Nonlinear Characteristics of Ensemble Perturbation Evolution and Their Application to Forecasting High-Impact Events

BRIAN C. ANCELL
Texas Tech University, Lubbock, Texas

(Manuscript received 4 September 2012, in final form 30 May 2013)

ABSTRACT

Ensemble forecasting is becoming an increasingly important aspect of numerical weather prediction. As ensemble perturbation evolution becomes more nonlinear as a forecast evolves, the ensemble mean can diverge from the model attractor on which ensemble members are constrained. In turn, the ensemble mean can become increasingly unrealistic, and although statistically best on average, it can provide poor forecast guidance for specific high-impact events. This study uses an ensemble Kalman filter to investigate this behavior at the synoptic scale for landfalling midlatitude cyclones. This work also aims to understand the best way to select “best members” closest to the mean that both behave realistically and possess the statistically beneficial qualities of the mean. It is found that substantial nonlinearity emerges within forecast times of a day, which roughly agrees with previous research addressing synoptic-scale nonlinearity more generally. The evolving nonlinearity results in unrealistic behavior of the ensemble mean that significantly underestimates precipitation and wind speeds associated with the cyclones. Choosing a single ensemble member closest to the ensemble mean over the entire forecast window provides forecasts that are unable to produce the relatively small errors of the ensemble mean. However, since different ensemble members are closest to the ensemble mean at different forecast times, the best forecast is composed of different ensemble members throughout the forecast window. The benefits and limitations of applying this methodology to improve forecasts of synoptic-scale high-impact weather events are discussed.

1. Introduction

Ensemble forecasting adds an important dimension to numerical weather prediction. Through the simultaneous evolution of forecasts that differ by their initial conditions and their model (using varied model configurations and/or stochastic forcing), an ensemble attempts to characterize the uncertainty of the current or future state of the atmosphere. This probabilistic information provides crucial decision-making support across a variety of important forecasting applications such as wind energy (Sloughter et al. 2010), the issuance of severe weather watches by the National Centers for Environmental Prediction’s (NCEP’s) Storm Prediction Center, and contaminant dispersion modeling (Lee et al. 2009; Peltier et al. 2010). In addition, ensemble mean forecasts have been shown to improve forecast skill, on average, over that of a single deterministic forecast, particularly once nonlinearity becomes prominent within the evolution of ensemble perturbations (discussed in Kalnay 2003). The use of ensembles has been found to be skillful from a data assimilation perspective; analyses and subsequent forecasts based on the ensemble Kalman filter (EnKF) technique (Evensen 1994) have been found to produce competitive results compared to other leading assimilation systems such as four-dimensional variational data assimilation (4DVAR) at a variety of scales (Caya et al. 2005; Buehner et al. 2010; Miyoshi et al. 2010; Zhang et al. 2011). Furthermore, the flow dependence that ensemble data assimilation provides can add value through hybrid techniques to assimilation systems that have historically used only climatological covariance structures (Hamill and Snyder 2000).

Although ensemble forecasting clearly provides these important benefits, it is still desirable from an operational forecasting perspective to generate a “best guess” prediction. This is particularly true when it becomes difficult to convey forecast uncertainty in a way end users can understand. Joslyn et al. (2009) provide a good
example of this difficulty, revealing a significant level of misunderstanding involved with probabilistic precipitation forecasts. Morss et al. (2008) and Mass et al. (2009) discuss the challenges in communicating probabilistic forecast information more generally, and Mass et al. (2009) underscores the importance of a most-likely forecast through its primary role in an experimental public, web-based probabilistic forecasting system (known as PROBCAST). Generating such a best-guess forecast is particularly motivated when, in addition to the difficulty in properly communicating forecast uncertainty, there is 1) a lack of available, accurate, and operational probabilistic forecast guidance and 2) difficulty in transitioning probabilistic guidance into appropriate risk management for specific end-user applications (National Research Council 2006). Generally, when ensemble forecasts are available, the ensemble forecast mean has been used as the most likely estimate of the future atmospheric state.

All ensemble members evolve on a forecast model attractor through the governing equations. However, the mean of the members is not constrained to be on that attractor (discussed in detail in section 2). In turn, the ensemble mean can diverge from the model attractor given a certain degree of nonlinear ensemble perturbation evolution. Given that the model is a highly simplified yet realistic simulation of the complex atmosphere, the ensemble mean can thus present a less likely, unphysical solution relative to the ensemble members as it drifts away from the model attractor. This presents an interesting forecasting issue: the statistically best forecast may be unrealistic. This challenge is only relevant once nonlinearity becomes significant, which on the synoptic scales has been shown to be on the 1–2-day range (Gilmour et al. 2001; Ancell and Mass 2006), but may be significantly shorter at the meso- and storm scales (Hohenegger and Schar 2007). It should be noted that this issue likely will become even more complicated within a multiphysics ensemble since different physics schemes are associated with different model attractors, resulting in an even higher likelihood of an unphysical ensemble mean. Here, we consider only initial condition perturbations, however, with the goal of understanding the degree to which an ensemble forecast mean can become unrealistic within the framework of a single model attractor.

To mitigate the issue of the ensemble mean becoming unrealistic, “best member” techniques have been developed. Such techniques are designed to identify an ensemble member closest in some way to the ensemble mean, thereby providing a best-guess forecast. Since the chosen member may be able to achieve a close fit to the ensemble mean but is still evolved with the forecast model, it potentially could possess both the ensemble mean’s statistically beneficial error qualities while still being a realistic solution. Whether or not it is possible for specific members on the attractor to achieve the errors possessed by the ensemble mean off the attractor is a hypothesis tested in this work. This approach takes advantage of the ability of the ensemble mean to filter out unpredictable scales of motion (Bellon and Zawadzki 1994; Germann and Zawadzki 2002; Hamill et al. 2004), resulting in reduced ensemble mean errors relative to single, deterministic forecasts. While such best-member techniques may be useful within a forecasting framework, it is still unclear how the best member should be chosen, particularly with regard to weather of sufficient intensity (e.g., significant accumulation of snow, widespread damaging winds) that negatively impacts the safety and/or productivity of a significant portion of society.

This study explores how nonlinear ensemble perturbation evolution on synoptic scales affects the use of the ensemble mean for forecasts of a particular high-impact weather event: landfalling midlatitude cyclones on the west coast of North America. These storms are significant in that they can produce high winds and heavy precipitation, yet studies suggest their prediction is still problematic (McMurdie and Mass 2004; Wedam et al. 2009; Charles and Colle 2009). This study also aims to determine an effective strategy for identifying the ensemble members that best forecast these cyclones, and comparing them to the ensemble mean forecast, a technique that could produce operational ensemble forecasting benefits for synoptic-scale, high-impact weather. The organization of this paper is as follows: section 2 gives background on the nonlinear perturbation evolution that plays a key role in this study, section 3 provides the methodology for this work, section 4 presents results and discussion, and section 5 gives a summary and conclusions.

2. Background

A nonlinear forecast model that represents how the $N$-dimensional atmospheric state vector $\mathbf{x}$ evolves in time can be written as

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}),$$

(1)

where $F(\mathbf{x})$ represents an $N$-dimensional vector-valued, nonlinear function of the state. Consider a second forecast trajectory separated from $\mathbf{x}$ by an $N$-dimensional vector perturbation $\Delta \mathbf{x}$ such that

$$\frac{d(\mathbf{x} + \Delta \mathbf{x})}{dt} = F(\mathbf{x} + \Delta \mathbf{x}).$$

(2)
Using a Taylor series expansion, Eq. (2) can be written as

\[ \frac{dx}{dt} + d(\Delta x)/dt = F(x) + (dF/dx)\Delta x + (1/2!)(d^2F/dx^2)\Delta x^2 + \text{higher order terms.} \] (3)

Neglecting quadratic and higher-order terms on the right-hand side of Eq. (3) yields an expression linearly approximating the continuous time evolution of the perturbation \( \Delta x \):

\[ d(\Delta x)/dt = (dF/dx)\Delta x. \] (4)

Finally, for discrete time steps the tangent linear model (TLM; Errico 1997) can be formulated, which describes how initial-time perturbations are mapped to forecast time:

\[ \Delta x_f = M\Delta x_o, \] (5)

where \( \Delta x_f \) and \( \Delta x_o \) are the initial-time and forecast-time \( N \)-dimensional vector perturbations about some reference forecast, respectively, and \( M \) is the tangent-linear model (\( N \times N \) matrix) valid over the specified forecast time interval that maps the initial-time perturbations to forecast time, and includes linearized terms from the right-hand side of Eq. (1). Unlike the solution of Eq. (1), which describes how the atmospheric state evolves in time (a forecast), the solution of the TLM describes how perturbations evolve about the control forecast and is a linear approximation. As discussed in Errico (1997), the accuracy of the TLM degrades as perturbations become larger, and as forecast times become longer due to the inherent nonlinearity of the original forecast model.

Consider an ensemble of \( V \) initial conditions with mean \( \bar{x} \), such that the mean of the perturbations from \( \bar{x} \) is zero. In addition, consider a reference forecast \( x_f \) that represents the deterministic forecast from the ensemble analysis mean \( \bar{x} \) [referred to as the centroid per Eckel and Mass (2005)] using the nonlinear forecast model in Eq. (1). The full set of initial-condition perturbations from the mean can be written as the \( N \times V \) matrix \( P_o \), where the columns of \( P_o \) each represent an initial-time perturbation from \( \bar{x} \), and the number of rows of \( P_o \) represents the state-space dimension \( N \). Using Eq. (5), the initial time perturbations in \( P_o \) can be mapped to forecast time to become \( P_f \) with the TLM:

\[ P_f = MP_o. \] (6)

Since Eq. (6) is linear, and the mean of the perturbations in \( P_o \) is zero, then the mean of the perturbations at forecast time in \( P_f \) (which are defined about \( x_f \)) must also be zero. This implies that as long as Eq. (6) well approximates the fully nonlinear perturbation evolution within the ensemble, the mean of the ensemble forecasts evolved with the nonlinear model \( (\bar{x}_f) \) will well approximate the centroid \( x_f \). In an opposite sense, \( \bar{x}_f \) and \( x_f \) can diverge in the presence of sufficient nonlinearity. In turn, for mostly linear perturbation evolution, \( x_f \) must be realistic as it must be very similar to \( x_f \) (and very close to the model attractor), but can become unrealistic with a certain degree of nonlinearity. The difference between \( \bar{x}_f \) and \( x_f \) can thus be viewed as a measure of the detrimental and unrealistic behavior growing nonlinearity can cause and is now examined in this work from a forecasting perspective.

3. Methodology

This study applies an 80-member ensemble prediction system developed at the University of Washington (Torn and Hakim 2008) using the Weather Research and Forecasting (WRF) model and ensemble initial conditions generated using an EnKF. The EnKF data assimilation technique, which is an ensemble square root filter (Whitaker and Hamill 2002), combines a short-term ensemble forecast (and forecast error statistics) with observations (and observation error statistics) to obtain an analysis (and analysis error statistics). A number of data assimilation parameters are used with the EnKF when assimilating observations, and the parameters used here are the same as those of Torn and Hakim (2008), who tuned these parameters to produce appropriate spread through the evaluation of the ratios of ensemble mean error to ensemble and observation error variance using radiosonde observations. These parameters include a 2000-km localization radius (Gaspari and Cohn 1999) and the inflation of background variance (Anderson and Anderson 1999), both of which are used to counteract the effects of small sample size. The inflation coefficient used is 1.17, which results in a fractional inflation of 0.83 of the variance removed during assimilation. As in Torn and Hakim (2008), the observations that are assimilated here include cloud-track wind; Aircraft Communication, Addressing, and Reporting System (ACARS) aircraft temperature and wind; radiosonde temperature, wind, and relative humidity; and surface temperature, wind, and altimeter setting. The WRF EnKF is run on a domain including the western United States and the northeast Pacific Ocean (Fig. 1), and has 38-vertical sigma levels at 36-km horizontal grid spacing. Ensemble boundary conditions are produced through perturbations about the Global Forecast System (GFS) forecast using the fixed covariance perturbation method of Torn et al. (2006).
The forecast model used in this study is the Advanced Research WRF model, version 3.0.1.1 (Skamarock et al. 2008). A single set of physics options is used that includes the Mellor–Yamada–Janjić (MYJ) planetary boundary layer scheme (Janjić 1990, 1996, 2002), the Kain–Fritsch cumulus parameterization (Kain and Fritsch 1990; Kain and Fritsch 1993), the Noah land surface model (Chen and Dudhia 2001), WRF single-moment three-class microphysics (Hong et al. 2004), the Rapid Radiative Transfer Model (RRTM) longwave radiation scheme (Mlawer et al. 1997), and the Dudhia shortwave radiation scheme (Dudhia 1989). The EnKF analysis is used for verification here instead of an analysis from an independent data assimilation–forecasting system. Thus, this methodology follows the strategy of other studies that use the same system to produce both the forecast and verifying analyses, a technique described and justified in Langland and Maue (2012). Although using the same system for both forecasts and analysis–verification may be problematic for ensemble/modeling configurations that do not contain appropriate spread or are biased, this issue likely is not significant here based on the analysis of the system’s statistical consistency and error properties with regard to the response function (see below).

Since a constant suite of model physics is used in this study, no physics uncertainty is accounted for, which could result in a poor characterization of the ensemble spread (typically underdispersion) of variables for which model error can dominate the forecast error [such as 2-m temperature or 10-m winds, as discussed in Ancell et al. (2011)]. However, this study is based on the analysis of sea level pressure, which was shown in Torn and Hakim (2008) to be unbiased and contain appropriate spread (through analysis of altimeter setting) through 24-h forecast time. Since this study involves forecasts to 48 h, the ratio of the RMS ensemble mean response function error to the ensemble response function standard deviation was calculated over all cyclone instances to demonstrate statistical consistency between ensemble response function spread and error at different forecast times. The values of these ratios were found to be 1.03 (12 h), 0.99 (24 h), 1.15 (36 h), and 1.07 (48 h). Since these values show no clear trend and fluctuate to a degree that is common within calibrated mesoscale EnKF configurations based on the author’s experience, it is assumed here that the current EnKF configuration contains appropriate response function spread through the 48-h forecast window through varying only the initial conditions with the current inflation scheme. In turn, any significant underdispersion based on the lack of model physics uncertainty is likely not an issue with regard to the oceanic cyclones investigated here.

The WRF EnKF is cycled every 6 h over a 6-month period from October 2009 through March 2010, and 48-h forecasts are produced from each of the 80 EnKF analyses to capture each landfalling cyclone during this time. An initial 1-week period is run at the end of September 2009 without extended forecasts, also on a 6-h assimilation cycle, to spin up flow-dependent covariances within the WRF EnKF that were initialized through WRF model variational data assimilation system (WRFVAR; Barker et al. 2012) climatological relationships. The centroid (48-h forecast begun from the mean EnKF initial conditions) is also produced using the GFS forecast for lateral boundary conditions for each case over the entire 6-month period, but is not used to compute any of the ensemble statistics.

Oceanic cyclones approaching the North American coast are identified within a coastal zone (shown in Fig. 1) through an algorithm that searches for minima in the sea level pressure field within the ensemble mean 6–48-h forecast. This algorithm identifies locations with a sea level pressure value that is less than that of all adjacent grid points. The full sea level pressure field was subsequently visually inspected for the identified cases to confirm the existence of a midlatitude cyclone, and cases were removed if local minima were not associated with a synoptic-scale system. Since a single cyclone may reside in the coastal zone for up to a day or more, it should be noted that multiple occurrences of the same cyclone are used in this study. Once cyclones are identified within the ensemble mean, the average sea level pressure in a 216 km × 216 km (7 × 7 grid point) model grid box centered on the cyclone minimum pressure (referred to as the response function) is calculated for the ensemble mean, the centroid, and each ensemble member. This box size incorporates both the location of the

![Fig. 1. The 36-km modeling domain used in this study. The area used to identify landfalling cyclones is shown by the thick black line.](image-url)
cyclone center and the surrounding pressure field (and associated pressure gradients), and is a convenient way to appropriately identify oceanic cyclones, based on the author’s experience (e.g., Ancell and Mass 2006). It is worth noting that the subsequent results presented here pertain to this specific response function only, and results may vary for response functions that address other features of oceanic cyclones beyond the sea level pressure field (such as fronts or precipitation).

4. Results and discussion

a. Nonlinearity and the ensemble mean

Figure 2 shows the absolute difference of the response function (the average sea level pressure surrounding the ensemble mean cyclone) between the ensemble mean and the centroid averaged over all cyclone occurrences for all forecast times. At analysis time this difference must be zero—the centroid and ensemble mean are identical at that time. However, as forecast time increases, so does the difference between the ensemble mean and the centroid at roughly a linear rate. Figure 2 also shows the absolute error of the response function, again averaged over all cyclone occurrences, for both the ensemble mean and the centroid (Student’s t test 90% confidence intervals are included). These forecast errors in the response function were measured against the EnKF analysis (assumed as truth) valid at each forecast time. Through forecast hour 12 the error is about the same for both the ensemble mean and centroid but begins to diverge beyond that time. The divergence between the errors of the ensemble mean and centroid grows with time until 48-h forecast time, where it reaches a maximum. As expected, errors in the ensemble mean are smaller at all forecast times beyond 12 h for the current sample, and are significant at the 90% confidence level using a one-sided Student’s t test at 24 h and beyond. Equivalently, the steady emergence of nonlinearity is substantial enough to result in higher skill within the ensemble mean relative to the centroid averaged over all cyclone occurrences at later forecast times. This provides a unique perspective on the nature of nonlinear perturbation evolution within an ensemble, and agrees well with the result found in Ancell and Mass (2006) that nonlinearity can become significant, even on synoptic scales, at forecast times less than a day in dynamic, cyclogenetic situations. Furthermore, this statistically justifies the ensemble mean as a good choice for a deterministic forecast.

Figure 3 depicts the absolute difference of the response function between the ensemble mean and the centroid for 12-, 24-, 36-, and 48-h forecast times for all cyclone occurrences identified at those times. From visual inspection, these plots reveal that over all cyclone occurrences, both the average difference and its variability increase as forecast times get longer.

Although these results averaged over hundreds of occurrences reveal interesting properties regarding nonlinearity within an ensemble for cyclones approaching the North American coast, it would be much more useful from a forecasting perspective to investigate the effects of nonlinearity within continuous cyclone forecasts. Thus, Fig. 4 shows the same results as Fig. 2 but for only those
cyclones that were found in the coastal zone at 48-h forecast time for which a minimum in the sea level pressure field could be tracked back to the initial time within the model domain. A total of 12 cyclones met this criterion. Aside from subtle differences (presumably due to the smaller sample used here), results averaged over these cyclones are very similar to those using all cyclone occurrences, and show an increasing difference

FIG. 3. The absolute difference of the response function (hPa) between the ensemble mean and centroid for each cyclone case at 12-, 24-, 36-, and 48-h forecast times.

FIG. 4. As in Fig. 2, but averaged over only the 12 traceable cyclones.
between the ensemble mean and centroid over the 0–48-h forecast window, as well as smaller forecast errors associated with the ensemble mean compared to the centroid for these 12 cyclone cases. A bootstrap resampling technique (Wilks 1995) applied to the mean and centroid was unable to show statistical significance between these small samples (90% confidence intervals are shown in Fig. 4b). Nonetheless, the smaller errors of the ensemble mean for this cyclone subset are confirmed to occur more generally through the much larger dataset shown through Fig. 2b.

Although the ensemble mean provides a better verifying forecast, the degree to which it has potentially become unrealistic is critical from a forecasting standpoint. Figure 5 shows the 0-, 24-, and 48-h centroid (forecast from the ensemble mean analysis) and the ensemble mean for one of the 12 cyclone cases initialized at 1200 UTC 17 January 2010. The ensemble mean and centroid are identical at initial time, but some important differences arise as the forecast evolves. At 24-h forecast time, a sharper trough extends to the southeast from the cyclone in the centroid, but the ensemble mean shows a broader trough in the same location. The pressure gradients to the north and northeast of the cyclone center are also stronger in the centroid than in the ensemble mean. More generally, the ensemble mean cyclone is essentially acquiring a smoothed-out nature, is losing the sharpness of its important features, and is beginning to appear by 24-h forecast time to be unrealistic. Since pressure gradients are directly proportional to wind speed, and since troughs can be associated with precipitation, the fact that these important weather features are becoming unrealistic is detrimental to the forecast if the ensemble mean was used to provide the best guidance. Interestingly, the one consistent characteristic relative to both the ensemble mean and centroid is the cyclone position.

To address more specifically the high-impact aspects of a cyclone that may be problematic within the ensemble mean, Fig. 6 shows the differences in the 24- and 48-h forecast between the ensemble mean and centroid for surface wind speed and 6-h precipitation, and is typical for all 12 cyclone cases examined. Wind speed differences of over 10 m s$^{-1}$ (lower values in the mean) are found over a large area surrounding the cyclone center at 48 h, whereas at 24 h the wind speed differences are slightly smaller and focused to the northwest. This is consistent with the weakening of pressure gradients around the cyclone shown in Fig. 5, which would result

---

**Fig. 5.** The 0-, 24-, and 48-h forecasts of sea level pressure (black contours, contour interval is 2 hPa); 10-m winds; and 925-hPa temperature (gray contours, contour interval is 2°C) of a specific cyclone by both the ensemble mean and centroid initialized at 1200 UTC 17 Jan 2010. The black arrows show the location of the cyclone in all forecasts.
in weaker winds. The 6-h precipitation differences reach nearly 25 mm (roughly an inch) in the same area at 48 h, but are again smaller at 24 h, reaching magnitudes of only about 6 mm. The larger discrepancies at later forecast time are consistent with the growth of nonlinearity shown in Fig. 4a. Since landfalling cyclones can produce wind speeds on the order of 10s of meters per second and 6-h precipitation amounts on the order of a few inches, these differences are substantial, and they represent the unrealistic behavior of the ensemble mean produced through nonlinearity that can occur within a 48-h forecast window. Although the ensemble mean is the best a priori choice statistically, its unrealistic behavior is clearly unfavorable from the perspective of predicting realistic, high-impact weather features.

b. Best-member techniques for high-impact events

The unrealistic behavior of the ensemble mean motivates an alternate choice in producing a best-guess forecast. Since the ensemble mean is the best choice statistically, but has been shown to be unrealistic as it was not evolved with the forecast model, it is reasonable to choose the forecast member closest to the ensemble mean averaged over the entire forecast. Since this member is evolved with the forecast model (and thus must be on the model attractor), yet is close to the ensemble mean, it should possess both realistic behavior and the relatively small errors associated with the ensemble mean. As the response function involving the sea level pressure surrounding the cyclone center is the forecast aspect we strive to best predict, the most effective method for choosing a best member is to identify the forecast with a response function value closest to that of the ensemble mean. There is no need to attempt to identify ensemble inliers based on dynamical aspects expected to be important (such as the geopotential height field aloft); here, we can directly calculate inliers based on exactly the forecast aspect we want to predict correctly. This is always possible as long as a specific forecast aspect has been chosen that can identify mobile flow features.

Table 1 shows the forecast member closest to the ensemble mean every 6 h, measured by the absolute difference in the response function, for all 12 cyclone cases. It also shows the member closest to the ensemble mean averaged over the entire 48-h forecast window. Interestingly, for all cases, a different forecast member is
closest to the ensemble mean at almost every 6-h increment as the forecast evolves, and no member remains closest to the ensemble mean for more than 6 h. This situation provides an interesting perspective on the notion that a Gaussian distribution characterizing the entire state will spread out over time through the solution of the Liouville equation (Ehrendorfer 1994): although the spread associated with the cyclone does indeed increase temporally (Fig. 7), specific forecast members are constantly repositioning themselves relative to the ensemble mean. This likely is a result of the cyclones representing only a small portion of the entire state. Figure 8 shows the absolute response function differences from the ensemble mean for four different members for the cyclone case initialized at 1200 UTC 17 January 2010 to illustrate this point. The member farthest from the ensemble mean at 48 h is closest to the ensemble mean from 0 to 6 h, and the members show different relative proximities to the ensemble mean as the forecast evolves. Such behavior explains why the members closest to the ensemble mean averaged over the entire forecast are not necessarily closest to the ensemble mean at any single forecast time. In 4 of the 12 cyclone cases, the centroid is closest to the ensemble mean averaged over the entire forecast, suggesting nonlinearity is less prominent for those storms.

Figure 9 summarizes the errors associated with the centroid, ensemble mean, and closest member to the ensemble mean averaged over all 12 cyclone cases. The best member calculated by averaging the errors of the member closest to the ensemble mean throughout the entire forecast period produces a 2.37-hPa average response function error for all cyclone cases. The same statistic for the centroid is 2.48 hPa. The average error for a randomly chosen member was also calculated for reference, and has a value of 2.90 hPa. As expected, the lowest average error of 2.19 hPa for all cases is produced by the ensemble mean. These results show that for the cyclones examined, the ensemble mean is able to provide the best a priori forecast relative to the centroid and the closest member to the mean averaged over all forecast times. Since bootstrap resampling techniques were again unable to provide statistical significance at a reasonable confidence level for the 12-cyclone subset (through the 90% confidence intervals shown in Fig. 9), these results can only be viewed as suggestive of the ensemble mean’s superior performance more generally.

Figure 9 also shows the best performing member, which is the single ensemble member with the smallest absolute response function error measured against the

![Fig. 7. The 0-, 24-, and 48-h forecasts of sea level pressure ensemble standard deviation (shaded) and ensemble mean sea level pressure (black contours, contour interval is 2 hPa) of the same cyclone in Fig. 5 initialized at 1200 UTC 17 Jan 2010. The black arrows show the location of the cyclone in all forecasts.](image-url)
analysis over the entire forecast period, averaged over all cyclone cases (average error of 0.19 hPa). As might be expected, ensemble members exist that are much better than the ensemble mean. The smaller errors associated with the best members appear very early in the forecast period (i.e., less than 6 h). Although there would be no way in real time to identify these members since the analyses against which they are measured are not available at forecast time, it is tempting to develop a method to identify such members. Since observations become available practically immediately, a potential technique to identify the best members would be to use sensitivity-weighted early forecast errors to identify the best members relative to the chosen forecast aspect. Since it is somewhat straightforward to produce such sensitivity with either an adjoint model or through an ensemble approach (both discussed in Ancell and Hakim 2007), such a technique might be viable to identify superior forecast members prior to the next data assimilation cycle and will be explored in future work.

Interestingly, studies using response functions similar to that here (Ancell and Mass 2006; Ancell and Hakim 2007) have shown that the most sensitive regions at earlier times are not necessarily collocated with the surface cyclone pressure field. Thus, sensitivity analysis suggests that the best ensemble members with regard to the cyclone at earlier times are not necessarily the best at later forecast times, reconciling well with the results here shown in Fig. 8.

Since the single, continuous member closest to the ensemble mean over the entire 48-h forecast period possesses higher errors than those of the ensemble mean (2.37 versus 2.19 hPa), this “best” member must diverge from the ensemble mean sufficiently enough at some point in the forecast to degrade its performance (as revealed in Fig. 8). It should be again noted that we cannot claim that this degradation of the closest continuous member relative to the mean applies more generally based on the confidence intervals shown in Fig. 9. Nonetheless, the degraded performance of this closest member within the sample here motivates selecting different ensemble members at different times that are closest to the ensemble mean and “patching” them together to create a forecast with reduced error. In fact, if the constraint is removed that the best member be a continuous, single forecast, and if such a patched-together forecast is created by choosing the ensemble member closest to the ensemble mean at each 6-h forecast time interval, the average error for all cyclone cases is the same as that of the ensemble mean both averaged over all 48 h (2.19 hPa) and at each individual forecast time (the patched forecast is shown in Fig. 9 by the black open circles). In turn, this patched-together forecast constructed from different ensemble members is essentially the same and statistically as good as the ensemble mean in predicting the oceanic midlatitude cyclones investigated here. Since the errors of the patched forecast are the same as those of the mean and its error bars would significantly overlap with that of the mean shown in Fig. 9, these two forecast products are indistinguishable more generally.
Although the patched forecast is not one continuously evolved with the forecast model, the forecast at any given time has been evolved within the model, ensuring its realistic behavior on the model attractor. It is likely that such members evolving on the attractor are able to achieve the same errors as the ensemble mean, which tends to average out more unpredictable and smaller-scale features, because the error is dominated by the more predictable broader-scale structure. This is not surprising as the response function was designed to capture a synoptic-scale phenomenon, and since smaller-scale features are dynamically linked to their synoptic counterparts, the finescale errors may also be relatively small. In any case, results may change if this methodology were applied to capture smaller-scale events such as severe convection.

One potential issue of using such a combined forecast in an operational setting is that the features within the forecast might appear somewhat erratic as they evolve in time since they are not temporally consistent as a single evolved forecast. However, the chosen forecast response function was forced to be the closest to the ensemble mean forecast response function at each time. This effectively anchors the forecast to a somewhat smoothly varying solution. Figure 10 depicts the 12-, 24-, 36-, and 48-h forecasts for the same cyclone case shown in Fig. 5, but depicts the closest ensemble members to the ensemble mean at each of those times. The evolution of the patched forecast appears smooth, and its track is typical for landfalling cyclones in the Pacific Northwest. The cyclone central pressure decreases slowly, something that occurs in both the ensemble mean and centroid. Frontal positions appear to evolve smoothly as well. At 12-h forecast time, a cold front extending south from the cyclone, marked by a wind shift, rotates around the cyclone in a typical manner without a clear warm frontal signature (common with oceanic cyclones). By 36-h forecast time, the cold front has reached the coast and, with a more pronounced warm front extending eastward from the cyclone at this time, resembles the “T-bone” structure characteristic of oceanic cyclones (Shapiro and Keyser 1990). Interestingly, the frontal bands appear to evolve smoothly even though, unlike the sea level pressure response function, they were not...
5. Summary and conclusions

This study explored the nonlinear nature of ensemble perturbation within an EnKf relative to midlatitude cyclones approaching the west coast of North America. This work is somewhat unique in that a Lagrangian approach was taken in this examination following only the cyclones over a 0–48-h forecast window (specifically the sea level pressure field near the cyclone center). It was demonstrated theoretically that as nonlinearity becomes more prominent during a forecast, the ensemble mean forecast and the forecast begun from the mean initial conditions (the centroid) can diverge. It was found that such divergence occurs at roughly a linear rate with respect to time for these cyclones, resulting in the ensemble mean possessing smaller cyclone forecast errors around 12-h forecast time and throughout the duration of the 48-h window.

Although the ensemble mean shows smaller errors statistically on average, the differences it acquires relative to the centroid during the forecast are associated with unrealistic behavior. Specific, high-impact features associated with the cyclones such as wind speed and precipitation were found to be reduced dramatically in the ensemble mean. This is a result of the smoothing of the sharp structures, such as tight pressure gradients and strong baroclinic zones associated with the high-impact features in the first place. Such unrealistic behavior can only occur in the presence of sufficient nonlinearity (which results in the ensemble mean going off the model attractor), and is noticeable with regard to the cyclones early in the forecast period (6–12 h), but becomes significant by 48-h forecast time. In turn, although the ensemble mean is the best a priori choice statistically for cyclone forecasts, it can provide poor guidance regarding specific, high-impact features.

The unrealistic behavior of the ensemble mean provides the motivation for choosing a different, more realistic forecast outcome. By choosing ensemble members (or the centroid) that are closest to the ensemble mean cyclone, one might expect both realistic behavior (since all ensemble members are forced to be on the model attractor) and small forecast errors. Although the member closest to the ensemble mean averaged over the entire forecast improves errors over that of the centroid, a “patched together” forecast involving different ensemble members at different times results in forecast errors indistinguishable from the ensemble mean in the cyclone sample investigated here. This cyclone forecast evolves smoothly, even with regard to features not included in the response function. Such a forecast provides an attractive ensemble forecast product that might add value to operational guidance aimed at identifying improved most-likely forecasts of high-impact events. A logical next step and planned extension of this work is to investigate the behavior of nonlinearity and the best-member techniques developed here on smaller scales, such as those involved with severe convection, where nonlinearity is likely to be significantly more important. In addition, the evaluation of patched-together forecasts for specific high-impact events by NWS forecasters would provide crucial feedback on their usefulness, and is another planned extension of this research.

Acknowledgments. This work is supported by the National Oceanic and Atmospheric Administration’s (NOAA) CSTAR program under Award NA11NWS4680001. The author also wishes to thank three anonymous reviewers for their thoughtful comments that have improved this manuscript, and Lynn McMurdie for her help in the initial identification of the cyclones used in this study.

REFERENCES


——, 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso Model. NCEP Office Note 437, 61 pp.


