Factors Influencing Ensemble Sensitivity–Based Targeted Observing Predictions at Convection-Allowing Resolutions

AARON J. HILL, a,b CHRISTOPHER C. WEISS, a and BRIAN C. ANCELL a

a Department of Geosciences, Texas Tech University, Lubbock, Texas

(Manuscript received 21 January 2020, in final form 31 August 2020)

ABSTRACT: Ensemble sensitivity analysis (ESA) is applied to select types of observations, in various locations and in advance of forecast convection, to systematically evaluate the effectiveness of ESA-based observation targeting for 10 convection forecasts. To facilitate the analysis, observing system simulation experiments and perfect models are utilized to generate synthetic targeted observations of temperature and pressure for future assimilation with an ensemble prediction system. Various observation assimilation experiments are carried out to assess the impacts of nonlinearity, covariance localization, and numerical noise on ESA-based observation-impact predictions. It is discovered that localization applied during data assimilation restricts targeted-observation increments onto the forecast responses of composite reflectivity and 3-hourly accumulated precipitation, making impact predictions poor. In addition, numerical noise introduced by nonlinear perturbation evolution tends to reduce the correlations between observed and predicted impacts; small, random-perturbation experiments often yielded similar impacts on forecasts as targeted observations. Nonlinearity also manifests in the observation impacts when comparing targeted observations with nontargeted, randomly chosen observations; random observations have seemingly the same impact on forecasts as targeted observations. The results, under idealized conditions and simplified ensemble configurations, demonstrate that ESA-based targeting for nonlinear convection forecasts may be most applicable at short time scales. Important implications for ESA-based targeting methods employed with real-time ensemble systems are also discussed.

KEYWORDS: Sensitivity studies; Ensembles; Data assimilation; Numerical weather prediction/forecasting

1. Introduction

Since the 1940s, targeted observing has been employed to improve numerical weather prediction (NWP) forecasts, particularly for weather phenomena that NWP models struggle to forecast accurately and for which errors may grow quickly. Over the last few decades, sophisticated algorithms have been developed and tested that objectively target new observations, such as adjoint sensitivity (e.g., Errico 1997; Bergot 1999) and singular vectors (e.g., Palmer et al. 1998; Buizza and Montani 1999). Adjoint sensitivity linearly relates changes in forecast states to initial perturbations by utilizing an adjoint model (LeDimet and Talagrand 1986), which mandates differentiable dynamics and linearization of the nonlinear model dynamics. Adjoint sensitivity has been applied extensively for predictability studies of tropical cyclones (e.g., Doyle et al. 2012), extratropical cyclones (e.g., Ancell and Hakim 2007; Doyle et al. 2014, 2019), and atmospheric rivers (e.g., Reynolds et al. 2019). Singular vectors describe the fastest growing initial perturbations relative to a forecast metric (e.g., Hoskins and Coutinho 2005; Kim and Jung 2009) and can be computed from a decomposition of the tangent linear model (Diaconescu and Laprise 2012).

These two methods intelligently account for forecast-error growth and uncertainties in the forecast and analysis, which make them ideal to determine where new observations should be sampled to reduce forecast and analysis errors. However, they are inherently limited to primarily linear dynamic regimes; only recently were moist dynamics included within an adjoint model (e.g., Doyle et al. 2014, 2019). In general, these aforementioned techniques and others have largely been applied in targeting programs for large-scale systems in which linearity is more valid, including extratropical (e.g., Joly et al. 1997, 1999; Langland et al. 1999; Szunyogh and Toth 2002; Langland 2005) and tropical cyclones (e.g., Aberson 2003; Reynolds et al. 2010). Langland (2005) summarized the results from a variety of targeting field programs and noted average short-range forecast error reductions of 10%, with maximum reductions in error as much as 50%, and improved forecast skill in 70% of cases.

Exploration of targeting methods for smaller-scale forecasts (e.g., isolated deep convection) has been less extensive. The lack of exploration is particularly important given the large economic impact from severe storms and associated hazards on society, with insured losses exceeding $14 billion (U.S. dollars) in recent years (Insurance Information Institute 2019). Majumdar (2016) reviewed recent targeting campaigns and additionally noted the lack of focus on mesoscale phenomena, suggesting targeted observing for high-impact phenomena should be the focus of future targeting programs. However, neither adjoint sensitivity nor singular vectors are particularly useful for high-impact applications when models approach convection-allowing resolutions (i.e., <4 km) and linear approximations over long time windows are strained (Ancell and Mass 2006). Ensemble sensitivity analysis (ESA; Ancell and Hakim 2007; Hakim and Torn 2008; Torn and Hakim 2008) is a targeting method easily applied to an ensemble of forecasts,
making it appealing for high-impact weather applications. ESA dynamically and statistically relates a forecast metric (e.g., area-averaged sea level pressure) to initial-condition perturbations, estimating how changes to prior states may influence a forecast. ESA has been utilized primarily to study dynamic forecast sensitivities of midlatitude cyclones (e.g., Torn and Hakim 2009; Garcies and Homar 2009, 2010; Chang et al. 2013; Zheng et al. 2013; McMurdie and Ancell 2014; Ancell 2016; Berman and Torn 2019) and tropical cyclones (e.g., Torn 2010; Torn and Cook 2013; Xie et al. 2013; Brown and Hakim 2015), but more recently has been applied to forecasts of convection (e.g., Bednarczyk and Ancell 2015; Torn and Romine 2015; Hill et al. 2016; Berman et al. 2017; Torn et al. 2017; Kerr et al. 2019; Kerr and Wang 2020) and wind (e.g., Zack et al. 2010a,b,c; Wile et al. 2015; Smith and Ancell 2017) with high-resolution forecast systems. ESA is capable of highlighting identifiable features (e.g., antecedent precipitation, midtropospheric troughs) that are dynamically related to forecast metrics (Hill et al. 2016). Coupling sensitivity with uncertainty estimates from an ensemble of forecasts, ESA can also predict how ensemble distributions will change from assimilated, targeted observations.

The ESA-based targeting method, derived from ensemble Kalman filter theory (EnKF; Kalman 1960; Evensen 1994), takes into account uncertainty of an ensemble forecast system analysis, targeted-observation error variance, perturbation evolution (i.e., sensitivity), and the data assimilation system that will be used to assimilate targeted observations (Ansell and Hakim 2007), all critical components to effective adaptive observing (Berliner et al. 1999; Langland 2005). Therefore, ESA-based targeting is capable of relating initial-condition uncertainty and potential error growth to a specific forecast response related to high-impact hazards (e.g., hail, precipitation, severe wind). Majumdar (2016) suggested that other targeting algorithms do not take into account data assimilation characteristics, lack error-variance information from the targeted observation, and do not estimate how assimilation increments will evolve with time; ESA takes into account all three factors.

ESA-based targeting procedures have been applied to tropical cyclone (e.g., Xie et al. 2013; Torn 2014) and extratropical cyclone (e.g., Ancell and Hakim 2007; Torn and Hakim 2008) forecasts. ESA-based targeting was retrospectively examined for extratropical cyclones in the Pacific Northwest by Torn and Hakim (2008), who found that for large-scale synoptic systems, ESA-based predictions of changes to forecast sea level pressure from a single assimilated buoy observation were accurate. Torn (2014) noted that for tropical cyclone applications, targeted dropsonde observations tended to improve forecasts relative to nontargeted observations. More recently, ESA was used to target observations for convection forecasts in the Great Plains of the United States (e.g., Torn and Romine 2015; Romine et al. 2016; Coniglio et al. 2016) during the Meso Scale Predictability Experiment (MPEX; Weisman et al. 2015). In contrast to the synoptic-based evaluation, Romine et al. (2016) determined that a number of 3-h accumulated precipitation forecasts during intensive observing periods from MPEX verified worse when targeted dropsonde observations were included in the assimilation process.

Various factors have been hypothesized that could influence targeted observation impacts and associated targeting predictions for all scales of flow, including nonlinear forecast evolution (Xie et al. 2013), localizing observations during assimilation and inflating ensemble distributions to reduce underdispersion (Hill et al. 2018), sampling error from limited ensemble members (Torn and Hakim 2008; Romine et al. 2016; Coniglio et al. 2019), insufficient sampling of dynamic target regions (Bergot 2001; Majumdar 2016), and suboptimal data assimilation systems (Romine et al. 2016). For example, the localization of observations—critical to reduce noise during assimilation—is not directly accounted for in the ESA-based method, thus its relative impact is not estimated prior to sampling and assimilation. Given the mixed performance of ESA-based targeting for convection forecasts in MPEX (Romine et al. 2016), and good algorithmic performance at larger scales (e.g., Torn and Hakim 2008), the efficacy of ESA-based targeting at convection-allowing resolutions warrants additional investigation.

The primary goal of this work is to examine ESA-based targeting for convection-allowing forecasts in an effort to understand what elements of the forecasts and assimilation procedure contribute to targeting success or failure. ESA-based targeting would accurately estimate observation impacts if applied to a linear model, with no localization of observations during EnKF data assimilation, and no model error nor numerical noise. Real-time convection-allowing NWP models are highly nonlinear, and model error and noise impacts are not easily quantified in real-time targeting applications. However, Observing System Simulation Experiments (OSSE) provide a framework to account for and remove model error; further discussion of OSSEs used in this work is provided in section 2. Various forecast and assimilation configurations to examine ESA-based targeting are also detailed in section 2. Section 3 evaluates ESA-based predictions after targeted observations are assimilated. Influences of numerical noise and nonlinearity are characterized in section 4. Section 5 offers a summary of the experiments and discussion of convective-scale targeted observing and implications for future applications.

2. Methodology

a. Ensemble sensitivity analysis–based targeting

ESA predicts optimal locations of targeted observations and their impact on forecasts by statistically linking forecast distributions to perturbations in the model state at earlier times. First, a scalar forecast metric \( R_f \) at time \( t \) is regressed back to the model state \( x_p \) at time \( p \) to describe how \( R_f \) will change due to a perturbation in \( x_p \) (i.e., sensitivity). Sensitivity is easily calculated given an ensemble forecast distribution and vector samples of \( R_f \) and \( x_p \):

\[
\frac{\partial R_f}{\partial x_p} = \frac{\text{covariance}(R_f, x_p)}{\text{variance}(x_p)} \tag{1}
\]

Sensitivity is simply the slope of the linear regression between \( R_f \) and \( x_p \).

ESA-based targeting estimates are derived by considering how an update to the analysis error-covariance matrix from a
These 10 cases are subjectively chosen from the 2011–13 spring season to evaluate ESA at convection-allowing resolutions. Based targeting algorithm estimates the change in variance of present by Ancell and Hakim (2007). Simply, the ESA-based targeting algorithm estimates the change in variance of present by Ancell and Hakim (2007).

\[ \delta \sigma^2 = \frac{\text{variance}(x_p) \times \text{sensitivity}(R_f, x_p)}{\text{variance}(x_p) + \text{variance}(\text{ob})} \]  

where “sensitivity” represents ensemble sensitivity as in (1). The only input value to (3), besides an ensemble of forecast output, is an estimate of targeted observation error variance \([\text{variance}(\text{ob})]\), typically prescribed from instrument error characteristics. Note that (3) is negative definite; mathematically, any new observation assimilated will always reduce forecast metric variance. The estimated variance reduction from a new observation is trivial to calculate over a gridded domain from an ensemble of forecasts. ESA-based estimates of forecast metric-variance change rely on the linear evolution of initial perturbations, and in particular how those perturbations evolve with the tangent linear model (Ancell and Hakim 2007). It should be expected this particular linear dependence will be challenged in the proposed convection-allowing forecast experiments and analysis described below.

Also note that collinearity can be introduced between variance-reduction estimates when independent calculations of (3) are made for different observation types. Collinearity is particularly problematic when observations are considered independently at the same point in space and time, which results in positive or negative results (good or bad predictions of response impacts, for example) for both variables that are a direct result of the two variables covarying with one another. Collinearity becomes less of a factor (but is not absent) when observations are selected independently in different locations, as is done in this work; more details with regard to observation selection are presented later.

### b. Cases

Ten dryline convection events in the southern plains are selected to evaluate ESA at convection-allowing resolutions. These 10 cases are subjectively chosen from the 2011–13 spring seasons (i.e., March–June). Each case produces convection within the southern Great Plains, which focuses the results and analyses on Texas and adjacent states. Each case is listed in Table 1, with corresponding date, location of primary convection, and approximate time of initiation, as determined by simulations discussed later in this section.

The cases are selected on the basis of cold-started simulations (i.e., no data assimilation) from 1 March to 30 June of each year, carried out with the Advanced Research core of the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008). Deterministic simulations were performed twice daily at 0000 and 1200 UTC with 4-km grid spacing from National Centers for Environmental Prediction Global Forecast System (GFS) initial conditions. Forecast simulated composite reflectivity and 2-m dewpoint temperature from each WRF simulation were visually scrutinized to select cases that produced convection along the dryline in the southern Great Plains. Dryline convection is exclusively chosen due to its relationship with the frequency of deep convection in this geographical region, but the described methodology can be applied to any convective regime. Forecasts of 2-m temperature were also inspected to verify that simulated convection was not initiated by a cold front (i.e., synoptically forced ascent).

### c. Observing system simulation experiments

To examine individual aspects that influence ESA-based targeting performance, an observing system simulation experiment (OSSE; Arnold and Dey 1986; Atlas 1997) configuration is used. OSSEs are valuable tools that have traditionally been used in the analysis of new observing systems before their deployment (e.g., Yussouf and Stensrud 2010), new data assimilation techniques (e.g., Liu et al. 2009; Sobash and Stensrud 2013; Sobash et al. 2015; Gao et al. 2016), observing network design (e.g., Khare and Anderson 2006), and observation-impact studies (e.g., Yussouf and Stensrud 2010; Lange and Craig 2014; Madaus and Hakim 2017). A perfect-model OSSE configuration is used herein, which synthetically generates observations from a model simulation rather than meteorological instrumentation (e.g., Zhang et al. 2004, 2006), and assimilates those observations with an ensemble prediction system to generate forecasts; the same dynamical model is used for both modeling systems. While the perfect model configuration often

<table>
<thead>
<tr>
<th>Case</th>
<th>Initiation time (UTC) and date</th>
<th>Initiating location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>2200 UTC 22 May 2011</td>
<td>Central Texas</td>
</tr>
<tr>
<td>Case 2</td>
<td>2300 UTC 24 May 2011</td>
<td>Central Oklahoma and North Texas</td>
</tr>
<tr>
<td>Case 3</td>
<td>1100 UTC 19 Jun 2011</td>
<td>West Texas</td>
</tr>
<tr>
<td>Case 4</td>
<td>0300 UTC 4 Apr 2012</td>
<td>Central Texas</td>
</tr>
<tr>
<td>Case 5</td>
<td>0100 UTC 1 May 2012</td>
<td>West Texas</td>
</tr>
<tr>
<td>Case 6</td>
<td>0100 UTC 5 May 2012</td>
<td>West Texas</td>
</tr>
<tr>
<td>Case 7</td>
<td>0000 UTC 10 Apr 2013</td>
<td>North Texas</td>
</tr>
<tr>
<td>Case 8</td>
<td>2200 UTC 15 May 2013</td>
<td>Central Texas</td>
</tr>
<tr>
<td>Case 9</td>
<td>2300 UTC 20 May 2013</td>
<td>Central Oklahoma and North Texas</td>
</tr>
<tr>
<td>Case 10</td>
<td>0100 UTC 29 May 2013</td>
<td>West Texas</td>
</tr>
</tbody>
</table>
leads to overly positive results (Masutani et al. 2010), it also removes the influence of model error on the forecast; observations generated from a model simulation will sample the model-error characteristics that are identical to the forecast model used for data assimilation and extended forecasts. It is advantageous for the proposed experiments to exploit this methodology and generate perturbed forecasts from targeted observations that are free of model error in order to isolate other system aspects that play a role in the impacts of targeted observations (e.g., localization) on convection forecasts.

The ensemble forecast system forms the basis for all experiment forecasts and targeted-observation evaluations; hereinafter, this ensemble is referred to as the control. The control ensemble uses version 3.8.1 of the WRF Model during data assimilation and the free-forecast period to advance the control ensemble members. In total, 50 ensemble members are generated through random perturbations to carry out cycled data assimilation and execute free forecasts. Three one-way nested domains are used with 27-, 9-, and 3-km horizontal grid spacing (Ens d01, d02, and d03 in Fig. 1) and 38 vertical levels; the 3-km domain is centered over the southern Great Plains (Ens d03 in Fig. 1). Initial and boundary conditions from the GFS are perturbed with the WRF three-dimensional variational data assimilation (Barker et al. 2004) and fixed-covariance boundary perturbation technique (Torn et al. 2006) for each ensemble member on the outermost domain. The innermost-domain members receive downscaled initial and boundary conditions from their respective parent members.

An ensemble adjustment Kalman filter (Anderson 2001) method within the Data Assimilation Research Testbed (DART; Anderson 2009) is used to develop flow-dependent covariances between model state variables during cycling, which is critical for ESA application. Cycling of synthetic observations begins 48 h prior to forecast initialization on the outermost domain, assimilating observations every 6 h. After 24 h of cycling, the inner domains are initialized by downscaling the outer domain. Following downscaling, another 24 h of cycled data assimilation is conducted on all three domains, with 6-hourly assimilation. Model parameterizations must be utilized so as to properly simulate subgrid-scale processes, which include the Rapid Radiative Transfer Model (Mlawer et al. 1997) and Dudhia (Dudhia 1989) radiation schemes, Yonsei University boundary layer scheme (Hong et al. 2006), Kain–Fritsch scheme (Kain 2004) for cumulus processes, Noah Land Surface Model (Chen and Dudhia 2001), and the Thompson microphysics scheme (Thompson et al. 2004). All listed schemes except for the cumulus parameterization are applied to the three domains; convection is explicitly simulated with 3-km grid spacing (Bryan et al. 2003; Kain et al. 2013) on the innermost domain. After cycling, ensemble members are integrated through a 36-h free-forecast period that spans the convective event.

To promote spread within the ensemble during cycling, a spatially and temporally adapting inflation mechanism (Anderson 2007, 2009) is applied. Inflation is used only on the posterior ensemble to inflate areas across the domain with too little spread after assimilation. Inflation counteracts the impact of undersampling the degrees of freedom, as a result of using only 50 ensemble members. Furthermore, to reduce spurious covariances across the domain that arise when a finite

---

**FIG. 1.** Outermost rectangles (Ref d01, d02, and d03) represent the WRF observation-simulation domain configuration, where observations are gathered strictly from the innermost domain for cycled data assimilation (Ref d03). The innermost domains (Ens d01, d02, and d03) are used in the OSSE control ensemble-forecast simulations.
ensemble is used to approximate the true error-covariance matrix, covariance localization is employed during cycling to reduce the impact of observations far from the observing location (Houtekamer and Mitchell 1998; Hamill 2001). A Gaspari–Cohn localization function (Gaspari and Cohn 1999) is employed on each observation with horizontal half-widths of 950 and 320 km, and vertical half-widths of 1.5 and 0.5 km, for the outermost and nested domains, respectively. At two times the half-width, the assimilated observation has no impact on state variables.

Synthetic observations used during cycling are gathered from a separate WRF simulation [i.e., synthetic-observation simulation (SOS)] by interpolating model state variables to observation locations. The SOS is integrated for 54 h with version 3.8.1 of WRF, beginning 6 h prior to data assimilation, in order to sample the atmosphere over the cycled data assimilation window. The simulation is cold-started from GFS boundary and initial conditions, with three one-way nested domains at 27-, 9-, and 3-km grid spacing (Ref d01, d02, and d03 in Fig. 1). The innermost domain covers the same geographical area as the outermost control ensemble domain; all synthetic observations are gathered from this large, high-resolution domain. The same parameterization schemes are used in the SOS as in the control ensemble, and no convection parameterization is used on the 3-km grid spacing domain.

Surface observation locations are obtained from the Meteorological Data Ingest System (MADIS). Observations of temperature, specific humidity, altimeter, and both horizontal components of wind are gathered from the SOS at surface station locations across the United States (e.g., Automated Surface Observing System network, Mesonets, and Citizen Weather Observer Program) and a handful of buoys (Fig. 2). Additionally, observations are gathered at active and inactive radiosonde locations across the United States (stars in Fig. 2), retrieving temperature, specific humidity, and wind components from all 38 vertical model levels, as well as station altimeter, from the SOS. Error is added to each observation to reduce effects of underdispersion within the ensemble during data assimilation. The error characteristics of each observation type are obtained from DART and are described in Table 2. Errors are added to observations by randomly selecting a value from a normal distribution with mean 0 and variance equal to the specified error statistic for each individual observation (Table 2). Combined, these two observation sets produce approximately 150 000 observations at any particular assimilation time during cycling.

d. Targeted observing

Area-averaged composite reflectivity (MDBZ) and 3-hourly accumulated precipitation (RAIN) are chosen as the forecast metrics ($R_f$) to represent deep convection along drylines (e.g., Fig. 3b) within the ensemble forecasts. They are calculated at prescribed times over regional areas (response regions) on the innermost control ensemble domain (e.g., Fig. 3a). Using the ESA-based targeting formula (described in section 2), the ensemble forecast estimates of $R_f$ are regressed back to state variables at earlier times (e.g., forecast hour 6) and combined with initial-condition error variance to determine where targeted observations should be gathered (i.e., target fields). Note that ESA-based targeting estimates are only computed on the innermost convection-allowing control ensemble domain, and therefore targeted observations are only assimilated on this high-resolution domain. A two-sided t test is performed on the sample distribution regression slope (i.e., ensemble sensitivity)
with a 90% confidence interval to subset sensitivities that are statistically significant and account for sampling errors with a limited ensemble size. Significant sensitivity rejects the null hypothesis that changes to the initial conditions will not affect the forecast metric. Statistical significance of sensitivity is then used to spatially subset potential targeted observations. Target fields are calculated at forecast hours 6, 12, and 18, providing 6-hourly spaced lead times (i.e., time between assimilation and calculated $R_f$) to examine how observation impacts vary with integration length and if nonlinearity influences are more prevalent with longer lead times. Initial conditions at each assimilation time are obtained from control ensemble member forecasts valid at that time.

For concision, only temperature and altimeter are used as initial condition variables against which $R_f$ is regressed and these act as the only targeted-observation types. Given prior ESA results (e.g., Bednarczyk and Ancell 2015; Hill et al. 2016), it is expected that forecasts of reflectivity and precipitation are sensitive to low-level and midlevel temperatures given their relationship to buoyancy. Moreover, Madaus et al. (2014) detailed the value of dense pressure observations for convection forecasts, demonstrating improved mesoscale analyses and forecasts when they were assimilated. To critically evaluate the ESA methodology, 10 gridpoint locations are chosen from the generated target fields at a single time in areas of high estimated MDBZ and RAIN variance reduction. The selected target locations are not necessarily at the maximum in variance reduction, but rather chosen in areas reflective of statistically significant sensitivity, which typically coincide with areas of high variance reduction. Five of the identified locations are used to select targeted temperature observations: one surface 2-m temperature observation (e.g., Figs. 4c,d) and four temperature observations aloft at 850, 700, 500, and 300 hPa. The other five locations are reserved for targeting altimeter observations using the same selection procedure (e.g., Figs. 4a,b). In addition to these 10 target locations, 10 other locations are chosen randomly for observing, at the same vertical levels, in areas designated as “nontargets” (i.e., areas that do not exhibit statistically significant sensitivity). In this manner, the ability of ESA to diagnose the best sampling regions will be judged. While this procedure can be automated, observation locations are selected manually to guarantee no observations are selected in mountainous regions, and to simulate a real-time application of ESA with human decision makers.

Once all 20 locations are determined, the respective temperature and pressure observations at the appropriate observing time are retrieved from a single ensemble member, which guarantees the assimilated target and nontarget observations fall within the ensemble distribution. Typically, the SOS would be used to generate targeted observations in an OSSE. However, it is critical that the chosen observations are not discarded by the filtering procedure, which could happen if the

<table>
<thead>
<tr>
<th>Observation type</th>
<th>Error variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface temperature</td>
<td>$6.25 \text{K}^2$</td>
</tr>
<tr>
<td>Altimeter</td>
<td>$1.0 \text{hPa}^2$</td>
</tr>
<tr>
<td>Surface specific humidity</td>
<td>$1.0 \text{g}^2 \text{kg}^{-2}$</td>
</tr>
<tr>
<td>Surface wind components</td>
<td>$12.25 \text{m}^2 \text{s}^{-2}$</td>
</tr>
<tr>
<td>Radiosonde temperature</td>
<td>$1.5 \text{K}^2$</td>
</tr>
<tr>
<td>Radiosonde altimeter</td>
<td>$1.0 \text{hPa}^2$</td>
</tr>
<tr>
<td>Radiosonde specific humidity</td>
<td>$2^{1/2} \text{g}^2 \text{kg}^{-2}$</td>
</tr>
<tr>
<td>Radiosonde wind components</td>
<td>$6.25 \text{m}^2 \text{s}^{-2}$</td>
</tr>
</tbody>
</table>

**Fig. 3.** Ensemble mean forecasts of (a) composite reflectivity (dBZ) and (b) 2-m dewpoint temperature (°C) for case 9 at 2300 UTC 20 May 2013. Barbs represent the 10-m wind speed and direction, and contours in (b) are the mean sea level pressure (hPa, contoured every 2 hPa). The black-outlined rectangle in (a) identifies the response region for this case.
independent SOS and control ensemble forecasts significantly diverge. The only significant configuration difference between the SOS and a single ensemble member is the domain size; the model parameterizations and grid spacing are identical.

All observations selected are assimilated twice (in parallel assimilation systems) to evaluate the impacts of localization on ESA predictions. The ESA-based targeting methodology assumes no localization for targeting prediction and EnKF assimilation. Every targeted and nontargeted observation is assimilated with localization employed in one instance (320 km horizontally and 0.5 km vertically) and localization turned off in a parallel and identical second EnKF assimilation. Brief tests were conducted on the magnitude of vertical and horizontal localization radii, with little sensitivity noted (not shown). It should also be mentioned that no posterior inflation is used in experiment assimilations. Forecasts are carried out using these two parallel analyses with boundary conditions for each ensemble member from their respective parent domain control ensemble forecasts. These side-by-side-experiment forecasts provide the basis for determining how localizing observations may alter the observation’s impact on the forecast distributions of MDBZ and RAIN.

All targeting experiment permutations amount to 120 observations individually assimilated per case and 1200 in total: (10 target observations and 10 nontarget observations) × 3 assimilation times × 2 assimilation configurations × 10 cases. Therefore, while the response functions and initial condition variables are limited, there are a sufficient number of simulations to reach conclusions regarding the effectiveness of ESA-based targeting and what factors are playing a role in its success or failure within convection-allowing forecast applications.

e. Analysis

To assess the performance of ESA predictions, comparisons are made between forecast impacts predicted by the ESA
methodology and realized impacts after targeted and nontargeted observations are assimilated and analyses are integrated forward. Comparisons are made between the control ensemble forecasts and forecasts from targeting experiments that vary assimilation times, assimilation configurations, response types, and observation types (i.e., targeted versus nontargeted observations).

The predicted change of the response mean ($\delta R_p$) and variance ($\delta \sigma^2_p$) is compared with the actual change of response mean ($\delta R_A$) and variance ($\delta \sigma^2_A$), where the subscripts $P$ and $A$ correspond to predicted and actual changes, respectively. The estimated change of the response mean,

$$\delta R_p = \frac{\partial R}{\partial x} \delta x,$$  

is simply calculated as the product of sensitivity ($\frac{\partial R}{\partial x}$) and the ensemble-mean analysis increment due to observation assimilation ($\delta x$). The predicted change in response variance ($\delta \sigma^2_p$) is simply the negative definite ESA-estimated target value from (3). Actual changes of the response distribution statistics are calculated as the difference between experiment (subscript $E$) and control ensemble (subscript $C$) forecasts:

$$\delta R_A = R_E - R_C \quad \text{and} \quad \delta \sigma^2_A = \sigma^2_E - \sigma^2_C.$$  

The agreement between predicted and actual changes to the response distributions is examined through linear regressions and associated correlation coefficients of the samples. Statistical significance is applied to the linear regressions when sample sizes are sufficiently large ($>30$) to determine when regression slopes are significant (i.e., different than zero). A Wald test (Oliphant 2007) is used to determine significance with a two-sided $p$ value $< 0.001$.

To examine how numerical noise may influence the targeted-observation evaluation, experiments are designed that randomly perturb forecast-generated analyses in remote locations and integrate the new analyses with WRF over the forecast window. The resulting changes in response distributions are quantified and compared against the targeted- and nontargeted observation experiments. New, perturbed analyses are generated from the forecast distribution at each experiment assimilation time (i.e., forecast hours 6, 12, and 18) and integrated forward, resulting in one noise experiment at each assimilation time per convective case. To prescribe a random perturbation that would reasonably have no influence on convection, other than through numerical noise, the perturbation is made to a WRF prognostic variable (i.e., perturbation potential temperature) well removed from the convection of interest on each ensemble member. Perturbations are made at the topmost vertical level in the northeastern-most corner of the domain at a single grid point. The magnitudes of the perturbations are drawn separately for each ensemble member from a normal distribution with mean 0 and variance 0.5 K. Other variance magnitudes were tested to draw perturbations and produced similar results (not shown).

To quantify the impact of numerical noise on ESA predictions, each targeting experiment is assigned a ratio to represent the proportion of response-change signal explainable by noise, such that noise is less impactful on predicted and actual observation impacts as the proportion increases:

$$\left| \frac{\delta \sigma^2_p}{\delta \sigma^2_A} \right| = C_M \quad \text{and} \quad \left| \frac{\delta \sigma^2_p}{\delta \sigma^2_A} \right| = C_V,$$

where $\delta \sigma^2_p$ and $\delta \sigma^2_A$ represent the change in response variance and mean in the noise experiments, respectively; and $C_M$ and $C_V$ are the ratios that describe the proportions for changes in response mean and variance, respectively. For example, a $C_M$ or $C_V$ of 1 would indicate a targeted observation induces the same magnitude change on the response distribution as an innocuous, random perturbation. $C_M$ and $C_V$ can then be assessed as a function of the sample-regression correlation coefficients (CCs) to determine the influence of noise on ESA predictions through a spectrum of ratios.

Additional analysis will be carried out in a companion paper to examine the spatial distribution of observation increments and resulting forecast increments, which is beyond the scope of this paper. Moreover, while the analysis herein will focus on ESA-based predictions of observations and single-observation impacts on forecast distributions, conclusions regarding forecast skill improvements (i.e., forecast error) will not be discussed, given the limited number of targeted observations, response metrics, and cases to reach robust conclusions about changes in forecast skill from ESA-based targeted observations.

3. Predicted versus actual impacts

a. Targeted and nontargeted observations

From an aggregate perspective, ESA has stronger predictive skill for targeted observations assimilated at shorter lead times (i.e., there is generally better association between $\delta R_p$ and $\delta R_A$). There is little association between $\delta R_p$ and $\delta R_A$ for targeted observations assimilated at forecast hour 6 (the longest lead time), as measured by $-0.28$ and $-0.29$ CCs for the MDBZ (Table 3) and RAIN (Table 4) responses, respectively; this association is similarly true at forecast hour 12, where predicted response-mean impacts relate to actual impacts with 0.1 and 0.06 CCs for the MDBZ and RAIN response variables, respectively. Prediction errors of $\delta R_A$ for MBDZ are largest for observations assimilated at hours 6 and 12, with linear regression slopes of $-0.27$ and $0.13$, respectively. At forecast hour 18 (the shortest lead time), $\delta R_p$ and $\delta R_A$ more strongly correlate, though the sample linear regressions suggest only a 0.44 CC for MBDZ and 0.7 CC for RAIN. Generally, despite weak correlations, response mean change predictions improve as lead time increases (i.e., linear regression slopes approach one). At forecast hours 6 and 18, the associated linear regressions of $\delta R_A$ are statistically significant (boldface values in Tables 3 and 4) for both response variables.
Table 3. Sample linear regression slope and correlation coefficient between predicted and actual changes of MDBZ response mean and variance at forecast hours 6, 12, and 18 for targeted and nontargeted observations. Sample size is 200 in each forecast-hour sample. Linear regression slopes that are statistically significant ($p$ value < 0.001) are shown with boldface type.

<table>
<thead>
<tr>
<th>Forecast hour</th>
<th>Targeted obs Mean</th>
<th>Variance</th>
<th>Nontargeted obs Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope  CC</td>
<td>Slope</td>
<td>CC</td>
<td>Slope</td>
</tr>
<tr>
<td>6</td>
<td>-0.27  -0.28</td>
<td>1.87</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.13</td>
<td>0.1</td>
<td>1.92</td>
<td>0.72</td>
</tr>
<tr>
<td>18</td>
<td>0.21</td>
<td>0.44</td>
<td>1.17</td>
<td>0.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Forecast hour</th>
<th>Targeted obs Mean</th>
<th>Variance</th>
<th>Nontargeted obs Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope  CC</td>
<td>Slope</td>
<td>CC</td>
<td>Slope</td>
</tr>
<tr>
<td>6</td>
<td>-2.48  -0.37</td>
<td>16.5</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.09</td>
<td>0.02</td>
<td>4.46</td>
<td>0.51</td>
</tr>
<tr>
<td>18</td>
<td>0.05</td>
<td>0.02</td>
<td>1.2</td>
<td>0.06</td>
</tr>
</tbody>
</table>

despite poor predictions. Conversely, nontargeted observations have almost zero correlation between $\delta R_p$ and $\delta R_d$, for forecast hours 18 and 12 (Tables 3 and 4). This result is in large part because mean-change predictions of MDBZ and RAIN are relatively small compared to actual changes (not shown); nontargeted observations are impacting the forecast at the same magnitude as target observations, but are not predicted by ESA to do so.

Targeted observations assimilated at forecast hour 18 act to reduce MDBZ response variance in 83% of experiments, compared to 58% and 62% for forecast hours 12 and 6, respectively; RAIN response variance is decreased in 63%, 41%, and 56% of experiments at forecast hours 12, 18, and 6, respectively (not shown). This result suggests that nonlinear perturbation evolution may be more prevalent as lead time increases, causing ESA-based predictions to degrade in skill. Linear regressions of $\delta R_p$ and $\delta R_d$ for targeted observations at all assimilation times are statistically significant for both response variables, with forecast-hour-18 assimilation impacts correlating best with ESA-based predictions for RAIN (Table 4); hour-12 impacts correlate best to predictions for MDBZ (Table 3). As was the case when comparing targeted- and nontargeted observation impacts on $\delta R$, nontargeted observations tend to produce a similar $\delta R_p$ as targeted observations (not shown). Moreover, nontargeted observations have statistically significant correlations at assimilation hours 6 and 12, but the linear regression slopes themselves are nearly vertical for the MDBZ response (large slopes); no regression slopes of the nontargeted observation impacts on response variance for RAIN are significant.

b. Lead time

ESA prediction-skill dependence on lead time is explored by further partitioning experiments by the assimilation configuration. For targeted observations assimilated with the EnKF and no localization, correlations between $\delta R_p$ and $\delta R_d$ increase and prediction errors largely decrease with decreasing lead time (Fig. 5). At forecast hour 6, the correlations between $\delta R_p$ and $\delta R_d$ for both responses are weak, and slightly negative (Figs. 5a,b). The MDBZ response has a 0.3 CC between $\delta R_p$ and $\delta R_d$ at forecast hour 12, which increases to 0.85 at forecast hour 18. The RAIN response has a similar increase in CCs, from 0.2 at forecast hour 12 to 0.98 at hour 18. At forecast hour 18, the sample-regression slopes for MDBZ and RAIN are statistically significant and more closely align with the perfect-prediction 1–1 line (Figs. 5e,f).

In comparison with response-mean correlations, changes and predictions of variance change are more correlated at forecast hours 6 and 12 (>0.5), although changes to the RAIN response at hour 12 exhibit a negative correlation (Fig. 6d), influenced by numerous positive changes to the variance of the rainfall response when predictions were negative. The targeted observations act to both increase and decrease response variance by as much as 30 dB$^2$ and 0.75 mm$^2$ at forecast hours 6 and 12, respectively. In contrast, at forecast hour 18, targeted observations primarily decrease variance for both response variables (Figs. 6e,f), correlating best to $\delta R_p$ with the RAIN response (0.83 CC). Predicted impacts of targeted observations on response variance generally span from $-20$ to 0 dB$^2$ and from $-0.5$ to 0 mm$^2$, whereas actual impacts span from $-40$ to 30 dB$^2$ and from $-0.5$ to 0.75 mm$^2$ for the MDBZ and RAIN responses, respectively (Fig. 6). It is hypothesized that the increases in MDBZ and RAIN variance are largely a result of nonlinear perturbation evolution that ESA cannot properly predict; nonlinearity is explored in section 4.

c. Localization

The assimilation configuration is also altered in the experiments to examine the influence of localization on ESA predictions of observation impacts. Localization negatively

Table 4. As in Table 3, but slopes and correlation coefficients are calculated with respect to the RAIN response variable.
impacts the assimilation and subsequent correlation between predicted and actual response impacts (Figs. 7 and 8). Localization narrows the impact of the targeted observation at assimilation, and subsequently reduces the impact on response mean (cf. Figs. 7 and 5). This result is particularly evident at forecast hour 18 when changes in the MDBZ and RAIN means hover near zero, while predicted changes span from $-4$ to $4$ $\text{dBZ}$ and from $-0.75$ to $0.5$ $\text{mm}$, respectively (Figs. 7e,f). Additionally, observations assimilated with localization employed have little impact on response variance (Fig. 8). While some observations induce a nonzero change in response variance, the majority of observations at all forecast times, and in particular hour 18 (Figs. 8e,f), produce little to no change in variance. Linear regressions applied to these sample subsets are often misleading with high CCs, since clustering of the samples is occurring (e.g., Figs. 7e and 8c,e). These results demonstrate that localization negatively impacts the prediction of response changes as determined by ESA.

4. Other sources of error

While the presented experiments were designed to control for a number of error sources, two main error sources remain: numerical noise and nonlinearity. Numerical noise may be contributing to poor correlation results through chaos seeding (Ancell et al. 2018). Additionally, nonlinear perturbation evolution through convective processes are unaccounted for in the presented experiments and may be contributing to the unfavorable experimental results related to ESA-based observation predictions. To some extent, this signal of nonlinearity is evident in the comparison between targeted

---

**Fig. 5.** Predicted changes vs actual changes of response mean for experiments of targeted observations with EnKF assimilation in all cases for (left) composite reflectivity (dBZ) and (right) accumulated rainfall (mm) responses at forecast hours (a),(b) 6; (c),(d) 12; and (e),(f) 18. The colored lines and the numbers in the bottom-right corners represent the linear regressions and regression correlation coefficients for the sample populations, respectively. Linear-regression slopes that are statistically significant ($p$ value < 0.001) are denoted with an asterisk.
and nontargeted observations as well as prediction-skill dependence on lead time. The ESA-based method fits a linear relationship between response and initial-condition variables, which may poorly describe the complex and potentially nonlinear relationships in the simulated cases, such as the relationship between surface temperatures and level of free convection attainment by parcels near the dryline. The impacts of numerical noise due to chaos seeding (i.e., perturbation evolution) and nonlinearity will be investigated further to enhance understanding of how targeted observations influence convection forecasts and the utility of ESA to target these observations.

a. Numerical noise

First, only the experiments that exhibit a $C_M$ or $C_V$ greater than 1 are considered; those experiments with a noise ratio below 1 are withheld from the samples. The correlations between $\delta R_p$ and $\delta R_A$ increase as a result of removing the “noisy” experiments—for both response variables—particularly at forecast hour 18. Observations assimilated at forecast hours 18 and 12 have stronger response-mean change correlations when noise is used to subset the sample (cf. Tables 5 and 3), indicating perhaps that numerical noise is a contributing factor to poor predictions of response mean change. Forecast hour 18 remains the best prediction time, with a correlation of $\delta R$ and $\delta R_A$ above 0.6 for both response variables, whereas forecast hours 12 and 6 have CCs < 0.2; the forecast-hour-6 sample still has a negative correlation in both response variables (Tables 5 and 6). Similar improvements in correlation are seen for nontargeted observation impacts on response mean when they are assimilated at forecast hour 18; the correlations remain small but do slightly improve from 0.02 to 0.19 (cf. Tables 5 and 6) for MDBZ and RAIN responses, respectively. Correlations of predicted and actual response-variance changes by targeted observations appear minimally impacted by noise (cf. Tables 5 and 3). Sample regression slopes are largely unchanged, except for MDBZ regressions between $\delta R_p$ and $\delta R_A$ (Table 5) which have all progressed closer to one.

FIG. 6. As in Fig. 5, but for samples of predicted changes vs actual changes of response variance (dBZ$^2$ and mm$^2$).
As $C_M$ and $C_V$ are increased, the sample correlations between predicted and actual changes of MDBZ response mean and variance increase and decrease, respectively, regardless of assimilation time (Fig. 9). The correlation coefficients that describe changes to response mean are a function of $C_M$ and increase by as much as 0.25 as the sample size of experiments drops from 200 to 75 for all times (Figs. 9a,c,e). Interestingly, the ESA-based predictions of response variance changes are negatively impacted by removing experiments that have a large component of noise in the experiment signal (Figs. 9b,d,f), which could hint at the rather complex variance distribution for a highly nonlinear response metric like MDBZ. Moreover, these same correlation changes as a function of noise ratio are found for the RAIN response (Fig. 10), but the variance correlations of RAIN are not affected as much as the MDBZ response (Figs. 10b,d,f); the response variance remains nearly constant despite removal of more than three quarters of the experiments (Figs. 10b,d,f). Furthermore, the correlation coefficient of predicted and actual impacts on response mean at assimilation hour 18 is nearly perfect when the only the largest noise ratios are considered (i.e., small proportion of experiment impact explained by noise). These results suggest that noisy observation increments play a role in ESA’s ability to predict observation impacts on response distributions.

b. Nonlinearity influences

Additional experiments are conducted to further elucidate the impact of nonlinearity in the assessment of ESA-based predictions. As was previously discussed, the ESA method is developed using linear assumptions that describe the relationship between response and initial-condition variables. This assumption may not be valid at longer lead times when convection is present, but may approach validity at shorter lead times as seen through the analysis of the targeted-observation experiments above. However, if convection was removed from the forecast, then perturbation evolution through the forecast period—as a result of observation assimilation—would approach more linear assumptions, even for high-resolution
forecasts. Therefore, a series of targeted and nontargeted observation experiments are conducted for a separate forecast case (23 January 2013) where little convection is present within the verification domain. Similar targeting procedures are followed as outlined in section 2, but the response variable is geopotential height at sigma level 25 in an area of high forecast spread. Temperature and altimeter observations are selected in target and nontarget locations as before, and assimilated at forecast hours 6, 12, 18 and 24. In this particular case, the response function is valid at hour 24 and an assimilation evaluation is attempted at the response time to understand how the EnKF assimilation spreads observational information directly onto the response function through state-variable covariances and without forecast integration. Such an evaluation is not

![Variance MDBZ](image1)

![Variance RAIN](image2)

**FIG. 8.** As in Fig. 6 but observations are assimilated with localization.

<table>
<thead>
<tr>
<th>Targeted observations</th>
<th>Nontargeted observations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>Variance</strong></td>
</tr>
<tr>
<td><strong>Forecast hour</strong></td>
<td>**Slope</td>
</tr>
<tr>
<td>6</td>
<td>-0.12</td>
</tr>
<tr>
<td>12</td>
<td>0.26</td>
</tr>
<tr>
<td>18</td>
<td>0.35</td>
</tr>
</tbody>
</table>

**TABLE 5.** As in Table 3, but only considering observations that induce changes in response variance that exceed the chaos-induced response. Sample sizes are included for reference.
possible for the MDBZ and RAIN responses since a number of the variables used to derive these responses are not contained within the WRF state vector.

Targeted observations demonstrate a strong relationship between predicted and actual changes of response mean at all forecast times (Fig. 11a) with sample regressions that deviate more from the one-to-one predicted relationship as the lead time increases. The linear regressions of the $\delta R$ samples suggest that the one-to-one predicted relationship is more valid for observations assimilated at shorter lead times, as the 24- and 18-h regressions have the largest slopes. Targeted observations assimilated at hour 6 produce nonmeaningful changes to geopotential height at hour 24 (Fig. 11a). Nontargeted observations have strong correlations at forecast hour 24 and 18 as well but exhibit smaller impacts on the response when compared with targeted observations (Fig. 11c). In addition, the changes to response variance for targeted observations assimilated at hours 24 and 18 are all below zero (Fig. 11b). Nontargeted observations have poorer correlations between predicted and actual changes of response variance, as response variance is increased for experiments at a number of assimilation hours (Fig. 11d). Nontargeted observations assimilated at hour 24 do not produce impacts on the response variance that are well predicted by ESA; they appear both positively and negatively biased. Moreover, the predicted change of response mean should best align with actual changes at forecast hour 24, since no model integration follows assimilation. This is indeed realized as both targeted and nontargeted observations have

### Table 6. As in Table 5, but for the RAIN response.

<table>
<thead>
<tr>
<th>Forecast hour</th>
<th>Slope</th>
<th>CC</th>
<th>Sample</th>
<th>Slope</th>
<th>CC</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-0.05</td>
<td>-0.1</td>
<td>144</td>
<td>0.6</td>
<td>0.51</td>
<td>144</td>
</tr>
<tr>
<td>12</td>
<td>0.12</td>
<td>0.12</td>
<td>135</td>
<td>-0.62</td>
<td>-0.4</td>
<td>133</td>
</tr>
<tr>
<td>18</td>
<td>0.34</td>
<td>0.74</td>
<td>145</td>
<td>0.56</td>
<td>0.52</td>
<td>136</td>
</tr>
<tr>
<td>Nontargeted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-2.49</td>
<td>-0.39</td>
<td>130</td>
<td>0.39</td>
<td>0.04</td>
<td>147</td>
</tr>
<tr>
<td>12</td>
<td>-0.68</td>
<td>-0.12</td>
<td>123</td>
<td>-0.04</td>
<td>-0.02</td>
<td>119</td>
</tr>
<tr>
<td>18</td>
<td>0.55</td>
<td>0.39</td>
<td>114</td>
<td>0.01</td>
<td>0.0</td>
<td>128</td>
</tr>
</tbody>
</table>

**FIG. 9.** Correlation coefficients between predicted and actual changes of MDBZ (left) mean and (right) variance as a consequence of increasing the noise ratios $C_M$ and $C_V$ (see the text for a description of each ratio) at assimilation hours (a),(b) 6; (c),(d) 12; and (e),(f) 18. Black lines denote the sample size used to compute the correlation coefficient.
nearly perfect prediction of observation impacts on the response mean when no model integration is done (Figs. 11a,c, purple stars and lines).

The nonconvection case presented illustrates the complexities of nonlinearity and its influence on targeted observing for convection. Observation impacts on a nonconvective response are more predictable than observations assimilated for the convection forecasts because the initial perturbations from observation assimilation evolve more linearly through the forecast period. Subjectively, the changes in response mean are strongly correlated to actual impacts, while changes to response variance are nearly always negative (except for nontargeted observations). This would indicate that, unsurprisingly, nonlinearity is negatively impacting the performance of targeted observations for the convection cases presented, contributing to unpredictable changes to the response distributions. Additionally, it should be noted that the GPH response is vastly different from the MDBZ and RAIN responses used in the convection cases. For one, GPH is a continuous field across the domain, whereas MDBZ and RAIN are inherently localized and have sharp horizontal gradients where convection occurs. Moreover, GPH is a simple conversion of a model state variable (perturbation potential), whereas the convection responses are derived from complex functions of hydrometeor variables. However, our nonconvection case results are largely independent of the response variable chosen; when GPH is used to target observations in a convection case (not shown), the observations impacts are poorly correlated to the predicted impacts, particularly when observations are assimilated at hours 6 and 12, as they were for the MDBZ and RAIN responses for the majority of convection experiments explored. These results further support prior evidence that ESA-based predictions are most accurate at short lead times, and therefore ESA-based targeting for nonlinear forecasts is most applicable at shorter time scales.

5. Summary and discussion

Ensemble sensitivity analysis (ESA) is applied to numerous convection cases across three spring seasons to critically evaluate the performance of an observation targeting algorithm on convection-allowing scales. ESA linearly relates chosen forecast metrics—composite reflectivity and 3-hourly accumulated rainfall—at various forecast times to prior model state variables, or functions of state variables, to evaluate how small initial-condition perturbations will evolve onto the forecast. Combined with this sensitivity information is an estimate of prior uncertainty, which yields optimal locations to target new observations that will reduce the variance of the chosen forecast metrics, along with estimations of the variance reduction. Observing system simulation experiments with perfect models are utilized to control for model error and simulate truth, which additionally allows for four-dimensional sampling to generate new observations. Targeted and non-targeted observations are selected in locations determined...
from the ESA-based algorithm and assimilated with ensemble forecast analyses. The assimilation configuration, assimilation time, observation type, and response variable are varied to evaluate different factors that lead to successful or failed predicted impacts on the response distributions.

Targeted and nontargeted observation impacts are compared to determine if the ESA methodology can effectively discriminate good and bad sampling regions. It is initially determined that targeted and nontargeted observations provided similar-magnitude impacts on the response distributions. However, when comparing the convective cases to a case in which convection is effectively removed, it is clear that the similarities in target and nontarget observation impacts in the convective cases are due in large part to nonlinear perturbation evolution. Both observations, targeted and nontargeted, spread increments into the moist dynamics that then grow nonlinearly and filter into MDBZ and RAIN, which are inherently tied to moist dynamic processes of the model.

Whereas changes to the response mean aligned with ESA-based predictions under certain experiment configurations, changes to response variance were less predictable for most experiment permutations. It was found that observations equally increased and decreased response variance, except when observations are assimilated at short lead times. These results further support the conclusions that the assimilation increments propagate nonlinearly onto the response variables, and in particular their variance. A comparison of the 10 convective cases with a nonconvective case supported these claims.

The data assimilation configuration that did not include localization yielded the best predictions of response distribution changes, particularly with respect to the response mean. This result is not too surprising, given that the derivation of the ESA-targeting methodology, and estimated impacts, follows that of an EnKF assimilation with no localization. The impact of localization is dramatic, as the assimilation tool truncates the

---

**Fig. 11.** Predicted changes vs actual changes of (left) response mean ($m$) and (right) response variance ($m^2$) for (a),(b) targeted observations and (c),(d) nontargeted observations in the nonconvection case referenced in the text. Red diamonds correspond to observations assimilated at forecast hour 6, blue triangles correspond to observations assimilated at forecast hour 12, green circles correspond to observations assimilated at forecast hour 18, and purple stars correspond to observations assimilated at forecast hour 24. Colored lines represent the linear regressions for their colored sample populations.
observation’s impact on the response distribution. It is plausible that allowing a single observation to update the entire state space is allowing the observational information to spread to the response metrics dynamically through the adjoint sensitivity. Given ensemble sensitivity is the projection of adjoint sensitivity onto the ensemble statistics (Ancell and Hakim 2007), allowing the observation to update the entire domain connects the observation, no matter the location, to the true dynamic sensitivity. When the analysis increment is restricted to a certain localization radius, this dynamic connection may be lost. Furthermore, it is entirely possible that the prescribed localization radius may be most optimal for the assimilation of a normal observation set (i.e., millions of observations), but is too severe for small observation sets or single-observation assimilation experiments as in this study.

To investigate any influence from numerical noise, additional experiments were conducted to provide a baseline change in response mean and variance from noise (i.e., random perturbations) and evaluate if targeted observations were making a perturbation on the response that was significant and larger than noise. When experiments were stratified by the ratio of targeting-experiment response change to noise-experiment response change, the response-mean correlations increased as the ratio increased, suggesting that when observation impacts are large, relative to numerical noise, they are more predictable. As the ratios increased, changes to MDBZ variance were less predictable, whereas predictability for RAIN-variance changes remained the same, highlighting the respective nonlinear and linear characteristics of the individual response types, respectively. Overall, while targeted and non-targeted observation assimilation and associated impacts were not completely predictable under any configuration, it was apparent that the best predictability in this EnKF assimilation framework arises under a very idealized modeling configuration with no model error, no localization, and for short lead times, supporting the conclusion that targeted observing estimates are more accurate for nonlinear convection forecasts at short time scales. Madaus and Hakim (2017) similarly showed that convection initiation prediction in the absence of strong forcing only benefited from observations within 1–2 h of initiation.

A number of caveats to these conclusions exist, however, given the idealized nature of the experimental design, that have significant implications for real-time targeting applications. First, these OSSEs only consider single-observation sampling. Real-data experiments would ideally target a location and sample either at multiple levels or at recurring times, providing more numerous observations that could constrict an analysis toward reality and produce large and meaningful impacts on the forecast. Ancell and Hakim (2007) describe the process of targeting two observations, and they suggest the calculation of variance reduction for a few observations would be trivial. However, they also stress the computational requirement to iteratively select numerous observations and the potential diminishing returns on variance reduction after a few observations are assimilated. Additionally, the single-observation assimilation experiments presented may artificially introduce gravity wave noise through the forecast period due to initial dynamic imbalances that reflect off the model domain due to inconsistencies with the analysis increment and boundary conditions (not shown). Supplementary analysis suggests that this noise gradually dampens with time before response variables are calculated (not shown).

Second, these OSSEs disregard regularly available observations from satellite platforms, radiosondes, etc., that would potentially influence the predicted value of targeted observations, particularly if there are redundant observations in target locations. It should be expected that targeted mesoscale observations would have even smaller impacts if considered within a real forecasting system since the background analysis would theoretically be improved with routine observations. In fact, this complication was mentioned specifically by Romine et al. (2016) as a limiting factor for targeting operations during MPEX, when routine analysis cycles were not accounted for between target selection and sampling, potentially removing the need for targeted observations. Additionally, the forecast metrics used are inherently tied to processes particularly difficult to model accurately (e.g., microphysics). Perhaps a metric tied to dryline position, for example, would yield better ESA predictions of observation impacts.

It is important to compare these idealized results with those obtained prior on synoptic-scale systems (e.g., Ancell and Hakim 2007; Torn and Hakim 2008). Torn and Hakim (2008) documented that both response mean and variance changes were predictable as the result of denying a single pressure observation, with much higher correlations than observed in these experiments. Particularly interesting is they found that changes to response variance were better predicted by ESA than the changes to response mean, which they attributed to sampling error under the pretense of no significant nonlinearity nor model error. Moreover, Torn and Hakim (2008) noted the ability of ESA to predict changes to response error (sea level pressure error) as well. This analysis step was not completed in this work due to the complexity of verifying rainfall and reflectivity in a gridpoint manner; object-based verification would be the proper method, which removes the applicability of gridpoint-based ESA-error reduction analyses. Additional tests could validate ESA on scales of < 3 km with analyses of continuous forecast parameters (e.g., surface pressure), which can be assessed appropriately on a grid.

The results of this work motivate additional research that could elucidate the validity of targeted observing for convection forecasts. Additional OSSEs could be implemented, either under perfect-model conditions or not, that assimilate multiple targeted observations and take into account additional analysis cycles between ensemble initiation and target sampling. These experiments would allow for a more robust examination of target observation types (e.g., moisture) and response variables, sampling all vertical levels of a radiosonde versus continuous sampling at the surface, or mobile observations at the surface (e.g., StickNet; Hill et al. 2020; mobile mesonets) and aloft (e.g., uncrewed aircraft systems). One of the more important avenues of exploration that needs to be considered is the application of multivariate regressions. Specifically, a multivariate approach that selects the most important state variable to sample (e.g., ANOVA test) could yield more
impactful adjustments to the response distribution than subjectively chosen observation types. As targeted observation studies go to smaller and smaller scales (e.g., Limpert and Houston 2018; Kerr et al. 2019; Kerr and Wang 2020), the utility of other statistical techniques may be necessary to capture the relationships of state variables and how they co-vary with one another. In particular, this work does not consider collinearity between targeted-observation variables, which would undoubtedly need to be addressed when observations are sampled at the same point in space. Although this work does not close the door on real-time applications of targeted observing for convection forecasts, it certainly brings to light some of the complexities and limitations for its use that motivate continued evaluation.

Acknowledgments. The authors acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR’s Computational and Information Systems Laboratory, sponsored by the National Science Foundation. We also acknowledge Dr. Glen Romine for thoughtful and insightful input into the methodology of this work during the first author’s Graduate Visitor stint at the National Center for Atmospheric Research. We appreciate the thoughtful comments and suggestions from Dr. Luke Madaus and two anonymous reviewers during the review process. This work was supported by NOAA Award NA17NWS4680003 and NSF Grant IIS-1527183.

REFERENCES


——, G. S. Romine, D. D. Turner, and R. D. Torn, 2019: Impacts of targeted AERI and Doppler lidar wind retrievals on short-term forecasts of the initiation and early evolution of...


Gaspari, G., and S. E. Cohn, 1999: Construction of correlation


Houtekamer, P., and Coauthors, 1997: The North Pacific Experiment (NORPEX-98): Targeted observations for improved North


