

Exploring Ensemble Forecast Sensitivity to Observations for a Convective-Scale Data Assimilation System over the Dallas–Fort Worth Testbed

NICHOLAS A. GASPERONI,^a XUGUANG WANG,^a KEITH A. BREWSTER,^b AND FREDERICK H. CARR^{a,b}

^a *School of Meteorology, University of Oklahoma, Norman, Oklahoma*

^b *Center for Analysis and Prediction of Storms, University of Oklahoma Norman, Oklahoma*

(Manuscript received 25 April 2023, in final form 14 December 2023, accepted 19 December 2023)

ABSTRACT: Forecast sensitivity to observation (FSO) methods have become increasingly popular over the past two decades, providing the ability to quantify the impacts of various observing systems on forecasts without having to conduct costly data denial experiments. While adjoint- and ensemble-based FSO are employed in many global operational systems, their use for regional convection-allowing data assimilation (DA) and forecast systems have not been fully examined. In this study, ensemble FSO (EFSO) is explored for high-frequency convective-scale DA for a severe weather case study over the Dallas–Fort Worth testbed. This testbed, originally established by the Collaborative Adaptive Sensing of the Atmosphere (CASA) project, aims to improve high-resolution DA systems by assimilating a variety of existing state and regional mesoscale observing systems to fill gaps of conventional observing networks. This study utilizes EFSO to estimate relative impacts of nonconventional surface observations against conventional observations, and further incorporates assimilated radar observations into EFSO. Results show that, when applying advected localization and a neighborhood upscale averaging technique, EFSO estimates remain correlated and skillful with the actual error reduction of all assimilated observations for the duration of 2-h forecasts. The ability for EFSO to verify against other metrics (surface T , u , v , q) beside energy norms is also demonstrated, emphasizing that EFSO can be used to evaluate impacts of specific parts of the forecast system rather than integrated quantities. Partitioned EFSO revealed that while conventional and radar observations contributed to most of the total energy, nonconventional observations contributed a significant percentage (up to 25%) of the total impact to surface thermodynamic fields.

KEYWORDS: Convective storms/systems; Short-range prediction; Data assimilation; Diagnostics; Ensembles; Regional models

1. Introduction

Current state-of-the-art data assimilation (DA) systems assimilate a wide variety of observations from many different platforms, from in situ surface stations, weather balloons, and aircraft data, to remote sensing satellite and radar platforms. This combination of observations improves the overall accuracy of analyses and subsequent forecasts. As such, advancements in DA methodologies along with the ever-increasing volume of observations being assimilated have aided in improving operational forecast accuracy by an average of about one day per decade over the last 40–50 years (Bauer et al. 2015). An essential component of continuing this improvement into the future is determining the added value or impact of existing and proposed observing systems on forecasts. It is of great interest to thus develop methods which efficiently quantify observational impact on forecasts by different types or platforms of observations.

The most straightforward method of determining observation impact for existing observation systems is to conduct sets of observing system experiments (OSEs). In the OSE approach, impact is determined by comparing a control experiment with an experiment that either withholds (data denial) or adds (data addition) a set of observations, and the differences among experiment forecasts directly quantify the impact of this set of observations. Such an approach has been used many times over the last few

decades (e.g., Zapotocny et al. 2002, 2007; Benjamin et al. 2010; Carlaw et al. 2015; James and Benjamin 2017; Gasperoni et al. (2018, hereafter G18; Degelia et al. 2019; Morris et al. 2021). However, it can be computationally demanding to run multiple sets of experiments in modern DA and forecast systems to determine observational impact. As such, OSE studies are typically limited to either case studies or small time periods and only a few subsets of observations.

A more computationally efficient approach to quantifying observation impact was developed by Langland and Baker (2004), who introduced an adjoint-based technique to calculate the forecast sensitivity to observation (FSO) impact within their 3D-variational DA system. Gelaro and Zhu (2009) found that the FSO technique generally agreed with results of data denial experiments. The FSO method has been used in many different forecasting systems since then (e.g., Cardinali 2009; Gelaro et al. 2010; Weissmann et al. 2012; Lorenc and Marriott 2014). While it is a valuable tool for diagnosing forecast impacts of observations without OSEs, FSO requires adjoint operators of the forecast model that are difficult to develop.

An ensemble-based alternative to the adjoint-based FSO technique was explored by Ansell and Hakim (2007), where sensitivity with respect to observation increments is defined as a function of linear regression of ensemble forecast perturbations for a defined forecast metric. Liu and Kalnay (2008) and Li et al. (2010) proposed another ensemble-based FSO technique by defining sensitivity as a function of the forecast error, similar to Langland and Baker (2004), with successful

Corresponding author: Nicholas Gasperoni, ngaspero@ou.edu

DOI: 10.1175/MWR-D-23-0091.1

© 2024 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

application to tropical cyclone forecasting by [Kunii et al. \(2012\)](#). [Kalnay et al. \(2012\)](#) derived a simpler formation, called ensemble FSO (EFSO), which utilizes readily available output from any ensemble Kalman filter (EnKF; [Evensen 1994](#)). [Ota et al. \(2013\)](#) successfully applied this EFSO formulation to the National Centers for Environmental Prediction (NCEP) operational global EnKF system. More recently, [Buehner et al. \(2018\)](#) combined the ensemble-based and adjoint-based FSO techniques to estimate observation impact within a hybrid ensemble-variational (EnVar) system.

While most global operational centers now use some flavor of FSO for monitoring of observation impact, their use in convection-permitting numerical weather prediction and DA systems have been limited. However, operational ensemble convection-allowing DA and forecast systems such as the NOAA Warn-on-Forecast System (WoFS; e.g., [Lawson et al. 2018](#)) and Rapid Refresh Forecast System (RRFS; [Carley et al. 2021](#)) are now feasible. As with global systems, it will be of interest to apply an FSO method to monitor observation impact within these high-resolution regional systems. Further, a report by the United States [National Research Council \(2009\)](#) recommended the integration of existing and future mesoscale observing networks into a Nationwide Network of Networks (NNoN). As shown in data denial case studies of [Carlaw et al. \(2015\)](#), [G18](#), and [Morris et al. \(2021\)](#), assimilating dense nonconventional mesoscale surface observations in a NNoN testbed can lead to improvements in convection initiation (CI) and prediction of severe hail- and tornado-producing storms. Given the large diversity of datasets involved with a NNoN approach, and added costs of running ensemble-based convective-scale DA and forecast systems such as RRFS, it is essential that a method such as EFSO can accurately and efficiently identify the impacts of different observations on a given high-resolution forecast.

Only a few studies so far have demonstrated the usefulness of EFSO within a convective-scale modeling system ([Sommer and Weissmann 2014, 2016](#); [Necker et al. 2018](#)). [Sommer and Weissmann \(2014\)](#) demonstrated that EFSO provides good agreement with data denial impact for a 2.8-km-horizontal-resolution model at 0-, 3-, and 6-h forecast valid times. Later, [Sommer and Weissmann \(2016\)](#) modified the EFSO method to verify against observations directly. [Necker et al. \(2018\)](#) further expanded the convective-scale implementation of EFSO to different verification observations, including radar-derived precipitation. They found that EFSO estimates are particularly sensitive to biases within the observations or model. Although these studies establish that EFSO can work with a convective-scale model, the DA system assimilated only conventional observations every 3 h.

Future convection-allowing ensemble forecast systems such as RRFS will use more-frequent DA and include assimilation of radar reflectivity. To this point, EFSO has not been tested with convective-scale assimilation at higher cycling frequencies, and application of EFSO impact of assimilating radar observations and newer nonconventional observing systems DA have only recently begun ([Casaretto et al. 2023a,b](#)). In this paper, the EFSO method will be applied to the same case study explored in [G18](#) over the Dallas-Fort Worth NNoN testbed ([National Research Council 2012](#)). Different from [Sommer and Weissmann \(2014\)](#), we test EFSO in a convective-scale

DA approach, which includes high-frequency 5-min cycling of conventional observations, radar reflectivity and radial velocity, and dense nonconventional surface observations from several platforms. As in [Necker et al. \(2018\)](#), it is expected that different verification metrics should be used for a convective-scale DA and forecast system than integrated energy norms, which may obscure impacts of smaller-scale observations. Here, given the emphasis of new surface observations, we are interested in verifying against surface variables directly. As a first step for convective-scale DA implementation, the EFSO estimate of impact is compared directly with actual overall forecast error reduction to identify its overall accuracy. The impact of 5-min DA is assessed on forecasts ranging from 0 to 2 h in length. Further, we test the impact of choosing advected or static localizations on the accuracy of EFSO metric. Although forecast durations are shorter for convective-scale studies, the sensitivity of EFSO accuracy to the localization time-forecast component has not been studied to this point at this scale. Additionally, we compare contributions of different observations by observing platform, variable, and verification metrics to assess impacts, with subjective comparisons to data denial experiments from [G18](#).

The rest of the paper is organized as follows. In [section 2](#), the EFSO methodology and verification metrics will be introduced. In [section 3](#), the experiment setup will be discussed, with a definition of diagnostics used for comparing EFSO with actual forecast error reduction. In [section 4](#), the results of the application of EFSO are shown, including analysis of different localizations, different verification metrics, and contributions of EFSO partitioned by observing variable and platform types. A summary and discussion are provided in [section 5](#).

2. EFSO methodology for convective scales

The adjoint-based FSO of [Langland and Baker \(2004\)](#) defines a cost function, J , as the squared forecast error reduction:

$$J = \mathbf{e}_{t_0}^T \mathbf{C} \mathbf{e}_{t_0} - \mathbf{e}_{t-n}^T \mathbf{C} \mathbf{e}_{t-n} = (\mathbf{e}_{t_0} - \mathbf{e}_{t-n})^T \mathbf{C} (\mathbf{e}_{t_0} + \mathbf{e}_{t-n}), \quad (1)$$

where \mathbf{e}_{t-n} and \mathbf{e}_{t_0} are arrays of forecast errors at each grid point and state variable valid at time t , initialized from consecutive DA analyses at times $-n$ and 0, respectively (n is the DA cycling time interval). The matrix of weights \mathbf{C} is typically included which defines the integrated energy norm transformation, allowing the whole modeling system to be evaluated within the impact metric. The most common energy norms used are the *dry* and *moist total energy norms* ([Ehrendorfer et al. 1999](#)).

Equation (1) is summarized as the total forecast impact of all observations assimilated at time 0. When J is negative the magnitude of error in \mathbf{e}_{t_0} is less than the magnitude of error in \mathbf{e}_{t-n} , which can be interpreted as *beneficial impact*. Conversely, when J is positive the observations at time 0 resulted in an increase in forecast error, thus a *detrimental impact*.

[Kalnay et al. \(2012\)](#) derived an ensemble-based FSO metric generalizable to any EnKF based on the above definition of observation impact (1). If we let $\bar{\mathbf{x}}_{t_0}^f$ define the deterministic forecast valid at time t initialized from the mean analysis at time 0 ($\bar{\mathbf{x}}_0^a$), and $\mathbf{x}_t^{\text{truth}}$ as the truth valid at time t , then the

error terms in (1) can be written as $\mathbf{e}_{i|n} = \bar{\mathbf{x}}_{i|n}^f - \mathbf{x}_i^{\text{truth}}$ and $\mathbf{e}_{i|0} = \bar{\mathbf{x}}_{i|0}^f - \mathbf{x}_i^{\text{truth}}$. With these definitions, Kalnay et al. (2012) used the Kalman gain update for the mean analysis $\bar{\mathbf{x}}_0^a$ to derive the EFSO form of observation impact:

$$J \approx \frac{1}{K-1} \delta \mathbf{y}_0^T \mathbf{R}^{-1} \mathbf{H} \mathbf{X}_0^a \mathbf{X}_{i|0}^{fT} (\mathbf{e}_{i|0} + \mathbf{e}_{i|n}), \quad (2)$$

where $\delta \mathbf{y}_0$ is the observation innovation vector for observations assimilated at time 0, \mathbf{H} is the linear observation operator that transforms ensemble perturbations to observation space, \mathbf{R} contains observation error covariances, and \mathbf{X}_0^a and $\mathbf{X}_{i|0}^f$ are $m \times K$ analysis and forecast ensemble perturbation matrices, respectively (m = size of model state, K = ensemble size). Each column in $\mathbf{X}_{i|0}^f$ can be computed using the full nonlinear model, although Kalnay et al. (2012) showed that a tangent linear approximation was used to derive (2).

Covariance localization is necessary for the EFSO method in (2) to limit sampling error, just as with any ensemble-based DA method that uses relatively small ensembles to estimate covariances for a model with high degrees of freedom. Localization of (2) is applied directly to the matrix product $\mathbf{Y}_0^a \mathbf{X}_{i|0}^{fT} = \mathbf{H} \mathbf{X}_0^a \mathbf{X}_{i|0}^{fT}$ of size $p \times m$ (p = number of observations), which estimates model error covariances between the analysis in observation space and forecast at valid time t . Defining the so-called impact localization matrix $\boldsymbol{\rho}_t$, the form of EFSO modified by localization can be written as

$$J \approx \frac{1}{K-1} \delta \mathbf{y}_0^T \mathbf{R}^{-1} [\boldsymbol{\rho}_t^a (\mathbf{Y}_0^a \mathbf{X}_{i|0}^{fT})] (\mathbf{e}_{i|0} + \mathbf{e}_{i|n}). \quad (3)$$

The localization matrix $\boldsymbol{\rho}_t$ defines a $p \times m$ matrix, where every state observation and gridpoint pair can have a unique localization weight.

Kalnay et al. (2012) and Ota et al. (2013) showed that the EFSO localization function should account for the time-forecast component. Gasperoni and Wang (2015) further explained how the proper choice of EFSO localization is inherently linked to the DA localization, in addition to other dependencies including the time-forecast component. Although an added layer of complexity, this link with the localization used during DA helps to constrain the problem. A simple method such as the advected localization of Ota et al. (2013) may work well even for convective-scale DA, since advection is an important component to signal propagation for many variables. Griewank et al. (2023) also demonstrated the need for accurate signal propagation methods may work well for ensemble sensitivity methods; however, for longer lead teams with more uncertain propagation, explicit covariance matrix inversion methods may be necessary for accurate impact estimates.

An additional consideration for the application of EFSO on the convective scale is the choice of cost function metric for quantifying impact. An integrated energy norm may not be appropriate to describe the impacts at convective scales, where we may be interested in only specific parts of the modeling system such as the performance of convection described by hydrometeor variables in the state (often represented as reflectivity). Such a need was similarly identified and tested by Necker et al. (2018) for convective-scale model

forecast impacts over Europe. For the CI case study examined here (described in section 3), we are mainly interested in the near-surface fields, given the strong dependence of CI on the resultant details of those fields. Using a total energy metric may hide some of these details within a high-resolution forecast. For these reasons, near-surface fields (u_s, v_s, T_s, q_s) will be studied in addition to energy norms to see what differences exist when applying EFSO to error measures other than integrated energy norms.

The localized EFSO form in (3) is a summed quantity, as with the actual squared error reduction of (1). However, this sum can be decomposed into contributions from each observation and model state pair, revealing the important utility of EFSO. We can partition impacts by any subset within the sum of (3), such as the impacts by individual observations or subsets of observations (e.g., by platform, observing variable, or observation type) as done in most EFSO studies.

3. Experiment setup

a. Control experiment overview

The model and DA settings for this study are equivalent to the GSI-based EnKF DA and WRF forecast system described in G18, specifically the control experiment that assimilates all observation sources. The 3 April 2014 case is used, a high-impact severe weather case featuring large damaging hail, wind, and a few small tornadoes in the testbed region between 1800 and 2200 UTC (NCDC 2014). A 43-member ensemble was initialized at 0300 UTC 3 April 2014 on a mesoscale grid with 12-km horizontal resolution, with conventional observations except radar (Table 1) assimilated every 3 h from 0600 to 1500 UTC. At 1500 UTC, a $351 \times 351 \times 50$ inner grid with 2.4-km horizontal resolution is initialized (Fig. 1), including two-way feedback with the 12-km outer mesoscale grid. After an hour to spinup convective-scale processes, 5-min cycling of all data sources—conventional, nonconventional, and radar—is conducted from 1600 to 1800 UTC on this inner grid (Fig. 2).

All conventional in situ observations were obtained from the NAM Data Assimilation System (NDAS) data stream in prepbufr format. Nonconventional surface observations were obtained through the Meteorological Data Assimilation Ingest System (MADIS). Radar reflectivity and radial velocity was obtained from NEXRAD and assimilated to capture pre-existing precipitation and help remove spurious convection that develops within the high-frequency cycling period. A full list of assimilated observing types is given in Table 1.

The Gaspari and Cohn (GC) (1999) function is used with an explicit cutoff radius for covariance localization. On the inner grid, assumed observation error standard deviation and horizontal localization radii are varied by observation network density and type (Table 1). Localization values for mesonet and radar observations are similar to past studies who have used them (e.g., Sobash and Stensrud 2015 and Johnson et al. 2015, respectively). All observation types employed a vertical localization of 0.55 during DA in natural log coordinates.

TABLE 1. List of observations assimilated in this study. C and NC refer to conventional and nonconventional. Observation errors are defined for surface pressure, temperature, moisture (RH), and wind observations, respectively, with the exception of radar (reflectivity and radial velocity). Observation errors for upper-air observations are taken from default table in GSI.

Observation category	Full name	C or NC?	Observation error	Localization scale (km)
ConvUpperAir	Upper-air observations (radiosondes, NOAA wind profilers, GPS precipitable water, aircraft)	C	GSI default	200
NEXRAD	Next-generation weather radar	C	5 dBZ, 2.0 m s^{-1} (V_r)	20
ASOS/AWOS	Automated surface observing system, Automated weather observing system (surface)	C	0.5426 hPa, 0.8°C , 2.705% , 1.5 m s^{-1}	200
OK/TX Mesonet	Oklahoma Mesonet (McPherson et al. 2007), West Texas Mesonet (Schroeder et al. 2005)	C	0.75 hPa , 1.0°C , 3.5% , 1.5 m s^{-1}	80
ERNET	Earth networks stations (https://www.earthnetworks.com)	NC	1.125 hPa , 1.5°C , 5.0% , 2.0 m s^{-1}	40
CWOP	Citizen weather observer program (http://www.wxqa.com/)	NC	1.5 hPa , 2.0°C , 7.0% , 2.5 m s^{-1}	40
MoPED	Mobile platform environmental data network (Heppner 2013)	NC	As in ERNET	40
MiscMeso	Miscellaneous surface stations from federal and state agencies (G18)	NC	As in ERNET	40

Further details of the case study overview, model, and DA settings are described in G18.

b. EFSO settings

The EFSO method is the same as that used in Ota et al. (2013), adapted to interface with WRF Model input data for this study. Multiple verification metrics are tested for EFSO accuracy. The moist total energy (MTE) norm is tested as with most other EFSO studies [i.e., Eq. (9) of Ota et al. 2013]. We also include dry total energy (DTE, no moisture component) and kinetic energy (KE, only horizontal wind components).

Other nonintegrated metrics will include near-surface model levels partitioned by state variable: u_s , v_s , T_s , q_s .

The control experiment setup described in section 3a includes 25 five-minute DA cycles over the 2-h period from 1600 to 1800 UTC on 3 April 2014. The 2-h free forecasts were run for each of the 25 ensemble analyses between 1600 and 1800 UTC (Fig. 2), and EFSO is computed at 10-min forecast intervals. Each DA cycle offers a different sample from which EFSO can be calculated for a given forecast verification time. In this study, the truth is assumed to be a verifying analysis. Since the latest valid time for EFSO evaluation is at 2000 UTC, we continue 5-min DA cycling from 1800 to

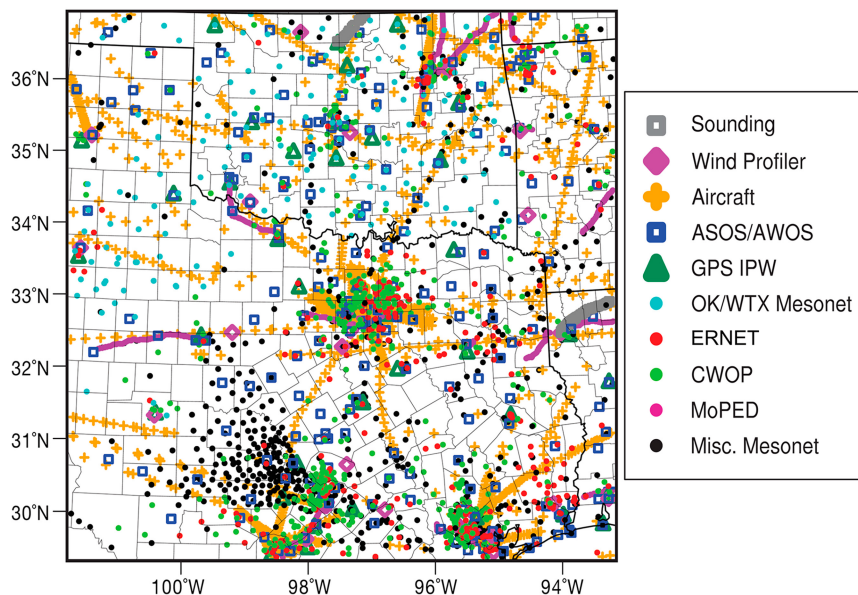


FIG. 1. Inner grid domain for DA experiments, with markers showing location and types of observations (listed in Table 1) assimilated between 1600 and 1800 UTC.

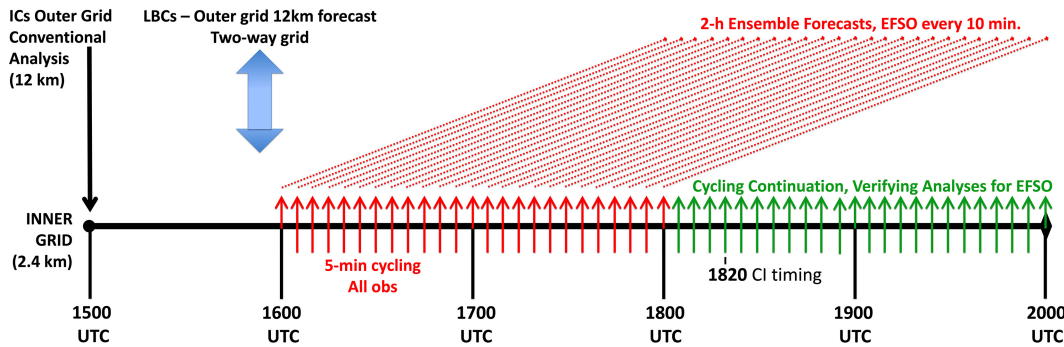


FIG. 2. The 5-min-cycled DA setup for EFSO experiments. Red indicates analyses and 2-h forecasts used for EFSO impact estimates, and green indicates further cycling to provide verifying analyses for the entire EFSO forecast evaluation period.

2000 UTC to have verifying analyses available for all forecasts (Fig. 2). One caveat of this study is that model biases and errors are correlated with the forecast, which may affect the EFSO estimates because the same DA system is used for the verifying analysis. However, given the frequent DA cycling in this study, the impact of near-surface model biases from the verifying analysis should be reduced (e.g., Sobash and Stensrud 2015). Further, Kotsuki et al. (2019) showed that better accuracy of EFSO is attained when no posterior inflation is applied to the analysis ensemble, such that the Kalman gain for the EFSO estimate is consistent with the Kalman gain used during EnKF to produce the analysis mean. Given the application of covariance inflation during DA cycling to maintain sufficient ensemble spread, another caveat in this study is that EFSO may overestimate impact especially in dense observing regions as a result of applying inflation to the ensemble analysis.

Multiple horizontal localization methods are tested for the EFSO metric and compared for accuracy against the actual impact, defined as the actual forecast error reduction caused by the assimilation of all observations in a given DA cycle [i.e., using Eq. (1)]. The first is static localization, applying the same GC localization as that used during DA for each observation type (EFSOnoadv). The next is advected localization, where the center of the GC localization is moved utilizing the average of the analysis and forecast horizontal wind at each vertical level, following Ota et al. (2013). EFSO estimates with two advected localization coefficients of 0.75 and 1.5 are tested (EFSOadv0.75 and EFSOadv1.5, respectively). The former is equivalent to the optimal value found in Ota et al. (2013), while the latter tests the sensitivity to this choice of coefficient. A visual demonstration of 200-km GC localization and advected localization for a 30-min forecast is shown in Fig. 3. The center of the advected localization has shifted eastward and morphed to a northeast-

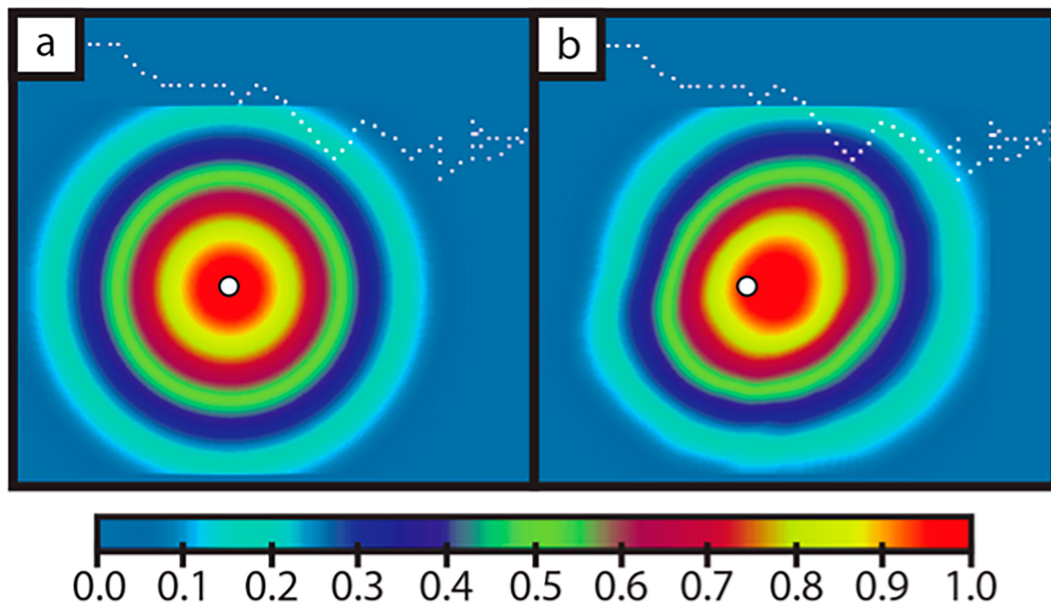


FIG. 3. (a) Static 200-km GC localization function. (b) Advected GC localization (coef = 0.75) using $t = 30$ -min forecast. White dot indicates location of observation.

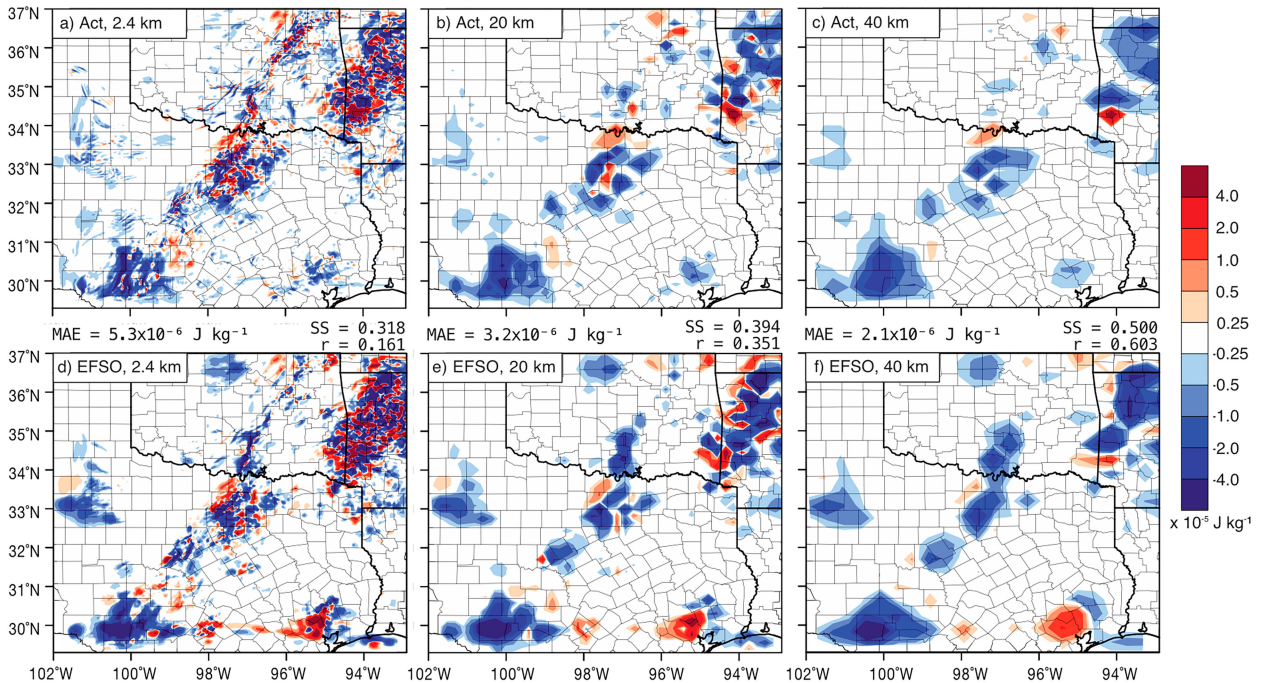


FIG. 4. Two-dimensional maps of (a)–(c) actual impact and (d)–(f) EFSOadv0.75-estimated impact in terms of moist total energy (J kg^{-1}), plotted on (a),(d) the original integration 2.4-km domain; (b),(e) the upscaled 20-km domain; and (c),(f) the 40-km domain. All panels depict 1-h forecast impact initialized by analysis at 1730 UTC 14 Apr 2014.

southwest-oriented elliptical shape, in accordance with the wind flow. Note that the vertical localization for EFSO follows the same settings as DA (0.55 in natural log pressure)

c. Diagnostics for evaluation

As described in section 2, the essence of EFSO is in its ability to diagnose the partial sums of the total impact in (3) by subsets of observations, making it a useful tool for monitoring and improving ensemble-based DA systems. However, before examining these partitioned impacts, we seek to assess the overall quality of EFSO estimates for high-frequency convective-scale DA and forecasts.

Another approach is to sum up the EFSO contributions of all observations on the impact at a given grid point. This approach can allow for direct comparisons of 2D maps of EFSO with the maps of actual error reduction of all assimilated observations in each 5-min DA cycle, since the domain-wide sum of actual error reduction [Eq. (1)] can similarly be partitioned by grid point. Two statistical verification metrics are used in this evaluation: pattern correlation and a skill score (SS) metric based on mean-absolute-error (MAE). Pattern correlation, r , is defined as the Pearson coefficient of linear correlation between values of two variables (EFSO and Actual) at equivalent horizontal gridpoint locations m on two different maps:

$$r = \frac{\sum_m (\text{EFSO}_m - \overline{\text{EFSO}})(\text{Actual}_m - \overline{\text{Actual}})}{\sqrt{\sum_m (\text{EFSO}_m - \overline{\text{EFSO}})^2 \sum_m (\text{Actual}_m - \overline{\text{Actual}})^2}}, \quad (4)$$

where overbars denote domain-wide averages.

The SS is defined as

$$\text{SS} = 1 - \frac{\text{MAE}}{\text{MAE}_{\text{ref}}}, \quad (5)$$

where $\text{MAE} = \sum_m |\text{EFSO}_m - \text{Actual}_m|$ and $\text{MAE}_{\text{ref}} = \sum_m (|\text{EFSO}_m| + |\text{Actual}_m|)$, and an ideal SS is 1. The reference MAE used for (5) is similar in concept to the reference used for computing fractions skill score (FSS; Roberts and Lean 2008). In this case, it defines zero skill as the hypothetical situation where nonzero locations of EFSO do not overlap the nonzero locations of actual impact. Unlike FSS, the SS here may go below zero if, for example, EFSO impact values are the opposite sign compared to the actual impact at many locations. MAE is chosen because we do not want to overemphasize extreme departures in the comparison.

As a final consideration, EFSO is compared both at gridpoint scale and at coarser scales. Verification at convection-permitting scales often employ neighborhood averaging techniques to prevent small displacement errors from causing outsized negative scores in the verifications (e.g., Roberts and Lean 2008). To avoid these issues, maps of EFSO and actual impact are re-gridded to 20- and 40-km NOAA grids.¹ This upscaling is performed by spatially averaging within each 20- or 40-km grid cell all impact values from the 2.4-km grid found within that cell,

¹ The 20- and 40-km grids are NCEP 212 and 215 grids, respectively, defined at <https://www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html>.

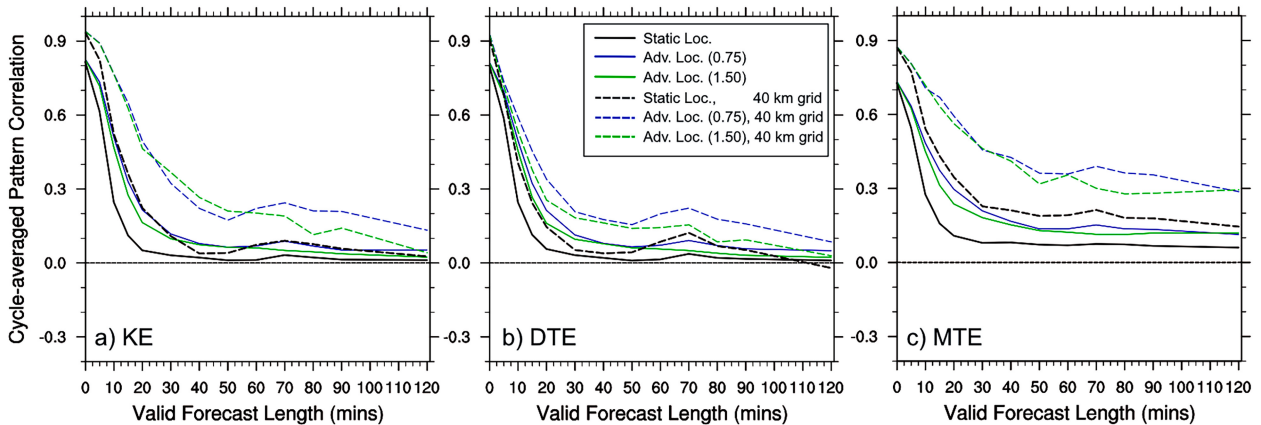


FIG. 5. Pattern correlation of EFSO estimate compared to actual impact, averaged over the number of DA cycles available (25) for (a) kinetic energy, (b) dry total energy, and (c) moist total energy. Black lines indicate static GC localization (EFSOadv), while blue and green lines indicate EFSO estimates using advected localization with weighting coefficients of 0.75 (EFSOadv0.75) and 1.5 (EFSOadv1.5), respectively. Solid and dashed lines indicate correlations computed on the original 2.4-km grid and upscaled 40-km grid, respectively.

similar to neighborhood averaging techniques described in Schwartz and Sobash (2017).

4. EFSO results

a. Comparison of EFSO estimated impact with actual error reduction

1) SUBJECTIVE EVALUATION

Two-dimensional patterns of impact are shown in Fig. 4, where each gridpoint value is a vertical integral of energy components from the MTE formula. At analysis time (not shown), the EFSO and actual impact patterns match very closely, with only minor differences. At 60-min forecast time (Figs. 4a,d), the patterns begin to diverge, though the general locations of large impact values are the same. Convective activity is forming ahead of the dryline in Texas and the cold front in Oklahoma, which manifests as strong areas of small-scale beneficial and detrimental impacts along those areas. Similar small-scale variations are seen in Arkansas, associated with ongoing convection in that area. The effect of upscaling to 20- and 40-km grids allows for a much clearer subjective evaluation of the regions with small-scale variations in Figs. 4a and 4d. For example, the Arkansas and DFW testbed regions had generally large-scale beneficial impacts from observations, while in northeast Oklahoma the impact regions are only at smaller scales. It is also apparent that the EFSO estimates in the bottom row match these large-scale variations of actual impact, with the exception of the region in southeast Texas.

2) OBJECTIVE EVALUATION—PATTERN CORRELATION AND SKILL

Though subjective results show general agreement of EFSO with actual error reduction, cycle-average statistics at different forecast times can summarize the comparison with actual impact for different verification metrics, grid scales, and localization settings. The first is DA cycle-averaged

pattern correlation in Fig. 5 for energy norm verification. On the 2.4-km grid, the correlation starts off high, around 0.8, at $t = 0$ but drops off quickly for each metric. The use of advected localization does improve the correlations at 10–20-min forecast times, but beyond that correlation is weak (below 0.3). However, the 40-km upscaling improves the pattern correlations by around 0.1–0.3. When combining the upscaling with advected localization in EFSOadv0.75, the pattern correlation remains above 0.3 for longer, out to 30–40 forecast minutes for KE and DTE and for the full 2-h forecast in MTE.

Correlations for surface metrics are shown in Fig. 6. In contrast to energy norms, the 2.4-km EFSO of surface variables have a slower drop off in correlation with forecast time. Using a weak correlation threshold of 0.3, the 2.4-km EFSO for surface variables is accurate up to 50 min, even as high as 80 min for surface moisture. Additionally, surface thermodynamic variables match for longer (60–70 min) than either wind component (20–30 min). Further, correlations on the 40-km grid are similarly higher than on the 2.4-km grid. For surface wind, this boosts correlations to above 0.3 out to 60–80 min, and for temperature correlation is above 0.3 out to 90 min (EFSOadv1.5). The best pattern correlation is seen in surface moisture, having moderate correlation (≥ 0.5) for the entire 2-h forecast.

Skill is assessed next to compare the magnitudes of error in EFSO estimates, shown for 40-km comparison in Figs. 7 and 8. For energy norms (Fig. 7), the skill of KE only remains above the baseline climatology for the first 20–30 min of the forecast; however, in DTE and MTE the skill remains higher than baseline by as much as 0.3 for the full forecast duration. This baseline is defined as the SS of using the domain- and cycle-averaged impact as the estimate at all grid points. Although MTE and DTE are similar in skill, MTE has slightly higher skill by the end of the forecast than DTE.

In terms of surface verification metrics (Fig. 8), skill in EFSO is similarly higher than the baseline for all metrics and forecast hours. The skill dropoff is faster in the first 20 min of the forecast compared to the energy norms. However, EFSO

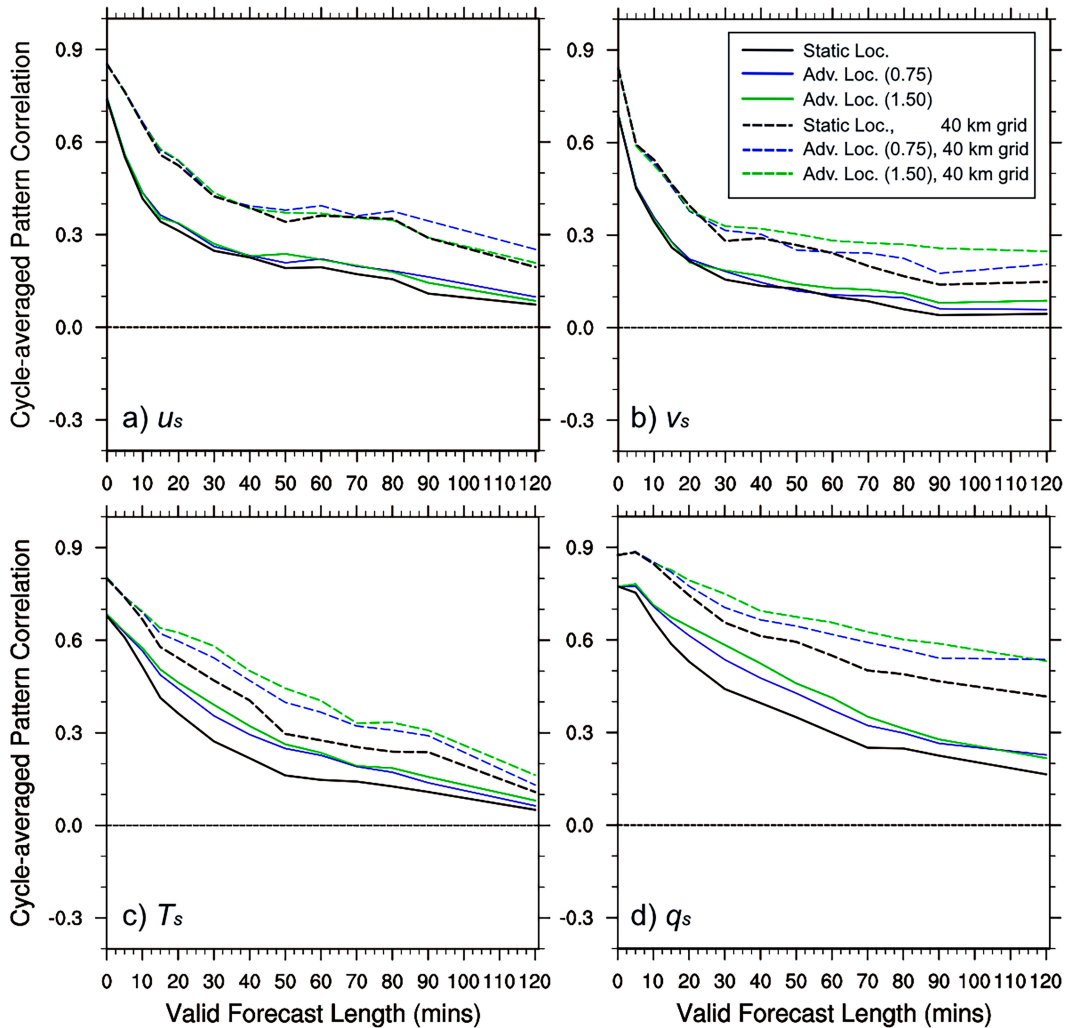


FIG. 6. As in Fig. 5, but for impact of surface verification fields of (a) zonal wind, (b) meridional wind, (c) temperature, and (d) specific humidity.

estimates in T_s and q_s have higher SS than surface wind or energy norm components for longer (≥ 40 min) forecast lengths, ending with the highest skill of 0.4 by the 120-min forecast (EFSOadv1.5).

In summary, pattern correlations on upscaled grid demonstrate a linear relationship out to 30 min for KE and DTE, 60–90 min for u_s , v_s , and T_s , and for the entire 120-min forecast for MTE and q_s , with q_s having the best overall correlation. The skill of EFSO in DTE, MTE, and surface evaluation metrics remains well above the baseline climatology throughout the 2-h forecast. In other words, EFSO is capable of matching actual patterns of impact on 5-min DA cycling for these durations, especially when upscaling is used. These results are consistent with Necker et al. (2018), who found substantial sensitivity of convective-scale EFSO accuracy when applying an upscaling technique within the verification metric up to 10×10 grid points ($28 \text{ km} \times 28 \text{ km}$ in their simulations). More accurate estimates are attained when upscaling

EFSO to at least $\sim 8\Delta x$ to avoid double-penalty issues of convective scales, where large magnitude errors in verification are caused by small displacements (e.g., a storm which is misplaced by 10 km). The results also suggest increased accuracy of EFSO for verifying MTE and surface moisture, which will be further explored in section 4c.

b. Impact of localization choice on EFSO accuracy

In addition to upscaling, the use of advected localization further boosts correlation values by as much as 0.3 when evaluating energy norms (Fig. 5), with slightly better correlations for EFSOadv0.75 compared to EFSOadv1.5 especially at longer forecast lengths. This improvement in correlation of up to 0.3 using a moving localization compared to static localization is similar in scale to the improvement seen in Ota et al. (2013) for EFSO application within the Global Forecast System. SS is similarly boosted by 0.1–0.2 (Fig. 7), with substantial differences as early as 10–20 min into the forecast. EFSOadv1.5 has

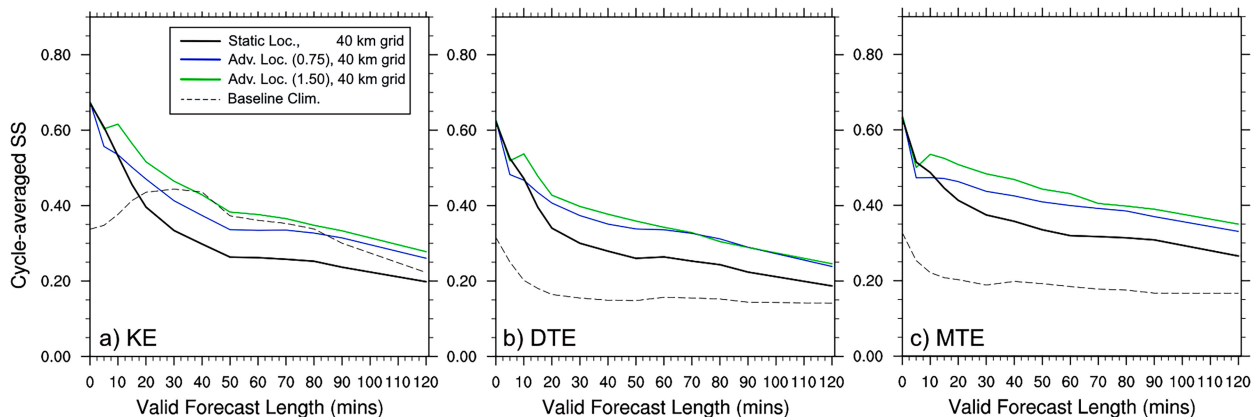


FIG. 7. MAE-based SS of 2D EFSO maps [Eq. (5)] averaged over all 25 DA cycles for EFSOadv (black), EFSOadv0.75 (blue), and EFSOadv1.5 (green). Skill using energy norm verification metrics (a) KE, (b) DTE, and (c) MTE, verified on the 40-km grid. Dashed black line is baseline climatology, the SS of using the domain- and cycle-averaged impact value over all grid points in the domain at each forecast hour.

higher skill than EFSOadv0.75, especially for the 20–60-min forecast period. An optimal coefficient greater than unity could suggest processes other than physical advection are contributing to the impact metric.

For surface verification, it is apparent that EFSOadv1.5 has better matching patterns than EFSOadv0.75 for all surface metrics except zonal wind (Fig. 6). However, the difference (0.05–0.1) between those two is smaller than their differences to EFSOadv or when comparing 2.4- to 40-km correlations. For all surface metrics except u_s , EFSO with advected localization has higher skill (by 0.1) than EFSOadv (Fig. 8). Further, EFSOadv1.5 has consistently better skill than EFSOadv0.75 for forecast lengths over 30 min (by about 0.05).

A subjective evaluation of EFSO for different localizations is shown in Fig. 9, zoomed into the DFW domain on the 20-km upscaled grid. For a shorter (30-min) forecast, it is difficult to see many subjective differences between EFSOadv and EFSOadv1.5; however, the error reduction of EFSOadv1.5 compared to EFSOadv (Fig. 9c) demonstrates that EFSOadv1.5 is more accurate than EFSOadv at most of the locations plotted. On the other hand, it is easier to see the subjective effect of advected localization for a 90-min forecast (bottom row, Fig. 9). For example, the beneficial impact region centered near 33°N, 98°W extends too far westward and does not cover the northernmost part of the actual impact region for EFSOadv (Fig. 9d). Conversely, EFSOadv1.5 properly removes this spurious westward beneficial impact and further extends the region to the northeast to better match the actual impact region (Fig. 9e). The error reduction (Fig. 9f) reflects that these subjective improvements are overall more accurate for EFSOadv1.5 than EFSOadv; however, these are not universal improvements at all location, which is an indication that the optimal coefficient for advected localization may vary by underlying flow. The EFSO estimate may benefit further from an adaptive localization technique that better reflects the time-forecast component for different flow and dynamical regimes. Note that we attempted the regression confidence factor (RCF) technique of Gasperoni and Wang (2015), but were not able to improve upon simple advected localization

(not shown). RCF was able to adaptively capture the time-forecast component adequately; however, worse EFSO estimates were likely due to having no tie with localization shape during DA, which was demonstrated as an important dependence for optimal localization choice in EFSO within previous idealized two-layer model experiments (Gasperoni and Wang 2015).

The results of this section reveals the importance of considering the time-forecast component to EFSO localization. Some previous studies on EFSO ignore the forecast component and apply static localization, arguing that it is unimportant for short forecast lengths (e.g., Sommer and Weismann 2014; Hotta et al. 2017; Necker et al. 2018). However, it is shown here for convective-scale, high-frequency 5-min DA cycling that the forecast component of localization is important for maximizing EFSO skill out to 2 h, and could have noticeable differences on EFSO skill for forecast lengths as short as 20–30 min depending on chosen evaluation metric.

c. Partitioned EFSO impact results by observation variable, platform, and evaluation metric

In this section, the DA cycle-averaged EFSO impact is partitioned by different observation platforms and variables to assess their relative importance on the ensemble DA and forecast system. Here, we will compare DA cycle-averaged partitioned sums for the 30-min impact of EFSOadv1.5, since we can be confident based on results of sections 3a and 3b that these estimates generally compare most favorably with the actual 30-min forecast impact.

1) PARTITIONED EFSO–ENERGY NORMS

Figures 10a, 10c, and 10d shows summed EFSO impacts of energy norms partitioned by observing type in four categories. Note that the well-sited and calibrated Oklahoma and West Texas mesonets are included within conventional surface observations. All sources have a negative summed EFSO, indicating beneficial impacts for all observation types. The largest sum for each energy

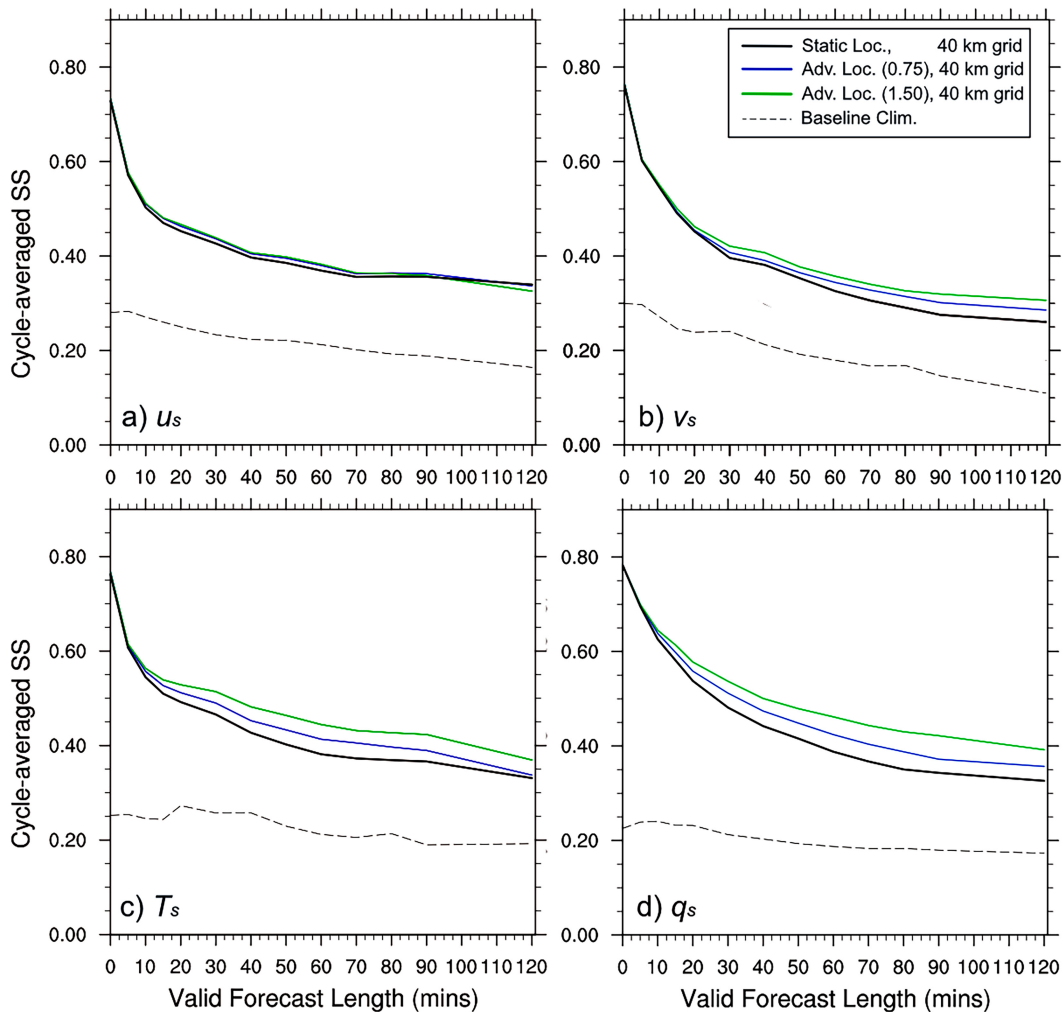


FIG. 8. As in Fig. 7, but for skill of EFSO using surface verification metrics of (a) zonal wind, (b) meridional wind, (c) temperature, and (d) specific humidity.

norm is from radar observations. Conventional surface observations and upper-air observations have roughly equal contributions to KE and DTE, about one-third of the impact for radar. However, for MTE, conventional surface observations contribute twice as much as upper-air observations, in part because there are very few moisture observations from upper-air sources. The nonconventional surface observation category shows the least amount of impact, with minimal contributions to KE and DTE and a small contribution to MTE that is roughly 15% of the size of the EFSO sum of conventional surface observations. Averaging EFSO by observation counts (Figs. 10b,d,f) reveals that radar reflectivity has the least impact per observation on energy norms, which ties its large, summed impact to the volume of observations. This is similar to other FSO works that show that the per-observation impact from remotely sensed satellite radiances is low despite having one of the largest sums (e.g., Ota et al. 2013; Kim et al. 2017). Conventional upper-air observations and surface stations are the most impactful per observation, with surface stations having a much larger influence in MTE than KE and

DTE. Nonconventional surface observations have a substantially higher impact per observation on MTE in particular, just under half of the impact of conventional surface observations. These results suggest a high sensitivity of beneficial forecast evolution in this case to the assimilation of moisture observations, echoing the results of data denial experiments in G18.

When partitioning by observing variable (Fig. 11), we can see the largest beneficial contributions come from radar reflectivity, radar radial velocity, and in situ wind observations to KE and DTE, in similar amounts. However, in terms of MTE, the reflectivity observations have by far the most contribution to impact. This further underscores the importance that the large volume of reflectivity observations have on moisture and related hydrometeors. Among thermodynamic observations, moisture impacts MTE more substantially than temperature. The only observing variable with detrimental impact are surface pressure observations for DTE and MTE, though the magnitude of this impact is small compared to the sum. Casaretto et al. (2023a) also reported on

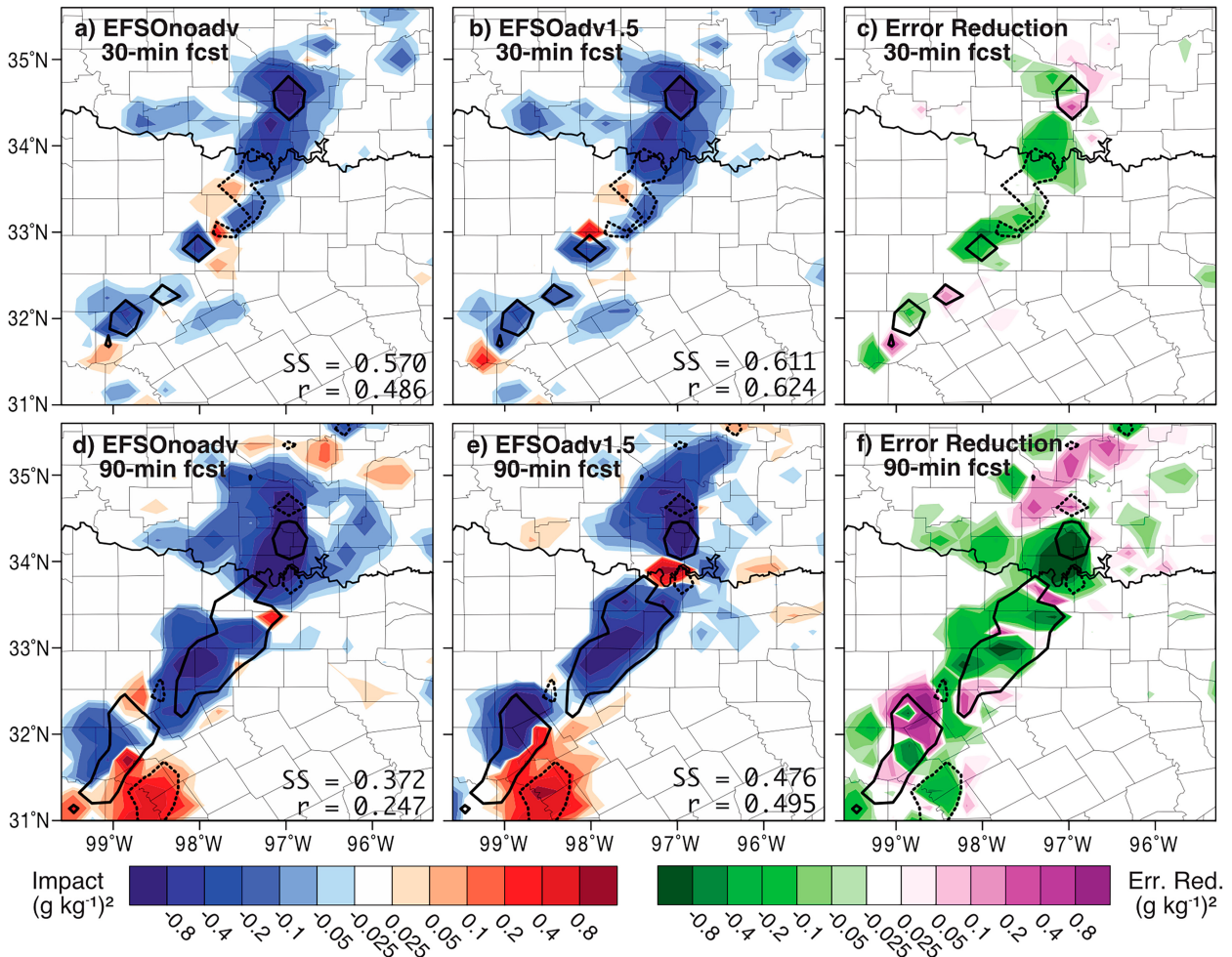


FIG. 9. (a),(d) EFSOnoadv-estimated impact in surface moisture; (b),(e) EFSOadv1.5-estimated impact in surface moisture; and (c),(f) error reduction of advected localization estimate, defined as $|EFSOadv1.5 - Actual| - |EFSOnoadv - Actual|$. Plots are valid for (top) the 30-min forecast impact of 1730 UTC analysis and (bottom) the 90-min forecast impact of 1630 UTC analysis on upscaled 20-km grid. Black solid and dashed contours show actual impact values of -0.25 and $0.25 \text{ g}^2 \text{ kg}^{-2}$, respectively.

average detrimental impacts from surface pressure observations, validated by data denial experiments. Limited impact is seen in precipitable water observations since they were only available for one cycle (1715 UTC) during the experimental DA period.

2) PARTITIONED EFSO–NEAR-SURFACE VERIFICATION FIELDS

Partitioning EFSO for near-surface evaluation metrics reveals some differences compared to energy norms (Fig. 12). While radar and conventional upper-air observations contributed the majority to the total energy norm impact, in terms of surface verification these sums are much smaller. This smaller influence makes sense because the former observations largely influence locations above the surface. Further, there is a detrimental impact for radar observations in terms of u_s , although this is only a small percentage of the sum (about 7%, Fig. 12e).

For surface wind verification, conventional surface observations have the highest impact, together representing over 85% of the sums (Figs. 12e,f). These sources also have the highest contributions to T_s and q_s , but interestingly ASOS observations have the biggest influence on q_s (47%) while the Oklahoma and Texas mesonets affect T_s most strongly (62%). The latter is a result of these observations having better coverage to analyze a frontal zone located across Oklahoma and Texas during the case (see Fig. 3 of G18)

In terms of nonconventional surface observations, each source has a small but beneficial impact, with larger percentage impacts in terms of thermodynamic verification metrics than wind verification metrics. Specifically, when summing up contributions from ERNET, CWOP, and miscellaneous mesonets, there is a near 25% contribution to T_s and q_s (Figs. 12g,h), compared to only 5%–10% for u_s and v_s (Figs. 12e,f). Among individual sources, the miscellaneous mesonet observations have the largest influence on q_s of 13%, while ERNET and CWOP

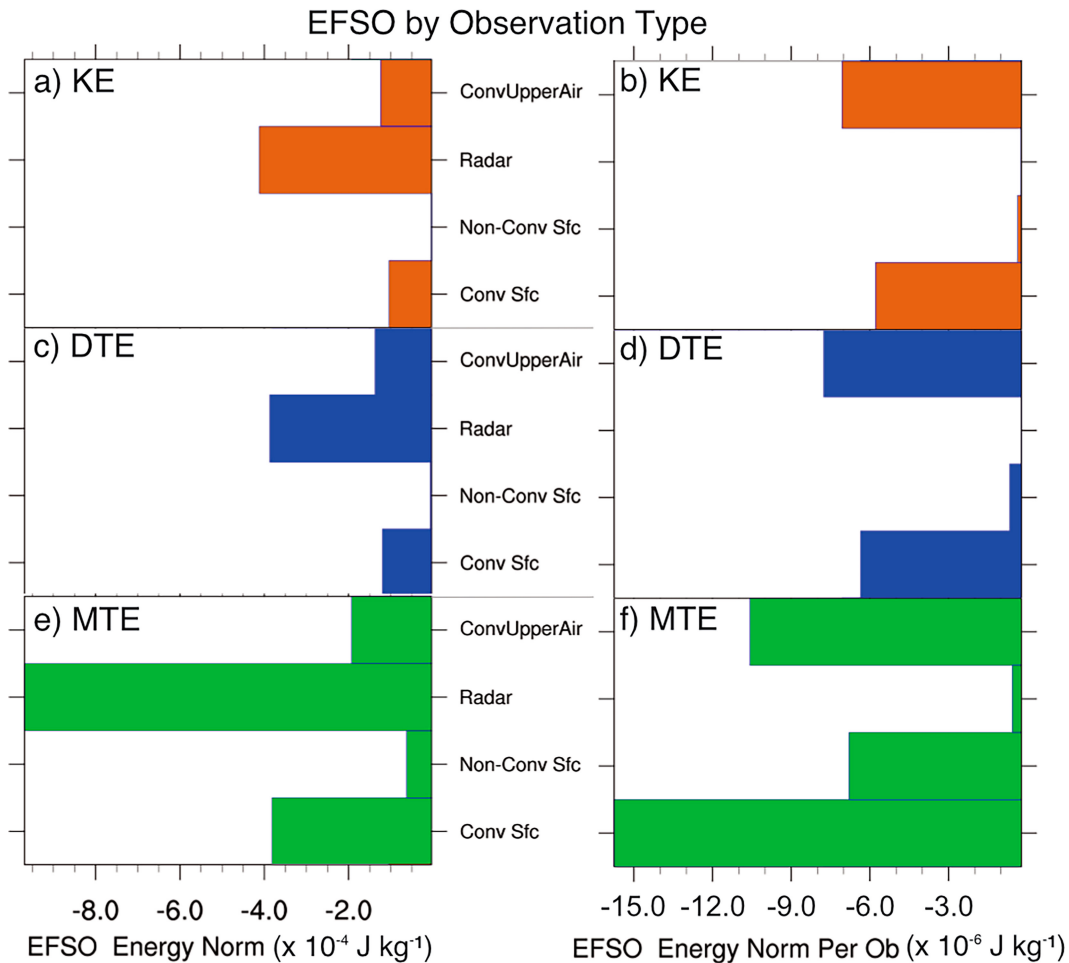


FIG. 10. (left) Summed and (right) per-observation average impact estimate of EFSOadv1.5 from 40-km upscaled grid, partitioned by observation types: conventional upper air, radar, conventional surface, and nonconventional surface. EFSO quantities shown for verification metrics (a),(b) KE; (c),(d) DTE; and (e),(f) MTE. Note the bars of radar EFSO in (b) and (d) are not visible due to their small magnitude ($> -0.01 \times 10^{-6}$).

are roughly equal and around 5%. This large influence from the miscellaneous mesonet may be tied to moisture observations from a hydromet network near the Colorado River (large cluster of black dots in southwest Texas, Fig. 1), which was also noted as having a large influence in data denial impact experiments from G18. In terms of T_s , ERNET observations have the largest influence, near 15%, while miscellaneous mesonet observations have a near 8% influence and CWOP only a small (~1%) percentage.

In the rightmost column of Fig. 13, the EFSO contributions are averaged per observation as well as per area of influence, defined using the localization radius for each type (Table 1). The latter averaging by area of influence is meant to take into account the different scales of observation impact—a single observation with 200-km localization will more likely have larger forecast influence than a single observation with 20-km localization. With this normalization, the observations from miscellaneous mesonets have the largest contributing magnitudes to all surface variables except meridional wind. When

compared to Fig. 1, it appears that these observations do well to fill in the gaps in more data sparse areas, especially in southern Texas in the moisture-rich region near the dryline for this case. The more mixed order of impacts per observation and per square kilometer suggests that part of the larger influence from conventional observations in the summed impacts is due to the larger localization radius applied during DA. It also demonstrates that impacts of nonconventional observations go beyond their larger observation counts and density.

Finally, in terms of observing variables (Fig. 13), the largest influence for each surface metric comes from observations of the same variable (i.e., moisture observations for q_s , temperature observations for T_s , etc.). Although this finding is not surprising, it does provide a good sanity check for EFSO estimates of these new verification metrics. Furthermore, comparing percentage contributions to total EFSO, we see that 80%–90% of verifying u_s and v_s comes from wind observations, while for thermodynamic verification (T_s and q_s),

EFSO Sum by Obs. Variable

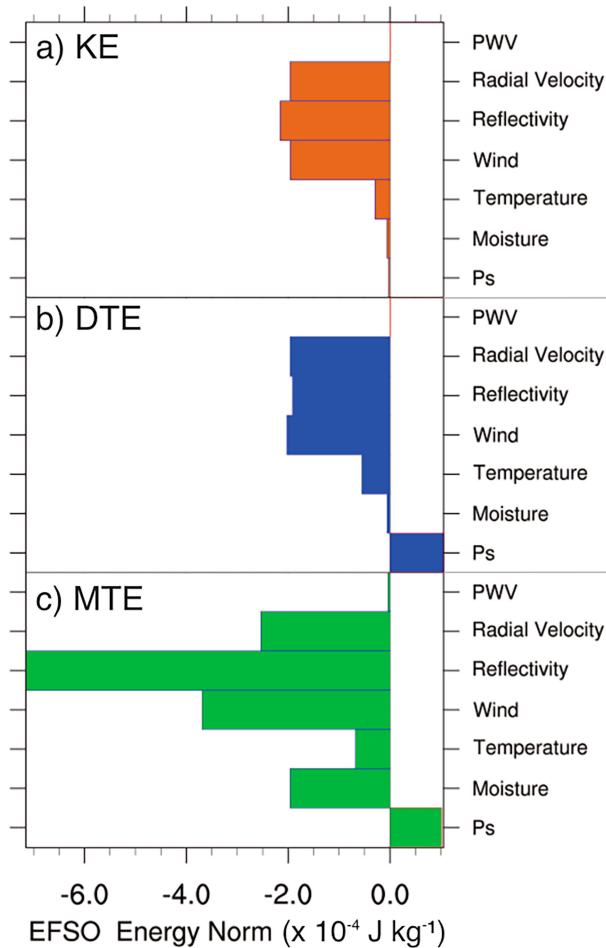


FIG. 11. As in Fig. 10, but for EFSO sums partitioned by observing variable.

smaller percentages (65% and 70%) come from temperature and moisture observations, respectively. The reduced percentage is accompanied by higher percentages from wind observations of 15%–25%. The higher wind observation percentage contribution for thermodynamic verification demonstrates the impact of advection on surface temperature and moisture forecasts, as well as the general enhanced sensitivity to wind near the dryline and cold-frontal boundaries that were prevalent in this case.

The results of partitioned EFSO suggest a reduced impact of nonconventional observations when compared to the impact of nonconventional observations within Casaretto et al. (2023a). Further, they suggest a larger influence from moisture observations than other similar studies (e.g., Casaretto et al. 2023a; Necker et al. 2018). This study covers just one case in a different region than the aforementioned studies – in particular a region where enhanced sensitivity to forecast precipitation is related to the dryline gradient and position, as well as the supply of moisture available south and east of the dryline that feeds the convection. Further, in the case of

Casaretto et al. (2023a) many of the impactful nonconventional observations were located in areas with less density of conventional observations near mountainous terrain. Here, the coverage of conventional stations is more uniform (Fig. 1) and most of the nonconventional stations are located in observation-dense metropolitan areas. Still, we show here that nonconventional surface observations have a relatively small but measurable impact to the forecast, a result that is shared subjectively by G18 for this case.

d. Effect of nonlinearity on EFSO accuracy

Surface zonal wind impact for longer lead times (60 and 120 min, Fig. 14) is used to demonstrate the effect of nonlinearity on the EFSO accuracy. The actual impact field shows extensive small gridscale areas of beneficial and detrimental impact. One such area in the western part of the domain appears to reflect convective rolls and cells found in the boundary layer. However, EFSO is unable to capture such fine-scale details, and instead shows broad areas of beneficial impact. This difference further grows for longer 120-min forecasts. Although upscaling may reduce some of these small-scale differences (e.g., Fig. 4), it cannot fully account for these differences in the resolution of gridscale impacts. These patterns suggest an inability for the tangent linear approximation to account for all nonlinear dynamics as the forecast time increase, this negatively affecting EFSO accuracy.

Model spin up is also a concern for another nonlinear variable, surface pressure, where the correlation is also low and at parts even negative (not shown). In fact, the only two observing variables to show detrimental summed EFSO impacts in this study were surface pressure for DTE and MTE (Figs. 11b,c), and radar reflectivity for u_s (Fig. 13a). A single observation experiment with a pressure observation demonstrated how extreme nonlinearity can negatively affect EFSO accuracy even just 10 min into the forecast (not shown). Complications of surface pressure observations have been noted in Necker et al. 2018, and both Chen and Kalnay (2020) and Casaretto et al. (2023a) found detrimental impacts from surface pressure observations. These examples demonstrate the need to further study EFSO accuracy for variables and verification metrics that are particularly sensitive to nonlinear processes.

5. Summary and discussion

This study explored the application of the ensemble-based forecast sensitivity to observations (EFSO) metric to a convective-scale DA and forecast system. Though a few previous studies (Sommer and Weissmann 2014, 2016; Necker et al. 2018) had applied EFSO to a convective-scale model, this is one of the first such studies to explore for convective-scale, high-frequency, 5-min DA, with the inclusion of nonconventional surface and radar observations. Furthermore, we explored the use of different verification metrics other than energy norms. The need for upscaling of EFSO estimates onto coarser 20- and 40-km grids was also explored, with direct comparisons in maps of EFSO estimate impact against maps of actual error reduction. This upscaling is similar to neighborhood averaging, where each coarse grid box

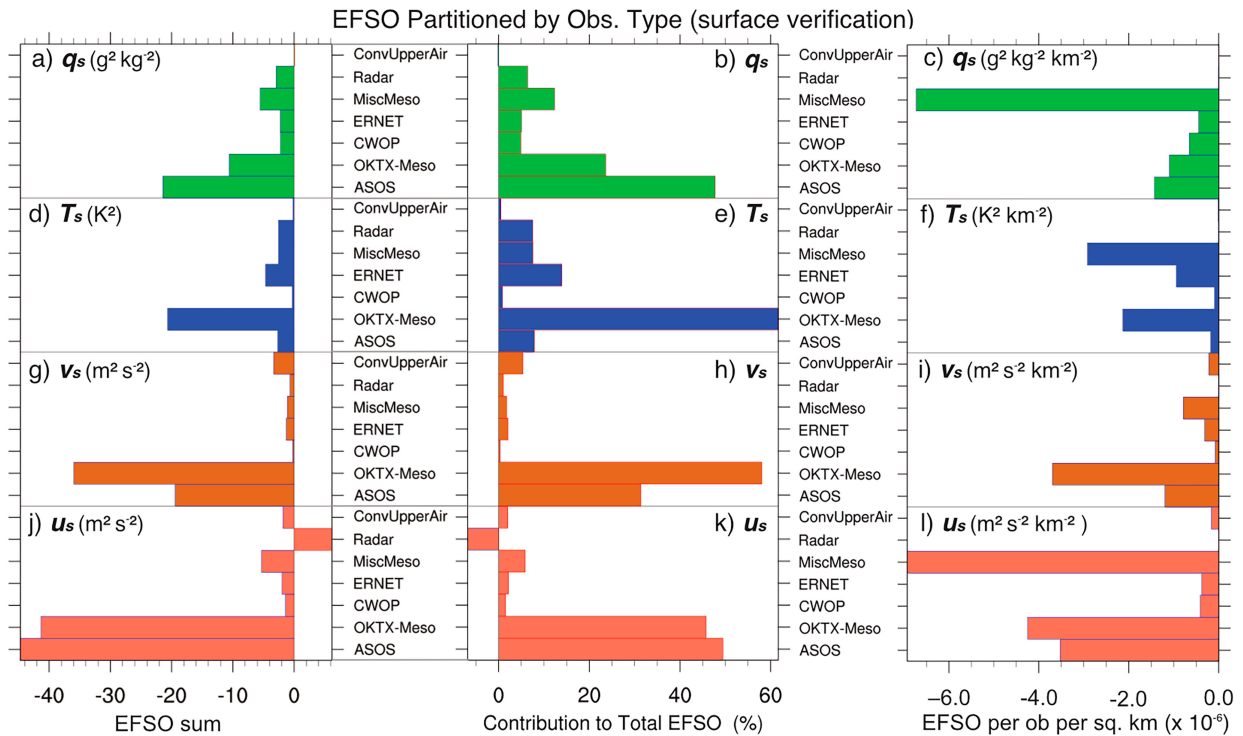


FIG. 12. (left) As in Fig. 10, but for EFSO estimates of surface verification metrics u_s , v_s , T_s , and q_s , with impacts further partitioned by individual surface observing platform types. (center) Percentage contribution of each observing platform to the total EFSO sum of each surface verification metric. (right) EFSO impact averaged by number of observations and area of influence (km^2), defined as circular area using localization scale for each observation type (Table 1).

represents the average of all values of EFSO from the original high-resolution (2.4 km) grid.

The control experiment from the 3 April 2014 CI case study of G18 was used for the EFSO experiments. Each 5-min DA cycle in the 2-h inner grid (1600–1800 UTC) allowed for 25 samples to compare EFSO estimates with the actual impact. Three EFSO estimates were tested with different localization settings for the EFSO estimate: a static localization equivalent to the DA localization (EFSO_{oadv}), and two advected localizations with coefficients of 0.75 and 1.5 (EFSO_{adv0.75} and EFSO_{adv1.5}), respectively. The advected localization method follows Ota et al. (2013) of proportionally moving the DA localization by the mean horizontal analysis and forecast wind, multiplied by the chosen coefficient.

A summary of findings is as follows:

- EFSO is feasible for convective-scale 5-min DA cycling. Cycle-average pattern correlation and MAE-based SS showed that skillful estimates can be attained for the full 2-h forecast period examined in this study. Furthermore, nearly all observation platforms and variable types had beneficial impacts for each verification metric.
- Upscaling and advected localization can substantially improve EFSO estimates. Although the 2.4-km gridscale EFSO had some correlation with the actual error reduction, this was limited to 30–60-min forecasts at best depending on

verification metric. However, upscaling to 20- or 40-km grids boosted correlation values by 0.3 or more and resulted in skillful estimates for the entire 2-h forecast duration. Advected localization could further boost pattern correlation and skill score by around 0.1–0.2, especially for forecast lengths above 30 min.

- Other verification metrics can be used for EFSO estimates. While many previous studies focus on energy norms, more appropriate verification metrics may be useful for examining a convective-scale DA and forecast system. This study demonstrated that surface variables (u_s , v_s , T_s , q_s) can be used as verification metrics for which similar more skillful EFSO estimates are found comparing to the OSE impact. In doing so, we found the best estimates related to moisture, due in part to the enhanced sensitivity of the forecast in this case study to near-surface moisture (as demonstrated in G18).
- All observations except for surface pressure resulted in beneficial impacts on energy norms. The largest contributing platform to energy norm impacts was from radar. This is the first study to apply assimilated radar observations within the EFSO estimate. On a per-observation basis, conventional surface and upper-air data had the largest contributions, but the volume of radar observations caused its sum to 2–3 times as high as the next highest contributing source.
- Nonconventional surface observations had small but important contributions to MTE and surface thermodynamic impacts. In

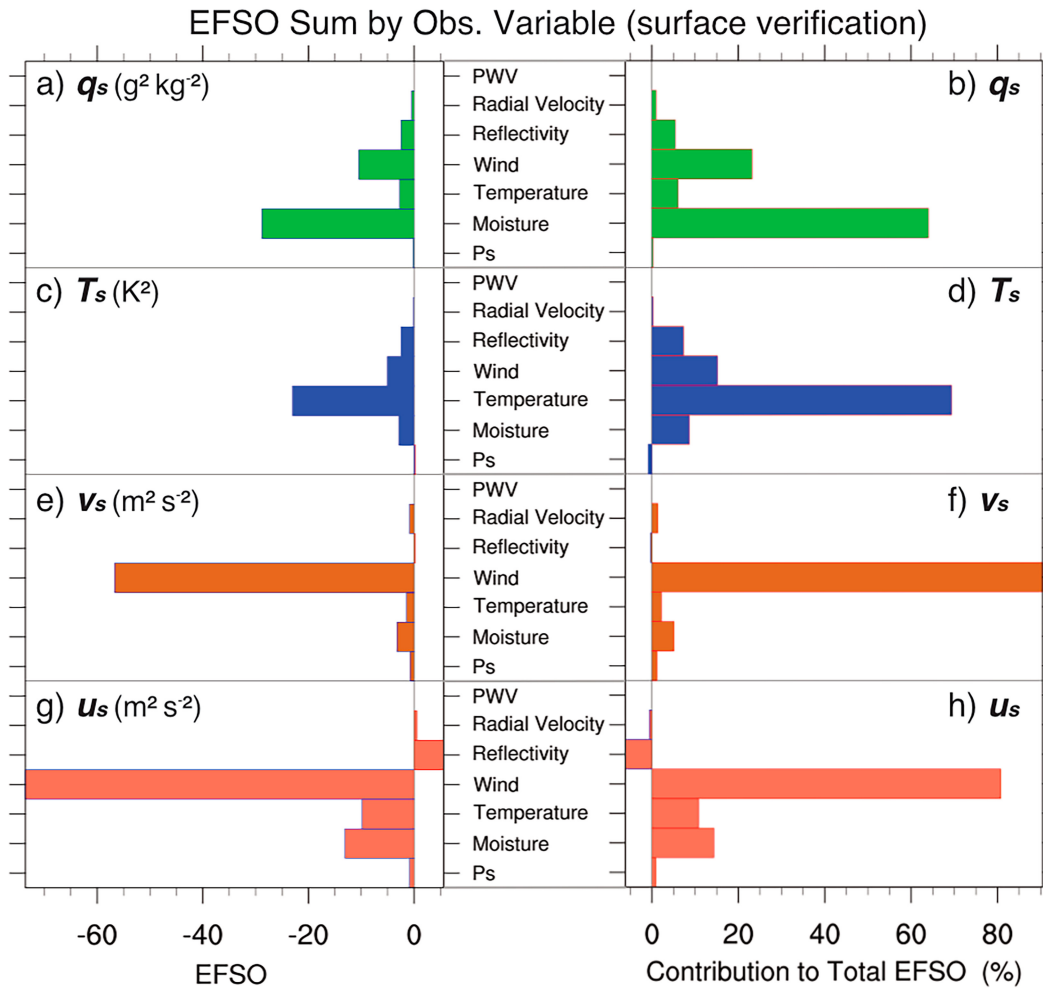


FIG. 13. As in Fig. 12, but for EFSO sums partitioned by observing variable.

particular, the contributions on q_s and T_s represented nearly 25% of the total EFSO sum, which is substantial considering the smaller DA 40-km localization radius of influence for these observations compared to 200 km for ASOS and 80 km for conventional Oklahoma and Texas mesonets. This is in broad agreement with data denial experiment results of G18, who found withholding nonconventional observations led to degradations in ensemble forecasts of storm intensity due to reduced accuracy of small-scale features within the dryline. On the other hand, less than 10% of total EFSO impact in u_s and v_s were from nonconventional sources, which reflects issues with data and station siting quality of nonconventional sources found in previous work. For example, G18 found that denying wind observations from CWOP and ERNET led to small improvements on the ensemble forecast over the control.

Since EFSOadv1.5 had the best overall skill in predicting impact for forecast lengths beyond 30 min in most verification metrics, it can be inferred that further improvement of EFSO can be attained with different localizations. A coefficient greater than 1 suggests that nonlinear processes play a role, and the

spatial variability of EFSO accuracy (e.g., Figs. 4, 9, and 14) further suggests that the optimal localization may be dependent on the underlying flow. An adaptive method may yield better results, but such a method would also need to account for the shape and spatial extent of localization that was applied during DA (Gasperoni and Wang 2015).

One caveat is that, although EFSO was computed over 25 cycles for stable statistics, the statistical significance of results in this study is inherently limited by the use of one case study and cannot be expected to generalize to all situations. The focus of CI along the dryline in the 3 April 2014 case may explain the enhanced sensitivity to moisture for the EFSO method, as also shown in G18. It is also possible that the enhanced EFSO sums and accuracy in terms of moisture may be an indication of biases within the observing system influencing EFSO, similar to Necker et al. (2018) who found largest sums in surface pressure observations were an indication of the presence of biased observations. More case studies are needed to determine the possible influences of biases on the results.

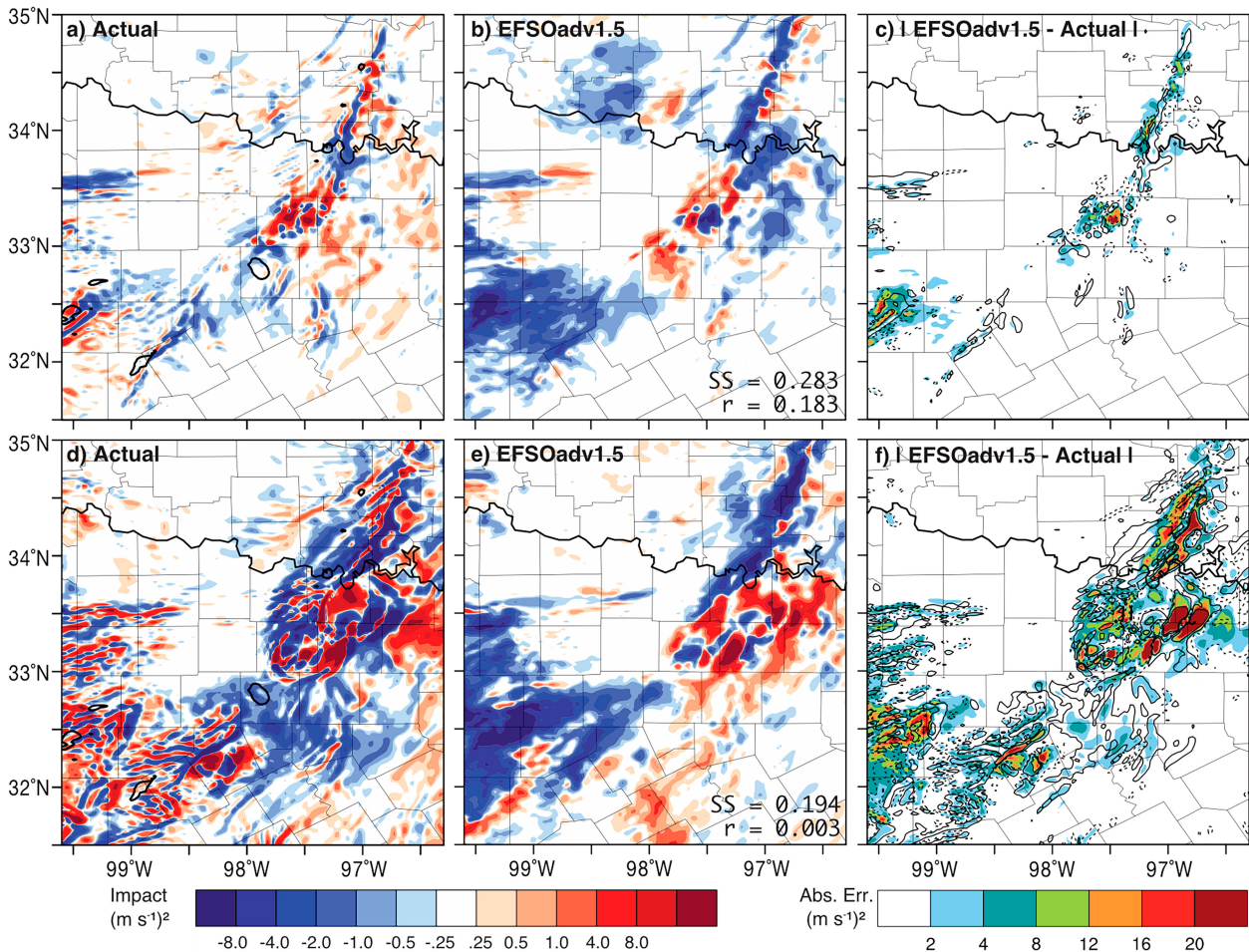


FIG. 14. (a),(d) Forecast gridscale 2.4-km actual impact (blue-to-red color fill); (b),(e) EFSOadv1.5 estimate (blue-to-red color fill); and (c),(f) error magnitude of EFSOadv1.5 estimate (rainbow color fill). (top) 60- and (bottom) 120-min forecast impact of 1800 UTC analysis in surface zonal wind. Black solid and dashed contours in (c) and (f) display actual impact at -1.0 and $1.0 m^2 s^{-2}$, respectively.

The biggest drawback to the application of EFSO to convective scales is the significant increase in nonlinear processes within model integration, which limited ensemble size covariances will not be able to adequately represent. The EFSO can work to identify impactful areas in developing storms, but the highly nonlinear nature of storm evolution limits the accuracy of EFSO in such cases. It may be the case, however, that more predictable scenarios such as linear storm systems ahead of cold fronts and mesoscale convective systems are better suited for EFSO, especially when verifying against other metrics such as precipitation or composite reflectivity. More study is needed in the application of the EFSO method to more diverse cases at convective scales and in identifying how EFSO may be used for more advanced verification metrics. In particular, an idealized study may be necessary to gain a complete understanding of the influences of nonlinearity of EFSO accuracy at convective scales, especially for radar and surface pressure observations. Upscaling prior to the computation of EFSO may further improve the estimates for these more nonlinear observations.

Acknowledgments. This work was supported by NOAA Grant NA19OAR4590138 and NOAA/NWS Grant 1305M220DNWWG0061 from the National Mesonet Program. The authors thank Aaron Johnson and Yongming Wang for providing the GSI-based EnKF assimilation system used for this work. We further thank Yoichiro Ota for providing the EFSO code for the GSI-based EnKF system. The computing for this project was performed at the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma (OU). The authors appreciate the diverse feedback from three anonymous reviewers, which helped to considerably improve the quality of this manuscript.

Data availability statement. Nonconventional surface observations were obtained through the NOAA Meteorological Data Assimilation Ingest System (MADIS). Model data, EFSO output, and observations are archived on tape at OSCER's OU and Regional Research Store (OURRstore), and can be made available upon request to the corresponding author. The specific model used in this study is version 3.7 of the Advanced Research WRF core (WRF-ARW;

Skamarock and Klemp 2008). The EFSO code is the same as used in Ota et al. (2013), but modified to interface with WRF Model output. The source code and system configurations used in this work can also be made available upon request to the corresponding author.

REFERENCES

- Ancell, B., and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, <https://doi.org/10.1175/2007MWR1904.1>.
- Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, <https://doi.org/10.1038/nature14956>.
- Benjamin, S. G., B. D. Jamison, W. R. Moninger, S. R. Sahn, B. E. Schwartz, and T. W. Schlatter, 2010: Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR, and mesonet observations via the RUC hourly assimilation cycle. *Mon. Wea. Rev.*, **138**, 1319–1343, <https://doi.org/10.1175/2009MWR3097.1>.
- Buehner, M., P. Du, and J. Bédard, 2018: A new approach for estimating the observation impact in ensemble-variational data assimilation. *Mon. Wea. Rev.*, **146**, 447–465, <https://doi.org/10.1175/MWR-D-17-0252.1>.
- Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135**, 239–250, <https://doi.org/10.1002/qj.366>.
- Carlaw, L. B., J. A. Brotzge, and F. H. Carr, 2015: Investigating the impacts of assimilating surface observations on high-resolution forecasts of the 15 May 2013 tornado event. *Electron. J. Severe Storms Meteor.*, **10** (2), <https://doi.org/10.55599/ejssm.v10i2.59>.
- Carley, J. R., and Coauthors, 2021: Status of NOAA's next generation convection-allowing ensemble: The Rapid Refresh forecast system. *Special Symp. on Global Mesoscale Models*, online, Amer. Meteor. Soc., 12.8, <https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/378383>.
- Casaretto, G., M. E. Dillon, Y. G. Skabar, J. J. Ruiz, and M. Sacco, 2023a: Ensemble Forecast Sensitivity to Observations Impact (EFSOI) applied to a regional data assimilation system over South-Eastern South America. *Atmos. Res.*, **295**, 106996, <https://doi.org/10.1016/j.atmosres.2023.106996>.
- , —, —, —, P. Maldonado, and M. Sacco, 2023b: Ensemble Forecast Sensitivity to Observations Impact (EFSOI) of a high impact weather event using a convection permitting data assimilation system. *EGU General Assembly 2023*, Vienna, Austria, European Geosciences Union, Abstract EGU23-415, <https://doi.org/10.5194/egusphere-egu23-415>.
- Chen, T.-C., and E. Kalnay, 2020: Proactive quality control: Observing system experiments using the NCEP global forecast system. *Mon. Wea. Rev.*, **148**, 3911–3931, <https://doi.org/10.1175/MWR-D-20-0001.1>.
- 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>.
- Ehrendorfer, M., R. M. Errico, and K. D. Raeder, 1999: Singular-vector perturbation growth in a primitive equation model with moist physics. *J. Atmos. Sci.*, **56**, 1627–1648, [https://doi.org/10.1175/1520-0469\(1999\)056<1627:SVPGIA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<1627:SVPGIA>2.0.CO;2).
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, <https://doi.org/10.1029/94JC00572>.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, <https://doi.org/10.1002/qj.49712555417>.
- Gasperoni, N. A., and X. Wang, 2015: Adaptive localization for the ensemble-based observation impact estimate using regression confidence factors. *Mon. Wea. Rev.*, **143**, 1981–2000, <https://doi.org/10.1175/MWR-D-14-00272.1>.
- , —, K. A. Brewster, and F. H. Carr, 2018: Assessing impacts of the high-frequency assimilation of surface observations for the forecast of convection initiation on 3 April 2014 within the Dallas–Fort Worth test bed. *Mon. Wea. Rev.*, **146**, 3845–3872, <https://doi.org/10.1175/MWR-D-18-0177.1>.
- Gelaro, R., and Y. Zhu, 2009: Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179–193, <https://doi.org/10.1111/j.1600-0870.2008.00388.x>.
- , R. H. Langland, S. Pellerin, and R. Todling, 2010: The THORPEX Observation Impact Intercomparison Experiment. *Mon. Wea. Rev.*, **138**, 4009–4025, <https://doi.org/10.1175/2010MWR3393.1>.
- Griewank, P. J., M. Weissmann, T. Necker, T. Nomokonova, and U. Löhnert, 2023: Ensemble-based estimates of the impact of potential observations. *Quart. J. Roy. Meteor. Soc.*, **149**, 1546–1571, <https://doi.org/10.1002/qj.4464>.
- Heppner, P., 2013: Building a national network of mobile platforms for weather detection. *29th Conf. on Environmental Information Processing Technologies*, Austin, TX, Amer. Meteor. Soc., 5.8, <https://ams.confex.com/ams/93Annual/webprogram/Paper222600.html>.
- Hotta, D., T.-C. Chen, E. Kalnay, Y. Ota, and T. Miyoshi, 2017: Proactive QC: A fully flow-dependent quality control scheme based on EFSO. *Mon. Wea. Rev.*, **145**, 3331–3354, <https://doi.org/10.1175/MWR-D-16-0290.1>.
- James, E. P., and S. G. Benjamin, 2017: Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble-variational data assimilation. *Mon. Wea. Rev.*, **145**, 2897–2918, <https://doi.org/10.1175/MWR-D-16-0398.1>.
- 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., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus*, **64A**, 18462, <https://doi.org/10.3402/tellusa.v64i0.18462>.
- Kim, M., H. M. Kim, J. Kim, S.-M. Kim, C. Velden, and B. Hoover, 2017: Effect of enhanced satellite-derived atmospheric motion vectors on numerical weather prediction in East Asia using an adjoint-based observation impact method. *Weather Forecasting*, **32**, 579–594, <https://doi.org/10.1175/WAF-D-16-0061.1>.
- Kotsuki, S., K. Kurosawa, and T. Miyoshi, 2019: On the properties of ensemble forecast sensitivity to observations. *Quart. J. Roy. Meteor. Soc.*, **145**, 1897–1914, <https://doi.org/10.1002/qj.3534>.
- Kunii, M., T. Miyoshi, and E. Kalnay, 2012: Estimating the impact of real observations in regional numerical weather prediction using an ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 1975–1987, <https://doi.org/10.1175/MWR-D-11-00205.1>.

- Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201, <https://doi.org/10.3402/tellusa.v56i3.14413>.
- 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>.
- Li, H., J. Liu, and E. Kalnay, 2010: Correction of ‘Estimating observation impact without adjoint model in an ensemble Kalman filter.’ *Quart. J. Roy. Meteor. Soc.*, **136**, 1652–1654, <https://doi.org/10.1002/qj.658>.
- Liu, J., and E. Kalnay, 2008: Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **134**, 1327–1335, <https://doi.org/10.1002/qj.280>.
- Lorenc, A. C., and R. T. Marriott, 2014: Forecast sensitivity to observations in the Met Office global numerical weather prediction system. *Quart. J. Roy. Meteor. Soc.*, **140**, 209–224, <https://doi.org/10.1002/qj.2122>.
- McPherson, R. A., and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **24**, 301–321, <https://doi.org/10.1175/JTECH1976.1>.
- Morris, M. T., K. A. Brewster, and F. H. Carr, 2021: Assessing the impact of non-conventional radar and surface observations on high-resolution analyses and forecasts of a severe hail storm. *Electron. J. Severe Storms Meteor.*, **16** (1), <https://ejssm.org/archives/2021/vol16-1-2021>.
- National Research Council, 2009: *Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks*. National Academies Press, 250 pp.
- , 2012: *Urban Meteorology: Forecasting, Monitoring, and Meeting Users’ Needs*. National Academies Press, 190 pp.
- NCDC, 2014: *Storm Data*. Vol. 56, No. 4, 652 pp.
- Necker, T., M. Weissmann, and M. Sommer, 2018: The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quart. J. Roy. Meteor. Soc.*, **144**, 1667–1680, <https://doi.org/10.1002/qj.3390>.
- Ota, Y., J. C. Derber, E. Kalnay, and T. Miyoshi, 2013: Ensemble-based observation impact estimates using the NCEP GFS. *Tellus*, **65A**, 20038, <https://doi.org/10.3402/tellusa.v65i0.20038>.
- 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>.
- Schroeder, J. L., W. S. Burgett, K. B. Haynie, I. Sonmez, G. D. Skwira, A. L. Doggett, and J. W. Lipe, 2005: The West Texas mesonet: A technical overview. *J. Atmos. Oceanic Technol.*, **22**, 211–222, <https://doi.org/10.1175/JTECH-1690.1>.
- 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 J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, <https://doi.org/10.1016/j.jcp.2007.01.037>.
- 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>.
- Sommer, M., and M. Weissmann, 2014: Observation impact in a convective-scale localized ensemble transform Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **140**, 2672–2679, <https://doi.org/10.1002/qj.2343>.
- , and —, 2016: Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus*, **68A**, 27885, <https://doi.org/10.3402/tellusa.v68.27885>.
- Weissmann, M., R. H. Langland, C. Cardinali, P. M. Pauley, and S. Rahm, 2012: Influence of airborne Doppler wind lidar profiles near Typhoon Sinlaku on ECMWF and NOGAPS forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 118–130, <https://doi.org/10.1002/qj.896>.
- Zapotocny, T. H., W. P. Menzel, J. P. Nelson III, and J. A. Jung, 2002: An impact study of five remotely sensed and five in situ data types in the eta data assimilation system. *Wea. Forecasting*, **17**, 263–285, [https://doi.org/10.1175/1520-0434\(2002\)017<0263:AISOFR>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0263:AISOFR>2.0.CO;2).
- , J. A. Jung, J. F. L. Marshall, and R. E. Treadon, 2007: A two-season impact study of satellite and in situ data in the NCEP global data assimilation system. *Wea. Forecasting*, **22**, 887–909, <https://doi.org/10.1175/WAF1025.1>.