BAMS Article

# New Perspectives on Ensemble Sensitivity Analysis with Applications to a Climatology of Severe Convection

Brian C. Ancell and Austin A. Coleman

**ABSTRACT:** Ensemble sensitivity analysis (ESA) is a statistical technique applied within an ensemble to reveal the atmospheric flow features that relate to a chosen aspect of the flow. Given its ease of use (it is simply a linear regression between a chosen function of the forecast variables and the entire atmospheric state earlier or simultaneously in time), ensemble sensitivity has been the focus of several studies over roughly the last 10 years. Such studies have primarily tried to understand the relevant dynamics and/or key precursors of high-impact weather events. Other applications of ESA have been more operationally oriented, including observation targeting within data assimilation systems and real-time adjustment techniques that attempt to utilize both sensitivity information and observations to improve forecasts. While ESA has gained popularity, its fundamental properties remain a substantially underutilized basis for realizing the technique's full scientific potential. For example, the relationship between ensemble sensitivity and the pure dynamics of the system can teach us how to appropriately apply various sensitivity-based applications, and combining sensitivity with other ensemble properties such as spread can distinguish between a fluid dynamics problem and a predictability one. This work aims to present new perspectives on ensemble sensitivity, and clarify its fundamentals, with the hopes of making it a more accessible, attractive, and useful tool in the atmospheric sciences. These new perspectives are applied in part to a short climatology of severe convection forecasts to demonstrate the unique knowledge that can gained through broadened use of ESA.

**KEYWORDS:** Numerical analysis/modeling; Regression analysis; Statistical techniques; Ensembles; Forecasting techniques; Probability forecasts/models/distribution

https://doi.org/10.1175/BAMS-D-20-0321.1

Corresponding author: Brian C. Ancell, brian.ancell@ttu.edu

In final form 18 October 2021

**AFFILIATIONS: Ancell and Coleman**—Texas Tech University, Lubbock, Texas

Ensemble sensitivity (Hakim and Torn 2008; Ancell and Hakim 2007; Torn and Hakim 2008) is a statistical tool applied within an ensemble of forecasts that reveals relationships between some forecast aspect of importance (e.g., magnitude of rotation within convection, or the strength of a tropical cyclone) and the atmosphere at times at, before, and even after the event. Specifically, ensemble sensitivity values are the slopes of linear regressions (regression coefficients) between a response function (a function of forecast variables that diagnoses the forecast aspect of interest) and the model state variables. Over the last 10 years or so the body of published work regarding ensemble sensitivity has become substantial and has involved several types of high-impact weather phenomena at a variety of scales. Such studies include investigations into large-scale blocking events (Parker et al. 2018; Quandt et al. 2019), synoptic-scale features such as midlatitude cyclones (Ancell and Hakim 2007; Hakim and Torn 2008; Torn and Hakim 2008; Zheng at al. 2013; Chang et al. 2013; Ancell 2016; Berman and Torn 2019), convective events (Hanley at al. 2013; Bednarczyk and Ancell 2015; Torn and Romine 2015; Hill et al. 2016; Berman et al. 2017; Limpert and Houston 2018; Kerr et al. 2019; Coleman and Ancell 2020; Hill et al. 2020), tropical cyclones (Torn and Hakim 2009; Torn 2010; Nystrom et al. 2018; Ren et al. 2019; Hu and Wu 2020), and flows in complex terrain with applications to wind power (Zack et al. 2010a,b,c; Wile et al. 2015; Smith and Ancell 2017). Ensemble sensitivity studies such as these primarily fall into three categories: 1) examining sensitivity fields to understand the relevant dynamics or predictability associated with a high-impact weather event (e.g., Nystrom et al. 2018), 2) using sensitivity within a data assimilation framework to understand the forecast value of targeted observations (e.g., Hill et al. 2020), and 3) supporting the development of operational tools that use sensitive regions to beneficially adjust probabilistic forecasts in a real-time environment (e.g., Coleman and Ancell 2020).

While prior ensemble sensitivity studies have contributed to a valuable and growing collection of research with key insights into dynamics and predictability, most of them have focused on applying the strict statistical formulation and interpreting subsequent experiments. However, ensemble sensitivity analysis (ESA) possesses the means to go well beyond a simple linear regression in ways that can substantially enhance its usefulness as a research tool. First and foremost, ensemble sensitivity's fundamental and explicit dependence on both the pure dynamics and the ensemble statistics of the atmospheric state (Ancell and Hakim 2007) provides a key concept that can provide a deeper understanding of the true nature of the ensemble sensitivity field. This fundamental property can also help us understand a main limitation of ESA—the inability to distinguish between direct and indirect dynamical processes (a key distinction when ESA is used for dynamical interpretation). The statistical–dynamical basis of ensemble sensitivity can even shed light on how multivariate regressions might, or might not, add additional value to ESA applications. Finally, simple manipulation of ensemble sensitivity fields with other ensemble parameters such as standard deviation can lead to new perspectives that, for example, can distinguish dynamics problems from those focusing on predictability. In short, ESA possesses far more potential as an important research and operational tool than has been realized to date.

The purpose of this study is to establish a comprehensive roadmap regarding how ensemble sensitivity can be used and interpreted within atmospheric sciences problems in ways that have gone mostly untapped in the past. We seek to provide a better understanding of basic ESA fundamentals, and ESA's strengths and weaknesses, toward enhancing its effectiveness as a research technique. We hope this "user's guide" makes ESA substantially more accessible

to researchers who are interested in atmospheric dynamics and predictability problems. To illustrate some of the new ESA perspectives discussed here, a brief analysis of ensemble sensitivity-related quantities is provided within a short climatology of severe storms.

## Expanding the scope of ensemble sensitivity research

At its core ensemble sensitivity is simply the slope of the linear regression, within an ensemble of forecasts, between a chosen scalar forecast response function and the entire atmospheric state. Statistically this amounts to the covariance between the response function $R$ (at response time, or the time the response is defined) and each state variable (at sensitivity time, or the time sensitivity is calculated) divided by the variance of the state variable (also at sensitivity time). Performing this calculation with respect to the entire atmospheric state (state vector $\mathbf{X}$) results in the full three-dimensional ensemble sensitivity vector $\partial R/\partial \mathbf{X}$:

$$\frac{\partial R}{\partial \mathbf{X}} = \frac{\mathrm{Cov}(R, \mathbf{X})}{\mathrm{Var}(\mathbf{X})}. \tag{1}$$

Plotting the ensemble sensitivity field reveals atmospheric features that are dynamically relevant to the response function, which is shown, for example, in Fig. 1. The response function in this case is the number of grid points exceeding 40 dB$Z$ in the green box in Florida between 30- and 36-h forecast time (summed over each hourly output time); the sensitivity to 500-hPa geopotential height at 9-h forecast time (points per meter) is shown in color here as an example (sensitivity to other atmospheric variables exists but is not shown). We see several primary sensitivity features outlined in yellow—negative values within the ridge in the western United States, a dipole over the subtropical jet over Mexico and the Caribbean, and a dipole surrounding the base of a shortwave trough near the Great Lakes—that tell us these features are related to the coverage of convection in Florida a day later. The negative sensitivities over the western U.S. ridge indicate a stronger ridge is associated with less convection in Florida 25 h later, and vice versa. The dipole signal near the Great Lakes suggests that a more intense trough (lower heights over negative sensitivity and higher heights over positive sensitivity resulting in a tighter geopotential height gradient) is related to enhanced coverage of Florida convection the following day, while a less intense trough is linked to a lower coverage of storms. Finally, a more pronounced trough–ridge couplet in the subtropical jet in the lower portion of the domain west of Florida (lower heights over the negative sensitivities and higher heights over the positive values) appears to be associated with more convection in Florida 25 h later. Interpreting these types of signals in a more detailed way has been one common, key use of ensemble sensitivity. Such interpretation usually becomes very involved since we could also view the sensitivity with respect to several other atmospheric variables at numerous different vertical levels.

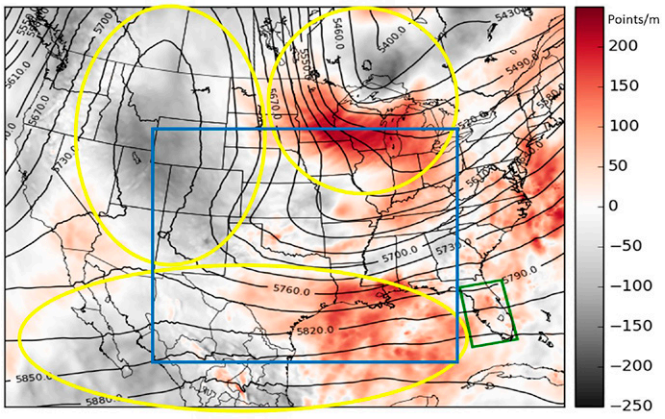


**Fig. 1. Ensemble sensitivity (shaded) of 30–36-h number of grid points exceeding 40 dB$Z$ in the green box to 9-h 500-hPa geopotential height and the ensemble mean 9-h 500-hPa geopotential height (black contours). The blue box depicts the 4-km nested domain used for the sensitivity climatology, while the outer domain represents the CONUS 12-km grid.**

If we stopped there, we would have found key, basic relationships across time between the atmosphere and the chosen response function. Interpreting ensemble sensitivity in this way is the subject of much of the ESA literature, and provides interesting insights and new knowledge regarding the relevant dynamics associated with high-impact weather. However, diving

deeper into the details of ensemble sensitivity cannot only enhance these interpretations, but can guide the development of ensemble sensitivity–based tools (and show why these tools either would or would not be expected to work) to improve atmospheric predictability. Several examples of an expanded viewpoint on ESA, starting with the fundamental properties that make them possible, are presented below.

***The fundamentals: Dynamics and statistics.*** While ensemble sensitivity results from a simple univariate linear regression as shown in Eq. (1), its fundamental nature can be expressed through a product of statistical and dynamical aspects (Ancell and Hakim 2007):

$$\frac{\partial R}{\partial \mathbf{X}_e} = \mathbf{D}^{-1}\mathbf{A}\frac{\partial R}{\partial \mathbf{X}_a}, \tag{2}$$

where $\partial R/\partial \mathbf{X}_e$ is the ensemble sensitivity vector, $\mathbf{D}$ is a diagonal matrix with the variance of the state variables on the diagonal, $\mathbf{A}$ is the symmetric analysis error covariance matrix (variance of state variables on the diagonal, covariances between each pair of state variables off the diagonal), and $\partial R/\partial \mathbf{X}_a$ is the adjoint sensitivity vector. The importance of the relationship in Eq. (2), which includes discrete statistical and dynamical pieces, is absolutely key for tapping into the potential of ESA. The product $\mathbf{D}^{-1}\mathbf{A}$ is purely a statistical term calculated through ensemble members describing the atmospheric state at sensitivity time, while adjoint sensitivity ($\partial R/\partial \mathbf{X}_a$) is purely dynamical and estimates the dynamical change to the response function due to any perturbation to the atmospheric state at sensitivity time within a deterministic simulation (Talagrand and Courtier 1987; Errico 1997).

More specifically, the product $\mathbf{D}^{-1}\mathbf{A}$ in Eq. (2) reduces to a matrix with the value of one on the diagonal, and the quotient of the covariance of state variable pairs and the variance of one of those variables off the diagonal. These off-diagonal terms, like ensemble sensitivity, represent the slopes of linear regressions between different state variables at sensitivity time. In turn, this matrix product simply contains the relationships among all pairs of sensitivity-time atmospheric state variables. The entire right-hand side of Eq. (2), and thus the building blocks that constitute ensemble sensitivity, represents a product of the relationships among different atmospheric variables and the pure dynamical sensitivity of the chosen response function to those same variables.

Figure 2 shows this fundamental difference through a conceptual schematic—both adjoint (orange) and ensemble (light turquoise) sensitivity are shown relative to a midlatitude cyclone (all valid at 0600 LT) for an area of thunderstorms later that day at 1800 LT. These hypothetical areas of sensitivity (shaded areas represent large sensitivity magnitudes) are shown with respect to the surface temperature field, and thus correspond closely to the location of the warm and cold front in the figure, indicating a sensitivity to the strength and/or position of those fronts. The localized area of adjoint sensitivity represents the only area where perturbations matter in a direct, dynamical way—perturbations there to temperature will change the nature of the thunderstorms 12 h later,

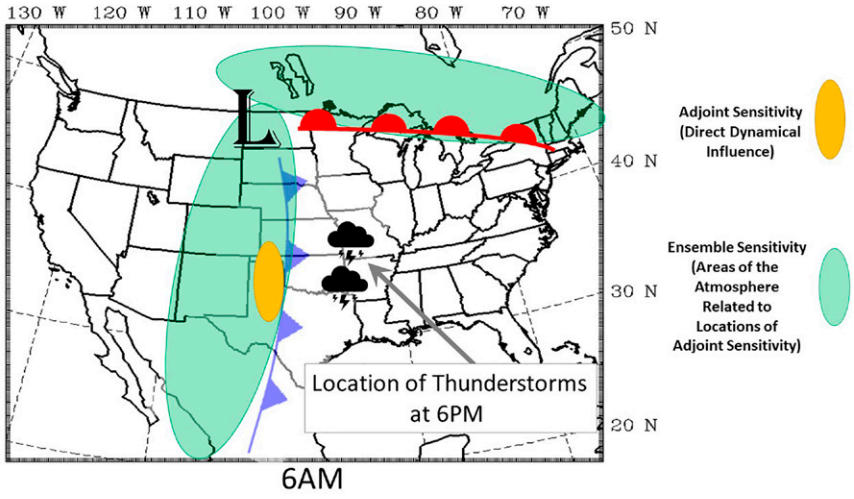


**Fig. 2. Schematic illustrating the fundamental differences between adjoint and ensemble sensitivity.**

perturbations elsewhere will not (assuming the areas not shaded possess zero sensitivity). Ensemble sensitivity, however, exists along both fronts because a change anywhere along and behind the fronts means a change in the area of adjoint sensitivity (the $\partial R/\partial \mathbf{X}_a$ term) since the fronts are all related to themselves and each other structurally (the $\mathbf{D}^{-1}\mathbf{A}$ term) within the cyclone. For ensemble sensitivity to be large at a point, it necessarily needs to be related to the areas of direct dynamical sensitivity. In other words, ensemble sensitivity is the result of mapping direct dynamical sensitivity onto all other areas and variables with covariance relationships. Since producing adjoint sensitivity is computationally intensive (more so than producing an ensemble since the model state must be saved and used at every model time step) and requires the use of an adjoint model which must contain tangent-linear versions of sometimes complex and highly nonlinear physics parameterizations (Errico 1997), a significant advantage is gained through ensemble sensitivity in that it intrinsically possesses the pure dynamical sensitivity without having to determine it explicitly. From another perspective, ensemble sensitivity estimates how statistically defined perturbations, if allowed to evolve, would affect the response function without the need to calculate adjoint sensitivity.

Mathematically, the change in response function ($\Delta R$) can be estimated by

$$\Delta R \approx \frac{\partial R}{\partial \mathbf{X}_e}\,\Delta \mathbf{X}, \tag{3}$$

where the ensemble sensitivity ($\partial R/\partial \mathbf{X}_e$) and state variable perturbation ($\Delta \mathbf{X}$) correspond to a single variable at a single model grid point. Equation (3) is the same method by which adjoint sensitivity would be used to estimate a change to a response function due to early forecast-time perturbations with the exception that ensemble sensitivity inherently incorporates a domainwide perturbation [through the statistical term in Eq. (2)] instead of a single point. This is why ensemble sensitivity is generally several orders of magnitude larger than adjoint sensitivity with respect to the same variable. Ancell and Hakim (2007) provide a more in-depth discussion on these key differences between adjoint and ensemble sensitivity.

The ability to separate the statistical and dynamical contributions to each ensemble sensitivity value provides an extremely important basis for interpretation of ensemble sensitivity. For conceptual ease, consider only three state variables at a single model grid point: temperature ($T$), pressure ($P$), and water vapor mixing ratio ($Q$). By expanding Eq. (2), the ensemble sensitivity to temperature ($\partial R/\partial T_e$) can be expressed as

$$\frac{\partial R}{\partial T_e} = \frac{\partial R}{\partial T_a} + \frac{\partial Q}{\partial T}\frac{\partial R}{\partial Q_a} + \frac{\partial P}{\partial T}\frac{\partial R}{\partial P_a}. \tag{4}$$

Here we see the ensemble sensitivity value is a sum of terms, the first the pure dynamical sensitivity ($\partial R/\partial T_a$), and the remaining consisting of the pure dynamical sensitivity with respect to all other variables (e.g., $\partial R/\partial Q_a$) multiplied by the relationships of temperature to those variables (e.g., $\partial Q/\partial T$). Ensemble sensitivity thus describes how a perturbation to a single state variable is spread throughout the rest of the atmospheric state space (through ensemble statistical relationships) and projected onto the entire dynamical sensitivity field [since the dynamical sensitivity to each state variable is represented within each term of the sum on the right-hand side of Eq. (4)]. In turn, ensemble sensitivity at a single point, with respect to a single variable, inherits both the domainwide relationships between that point/variable and all other points and variables, as well as the dynamical sensitivity with respect to every point and variable. These autocorrelations (the "domainwide relationships") at sensitivity time are a basic aspect of any single ensemble sensitivity value. Equation (2) was the basis of this interpretation, and that equation thus reveals both the fundamental difference

between ensemble and pure dynamical sensitivity and the multivariate nature of a univariate regression (more on this later). These fundamentals of ensemble sensitivity provide a greater perspective on several aspects of ESA which will be discussed now.

***Direct or indirect dynamics?*** A key use of ESA is to understand the dynamics relevant to high-impact weather events. For example, one might ask, "What atmospheric features control the coverage of rotating thunderstorms within a predicted area of convection 24 hours later?" ESA can provide one quick, potential answer to this question since it would show any significant relationships across the entire modeling domain between the atmospheric state a day prior and the chosen rotating convection coverage response function (although spurious correlations must be considered as discussed in the "ESA limitations" section). However, the answer ensemble sensitivity provides cannot distinguish between direct dynamical causation and indirect relationships. This reveals a major limitation of ESA for dynamical interpretation purposes, which can be completely understood through Eq. (2).

Consider the situation in Fig. 1 for example. In this case, one may aim to understand how storm coverage is controlled dynamically through the nature of different atmospheric features, several of which are shown to be sensitive through the ensemble sensitivity to 500-hPa geopotential height field. However, whereas pure dynamical sensitivity (as estimated by an adjoint model) would reveal any direct dynamical influence from a change to a given atmospheric feature, ensemble sensitivity shows us those dynamical features with a direct effect [the adjoint sensitivity term $\partial R/\partial \mathbf{X}_a$ in Eq. (2)] in addition to any other features related to those dynamical features [the statistical term $\mathbf{D}^{-1}\mathbf{A}$ in Eq. (2)]. In turn, far-off features like the ridge in the northwestern United States are shown to be sensitive in Fig. 1 because the ensemble statistics possess relationships to the pure dynamical sensitivities that directly control Florida storm coverage a day later. Modifying the northwestern U.S. ridge itself, however, would likely have no direct dynamical effect on the Florida convection. Unfortunately, this issue does not go away with respect to sensitive features much closer to, or upstream of, the response function location, illustrating the inability more generally of ensemble sensitivity to reveal whether direct or indirect dynamical processes are in play.

Overcoming this limitation simply involves additional analysis. Whereas ensemble sensitivity fields alone cannot distinguish between direct (causation) or indirect (association) dynamics themselves, they do reveal sensitive regions, that if perturbed independently, will subsequently reveal any direct dynamical influence. This type of experimentation effectively removes the statistical term from the fundamental building blocks of ensemble sensitivity, indicating whether only direct dynamics plays a role (as adjoint sensitivity fields would show). This allows one to use ensemble sensitivity as guidance to isolate the relevant direct dynamical processes affecting a chosen response function without the use of an adjoint model. These concepts illustrate why far-off features, sometimes well downstream, are very common throughout the domain within ensemble sensitivity fields, showing the interesting interplay across vast areas ultimately linked through large-scale dynamical evolution.

***How ensemble sensitivity fundamentals help us design ensemble sensitivity–based forecast tools.*** The fundamental basis of ensemble sensitivity in Eq. (2) can pave the way for the development of ensemble sensitivity–based tools for improving high-impact weather forecasts. Forecast sensitivity in any form can be valuable since it highlights areas where errors most degrade the prediction of a chosen forecast response. In turn, ensemble sensitivity has been leveraged in different ways as an operational tool to improve forecasts. One way has been to use sensitivity (in conjunction with initial condition uncertainty) to target observations that would produce the largest reduction in forecast response function uncertainty (e.g., Ancell and Hakim 2007; Hill et al. 2020). Another approach has been to

adjust ensemble forecasts in some way using ensemble sensitivity as a guide. For example, Madaus and Hakim (2015) show how forecasts can be directly adjusted once observations become available at early forecast times by changing later forecast variables based on covariances between the ensemble estimate of the observations and those later forecast variables. Ancell (2016) demonstrates a method that chooses subsets of the ensemble based on retaining members with the smallest errors in sensitive regions (both studies demonstrated forecast improvements). These adjustment techniques, whether objectively calculated or performed subjectively by forecasters, benefit from the numerous long-distance relationships to flow features (e.g., troughs, jet streaks, or gradients) revealed trough ESA. Unlike adjoint sensitivity, which tends to exist at smaller scales and is less obviously associated with discernable flow features (Ancell and Hakim 2007), ensemble sensitivity provides ample opportunity to make beneficial adjustments through the way it maps dynamical sensitivity across the domain with ensemble statistics.

Interestingly, what gives an advantage for adjustment techniques becomes a disadvantage for ensemble sensitivity–based observation targeting methods. Equation (2) reveals a substantial obstacle—since ensemble sensitivity cannot include localization around the pure dynamical sensitivity field at data assimilation time (since it is unknown explicitly), the impacts of targeted observations can be significantly overestimated since localization is typically applied when they are assimilated. This issue was found to be a major reason why ESA-based targeting might be problematic (Hill et al. 2020) and should be kept under consideration for any ensemble-based targeting scheme. While estimating the location of the pure dynamical sensitivity field may help mitigate this problem by applying a localization around that estimated location to reduce the ensemble sensitivity field, the success of such a technique is unclear given the likely errors in estimating pure dynamical sensitivity in the first place. Nonetheless, knowing the reasons behind this issue as expressed through Eq. (2) is the first step at improving ensemble sensitivity–based targeting.

***A tale of two sensitivities: Raw versus standardized.*** In addition to the raw ensemble sensitivity field (the regression coefficients between a chosen response function and the atmospheric state), the raw sensitivity can be multiplied by the ensemble standard deviation at each model grid point to produce "standardized sensitivity":

$$\frac{\partial R}{\partial \mathbf{X}_{\text{standardized}}} = \frac{\partial R}{\partial \mathbf{X}_{\text{raw}}} \text{Stdev}(\mathbf{X}), \tag{5}$$

where Stdev($\mathbf{X}$) represents the state variable standard deviation within the ensemble (for the same state variable to which sensitivity is calculated). Note that in some previous ESA studies (e.g., Torn and Romine 2015) this is referred to as "normalized sensitivity," although it is the same quantity and is hereafter referred to as "standardized sensitivity" since it is associated with a product and not a quotient. Mathematically, standardized sensitivity (which possesses the same units as the response function) in an estimate of the change in response function $\Delta R$ as a product of raw sensitivity and state variable spread. Since the same quantity ($\Delta R$) is produced at every point, standardized sensitivity allows for a fair comparison across all model variables of the estimated change in the response function due to a combination of sensitivity and the expected size of perturbations that project onto that sensitivity. Thus, unlike raw sensitivity that estimates the change in response from any arbitrary early forecast perturbation, standardized sensitivity considers the expected size of those early perturbations to estimate changes to the response in tune with the ensemble's early forecast uncertainty. Standardized sensitivity presents itself more as a predictability quantity as it considers how expected errors evolve dynamically, while raw sensitivity represents more of a fluid dynamics problem in that it shows sensitivity without any estimate of early forecast error.

To illustrate the differences between raw and standardized sensitivity, both quantities are shown in Fig. 3 with respect to the 6-h forecast 2-m dewpoint (this is a different case than shown in Fig. 1). The response function in this example is the 30–36-h number of grid points of 2–5-km updraft helicity greater than 50 $m^2 s^{-2}$ in the green box (a location where ensemble members forecast a range of helicity values, not shown). The highlighted area in the southwest corner of the raw sensitivity plot (left panel) shows very little sensitivity to the dryline shown by the strong gradient along the Texas–New Mexico border or the air mass behind it, indicating that changes to these features are not associated with large changes in rotating storm coverage the following day (particularly relative to much more sensitive features in north central Texas). The same dryline and trailing air mass in the standardized sensitivity field, however, possess the largest magnitudes in the field, revealing that the uncertainty in those features, combined with their sensitivity, is much more important than that in other areas. Thus, while the coverage of rotation (in the form of 2–5-km updraft helicity) within convection the next day in Missouri is less sensitive dynamically to the dryline and its trailing air mass in southwest Texas, it is those features in the 2-m dewpoint field which most contribute to the uncertainty of the rotating convection coverage response.

Statistically, standardized sensitivity can be written as

$$\frac{\partial R}{\partial \mathbf{X}_{\text{standardized}}} = \left[\frac{\text{Cov}(R,\mathbf{X})}{\text{Stdev}(\mathbf{X})\text{Stdev}(R)}\right]\text{Stdev}(R), \tag{6}$$

which is obtained by combining Eqs. (1) and (5) and multiplying both numerator and denominator by the response function standard deviation [Stdev($R$)]. This quantity is simply the correlation coefficient between the state variable and the response function (in brackets) multiplied by the standard deviation of the response function (a constant for an existing ensemble). In turn, when performing a linear regression between a response function and the atmospheric state, the predictability of the response is linked to the correlation coefficient (found through standardized sensitivity) while the sensitivity (with its more fluid dynamics perspective produced through raw sensitivity) is linked to the slope (regression coefficient). In this interesting way, one can examine with different statistical parameters of a regression how atmospheric features preceding a high-impact weather event might be associated in different ways with its predictability (correlation coefficient) or its dynamical sensitivity (regression coefficient) as shown in Fig. 3.

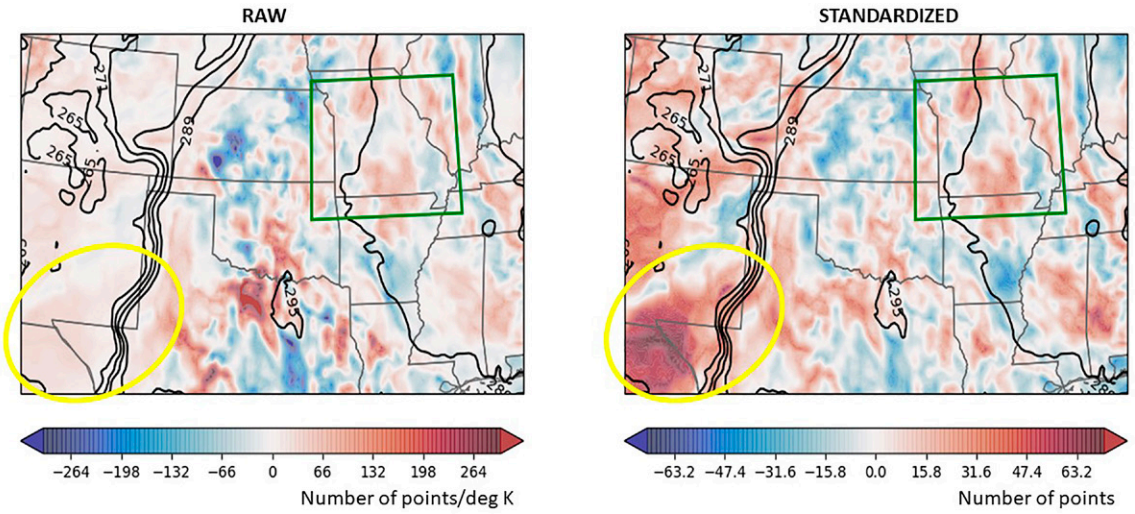


**Fig. 3. Raw and standardized ensemble sensitivity (shaded) of 30–36-h number of grid points exceeding 50 $m^2 s^{-2}$ 2–5-km updraft helicity in the green box to 6-h 2-m dewpoint and the ensemble mean 6-h 2-m dewpoint (black contours).**

***Can multivariate regression add value?*** Ensemble sensitivity is generated through a univariate regression of a response function onto each atmospheric state variable. Since the response is a function of forecast variables, which in turn are a function of atmospheric state variables earlier in time through the model governing equations, it is natural to suspect that information is being left out of ensemble sensitivity given its univariate nature, and that perhaps a multivariate regression would incorporate more data and usefulness. Both Hacker and Lei (2015) and Ren et al. (2019) explored this possibility. Hacker and Lei (2015) found some improvements in a low-dimensional model using multivariate regression, and Ren et al. (2019) showed very similar spatial patterns (but overestimations from traditional, univariate sensitivity) between the univariate and multivariate sensitivity fields within a more realistic typhoon simulation. Both studies, however, use a localization technique with the multivariate sensitivities which may be contributing to the differences presented.

Equations (1) and (2) here show that ensemble sensitivity, while calculated through a simple univariate regression, is actually a multivariate quantity by nature. In particular, Eq. (2) reveals that ensemble sensitivity simultaneously considers the covariance relationships between all other variables as well as the pure dynamical sensitivity of the entire state. In other words, the slope of a univariate linear regression valid at a single point is equivalent to the dynamical sensitivity of a domainwide perturbation. It would thus be expected that multivariate regression could add only redundant information to a traditional ensemble sensitivity value, and that the univariate nature of a single ensemble sensitivity value provides an appropriate, domainwide perspective with little expected additional value from multivariate techniques. This situation may explain the strong similarities found in Hacker and Lei (2015) and Ren et al. (2019) between univariate and multivariate sensitivities, although further investigation is needed to fully understand the differences that do occur. As in the prior discussion on targeting and its overestimation of the impacts of early perturbations on later forecasts due to localization, we speculate the localization in the studies investigating multivariate sensitivity may have caused the demonstrated similar overestimations.

***ESA limitations.*** Given the purpose of this article is to provide a complete guide to ESA, its limitations must be discussed. The most obvious ESA limitation is nonlinearity, which can manifest itself in different ways. If perturbation evolution is nonlinear (e.g., the adjoint model, based on tangent-linear principles, would poorly estimate how perturbations evolve), then the equality in Eq. (2) becomes an approximation, and ensemble sensitivity must be thought of as a term that varies based on the value of the atmospheric state (i.e., it is no longer a constant sensitivity). In general, nonlinearity becomes more significant as forecast time increases, but also depends on the response function and the scale of the structure it is designed to diagnose. Ancell and Mass (2006) discuss how the linear approximation holds for periods of 1–2 days at synoptic scales (and for ESA using response functions like midlatitude cyclone central pressure) but for convection (and response functions like storm rotation) last only hours. While these time scales provide some beneficial guidance on applying ESA for different weather phenomena, ensemble sensitivity can still provide useful insights when nonlinearity arises at longer forecast times—both at synoptic scales (e.g., Chang et al. 2013) and for convection (e.g., Coleman and Ancell 2020). In cases of convection, Coleman and Ancell (2020) show how nonlinearity can be less significant when larger response functions that diagnose rotation, for example, over potentially stormy areas are used instead of those that diagnose specific aspects of individual storms.

Ultimately the presence of significant nonlinear ensemble perturbation evolution should lead to weak linear relationships within the regressions that produce ensemble sensitivity values. If some nonlinear dependence is strong (e.g., a scatterplot that exhibits a parabola shape

when plotting response function/sensitivity variable pairs), nonlinear relationships might be apparent which could be detected with nonlinear regressions, adding a higher-order correction to ESA. This type of analysis may help particularly when fine-scale processes like convection are considered that possess substantial nonlinearity (Limpert and Houston 2018; Hill et al. 2020) and is suggested as an avenue for future ESA enhancement. Such higher-order regressions might identify early atmospheric dependencies that reveal "Goldilocks" behavior for severe weather like tornadoes (Markowski and Richardson 2014), where neither high nor low temperature values in some portion of the storm's evolution, for example, result in tornadoes, but temperatures in-between are associated with them. Nonlinearity also becomes an issue when the response function is not continuously distributed such as when a bimodal ensemble distribution may show members with strong convection and those with no convection. In this case, while ESA will show the correct trend (Hill et al. 2016), it is unrealistic to interpret changes to the state producing a linear response that results in weak convection (since only strong convection or the lack of convection altogether were forecast as possibilities). While it is difficult to assess the precise degree of nonlinearity in any ESA application, inspection of the response function distribution is the best way to determine the validity of the linear assumption—continuous, unimodal distributions that show significance in the regression coefficient are likely good candidates for ESA and its applications discussed here.

Another issue with ensemble sensitivity occurs within an ensemble that suffers from too much or too little spread. Since the calculation of ensemble sensitivity involves the variance of the sensitivity-time state variable in the denominator [Eq. (1)], very small variance values can inflate raw sensitivities substantially. The authors have witnessed on numerous occasions that sensitivities to surface variables such at 2-m temperature or 10-m winds within a Weather Research and Forecasting (WRF; Skamarock et al. 2008) Model ensemble are several orders of magnitude larger than the same types of variables slightly aloft (corresponding to too little ensemble variance in those surface variables; not shown). Tiny spread at the surface associated with underdispersiveness there relates to much larger spread aloft (where spread is appropriate) through the ensemble statistics, and thus a unit change in surface variables, which may be much larger than the ensemble spread, is associated with massive perturbations aloft. These massive perturbations are then the cause of huge changes to the response function through the ensemble sensitivity value in Eq. (3). Turning this argument around, overdispersive areas will experience very small raw sensitivity values.

From the perspective of the ensemble, which has no knowledge of whether or not it is appropriately calibrated, none of this is a problem—the ensemble expects only tiny perturbations at the surface, for example, such that the seemingly overinflated raw sensitivity values there still estimate reasonable response function perturbations. It is only when ensembles are truly over- or underdispersive when very large (or too small) raw sensitivities become an issue. For targeted observing, for example, if near-surface observations deviated from the ensemble mean to a degree that was much larger than the underdispersive ensemble spread in that variable, the ensemble expected perturbation size is violated, leading to truly overinflated estimates to a change in response function through the raw sensitivity field. In turn, ESA performed with poorly calibrated ensembles will suffer if applied outside the ensemble more generally. Thus, ensemble spread characteristics and the general quality of the ensemble must be scrutinized when developing ESA-based tools. It should be noted that ESA applies most appropriately within ensembles that vary by their initial conditions since sensitivity itself relates to how initial or early forecast perturbations relate to differences in the response later in time. If an ensemble was generated by different physics schemes, all with identical initial conditions, ensemble sensitivity at initial time loses its meaning since a regression onto all members with the same value provides no useful relationship. While

ESA within ensembles based purely on physics variability may provide useful sensitivity values once ensemble members have evolved in time, it remains most appropriate to apply ESA within an ensemble driven by differences in initial conditions.

Last, spurious correlations are always possible when performing millions of linear regressions of a chosen response onto a large atmospheric state space. It is likely that some of the very small-scale variability that is common in sensitivity fields (as can be seen in Fig. 1 in the Gulf of Mexico, for example) exist due to sample size limitations of the ensemble in the form of these spurious correlations. These issues can be objectively addressed through statistical testing of the regression coefficient (as shown in Ancell and Hakim 2007), a process that will reveal whether the regression slope is significantly different than zero at some confidence level. Whether sensitivities are significant depends partly on the ensemble size, which is usually practically limited due to computational constraints to around 100 members or less (e.g., Torn and Romine 2015; Limpert and Houston 2018; Bednarczyk and Ancell 2015; a number much less than the millions of state variables sampled). Thus, general practice usually involves running the largest ensemble allowed by the available computational resources and assessing the resulting sensitivities for significance. In addition to objective statistical significance testing [as performed in Torn and Romine (2015), for example], the temporal evolution of sensitivity signals can be examined subjectively to reveal any continuity of the signal, which for several adjacent grid points is very unlikely to be associated with random spurious correlations.

## Application of new perspectives to a climatology of severe storms

To demonstrate some of the new perspectives discussed here, ESA is applied to a climatology of severe storms over a period from late April to early June 2016. This application to severe storms is made here because 1) ESA studies involving convection nearly all involve a small number of cases and demonstrating unique ESA perspectives within many cases of severe storms can provide new insights, and 2) adjoint sensitivity, which lies at the heart of ESA interpretation via Eq. (2), is not computationally feasible at convective scales and thus ESA can provide a framework to answer research questions involved with forecast sensitivity. For each 0000 and 1200 UTC ensemble initialization from 27 April to 3 June, response function rectangles were chosen for 6-h periods within the 48-h ensemble forecasts within which at least 10%–20% probabilities of 2–5-km updraft helicity exceeding 25 $m^2 s^{-2}$ occurred over the 6-h window (response windows started no earlier than 12-h forecast time such that they always exist after the 6-h sensitivity time examined below). These response boxes were subjectively chosen by the authors and varied in size to capture the probability signals associated with rotating convection. Figure 4 shows an example of one such response box, which was chosen to contain significant probabilities of neighborhood maximum hourly 2–5-km updraft helicity exceeding 25 $m^2 s^{-2}$ throughout the 6-h response time. Multiple response boxes existed for some initializations as there sometimes were several separate areas of convection, and these areas may exhibit different sensitivities. Other studies, such as the Mesoscale Predictability Experiment (MPEX; Weisman et al. 2015), also included ESA involved with numerous cases of convection and might provide additional insights on case-to-case variability of ensemble sensitivity fields.

Here we chose large response boxes that encompassed the entire probability signal over the 6-h time frame (as in Fig. 4, thus diagnosing the event as a whole), although in principle the choice of response location goes hand in hand with the problem of interest. For example, a very small response box placed in an area where convection has different forecast timing among ensemble members will show a sensitivity field that demonstrates a positional signal of early atmospheric features (e.g., a dipole of sensitivity with maxima on either side of a geopotential height trough). This is because the small response box contains

less convection from the members further downstream, and thus the range of response values reflects the timing of convection. A much larger response box (in space and perhaps time) would capture the convection for all members irrespective of the timing details and would demonstrate a sensitivity signal relevant to the magnitude of the event (e.g., a single maximum of sensitivity over a geopotential height trough). Bednarczyk and Ancell (2015) provide a detailed discussion of how the response box size and location influences the sensitivity field in this way.

The ensemble forecasts (42 total) in this study came from the real-time Texas Tech University prediction system, which utilizes the WRF Model with 38 vertical levels and the Data Assimilation Research Testbed (DART; Anderson et al. 2009) 42-member ensemble adjustment Kalman filter with the parameters described in Ancell et al. (2015). Four response functions were considered on a nested 4-km domain (4-km domain shown by the blue box in Fig. 1) encompassing the U.S. South Plains and Midwest: 1) maximum surface simulated reflectivity, 2) number of grid points exceeding 40-dB*Z* surface simulated reflectivity, 3) maximum 2–5-km updraft helicity, and 4) number of grid points exceeding 50 $m^2 s^{-2}$ 2–5-km updraft helicity. Ensemble sensitivity of these response functions were calculated with respect to the 6-h forecast three-dimensional (all model grid points) temperature, geopotential height, wind speed, and water vapor mixing ratio fields on the parent 12-km domain over the continental United States (the larger domain shown in Fig. 1). This strategy takes advantage of WRF's nesting capability within a sensitivity framework, allowing response functions that diagnose convection at convection-allowing scales (e.g., the 4-km domain) while also permitting sensitivity to be calculated on a much larger, coarser grid (the 12-km domain). The 6-h sensitivity time is used here as it produces sensitivities at an early forecast time when they could be used to potentially improve forecasts in a real-time framework.

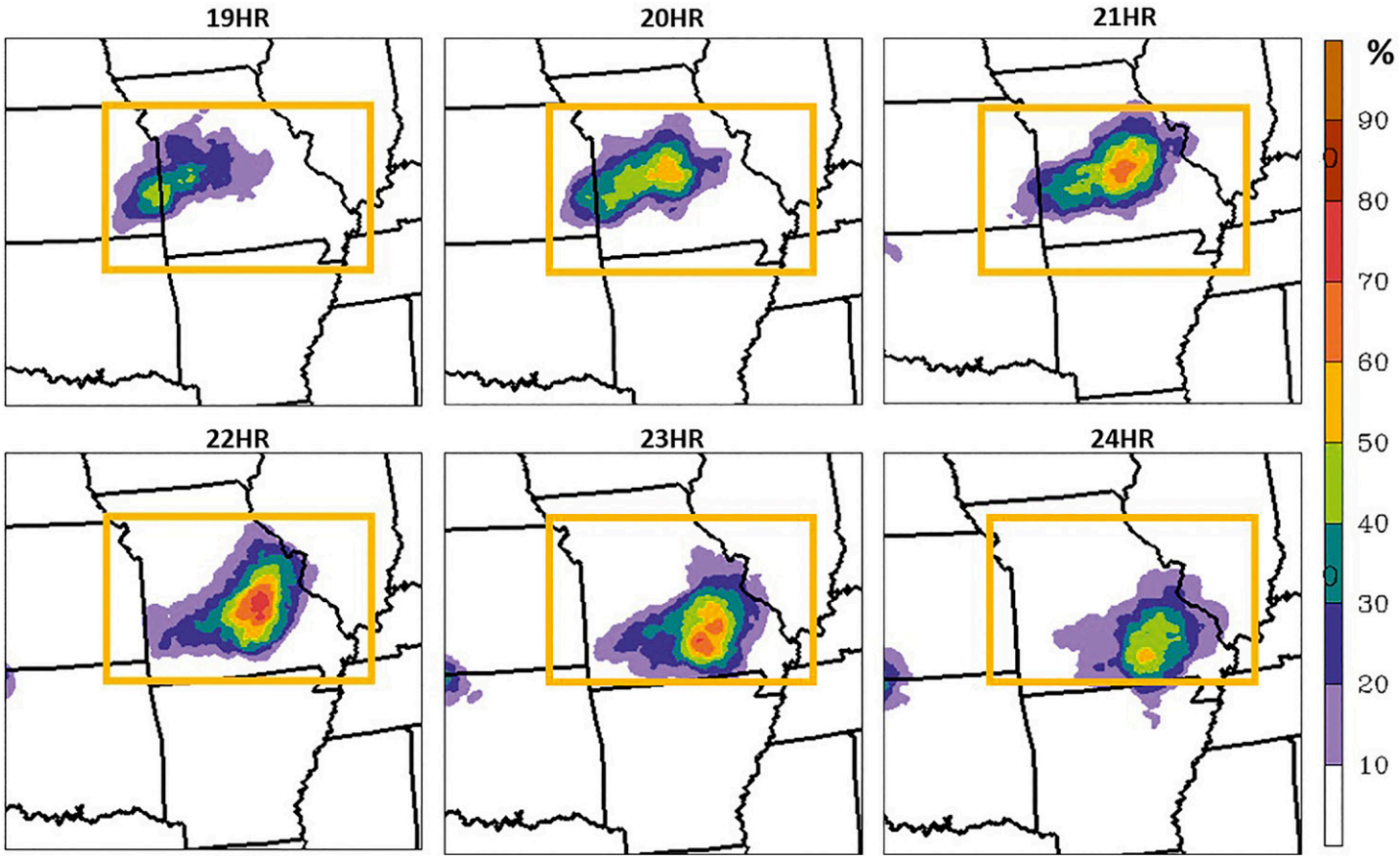


**Fig. 4. Probability of maximum hourly 2–5-km updraft helicity exceeding 25 $m^2 s^{-2}$ within 20 miles of a point over a 6-h forecast period with the chosen response function box shown in orange.**

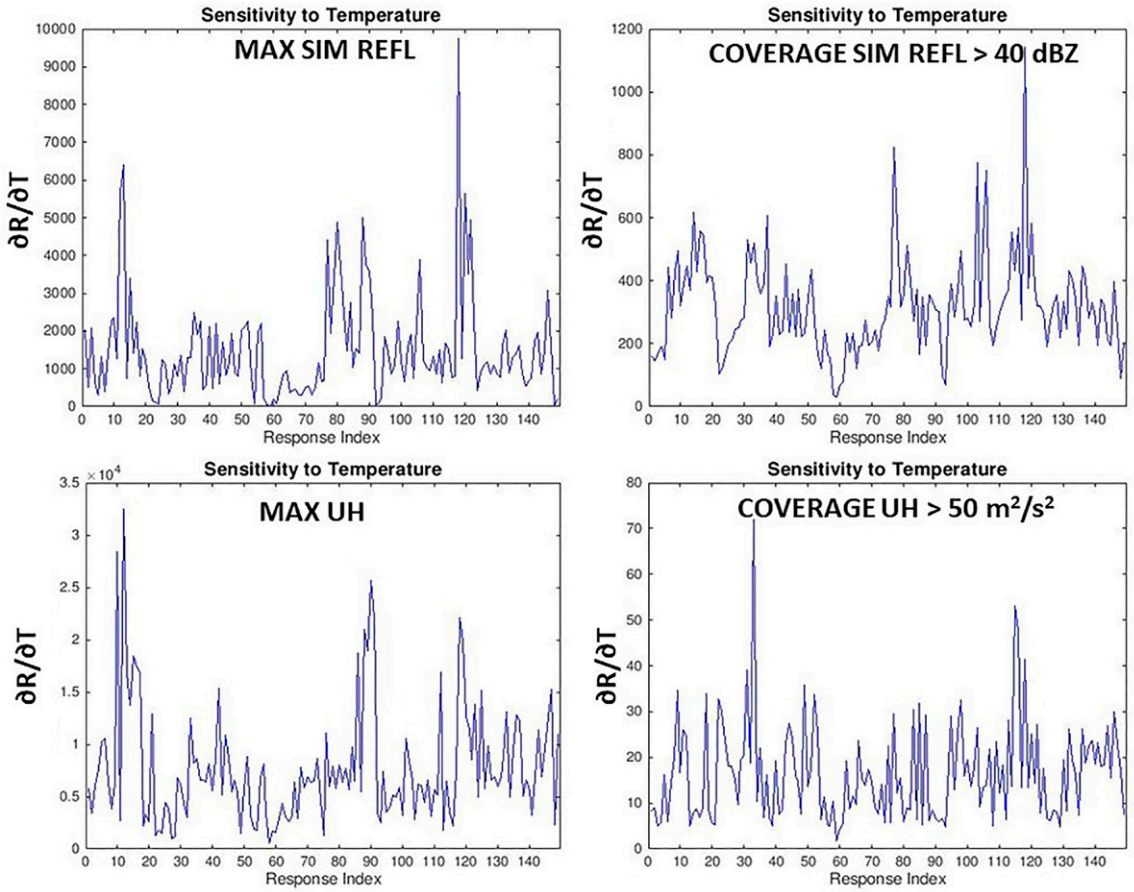


**Fig. 5. Maximum magnitude of the raw ensemble sensitivity field with respect to temperature over all cases for all four response functions.**

Figure 5 shows the maximum magnitude of the raw ensemble sensitivity field across the entire three-dimensional 12-km domain for all four response functions with respect to temperature for all response boxes (about 150 total). Sensitivity values within 10 grid points of the horizontal boundaries are not considered (here and throughout the study), and the values are not tested for statistical significance. These plots show how sensitivity of any response varies significantly across many cases, providing the range of sensitivities for which convective cases can be judged as highly or weakly sensitive. Picking out the cases with the highest and lowest sensitivity values can further reveal whether any specific characteristics of the flow, such as the jet stream pattern, are associated with the largest or smallest dynamic sensitivity or predictability (depending on whether raw or standardized sensitivity is used to do this). For example, Figs. 6 and 7 show the 500-hPa geopotential height field at 6-h forecast time for the six largest (Fig. 6) and smallest (Fig. 7) values of mean raw three-dimensional sensitivity to temperature for the updraft helicity coverage response. Clearly a consistent pattern with an upstream trough emerges for the most sensitive cases, while the least sensitive cases reveal a more mixed collection of flows. Note that these results are presented from a response perspective in that the most or least sensitive cases may exist for the same 48-h forecast periods (although different response boxes still identify different signals and 6-h time windows within those periods), and results may change if the 48-h forecast periods were not allowed to overlap. As in McMurdie and Ancell (2014), this type of analysis may reveal the first semblance of a link between the general flow characterized by 500-hPa geopotential height and the degree of domainwide sensitivity. If standardized sensitivity were instead examined, different flows may be associated with the strongest and weakest predictability of severe convection. Since all cases in this example use a response function that diagnoses

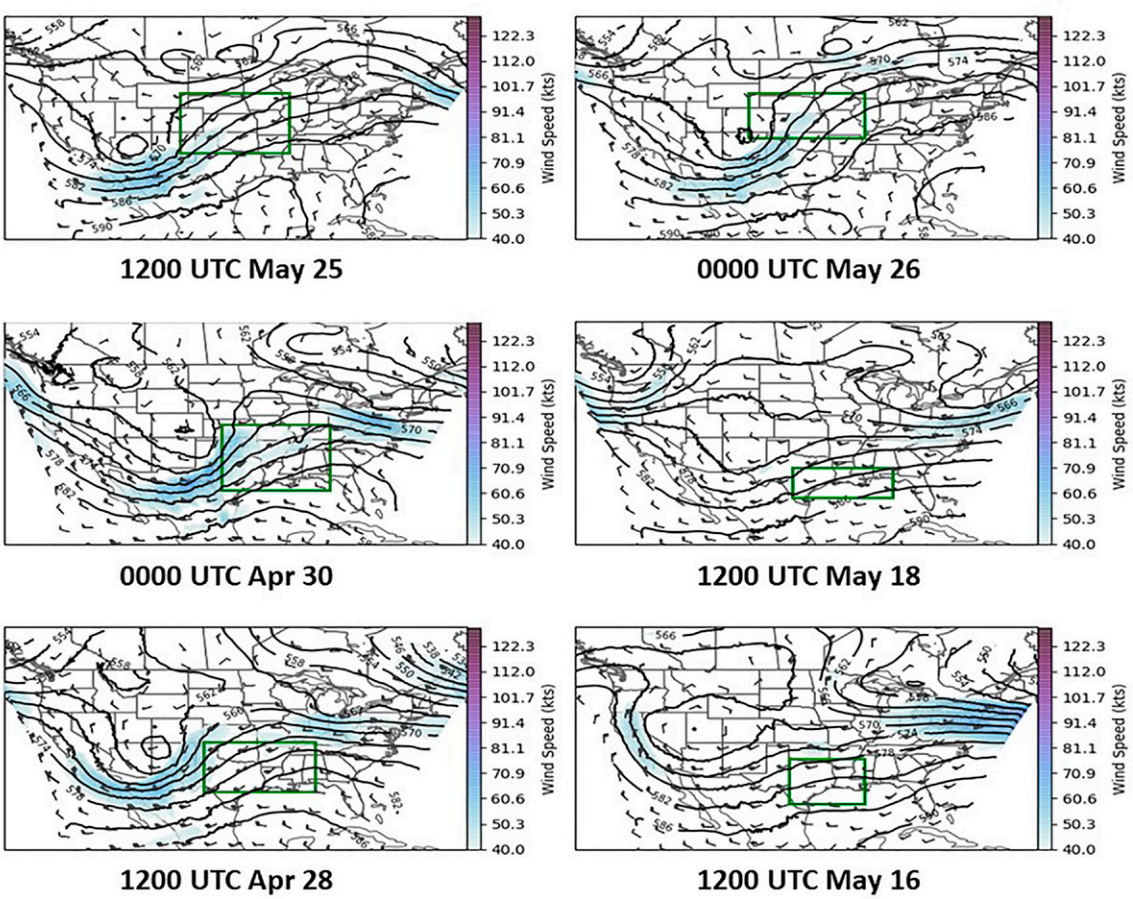


**Fig. 6. The 500-hPa ensemble mean geopotential height field (black contours), wind vectors, and wind speed (shaded) valid at 6 h for the six most sensitive cases (response function location is shown by the green box).**

rotating convection (the response boxes were all chosen to surround substantial probabilities of updraft helicity), a deeper dive with this type of analysis can help us understand differences in the dynamics and predictability among various flow types for specific convection characteristics like rotation.

Interestingly, when different response functions are examined together, relatively loose correlations exist. While some response box instances show similarities for all response functions (such as that around response 59 in Fig. 5 where all response functions reveal a low sensitivity), it is not hard to notice the numerous other response instances where relatively large or small values do not occur together. In fact, if we look more closely at how sensitivity of different response functions (coverage of simulated reflectivity versus updraft helicity) for the same event correlate (Fig. 8, which in this case shows standardized sensitivity to wind speed) we see a very weak relationship for both maximum magnitude and mean absolute value of the full three-dimensional sensitivity fields. These results are fairly universal across the climatology—they are consistent whether or not we consider raw or standardized sensitivity, different response functions, different variables to which the sensitivity is calculated, or whether we examine the maximum or mean (not shown). Analyzing the relationship between mean and maximum sensitivity is yet another possible way to analyze sensitivity across many cases that may provide insight into whether large, localized sensitivity can exist without more extensive domainwide sensitivity (which would be characterized by the mean value).

These results indicate that it is common for different characteristics of the same storms to be sensitive to features earlier in time in different ways. While this might be obvious at the time of convection (e.g., more vertical shear might increase the number of rotating storms

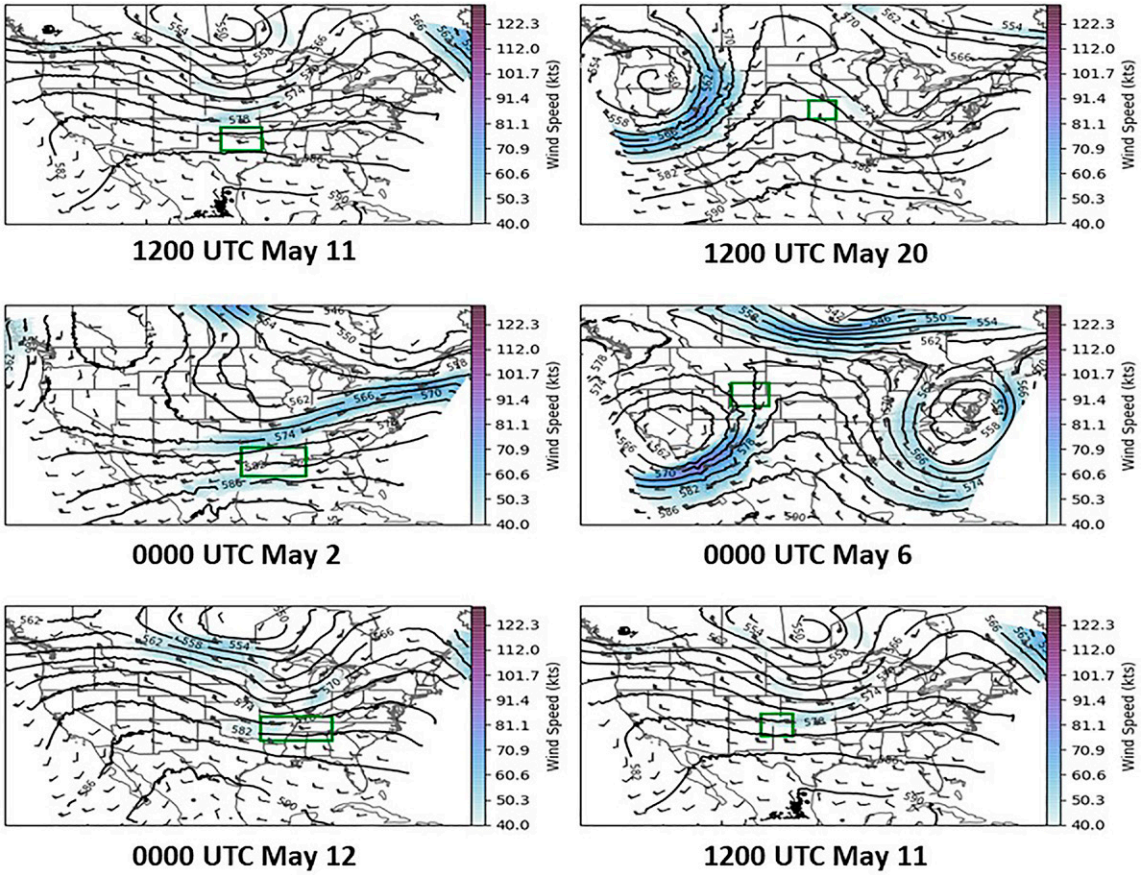


**Fig. 7. The 500-hPa ensemble mean geopotential height field (black contours), wind vectors, and wind speed (shaded) valid at 6 h for the six least sensitive cases (response function location is shown by the green box).**

while the total number of storms—rotating or not—remains the same), this suggests that cleverly designed ESA may reveal new knowledge about how the dynamics and predictability of individual severe hazards relate to different precursors. This idea is supported by Fig. 9, which shows histograms of mean standardized sensitivity for all four response functions with respect to the four sensitivity variables considered here (temperature, geopotential height, wind speed, and water vapor mixing ratio). The shapes of the histograms look fairly similar for a given response function across all sensitivity variables, but exhibit obvious differences for any single variable across different response functions, suggesting fundamental differences in the way sensitivity manifests itself for different storm characteristics (but not so much across different sensitivity variables).

Figure 10 provides an interesting perspective across the climatology of events on both raw and standardized sensitivity, and how, as described above, they pertain to a dynamics problem and a predictability problem, respectively. Raw and standardized sensitivity of 2–5-km updraft helicity coverage to 6-h wind speed is shown (both the mean and max value) for all cases. It is clear that the mean sensitivity fields correlate well across all events. However, when considering the maximum in the standardized and raw sensitivity fields, this relationship is substantially weaker (correlation coefficient of 0.76 compared to 0.98 when considering mean sensitivity), although a general positive relationship still exists. This reveals that the largest intrinsic potential for perturbation growth is not always associated with low predictability (and vice versa), which must be a result of small ensemble spread where the dynamical perturbation growth is fastest (or large uncertainty where perturbation growth is small). In any case, this shows that frequently the potential for perturbation growth highlighted by

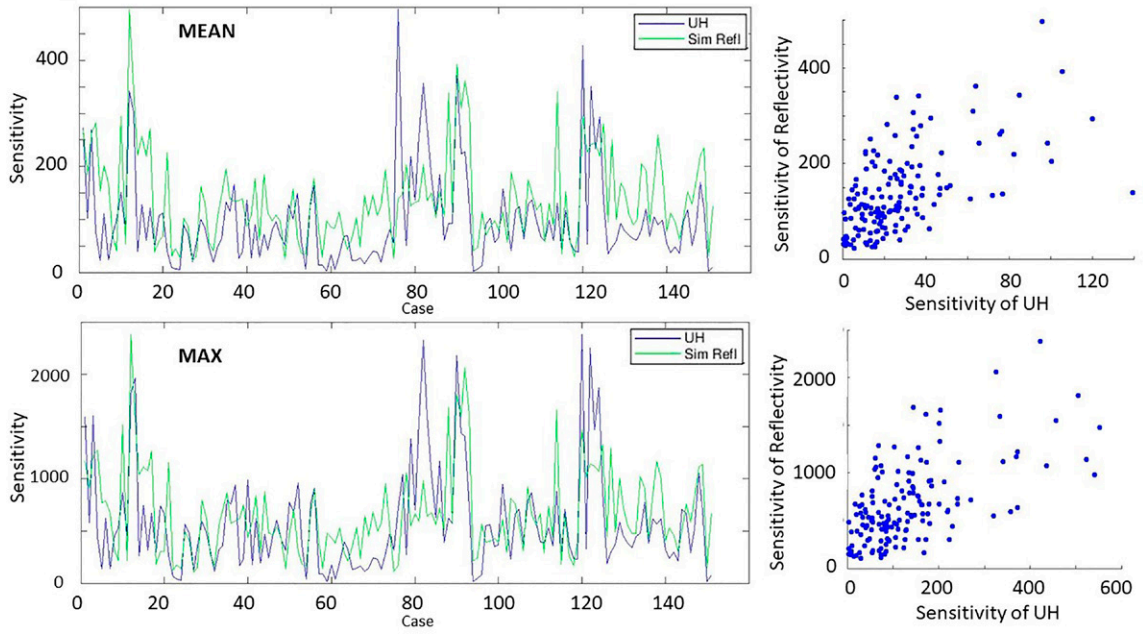


**Fig. 8. (left) Mean and maximum standardized sensitivity to wind speed for both the updraft helicity and simulated reflectivity coverage response functions (for visual purposes updraft helicity sensitivity is normalized to produce the same maximum value as simulated reflectivity sensitivity) and (right) the scatterplot of the same data for all cases.**

raw sensitivity (the dynamics problem) and the actual perturbation growth revealed by the way expected perturbations interact with perturbation growth potential (the predictability problem) are at odds at least for a response function diagnosing rotation and a wind speed sensitivity variable (it would be interesting to examine whether this holds for other response functions and sensitivity variables). In other words, simultaneous analysis of the raw and standardized sensitivity fields can reveal the frequency and degree to which specific atmospheric features tap into their potential to affect high-impact weather. Such analysis might reveal how our practical predictability (Melhauser and Zhang 2012) could be improved for certain types or regimes of convection, something that was explored for midlatitude cyclones in McMurdie and Ancell (2014).

## Summary

Our primary goal here was to establish a complete guide to the use of ensemble sensitivity analysis (ESA) toward enhancing its value as a research and operational tool. We felt this effort to be beneficial to the field of atmospheric sciences since many studies have employed some form of ESA with interesting results, but the broader context of how, when, and why to apply ESA, particularly in light of its fundamental principles, has been less clear. In turn, we hope the ESA roadmap presented here will help a powerful tool become more commonplace. Several key aspects and new perspectives of ESA have been demonstrated in this study:

- Ensemble sensitivity is a product of two fundamental building blocks: the pure dynamics of the system and ensemble statistics describing the relationships of atmospheric variables to each other.
- Understanding the fundamental building blocks is a key first step to guide the design of ensemble sensitivity–based research and operational tools.
- Raw ensemble sensitivity (the regression coefficient) can be compared to standardized ensemble sensitivity (raw sensitivity multiplied by the ensemble standard deviation,

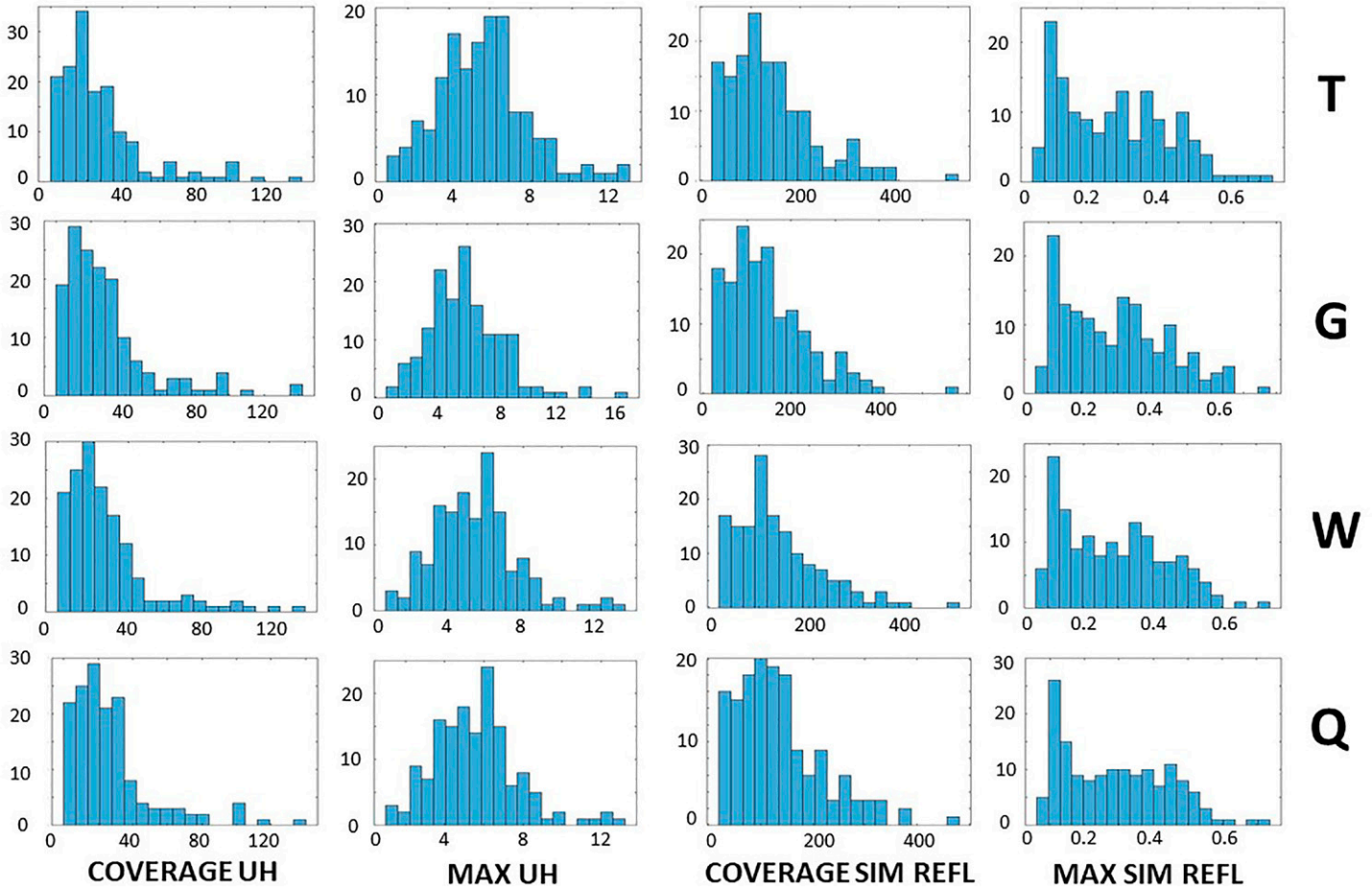


**Fig. 9. Histograms of the mean standardized sensitivity (*x* axis is magnitude, *y* axis is number of occurrences) to the four sensitivity variables used in this study (*T* = temperature, *G* = geopotential height, *W* = wind speed, *Q* = mixing ratio) for all four response functions.**

proportional to the correlation coefficient) to reveal differences between dynamical perturbation growth characteristics and the predictability associated with chosen response functions.
- Ensemble sensitivity, while a univariate regression, is multivariate in nature, likely reducing the value of multivariate regression techniques within sensitivity analysis.

We also examined ensemble sensitivity within a short climatology of severe convection to help demonstrate some practical applications of ESA that relate to the aforementioned key principles and new perspectives. We showed that

- substantial variability of mean and maximum sensitivity exists across a large number of convective events with the most sensitive cases showing a preferential flow pattern aloft,
- different aspects of the same convective event frequently exhibit different sensitivity characteristics, and
- maximum raw and standardized sensitivity across many convective cases showed a weak relationship, revealing for each case how the resulting predictability depends on how ensemble uncertainty taps into the potential for dynamical error growth.

While these initial results are intentionally broad to allow the presentation of several examples, we hope they highlight some of the unique ways ESA can be subsequently applied more thoroughly to help advance knowledge within atmospheric sciences dynamics and

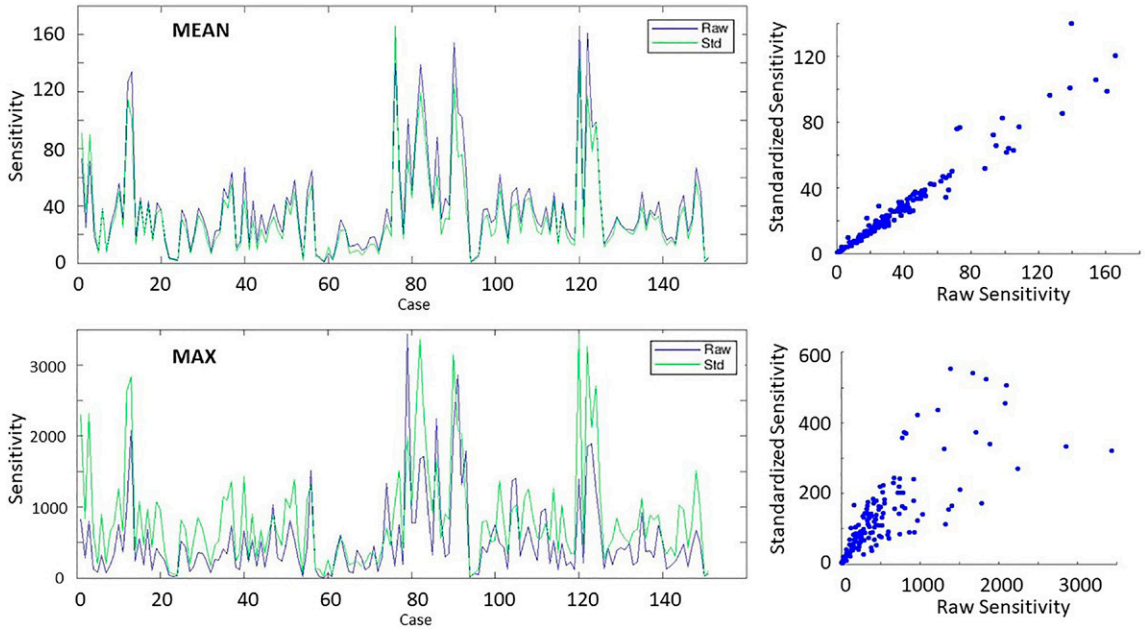


**Fig. 10. (left) Mean and maximum raw and standardized sensitivity of 2–5-km updraft helicity to wind speed (for visual purposes standardized sensitivity is normalized to produce the same maximum value as raw sensitivity) and (right) the scatterplot of the same data for all cases.**

predictability problems. Ultimately ESA is a form of data mining, and other more complex data mining schemes such as those based on machine learning algorithms could potentially use more sophistication to achieve some of the same goals of ESA. As a start, more sophisticated methods could allow for nonlinearity in a way ESA cannot. However, ESA's fundamental properties owing to the interaction of covariance relationships and dynamical evolution (not to mention its lack of reliance on a training dataset) is an attractive characteristic that will always keep ESA and its wide array of uses demonstrated here well grounded in core principles.

**Acknowledgments.** The authors wish to thank the Texas Tech High Performance Computing Center which maintained the computing cluster on which this work was performed. This study was supported by NOAA Grant NA20NWS4680045.

## References


Ancell, B. C., 2016: Improving high-impact forecasts through sensitivity-based ensemble subsets: Demonstration and initial tests. *Wea. Forecasting*, **31**, 1019–1036, https://doi.org/10.1175/WAF-D-15-0121.1.

——, and C. F. Mass, 2006: Structure, growth rates, and tangent linear accuracy of adjoint sensitivities with respect to horizontal and vertical resolution. *Mon. Wea. Rev.*, **134**, 2971–2988, https://doi.org/10.1175/MWR3227.1.

——, and G. J. Hakim, 2007: Comparing ensemble and adjoint sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, https://doi.org/10.1175/2007MWR1904.1.

——, E. Kashawlic, and J. L. Schroeder, 2015: Evaluation of wind forecasts and observation impacts from variational and ensemble data assimilation for wind energy applications. *Mon. Wea. Rev.*, **143**, 3230–3245, https://doi.org/10.1175/MWR-D-15-0001.1.

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

Bednarczyk, C. N., and B. C. Ancell, 2015: Ensemble sensitivity analysis applied to a Southern Plains convective event. *Mon. Wea. Rev.*, **143**, 230–249, https://doi.org/10.1175/MWR-D-13-00321.1.

Berman, J. D., and R. D. Torn, 2019: The impact of initial condition and warm conveyor belt forecast uncertainty on variability in the downstream waveguide in an ECWMF case study. *Mon. Wea. Rev.*, **147**, 4071–4089, https://doi.org/10.1175/MWR-D-18-0333.1.

——, ——, G. S. Romine, and M. L. Weisman, 2017: Sensitivity of northern Great Plains convection forecasts to upstream and downstream forecast errors. *Mon. Wea. Rev.*, **145**, 2141–2163, https://doi.org/10.1175/MWR-D-16-0353.1.

Chang, E. K. M., M. Zhang, and K. Raeder, 2013: Medium-range ensemble sensitivity analysis of two extreme Pacific extratropical cyclones. *Mon. Wea. Rev.*, **141**, 211–231, https://doi.org/10.1175/MWR-D-11-00304.1.

Coleman, A. A., and B. C. Ancell, 2020: Toward the improvement of high-impact probabilistic forecasts with a sensitivity-based convective-scale ensemble subsetting technique. *Mon. Wea. Rev.*, **148**, 4995–5014, https://doi.org/10.1175/MWR-D-20-0043.1.

Errico, R. M., 1997: What is an adjoint model? *Bull. Amer. Meteor. Soc.*, **78**, 2577–2591, https://doi.org/10.1175/1520-0477(1997)078<2577:WIAAM>2.0.CO;2.

Hacker, J. P., and L. Lei, 2015: Multivariate ensemble sensitivity with localization. *Mon. Wea. Rev.*, **143**, 2013–2027, https://doi.org/10.1175/MWR-D-14-00309.1.

Hakim, G. J., and R. D. Torn, 2008: Ensemble synoptic analysis. *Synoptic–Dynamic Meteorology and Weather Analysis and Forecasting: A Tribute to Fred Sanders*, *Meteor. Monogr.*, No. 55, Amer. Meteor. Soc., 147–162, https://doi.org/10.1007/978-0-933876-68-2_7.

Hanley, K. E., D. J. Kirshbaum, N. M. Roberts, and G. Leoncini, 2013: Sensitivities of a squall line over central Europe in a convective-scale ensemble. *Mon. Wea. Rev.*, **141**, 112–133, https://doi.org/10.1175/MWR-D-12-00013.1.

Hill, A. J., C. C. Weiss, and B. C. Ancell, 2016: Ensemble sensitivity analysis for mesoscale forecasts of convection initiation. *Mon. Wea. Rev.*, **144**, 4161–4182, https://doi.org/10.1175/MWR-D-15-0338.1.

——, ——, and ——, 2020: Factors influencing ensemble sensitivity–based targeted observing predictions at convection-allowing resolutions. *Mon. Wea. Rev.*, **148**, 4497–4517, https://doi.org/10.1175/MWR-D-20-0015.1.

Hu, C.-C., and C.-C. Wu, 2020: Ensemble sensitivity analysis of tropical cyclone intensification rate during the development stage. *J. Atmos. Sci.*, **77**, 3387–3405, https://doi.org/10.1175/JAS-D-19-0196.1.

Kerr, C. A., D. J. Stensrud, and X. Wang, 2019: Diagnosing convective dependencies on near-storm environments using ensemble sensitivity analyses. *Mon. Wea. Rev.*, **147**, 495–517, https://doi.org/10.1175/MWR-D-18-0140.1.

Limpert, G. L., and A. L. Houston, 2018: Ensemble sensitivity analysis for targeted observations of supercell thunderstorms. *Mon. Wea. Rev.*, **146**, 1705–1721, https://doi.org/10.1175/MWR-D-17-0029.1.

Madaus, L. E., and G. J. Hakim, 2015: Rapid, short-term ensemble forecast adjustment through offline data assimilation. *Quart. J. Roy. Meteor. Soc.*, **141**, 2630–2642, https://doi.org/10.1002/qj.2549.

Markowski, P. M., and Y. P. Richardson, 2014: What we know and don't know about tornado formation. *Phys. Today*, **67**, 26–31, https://doi.org/10.1063/PT.3.2514.

McMurdie, L. A., and B. C. Ancell, 2014: Predictability characteristics of landfalling cyclones along the North American west coast. *Mon. Wea. Rev.*, **142**, 301–319, https://doi.org/10.1175/MWR-D-13-00141.1.

Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, https://doi.org/10.1175/JAS-D-11-0315.1.

Nystrom, R. G., F. Zhang, E. B. Munsell, S. A. Braun, J. A. Sippel, Y. Wang, and K. A. Emanuel, 2018: Predictability and dynamics of Hurricane Joaquin (2015) explored through convection-permitting ensemble sensitivity experiments. *J. Atmos. Sci.*, **75**, 401–424, https://doi.org/10.1175/JAS-D-17-0137.1.

Parker, T., T. Woollings, and A. Weisheimer, 2018: Ensemble sensitivity analysis of Greenland blocking in medium-range forecasts. *Quart. J. Roy. Meteor. Soc.*, **144**, 2358–2379, https://doi.org/10.1002/qj.3391.

Quandt, L.-A., and Coauthors, 2019: Ensemble sensitivity analysis of the blocking system over Russia in summer 2010. *Mon. Wea. Rev.*, **147**, 657–675, https://doi.org/10.1175/MWR-D-18-0252.1.

Ren, S., L. Lei, Z.-M. Tan, and Y. Zhang, 2019: Multivariate ensemble sensitivity analysis for Super Typhoon Haiyan (2013). *Mon. Wea. Rev.*, **147**, 3467–3480, https://doi.org/10.1175/MWR-D-19-0074.1.

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

Smith, N. H., and B. C. Ancell, 2017: Ensemble sensitivity analysis of wind ramp events with applications to observation targeting. *Mon. Wea. Rev.*, **145**, 2505–2522, https://doi.org/10.1175/MWR-D-16-0306.1.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328, https://doi.org/10.1002/qj.49711347812.

Torn, R. D., 2010: Ensemble-based sensitivity analysis applied to African easterly waves. *Wea. Forecasting*, **25**, 61–78, https://doi.org/10.1175/2009WAF2222255.1.

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

——, and ——, 2009: Initial condition sensitivity of western Pacific extratropical transitions determined using ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **137**, 3388–3406, https://doi.org/10.1175/2009MWR2879.1.

——, and G. S. Romine, 2015: Sensitivity of central Oklahoma convection forecasts to upstream potential vorticity anomalies during two strongly forced cases during MPEX. *Mon. Wea. Rev.*, **143**, 4064–4087, https://doi.org/10.1175/MWR-D-15-0085.1.

Weisman, M. L., and Coauthors, 2015: The Mesoscale Predictability Experiment (MPEX). *Bull. Amer. Meteor. Soc.*, **96**, 2127–2149, https://doi.org/10.1175/BAMS-D-13-00281.1.

Wile, S. M., J. P. Hacker, and J. H. Chilcoat, 2015: The potential utility of high-resolution ensemble sensitivity analysis for observation placement during weak flow in complex terrain. *Wea. Forecasting*, **30**, 1521–1536, https://doi.org/10.1175/WAF-D-14-00066.1.

Zack, J., E. Natenberg, S. Young, J. Manobianco, and C. Kamath, 2010a: Application of ensemble sensitivity analysis to observation targeting for short-term wind speed forecasting. Lawrence Livermore National Laboratory Tech. Rep. LLNL-TR-42442, 32 pp.

——, ——, ——, G. V. Knowe, K. Waight, J. Manobianco, and C. Kamath, 2010b: Application of ensemble sensitivity analysis to observation targeting for short-term wind speed forecasting in the Washington-Oregon region. Lawrence Livermore National Laboratory Tech. Rep. LLNL-TR-458086, 65 pp.

——, ——, ——, ——, ——, ——, and ——, 2010c: Application of ensemble sensitivity analysis to observation targeting for short-term wind speed forecasting in the Tehachapi region winter season. Lawrence Livermore National Laboratory Tech. Rep. LLNL-TR-460956, 57 pp.

Zheng, M., E. K. M. Chang, and B. Colle, 2013: Ensemble sensitivity tools for assessing extratropical cyclone intensity and track predictability. *Wea. Forecasting*, **28**, 1133–1156, https://doi.org/10.1175/WAF-D-12-00132.1.