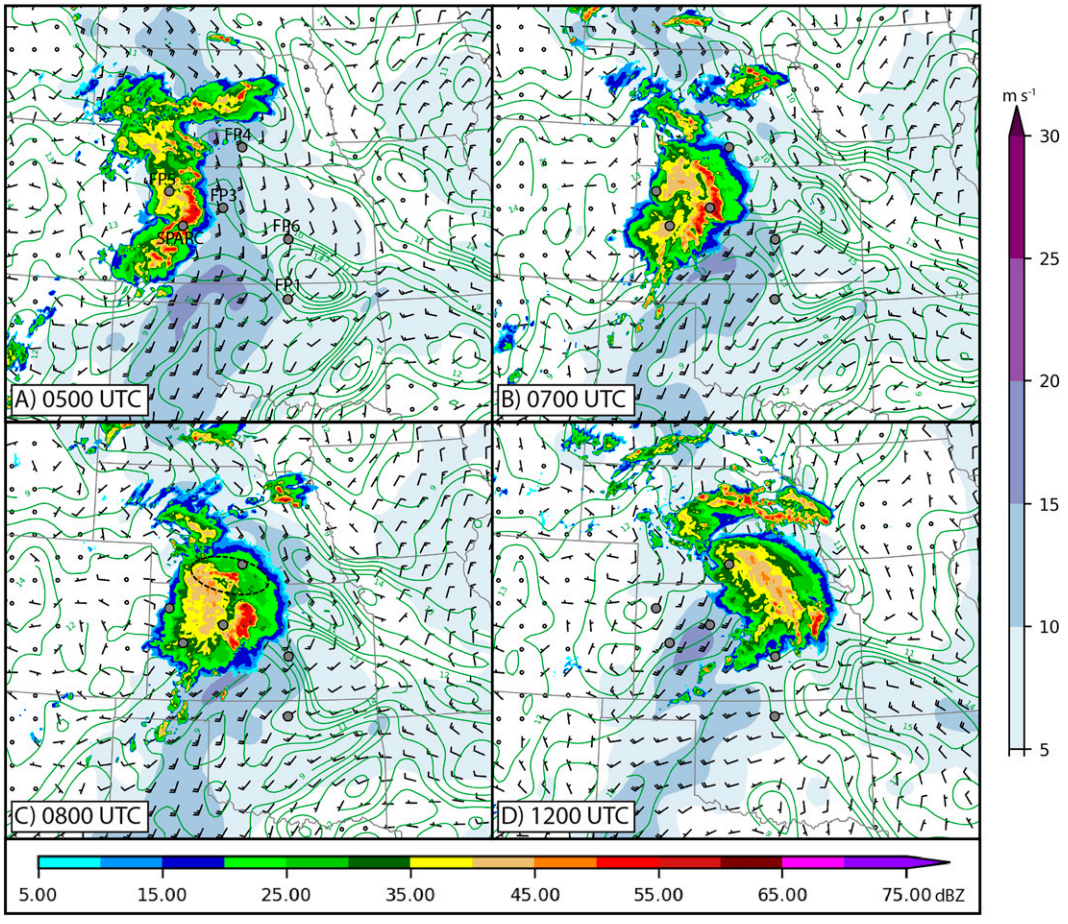

FIG. 4. RAP analyses for the 15 Jul 2015 nocturnal MCS event including composite reflectivity (dBZ; bottom color scale), 850-hPa winds (m s^{-1}; right color scale), and 850-hPa water vapor mixing ratio (g kg^{-1}; green contours). The location of the AERI platforms assimilated here are overlaid in each panel and labeled in (a).

degrade analyses (e.g., Bormann et al. 2016). These issues can be at least partially alleviated by applying a multiplicative inflation to the error variances (Stewart et al. 2014; Bormann et al. 2016), indicated by C in the Eq. (8). To be consistent with the AERIoe inflation methods and select the optimal observation error variance magnitudes for the D05 method, we also tune C using the same, 25 June 2015 case used to determine α. By varying C in intervals of 0.25, we find that $C = 1.25$ produces the highest values of FSS for our implementation of the D05 method.

As described previously, Eqs. (3) and (8) require many observation samples over which to apply the expectation operator. For the method applied here, these samples are obtained from the 13 cases presented in DWS20. We use the ensemble mean background and analysis residuals from each cycle to be consistent with the original derivation in D05. While multiple AERI retrievals were often assimilated during a single DA cycle in DWS20, we only select the retrieval closest to the center of the observation window so that only the most representative residuals are included. For this implementation, we note that we also compute unique error variances for both the outer and inner domains (Fig. 3) given that representation errors are highly related to model grid spacing (Fielding and Stiller 2019). The total number of samples differs per site and vertical level but averages to $n = 35$ retrievals on the outer domain and $n = 46$ retrievals on the inner domain.

c. Description of experiments

To evaluate the convective-scale impact of these novel methods, we design a set of experiments that differ only in their procedures used to assign the observation error statistics when assimilating AERI retrievals (Table 1). First, we perform a reference experiment where no AERI data are assimilated (NOAERI) to ensure that the above methods do not overinflate the observation error variances and result in little impact on the analysis. Next, we perform two experiments that inflate the error variances provided by AERIoe, including using the static inflation from Eq. (1) (AERIOE), and the modification in Eq. (2) wherein the error variances are adaptively inflated using the vertical resolution profile (AERIOE_VRES). Finally, we also apply a method that directly estimates the observation error variance using the Desroziers diagnostic (DESROZIERS). We hypothesize that these two novel methods (AERIOE_VRES and DESROZIERS) are likely

more impactful in certain situations. For example, we hypothesize that AERIOE_VRES can improve analysis results by inflating error statistics below cloud base where the retrieval often becomes very smooth (large representation errors). DESROZIERS could instead limit observation impacts in situations where the retrieval smooths through an inversion or shallow moist layer, resulting in large representation errors and a large innovation.

d. Case description

We evaluate each experiment for the nocturnal MCS observed during PECAN on 15 July 2015 (IOP30). Grasmick et al. (2018) provide a detailed overview of this event which we summarize here. On the late afternoon of 14 July, a cluster of disorganized convective cells developed in eastern Colorado along a group of shortwave troughs embedded within an upper-level ridge. Additional convection initiated along outflow boundaries produced by these storms shortly after sunset, eventually growing upscale into the MCS of interest (Fig. 4a). By 0700 UTC, this MCS began to develop the traditional leading-line, trailing stratiform structure with the strongest precipitation located along the northern half of the MCS (Fig. 4b). An additional band of convection, oriented primarily east–west, developed just north of the main convective line near the Nebraska border at 0800 UTC (Fig. 4c). This secondary convective line will be discussed later throughout the paper. The main MCS continued to propagate eastward until decaying around 1200 UTC (Fig. 4d). Though not considered a severe MCS (Grasmick et al. 2018), the NWS reported three heavy wind events (30 m s^{-1} or 60 kt) and one large hail event during its evolution.

Throughout this event, scientists collected a dense dataset of profiling and in situ observations from PECAN Integrated Sounding Arrays (PISAs). AERI instruments were located at five of the fixed PISAs (FP) and one mobile PISA (MP) site (gray dots in Fig. 4). Most AERIs collected data continuously throughout the experiment, though the MP site (SPARC) only operated between 0000 and 0500 UTC 15 July. DWS20 assimilate these AERI data in their systematic evaluation of the impact of remote sensing profilers for nocturnal convective forecasts. Though their baseline forecasts perform well, they find a reduction in skill when assimilating AERI data, especially between 0700 and 0900 UTC when the MCS was most organized and when the second, east–west convective band developed to the north (Fig. 4c). DWS20 hypothesize that the forecast degradations found in their study could partially be a result of underestimated observation error variances for some AERI platforms. Given these findings, we select the 15 July MCS to evaluate the impact of the novel methods proposed here.

4. Results: Diagnosed observation error statistics

To examine the structure of the observation error statistics diagnosed by each method, we first present time-averaged profiles of observation error standard deviation computed over all DA cycles (Fig. 5). We note that the plots in Figs. 5b and 5d represent the standard deviation across the distribution of σ_{Pf} and should not be confused with the error standard deviation itself.

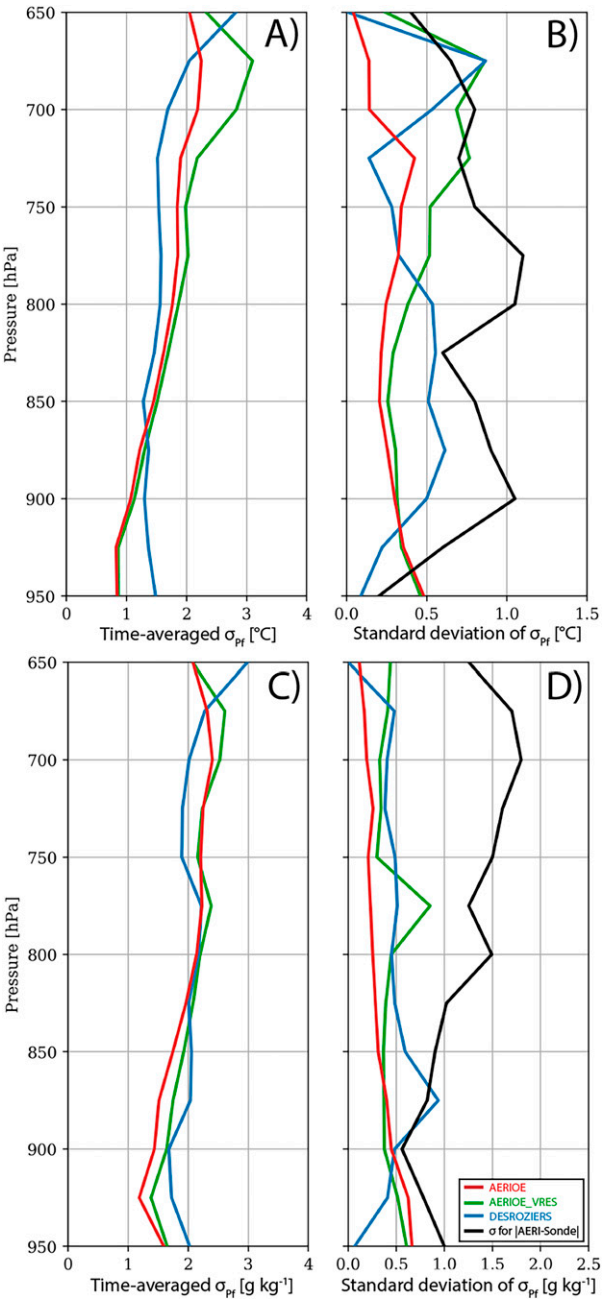

FIG. 5. (a),(c) Time-averaged and (b),(d) standard deviation profiles of the σ_{Pf} for (top) temperature (°C) and (bottom) water vapor mixing ratio (g kg^{-1}). Each profile is computed using observation error profiles from all sites. When computing the statistics, each profile is interpolated onto a standard vertical grid with 25-hPa spacing. Also overlaid in (b) and (d) are the one standard deviation profiles of the absolute differences between the retrievals and raw-insondes assimilated in DWS20 (black error bars in Fig. 1).

In general, each method estimates similar σ_{Pf} magnitudes for midlevel (~900–750 hPa) temperature and moisture

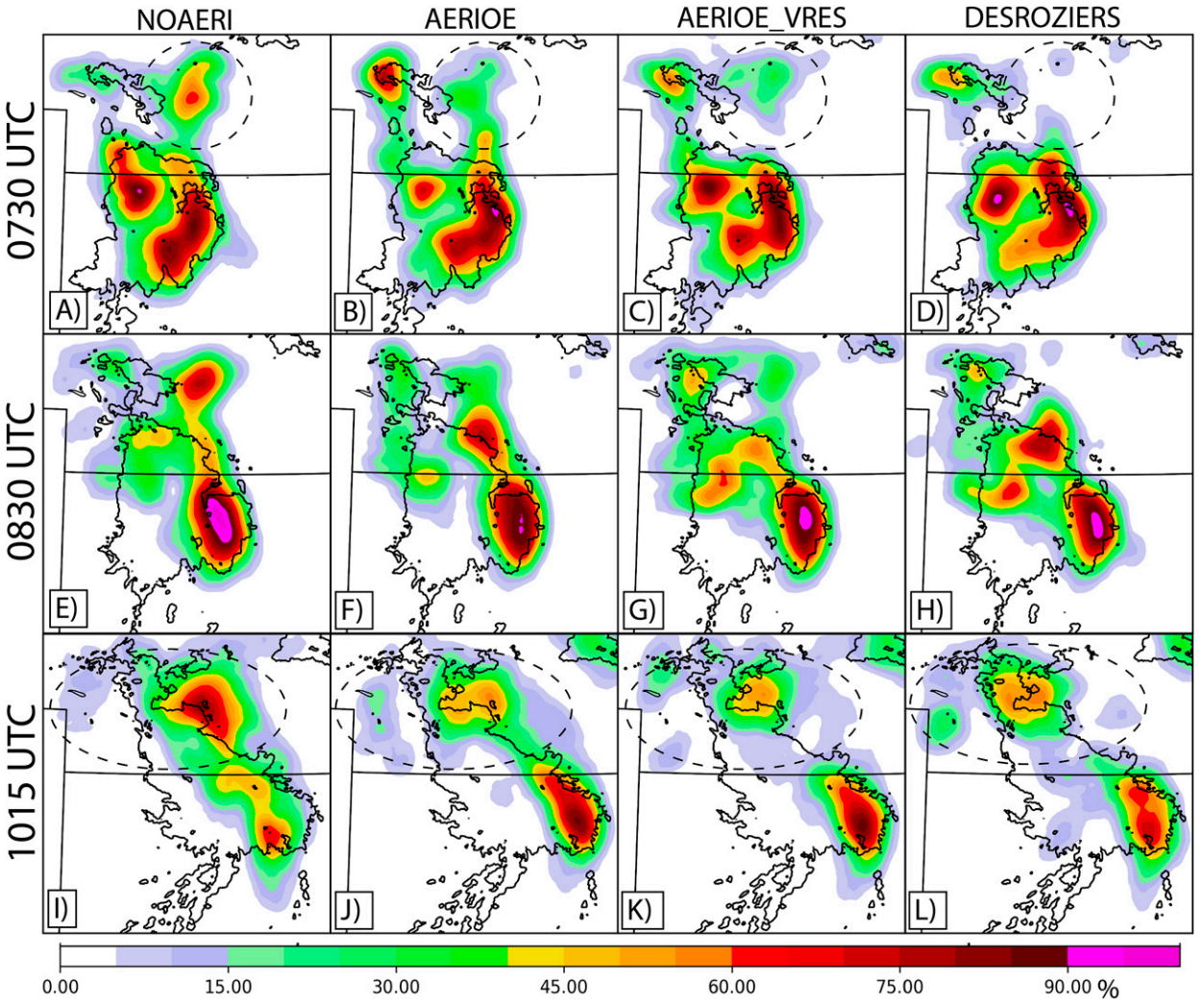

FIG. 6. Neighborhood ensemble probabilities (NEP) for each experiment and valid at (a)–(d) 0730 UTC, (e)–(h) 0830 UTC, and (i)–(l) 1015 UTC 15 Jul. The probabilities are computed for compositive reflectivity exceeding 30 dBZ. Also overlaid in black are the 30-dBZ contours of observed composite reflectivity from MRMS. See text for a description of the circled regions.

observations with larger differences occurring at the bottom and top of the retrievals. These differences primarily occur between the two main experiment sets (i.e., AERIOE and AERIOE_VRES versus DESROZIERS). When averaged across all sites, the methods which inflate the AERI error statistics produce vertically increasing σ_{Pf} profiles up to 700 hPa for both temperature (Fig. 5a) and moisture (Fig. 5c). Conversely, the DESROZIERS method shows a local maximum in σ_{Pf} near the surface that decreases through the planetary boundary layer (PBL) before increasing again around 700 hPa. This result indicates that the innovation and analysis residuals are typically larger near the surface compared to aloft, potentially due to local circulations sampled by AERIs that cannot be resolved in the 12- and 4-km simulations (i.e., larger representation errors).

The adaptive inflation extension in AERIOE_VRES follows a generally similar shape to its static counterpart. AERIOE and AERIOE_VRES both estimate similar σ_{Pf} profiles below 750 hPa due to β in Eq. (2) being tuned to produce the lowest RMSD between the two experiments. However, σ_{Pf} for AERIOE_VRES becomes larger than AERIOE in the upper

TABLE 3. Fractions skill score (FSS) values computed for accumulated precipitation exceeding 2.54 and 6.35 mm h^{-1}. Each score is computed across the entire forecast domain (d02 in Fig. 3) using a 16-km neighborhood and then averaged over the 7-h forecast period. The verifying precipitation observations are obtained from MRMS.

Experiment	FSS (2.54 mm h^{-1})	FSS (6.35 mm h^{-1})
NOAERI	0.485	0.536
AERIOE	0.472	0.596
AERIOE_VRES	0.496	0.613
DESROZIERS	0.496	0.605

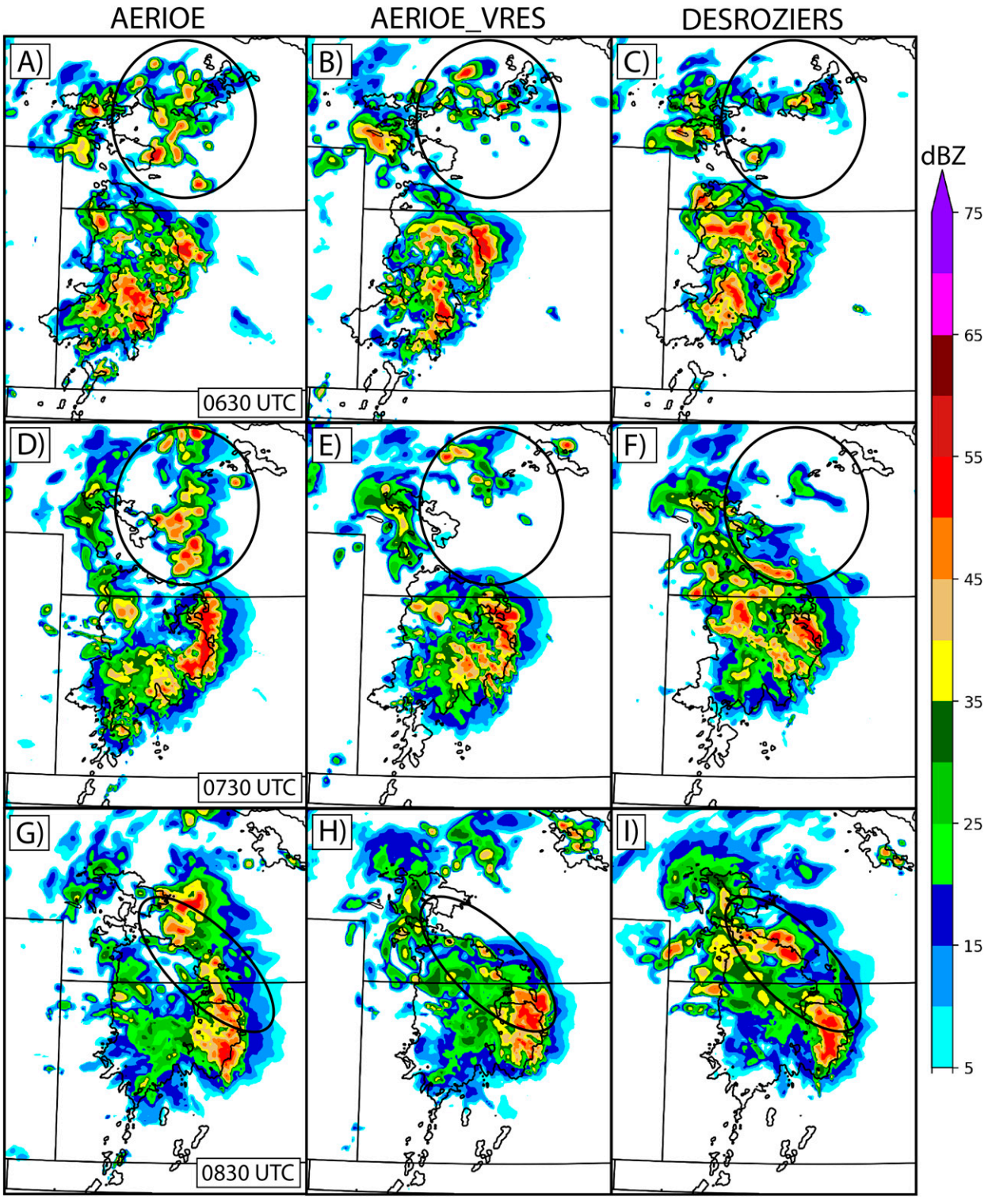

FIG. 7. Simulated composite reflectivity from member 1 of the AERIOE, AERIOE_VRES, and DESROZIERS experiments valid at (a)–(c) 0630, (d)–(f) 0730, and (g)–(i) 0830 UTC. Also overlaid in black are the 30-dBZ contours of observed composite reflectivity from MRMS. See text for a description of the circled regions.

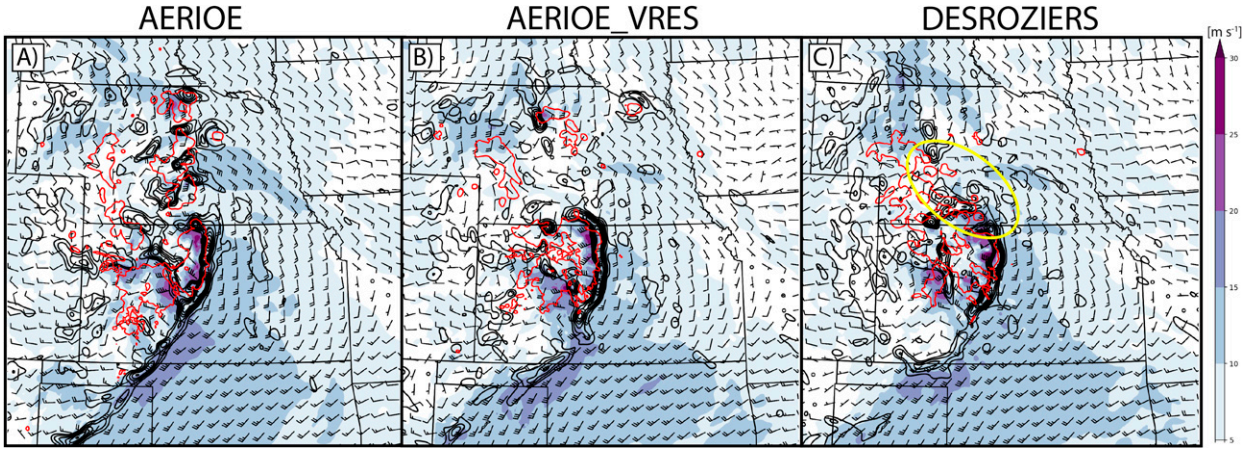

FIG. 8. Simulated 850-hPa winds (m s^{-1}; shading and barbs), horizontal convergence (contoured in black every $+5^{-6}$ s^{-1}), and composite reflectivity (30-dBZ contours in red) from member 1 of the AERIOE, AERIOE_VRES, and DESROZIERS experiments. Each plot is valid at 0730 UTC 15 Jul. The half barbs represent wind speeds of 2.5 m s^{-1}, and the full barbs represent wind speeds of 5 m s^{-1}. The yellow circle in (c) indicates convergence band that results from northerly winds converging with outflow.

portions of the profile, especially for temperature (Fig. 5a). This increase partially results from γ in Eq. (2) often increasing exponentially with height compared to the linearly increasing inflation in Eq. (1). Additionally, the presence of clouds often contributes to increased γ, resulting in larger σ_{Pf} in AERIOE_VRES compared to AERIOE.

A larger impact of the two new methods evaluated here (AERIOE_VRES and DESROZIERS) results from the increased temporal variability of the observation error statistics at all sites (Figs. 5b,d). This increased variability occurs for both moisture and temperature errors and better matches the observed error statistics (Fig. 1 and black lines in Figs. 5b,d). AERIOE_VRES features more variability than AERIOE throughout most of the profile due to the addition of γ in Eq. (2) that changes for each retrieval. AERIOE_VRES also features much larger variability in σ_{Pf} near the top of the profile, likely due to the presence of clouds increasing the vertical resolution of some retrievals. Except near the surface, DESROZIERS also features larger variations in σ_{Pf} compared to AERIOE. The DESROZIERS method also estimates much more site-to-site variability due to its relationship to the background and analysis states (not shown). These findings suggest that advanced methods can better capture larger error variations wherein the observation uncertainty increases in specific meteorological conditions or at specific locations.

5. Impact of observation error statistics on the prediction of the 15 July 2015 MCS

Next, we compare the forecast impacts of assimilating AERI data when using different methods for assigning the observation error variances. We primarily analyze the evolution of the 15 July MCS through neighborhood ensemble probabilities (NEP; Schwartz and Sobash 2017) and FSS computed over a 16-km neighborhood. We note that the interpretation of these results does not change when verifying over different neighborhoods or forecast products (e.g., reflectivity or precipitation).

Without assimilating any AERI data, an MCS develops in the correct location with maximum probabilities located along the main convective line and in the trailing stratiform region in northwestern Kansas (Fig. 6a). Although NOAERI correctly predicts the early structure of the MCS, a large region of spurious precipitation develops to the north of the primary convective line in Nebraska (dashed circle in Fig. 6a). This spurious convection, which we discuss at length throughout the rest of this paper, eventually expands into a large area of precipitation and results in an apparent northern shift of the MCS relative to observations. Additionally, we note that NOAERI fails to maintain the MCS in Kansas such that ensemble probabilities along the primary convective line weaken to ~60% by 1030 UTC (forecast hour 4; Fig. 6i).

Assimilating the AERI data using the static inflation method (AERIOE; Figs. 6b,f,j) slightly degrades the forecast quality during forecast hours 0–3. Primarily, the observation error statistics in AERIOE reduce the convective probabilities within the stratiform region of the MCS by ~30% (Figs. 6a,b). AERIOE also shifts, but does not suppress, much of the spurious convection further south such that it connects with the main convective line of the MCS (Fig. 6b). This shift causes the MCS to orient increasingly north–south and again extend too far north in AERIOE. By 1015 UTC (Figs. 6i,j), AERIOE eventually begins to indicate a better forecast than NOAERI due to improvements in both the spurious precipitation (reduced probabilities by ~20%) and better maintenance of the convective line in the MCS (increased probabilities by ~20%). However, the early forecast degradations within the trailing stratiform region, along with little improvement in the spurious precipitation, contribute to AERIOE producing a lower mean FSS than NOAERI for the weaker precipitation threshold (Table 3).

While the static inflation method slightly degrades the forecast relative to the baseline simulation, its adaptive extension

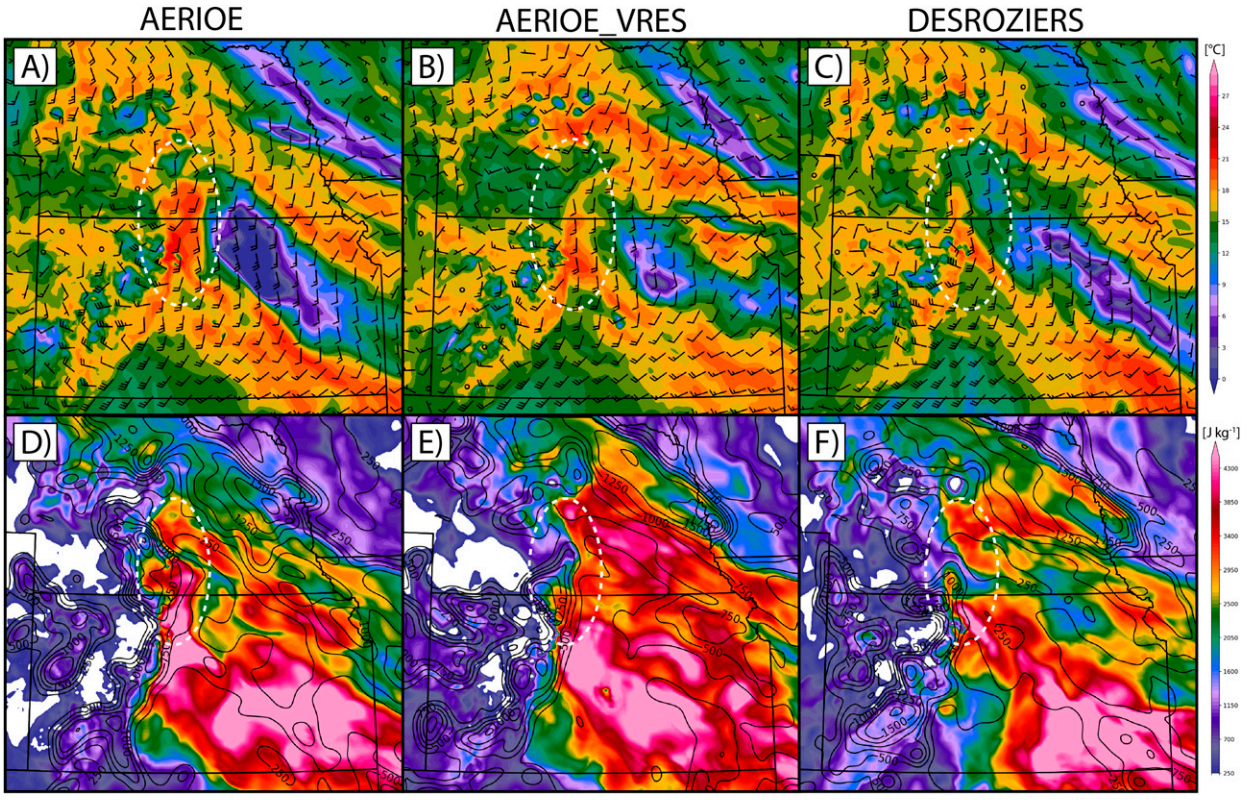

FIG. 9. Simulated pre-convective thermodynamic fields from member 1 of the AERIOE, AERIOE_VRES, and DESROZIERS experiments, including (a)–(c) 850-hPa dewpoint temperatures (°C; shading) and winds (m s^{-1}; barbs) and (d)–(f) most unstable CAPE (J kg^{-1}; shading) and lifted parcel levels for the most unstable parcels (m AGL; black contours). Each plot is valid at 0730 UTC 15 Jul. In (a)–(c), the half barbs represent wind speeds of 2.5 m s^{-1}, and the full barbs represent wind speeds of 5 m s^{-1}. The white circles represent regions corresponding to the spurious convection discussed in the text.

(AERIOE_VRES; Figs. 6c,g,k) improves upon many of the abovementioned problems. Primarily, we find that AERIOE_VRES results in reduced probabilities for the spurious precipitation across most of southern Nebraska (Figs. 6c,g). AERIOE_VRES also better maintains the trailing stratiform region of the MCS as indicated by increased probabilities in northwestern Kansas from 0730 to 0830 UTC (Figs. 6c,g). These results contribute to an overall more skillful forecast in AERIOE_VRES compared to NOAERI and AERIOE (Table 3) and suggest that some of the forecast degradations discussed in DWS20 can be alleviated by adaptively inflating the observation error statistics for ground-based profilers. However, the forecast in AERIOE_VRES eventually converges with AERIOE after 1030 UTC (not shown), indicating that the impact from modifying the observation error statistics only lasts ~4 h.

In addition to inflating the instrument errors, we also evaluate using the D05 method to directly estimate observation error variances based on observation-space statistics. The DESROZIERS experiment (Figs. 6d,h,l) similarly improves upon many of the issues discussed above. During the early forecast period, DESROZIERS simulates less precipitation in central Nebraska to the point where almost no spurious convection is seen at 0730 UTC (Fig. 6d). DESROZIERS also enhances probabilities within the trailing stratiform region to >90% compared to both AERIOE and AERIOE_VRES (Figs. 6d,h). However, DESROZIERS still develops low probabilities for unobserved convection along the Nebraska–Kansas border at later lead times that leads to a slight misorientation of the MCS at later lead times (Fig. 6h). This additional convection that develops at 0830 UTC will be discussed further in the next section.

Table 3 summarizes this NEP analysis and reveals that despite featuring larger mean error variances throughout much of the profile, the AERIOE_VRES and DESROZIERS experiments now show higher skill and thus more impact from assimilating the thermodynamic retrievals compared to AERIOE. This result is primarily related to improved suppression of spurious convection, with a small contribution stemming from better maintenance of the MCS at 0830 UTC. AERIOE_VRES and DESROZIERS produce similar FSS values, with AERIOE_VRES performing slightly better for higher precipitation rates.

Analysis of spurious precipitation

Given that modifications to the spurious precipitation in Nebraska are responsible for much of the skill differences between the experiments, we perform an additional

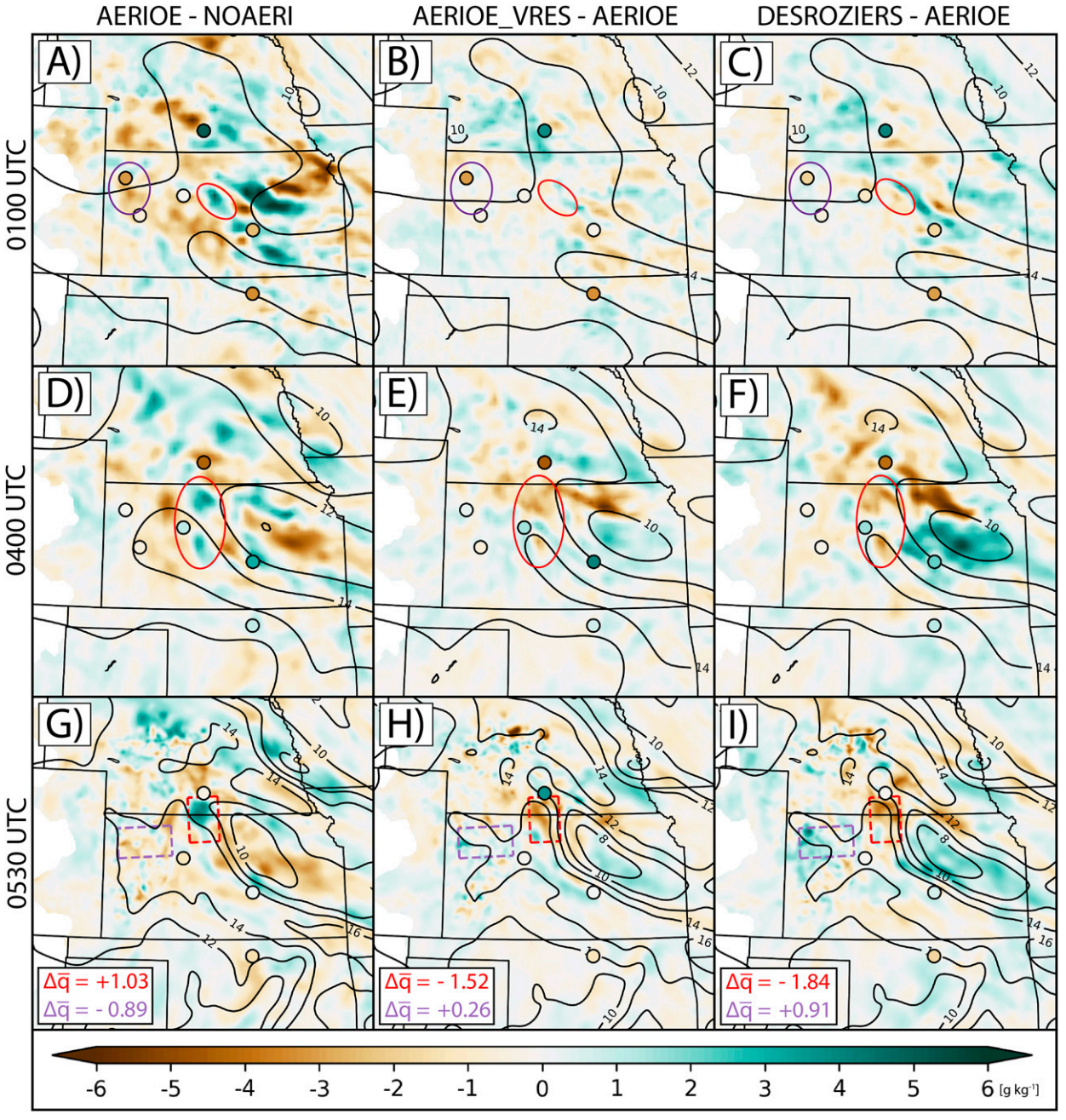

FIG. 10. Ensemble mean differences of 850-hPa water vapor mixing ratio (g kg^{-1}) for (a),(d),(g) AERIOE minus NOAERI; (b),(e),(h) AERIOE_VRES minus AERIOE; and (c),(f),(i) DESROZIERS minus AERIOE. The black contours indicate the ensemble mean 850-hPa water vapor mixing ratio simulated by AERIOE. Also overlaid are the innovations for AERI moisture retrievals closest to 850 hPa and assimilated at each cycle using the same color scale. The red regions correspond to positive moisture increments in AERIOE that enhance spurious convection in southern Nebraska, and the purple regions correspond to negative moisture increments in AERIOE that degrade the trailing stratiform region of the 15 Jul MCS. In (g)–(i), a mean difference value corresponding to the dashed rectangles are annotated.

analysis to determine what mechanisms lead to this convection. We compare member 1 from each experiment except for NOAERI, as this member is representative of the impacts described in the previous section. Primarily, member 1 from AERIOE generates a north–south-oriented line of unobserved convection (circles in Figs. 7a,d) whereas AERIOE_VRES and DESROZIERS produce a small amount of east–west-oriented convection that is observed at earlier lead times (circles in Figs. 7b,c,e,f). Additionally, member 1 from DESROZIERS generates a strong, northwest–southeast-oriented band of

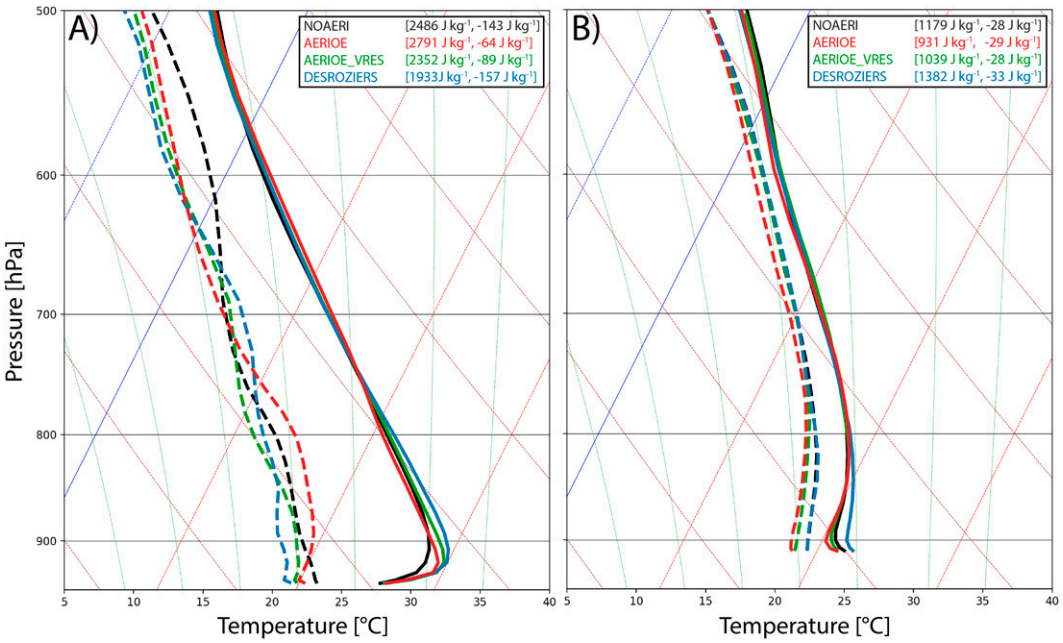

FIG. 11. Ensemble mean soundings computed over the (a) dashed red box and (b) dashed purple box in Figs. 10g–10i at 0530 UTC 15 Jul. The solid lines indicate temperature while the dashed lines indicate dewpoint temperature. Also annotated on each sounding are the most unstable CAPE (J kg^{-1}; first metric in brackets) and convective inhibition (CIN) associated with the most unstable parcel (J kg^{-1}; second metric in brackets).

convection along the northern edge of the MCS between 0730 and 0830 UTC (Figs. 7f,i). While this secondary convection simulated by DESROZIERS is displaced further north than observed, it still closely resembles the observed secondary CI event shown in Fig. 4c. AERIOE_VRES also generates some convection along the northern boundary of the MCS but of weaker magnitude (Fig. 7h). This secondary CI event in DESROZIERS occurs due to northwesterly winds converging with outflow associated with the MCS (yellow circle in Fig. 8c). Due to the large amount of spurious convection inhibiting the northwesterly winds in southwestern Nebraska (Fig. 8a), AERIOE does not simulate the convergence and associated convective band. As such, we hypothesize that the improved structure of the MCS in AERIOE_VRES and DESROZIERS also results from improved suppression of the spurious convection.

We find that the spurious convection develops within a north–south-oriented midlevel moist layer that extends into southern Nebraska in AERIOE (Fig. 9a). This moisture bulge results in enhanced midlevel instability, as denoted by air parcels at 1 km AGL featuring ~3000 J kg^{-1} of convective available potential energy (CAPE; Fig. 9c). In addition to the enhanced instability, this region also features low-level convergence due to the strong westerly winds associated with the MCS outflow (Fig. 9a). The spurious convection of interest is primarily supported by this enhanced instability and convergence. Although the convergence ahead of the MCS is predicted similarly in AERIOE_VRES and DESROZIERS, these novel experiments instead produce a weaker moist layer that does not extend as far north (Figs. 9b,c). We note that a similar moisture bulge is seen in RAP analyses at 0500 UTC (Fig. 4a) but does not extend as far north as is simulated in AERIOE (Fig. 9a), further supporting the simulations shown in AERIOE_VRES and DESROZIERS. As a result of the lower moisture, AERIOE_VRES and especially DESROZIERS predict less most unstable CAPE (MUCAPE) in southern Nebraska, leading to a suppression of the spurious convection relative to AERIOE (Fig. 9d).

6. Impact of observation error statistics on DA cycling

To better understand why the novel methods evaluated here improve the convective forecast, we use this final section of results to trace modifications to the convective ingredients into individual profiles of σ_{Pf}. We present difference and ensemble correlation plots to evaluate the impacts specifically for the spurious convection and the trailing stratiform region. Note that the differences in Fig. 10, apart from the first column, are plotted relative to AERIOE as that technique has been applied for previous studies (DWS19; DWS20) and can be treated as a baseline method for estimating observation error variances.

a. Impacts of observation error statistics on spurious precipitation

The difference plots reveal two moisture surges during DA that contribute to the northern extension of the elevated moist layer and the spurious convection simulated in AERIOE. The first of these surges occurs at 0100 UTC just east of FP3 when AERIOE increases the 850-hPa mixing ratio by ~3 g kg^{-1}

relative to NOAERI (red circle in Fig. 10a). This region of enhanced moisture is advected to the northwest between 0100 and 0530 UTC. At 0400 UTC, a second region of enhanced moisture develops in AERIOE, again just east of FP3 (Fig. 10d). These two moisture surges merge during later DA cycles, eventually producing a convective-scale region of enhanced moisture compared to NOAERI ($+1.03$ g kg^{-1}, red box in Fig. 10g). The final location of this positive moisture increment corresponds to the enhanced MUCAPE shown in Figs. 9a and 9d.

Assimilating the retrievals in the two novel experiments increasingly reduces the impact of both moisture surges (see negative values in the red boxes of Figs. 10h,i). The moistening is more reduced in DESROZIERS (-2.21 g kg^{-1}) and least reduced in AERIOE_VRES (-1.52 g kg^{-1}). We note that in general, the difference fields for (AERIOE_VRES − AERIOE) and (DESROZIERS − AERIOE) feature the same spatial structure as (AERIOE − NOAERI) except for the opposite sign, indicating that these two novel methods primarily modify the magnitude of the observation impacts instead of their shape. We also summarize the moisture impacts by computing average soundings during the final DA cycle (Fig. 11a). Here, each experiment reduces MUCAPE and enhances the most unstable convective inhibition (MUCIN) compared to AERIOE. Again, the impacts for DESROZIERS are larger than those in AERIOE_VRES. Given the modifications to the convective indices, these relative differences likely explain why AERIOE_VRES and DESROZIERS reduce the amount of spurious convection compared to AERIOE.

Next, we perform an ensemble correlation analysis to determine which retrievals contribute most to the positive moisture increments in AERIOE. While both moisture surges occur near FP3, Fig. 12 reveals that the impacts are likely related to AERI observations assimilated from FP4 (Fig. 12a; 0100 UTC surge) and FP6 (Fig. 12b; 0400 UTC surge). Both sites feature large, positive innovations ($\mathbf{d}_b^o > 3$ g kg^{-1}), meaning that assimilating these data results in strong, positive increments in regions of positive correlation and strong, negative increments in regions of negative correlation. Though the analysis increments are a superposition of all sites and observation levels assimilated, the correlation structures at FP4 and FP6 roughly correspond to the moisture surges highlighted above (yellow circles in Fig. 12).

Finally, we present the AERI retrievals assimilated at FP4 and FP6 to better understand how the observation error statistics relate to these moisture impacts (Figs. 13 and 14). We note that the background and analysis profiles are only shown for AERIOE as the corresponding profiles for other experiments do not significantly differ. As with Fig. 12, both retrievals indicate positive innovations above 900 hPa due to the very moist retrievals relative to the background. AERIOE estimates the lowest σ_{Pf} at both sites (Fig. 13c) and thus assigns the largest weight to the moist retrievals during DA. AERIOE_VRES features only slightly larger σ_{Pf} than AERIOE due to a marginally larger vertical resolution compared to the median (Figs. 13b, 14b). Despite this, the small error increases in AERIOE_VRES appear to have

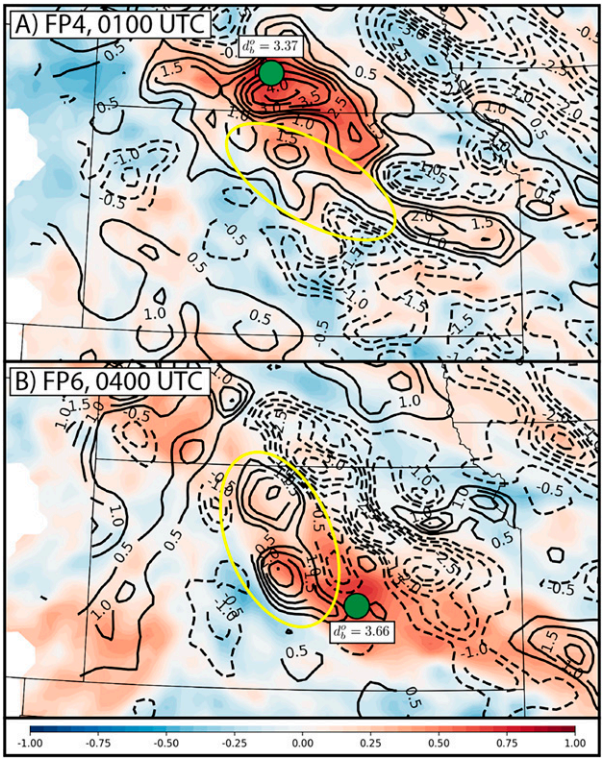

FIG. 12. Background ensemble correlations between 850-hPa water vapor mixing ratio at an AERI platform and the rest of the domain. The plots are computed at (a) FP4 valid at 0100 UTC and (b) FP6 valid at 0400 UTC. The location of the FP4 and FP6 sites are annotated with the green dots, and the mean values of their 850-hPa moisture innovations are overlaid (g kg^{-1}). Also overlaid in black contours are the analysis increments for the AERIOE experiment. See text for a description of the yellow circles.

been significant enough to reduce the impact of the moistening. Conversely, DESROZIERS produces much larger σ_{Pf} between 900 and 850 hPa at both sites. The observation standard deviation peaks at 2.9 g kg^{-1} at FP4 (Fig. 13c) and 3.5 g kg^{-1} at FP6 (Fig. 14c), indicating systematic peaks in the innovation and analysis residuals for the midlevel retrievals at FP4 and FP6. These larger values of σ_{Pf} reduce the weight assigned to the moist AERI retrievals and thus support the reduction in spurious convection. Apart from the improved forecast results when assimilating AERIs using the larger uncertainties, the increased σ_{Pf} are further supported by a collocated rawinsonde launched at FP6 that indicates a large moisture error in the AERI retrievals (Fig. 14a).

b. *Impacts of observation error statistics on trailing stratiform precipitation*

Finally, we also analyze the impacts of the observation error statistics on the low-level moisture near the trailing stratiform region (purple regions in Figs. 10g–i). Although the small NEP increases in this region do not impact the skill as much as the improvements to the spurious

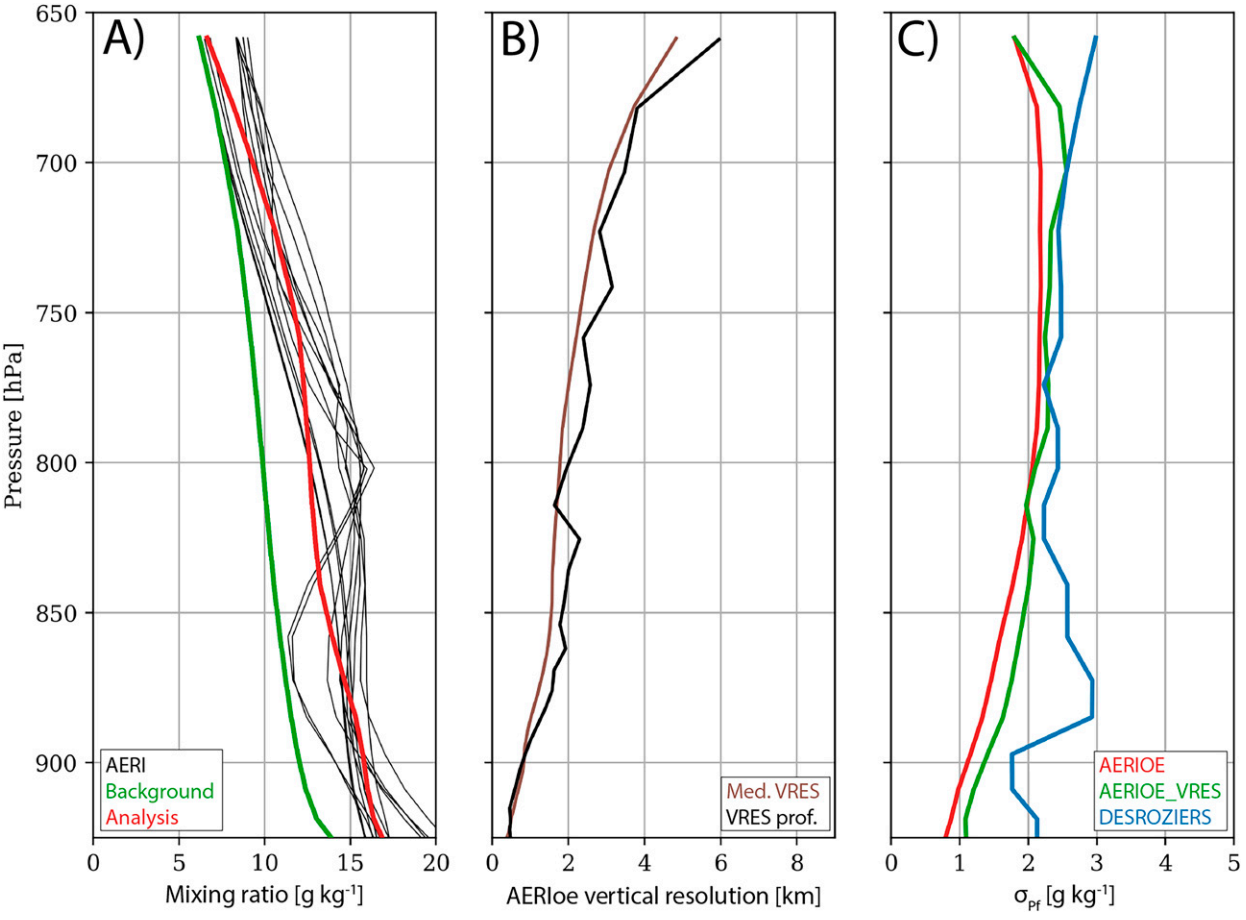

FIG. 13. (a) AERI water vapor mixing ratio (g kg^{-1}) observations assimilated at FP4 during the 0100 UTC cycle, along with corresponding background and analysis profiles from the AERIOE experiment. (b) Median effective vertical resolution (km) profiles output by the AERIoe retrieval at the same time. The black profile indicates the median computed over all PECAN retrievals, while the brown profile indicates the median computed only over the profiles assimilated during this cycle. (c) Observation error standard deviation (g kg^{-1}) profiles for these data assigned in each experiment.

precipitation, each method still shows a measurable benefit compared to the original static inflation method applied in AERIOE.

In general, we find that assimilating the AERI retrievals in the baseline AERIOE experiment produces moderate drying in northwestern Kansas, especially during the 0100 UTC cycle near FP5 (Fig. 10a) and the 0415 UTC cycle near SPARC (not shown). These moisture impacts are mostly stationary and eventually lead to moderate drying underneath the ongoing MCS during the final DA cycle (−0.89 g kg^{-1}, Fig. 11k). The moisture impacts are also illustrated by computing an average sounding that reveals reduced dewpoint temperatures below 600 hPa and 200 J kg^{-1} less of MUCAPE in AERIOE compared to NOAERI (Fig. 11b). The reduced instability likely explains why convection in the trailing stratiform region decays in AERIOE compared to other experiments. Conversely, assigning σ_{Pf} using the other two methods reduces the impact of this drying (Figs. 10h,i) and correspondingly increases the MUCAPE (Fig. 11b). As before, the relative magnitude of these differences is larger in DESROZIERS (Fig. 10i) compared to AERIOE_VRES (Fig. 10h).

We do not present a correlation analysis for this section, as the moisture impacts near the trailing stratiform region occur directly above FP5 and SPARC (Figs. 15 and 16). Unlike the previous section which details moist biases, these AERI observations instead feature dry layers that are not supported by collocated rawinsondes. Specifically, the 0100 UTC retrieval at FP5 is too dry between 875 and 750 hPa (Fig. 15a), while the 0415 UTC retrieval at SPARC is too dry above 850 hPa (Fig. 16a). At FP5, AERIOE and AERIOE_VRES estimate similar, lower σ_{Pf} and thus assign the heaviest weight to these dry retrievals (Fig. 15c). Conversely, DESROZIERS again estimates a midlevel peak in σ_{Pf} that reduces the impact of these dry retrievals (Fig. 15c). At SPARC, the dry error occurs just below cloud base, indicating that it could potentially be related to the emission issues detailed in Turner and Blumberg (2019). The issue is represented in the vertical resolution profile above 850 hPa (Fig. 16b) and as such, σ_{Pf} from SPARC is correspondingly inflated in AERIOE_VRES (Fig. 16c).

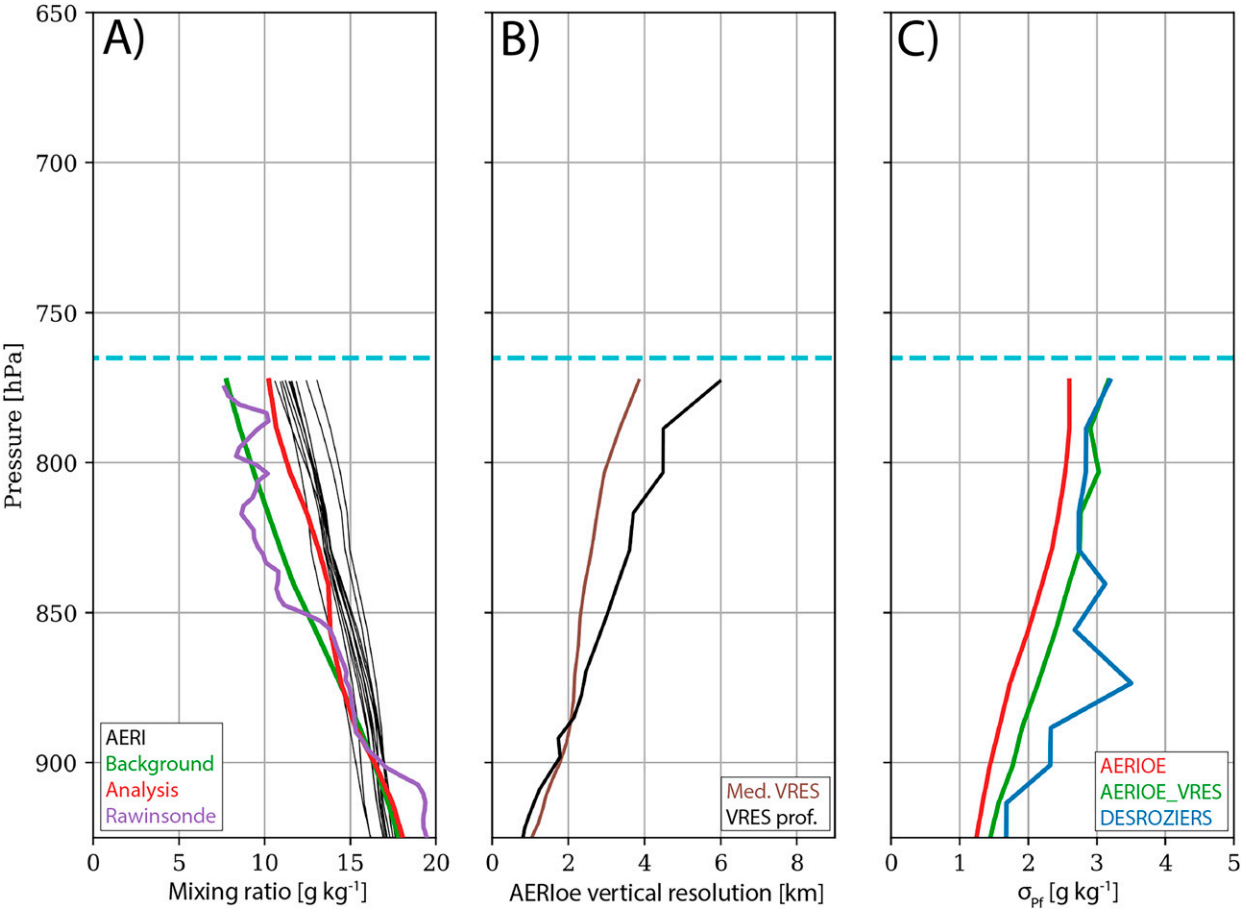

FIG. 14. As in Fig. 13, but for the FP6 moisture observations assimilated at 0400 UTC. Also overlaid now are the cloud base height indicated by a collocated lidar or ceilometer (dashed cyan line), and a corresponding moisture profile collected by a collocated rawinsonde [purple line in (a)].

7. Discussion

This study is among the first work to propose and evaluate the impact of using novel methods to estimate observation error statistics for ground-based thermodynamic profilers in a convective-scale DA system. Using high-frequency retrievals collected by AERIs, we compare two methods to a static inflation technique applied in earlier studies. These two novel methods include one that adaptively inflates retrieval error statistics, and another that directly estimates the full observation error variance, including representation components, based on the D05 diagnostic. Each of these methods is shown to produce more variant error statistics that better match the wide distribution of true observation error statistics. The adaptive inflation method primarily produces larger error variances below cloud base, thereby accounting for retrieval issues described in Turner and Blumberg (2019). The direct estimation method based on the D05 diagnostic instead produces larger error variances throughout most of the retrieval profile.

When implementing these methods for an MCS observed during PECAN, we find that a static error inflation applied in previous studies can reduce forecast skill compared to not assimilating the profiling data at all. Specifically, the static error inflation results in reduced probabilities within the trailing stratiform region due to large weights assigned to retrievals that are too dry. When assimilating AERI data using the adaptive inflation method, these dry retrievals are deweighted such that we find improved probabilities within the trailing stratiform region and better suppression of spurious convection developing near the MCS. These results suggest that constant observation error variances or static inflation methods are likely insufficient when assimilating remote sensing data at convective scales.

Similar to the adaptive inflation results, we also find that the direct estimation of observation error variances from the D05 method performs better than the static inflation method applied in DWS19 and DWS20. The experiment that applied the D05 method performed similarly to the adaptive inflation method in terms forecast quality for the nocturnal MCS with only a slight decrease in skill for higher precipitation thresholds. However, while the inflation methods are designed specifically for AERI retrievals, the D05 method can instead be generalized for any observation type. Additionally, the D05

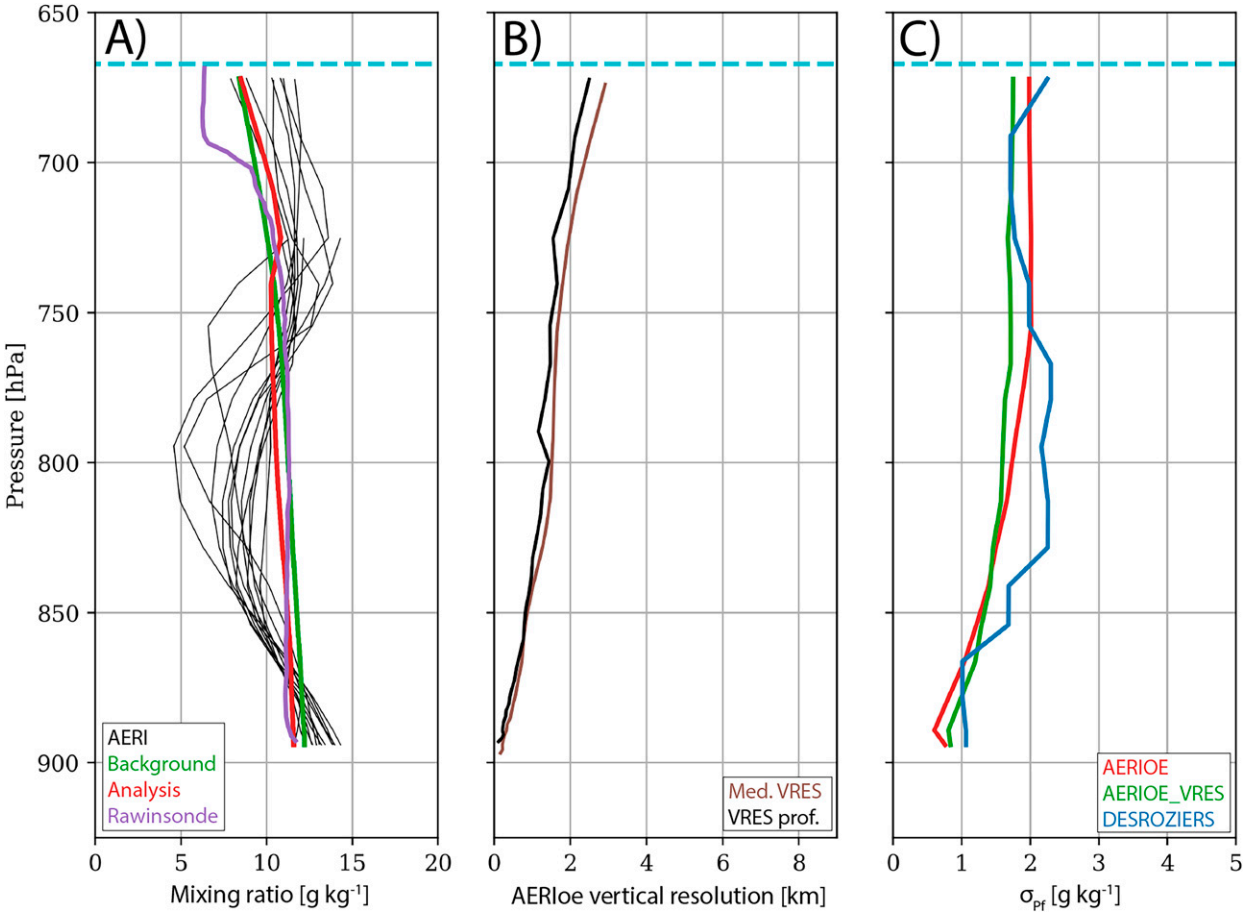

FIG. 15. As in Fig. 14, but for the FP5 moisture observations assimilated at 0100 UTC.

method can account for additional sources of uncertainty that cannot be easily represented by empirical inflation methods (e.g., large residuals and corresponding error variances near sharp gradients). As such, we recommend that future studies further explore the capabilities of using the D05 method for convective-scale applications.

The improvements shown here can also be interpreted as illustrating the benefits of accounting for flow-dependent observation error statistics. For example, the adaptive inflation method can account for additional observation uncertainty in the presence of clouds. The D05 method, despite estimating temporally constant error statistics, can instead account for increased uncertainty through various features such as an inversion. While static methods might perform well for coarser, synoptic applications, our analysis of the 15 July 2015 case highlights how observation error variances can become underestimated at the convective scale, and that these deficiencies can be propagated throughout the assimilation and forecast period to degrade precipitation forecasts. This result suggests that adaptive or flow-dependent methods are needed to account for observing times that feature larger errors (i.e., increased representation errors in certain conditions) and that such methods can greatly improve the impact of assimilating various datasets. As an initial effort to demonstrate and to understand the two proposed methods, a single case study is adopted. However, given that the components of **R** vary considerably by case and observing type, we recommend evaluating these methods for more cases and using additional remote sensing instruments including satellite-based profilers.

Future studies should also evaluate the potential impact of observation biases when using the D05 method. As mentioned in section 2, the D05 method assumes that the innovation and analysis residuals are unbiased. However, separating observation and model biases can be challenging (Chandramouli et al. 2021) especially given that the mean error variances for AERIoe vary significantly in the presence of clouds (Turner and Löhnert 2014). Therefore, as with other published studies (e.g., Satterfield et al. 2017), we cannot eliminate the influence of bias on the diagnosed error statistics. These untreated biases could result in overestimated observation error variances which could also lead to larger apparent impacts when applying the D05 method. We recommend that future studies either explore optimal bias correction procedures for AERIoe retrievals or apply methods that modify the D05 method to account for biases (e.g., Waller et al. 2017).

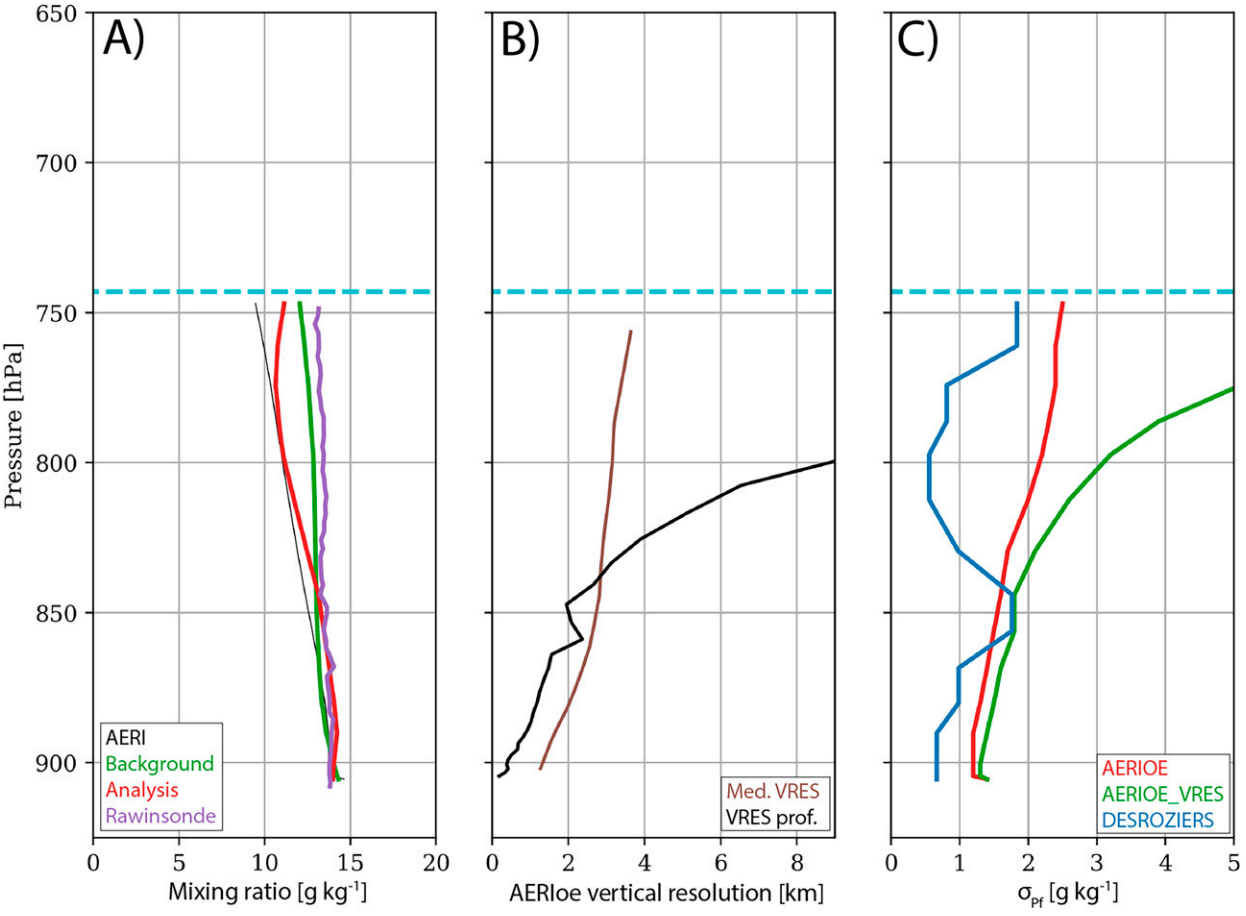

FIG. 16. As in Fig. 14, but for the SPARC moisture observations assimilated at 0415 UTC.

Finally, a large advantage of the D05 method results from its ability to diagnose a full observation error covariance matrix including observation error correlations. Though horizontal error correlations (i.e., site to site correlations) are likely small for profiling instruments, vertical correlations can be large (Turner and Löhnert 2014). Thus, our future work will also adopt other DA methods such as EnVar (Wang et al. 2013; Wang and Lei 2014) to evaluate the impact of assimilating the full **R** diagnosed by the D05 method instead of only inflating the observation error variances.

Acknowledgments. This project is primarily supported by the National Science Foundation (NSF) Award AGS-1359703 and the Awards NA16OAR3420115, NA16OAR4320115, and NA11OAR4320072. We would also like to acknowledge the high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research and sponsored by the NSF. Finally, the authors also wish to acknowledge David Stensrud for his helpful discussions in planning this project, along with the members of the Multi-scale data Assimilation and Predictability Laboratory (MAP; http://weather.ou.edu/~map/index.html) at OU.

Data availability statement. All AERI retrievals are available on the PECAN field catalog at http://catalog.eol.ucar.edu/pecan. The model output data for the assimilation experiments are archived locally and available upon request to the corresponding author.

REFERENCES

Bormann, N., M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally, 2016: Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quart. J. Roy. Meteor. Soc.*, **142**, 1767–1780, https://doi.org/10.1002/qj.2774.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550–1566, https://doi.org/10.1175/2009MWR3157.1.

Campbell, W. F., E. A. Satterfield, B. Ruston, and N. L. Baker, 2017: Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. *Mon. Wea. Rev.*, **145**, 1019–1032, https://doi.org/10.1175/MWR-D-16-0240.1.

Chandramouli, K., X. Wang, A. Johnson, and J. Otkin, 2021: Online nonlinear bias correction in ensemble Kalman filter to assimilate GOES-R all-sky radiances for the analysis and prediction of rapidly developing supercells. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002711, https://doi.org/10.1029/2021MS002711.

Chipilski, H. G., X. Wang, and D. B. Parsons, 2020: Impact of assimilating PECAN profilers on the prediction of bore-driven nocturnal convection: A multi-scale forecast evaluation for the 6 July 2015 case study. *Mon. Wea. Rev.*, **148**, 1147–1175, https://doi.org/10.1175/MWR-D-19-0171.1.

Cordoba, M., S. L. Dance, G. A. Kelly, N. K. Nichols, and J. A. Waller, 2017: Diagnosing atmospheric motion vector observation errors for an operational high-resolution data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **143**, 333–341, https://doi.org/10.1002/qj.2925.

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. *Mon. Wea. Rev.*, **147**, 2739–2764, https://doi.org/10.1175/MWR-D-18-0423.1.

——, ——, ——, and D. D. Turner, 2020: Systematic evaluation of the impact of assimilating a network of ground-based remote sensing profilers for forecasts of nocturnal convection initiation during PECAN. *Mon. Wea. Rev.*, **148**, 4703–4728, https://doi.org/10.1175/MWR-D-20-0118.1.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, https://doi.org/10.1256/qj.05.108.

Du, J., and Coauthors, 2014: NCEP regional ensemble update: Current systems and planned storm-scale ensembles. *26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction*, Atlanta, GA, Amer. Meteor. Soc., J1.4, https://ams.confex.com/ams/94Annual/webprogram/Paper239030.html.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, https://doi.org/10.1029/2002JD003296.

Feltz, W. F., W. L. Smith, H. B. Howell, R. O. Knuteson, H. Woolf, and H. E. Revercomb, 2003: Near-continuous profiling of temperature, moisture, and atmospheric stability using the atmospheric emitted radiance interferometer (AERI). *J. Appl. Meteor.*, **42**, 584–597, https://doi.org/10.1175/1520-0450(2003)042<0584:NPOTMA>2.0.CO;2.

Fielding, M. D., and M. Janiskova, 2018: Assimilating profiles of cloud radar and lidar observations into the ECMWF 4D-Var system. *Sixth Int. Symp. on Data Assimilation (2018)*, Munich, Germany, Universität München Schellingstraße, 3.6, https://isda2018.wavestoweather.de/.

——, and O. Stiller, 2019: Characterizing the representativity error of cloud profiling observations for data assimilation. *J. Geophys. Res. Atmos.*, **124**, 4086–4103, https://doi.org/10.1029/2018JD029949.

Fowler, A., and P. J. Van Leeuwen, 2013: Observation impact in data assimilation: The effect of non-Gaussian observation error. *Tellus*, **65A**, 20035, https://doi.org/10.3402/tellusa.v65i0.20035.

Geer, A. J., and P. Bauer, 2011: Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteor. Soc.*, **137**, 2024–2037, https://doi.org/10.1002/qj.830.

——, and Coauthors, 2018: All-sky satellite data assimilation at operational weather forecasting centres. *Quart. J. Roy. Meteor. Soc.*, **144**, 1191–1217, https://doi.org/10.1002/qj.3202.

Geerts, B., and Coauthors, 2017: The 2015 plains elevated convection at night field project. *Bull. Amer. Meteor. Soc.*, **98**, 767–786, https://doi.org/10.1175/BAMS-D-15-00257.1.

Grasmick, C., B. Geerts, D. D. Turner, Z. Wang, and T. M. Weckwerth, 2018: The relation between nocturnal MCS evolution and its outflow boundaries in the stable boundary layer: An observational study of the 15 July 2015 MCS in PECAN. *Mon. Wea. Rev.*, **146**, 3203–3226, https://doi.org/10.1175/MWR-D-18-0169.1.

Grell, G. A., and S. R. Freitas, 2013: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, **13**, 23 845–23 893, https://doi.org/10.5194/acpd-13-23845-2013.

Hodyss, D., and N. Nichols, 2015: The error of representation: Basic understanding. *Tellus*, **67A**, 24822, https://doi.org/10.3402/tellusa.v67.24822.

——, and E. Satterfield, 2017: The treatment, estimation, and issues with representation error modeling. *Data Assimilation for Atmospheric, Oceanic, and Hydrologic Applications,* S. K. Park and L. Xu, Eds., Vol. III, Springer International Publishing, 177–194.

Hong, S.-Y., and J. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.

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. *Wea. Forecasting*, **34**, 1453–1476, https://doi.org/10.1175/WAF-D-18-0200.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, 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.

Janjić, T., and Coauthors, 2018: On the representation error in data assimilation. *Quart. J. Roy. Meteor. Soc.*, **144**, 1257–1278, https://doi.org/10.1002/qj.3130.

Johnson, A., X. Wang, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108, https://doi.org/10.1175/MWR-D-14-00345.1.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press, 368 pp.

Lakshmanan, V., T. Smith, G. Stumpf, and K. Hondl, 2007: The Warning Decision Support System–Integrated Information. *Wea. Forecasting*, **22**, 596–612, https://doi.org/10.1175/WAF1009.1.

Lange, H., and T. Janjić, 2016: Assimilation of mode-S EHS aircraft observations in COSMO- KENDA. *Mon. Wea. Rev.*, **144**, 1697–1711, https://doi.org/10.1175/MWR-D-15-0112.1.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2.

Ménard, R., 2016: Error covariance estimation methods based on analysis residuals: Theoretical foundation and convergence properties derived from simplified observation networks. *Quart. J. Roy. Meteor. Soc.*, **142**, 257–273, https://doi.org/10.1002/qj.2650.

——, Y. Yang, and Y. Rochon, 2009: Convergence and stability of estimated error variances derived from assimilation residuals

in observation space. *ECMWF Workshop on Diagnostics of Data Assimilation System Performance*, Reading, United Kingdom, ECMWF, 133–143, https://www.ecmwf.int/sites/default/files/elibrary/2009/75646-convergence-and-stability-estimated-error-variances-derived-assimilation-residuals-observation_0.pdf.

Minamide, M., and F. Zhang, 2017: Adaptive observation error inflation for assimilating all-sky satellite radiance. *Mon. Wea. Rev.*, **145**, 1063–1081, https://doi.org/10.1175/MWR-D-16-0257.1.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, https://doi.org/10.1007/s10546-005-9030-8.

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

Satterfield, E., D. Hodyss, D. D. Kuhl, and C. H. Bishop, 2017: Investigating the use of ensemble variance to predict observation error of representation. *Mon. Wea. Rev.*, **145**, 653–667, https://doi.org/10.1175/MWR-D-16-0299.1.

Schwartz, C. S., and R. A. Sobash, 2017: Generating probabilistic forecasts from convection- allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, https://doi.org/10.1175/MWR-D-16-0400.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Stewart, L. M., S. L. Dance, N. K. Nichols, J. R. Eyre, and J. Cameron, 2014: Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quart. J. Roy. Meteor. Soc.*, **140**, 1236–1244, https://doi.org/10.1002/qj.2211.

Tandeo, P., P. Ailliot, M. Bocquiet, A. Carrassi, T. Miyoshi, M. Pulido, and Y. Zhen, 2020: A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation. *Mon. Wea. Rev.*, **148**, 3973–3994, https://doi.org/10.1175/MWR-D-19-0240.1.

Tao, W.-K., and Coauthors, 2003: Microphysics, radiation and surface processes in the Goddard cumulus ensemble (GCE) model. *Meteor. Atmos. Phys.*, **82**, 97–137, https://doi.org/10.1007/s00703-001-0594-7.

Turner, D. 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). *J. Appl. Meteor. Climatol.*, **53**, 752–771, https://doi.org/10.1175/JAMC-D-13-0126.1.

——, and W. G. Blumberg, 2019: Improvements to the AERIoe thermodynamic profile retrieval algorithm. *IEEE Sel. Top. Appl. Earth Obs. Remote Sens.*, **12**, 1339–1354, https://doi.org/10.1109/JSTARS.2018.2874968.

Vaisala, 2017: Vaisala radiosonde RS41 measurement performance. Vaisala white paper, Ref. B211356EN-B, 28 pp., https://www.vaisala.com/sites/default/files/documents/WEA-MET-RS41-Performance-White-paper-B211356EN-B-LOW-v3.pdf.

Waller, J. A., S. L. Dance, and N. K. Nichols, 2016a: Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Quart. J. Roy. Meteor. Soc.*, **142**, 418–431, https://doi.org/10.1002/qj.2661.

——, D. Simonin, S. L. Dance, N. K. Nichols, and S. P. Ballard, 2016b: Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Mon. Wea. Rev.*, **144**, 3533–3551, https://doi.org/10.1175/MWR-D-15-0340.1.

——, S. P. Ballard, S. L. Dance, G. Kelly, N. K. Nichols, and D. Simonin, 2016c: Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sens.*, **8**, 581, https://doi.org/10.3390/rs8070581.

——, S. L. Dance, and N. K. Nichols, 2017: On diagnosing observation-error statistics with local ensemble data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 2677–2686, https://doi.org/10.1002/qj.3117.

——, E. Bauernschubert, S. L. Dance, N. K. Nichols, R. Potthast, and D. Simonin, 2019: Observation error statistics for Doppler radar radial wind superobservations assimilated into the DWD COSMO-KENDA system. *Mon. Wea. Rev.*, **147**, 3351–3364, https://doi.org/10.1175/MWR-D-19-0104.1.

Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. *Mon. Wea. Rev.*, **142**, 3303–3325, https://doi.org/10.1175/MWR-D-13-00303.1.

——, C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble–3DVAR hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222–227, https://doi.org/10.1175/MWR3282.1.

——, D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116–5131, https://doi.org/10.1175/2008MWR2444.1.

——, ——, ——, and ——, 2008b: A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. Part II: Real observation experiments. *Mon. Wea. Rev.*, **136**, 5132–5147, https://doi.org/10.1175/2008MWR2445.1.

——, T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2009: A comparison of the hybrid and EnSRF analysis schemes in the presence of model errors due to unresolved scales. *Mon. Wea. Rev.*, **137**, 3219–3232, https://doi.org/10.1175/2009MWR2923.1.

——, D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble- variational hybrid data assimilation for NCEP global forecast system: Single resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, https://doi.org/10.1175/MWR-D-12-00141.1.

Wang, Y., and X. Wang, 2017: Direct assimilation of radar reflectivity without tangent linear and adjoint of the nonlinear observation operator in the GSI-Based EnVar system: Methodology and experiment with the 8 May 2003 Oklahoma City tornadic supercell. *Mon. Wea. Rev.*, **145**, 1447–1471, https://doi.org/10.1175/MWR-D-16-0231.1.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79, https://doi.org/10.1111/j.1600-0870.2007.00273.x.

Weston, P. P., W. Bell, and J. Eyre, 2014: Accounting for correlated error in the assimilation of high-resolution sounder data. *Quart. J. Roy. Meteor. Soc.*, **140**, 2420–2429, https://doi.org/10.1002/qj.2306.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.