

The Effects of Sampling Errors on the EnKF Assimilation of Inner-Core Hurricane Observations

JONATHAN POTERJOY, FUQING ZHANG, AND YONGHUI WENG

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 30 September 2013, in final form 6 December 2013)

ABSTRACT

Atmospheric data assimilation methods that estimate flow-dependent forecast statistics from ensembles are sensitive to sampling errors. This sensitivity is investigated in the context of vortex-scale hurricane data assimilation by cycling an ensemble Kalman filter to assimilate observations with a convection-permitting mesoscale model. In a set of numerical experiments, airborne Doppler radar observations are assimilated for Hurricane Katrina (2005) using an ensemble size that ranges from 30 to 300 members, and a varying degree of covariance inflation through relaxation to the prior. The range of ensemble sizes is shown to produce variations in posterior storm structure that persist for days in deterministic forecasts, with the most substantial differences appearing in the vortex outer-core wind and pressure fields. Ensembles with 60 or more members converge toward similar axisymmetric and asymmetric inner-core solutions by the end of the cycling, while producing qualitatively similar sample correlations between the state variables. Though covariance relaxation has little impact on model variables far from the observations, the structure of the inner-core vortex can benefit from a more optimal tuning of the relaxation coefficient. Results from this study provide insight into how sampling errors may affect the performance of an ensemble hurricane data assimilation system during cycling.

1. Introduction

Accurate tropical cyclone forecasting continues to be one of the greatest challenges for operational weather prediction models. One problem lies in our inability to initialize the tropical cyclone core structure accurately in real time, owing to limitations in model resolution, data coverage, and our current data assimilation procedures (Houze et al. 2007; Zhang et al. 2009). The data assimilation process combines two state estimates: the atmosphere as forecasted by a model (denoted prior) and the atmosphere as depicted by the observations to generate an analysis state (denoted posterior) to be used as initial conditions for a forecast (Talagrand 1997).

One approach called the Kalman filter propagates the multivariate state vector and its error covariance forward in time from the previous assimilation cycle using a forecast model. The new covariance is used to find the least squares estimate of the posterior state under the

assumptions of Gaussian errors and linear model dynamics (Kalman and Bucy 1960). The ensemble Kalman filter (EnKF) approximates the standard Kalman filter by using an ensemble of model forecasts to estimate the prior and posterior error covariance, thus providing an affordable means of applying the filter for nonlinear models (Evensen 1994). The EnKF has proven to be an effective data assimilation option for a wide range of weather applications (e.g., Snyder and Zhang 2003; Zhang et al. 2004; Dowell et al. 2004; Tong and Xue 2005; Zhang et al. 2006; Torn et al. 2006; Meng and Zhang 2007; Fujita et al. 2007; Meng and Zhang 2008b; Zhang et al. 2009; Weng and Zhang 2012). Its largest benefit comes from the use of a forecast ensemble to estimate the prior error covariance at each assimilation cycle, rather than relying on a climatological error covariance as is done in alternative data assimilation methods (i.e., three- and four-dimensional variational systems). For the case of tropical cyclones, Poterjoy and Zhang (2011) show that variance and correlations calculated from a storm-relative ensemble of hurricane forecasts can represent relationships in the storm core that are physically consistent with our understanding of tropical cyclone dynamics. Their results

Corresponding author address: Dr. Fuqing Zhang, Department of Meteorology, The Pennsylvania State University, 503 Walker Building, University Park, PA 16802.
E-mail: fzhang@psu.edu

also support the notion that a cycling EnKF can benefit from short data assimilation cycles that alleviate the largely non-Gaussian forecast errors that may result from vortex position spread (Chen and Snyder 2007).

While ensembles provide a flow-dependent prior error estimation for each data assimilation cycle, the statistics are often sensitive to sampling errors. The computational cost of integrating an ensemble of high-dimensional models after each consecutive cycle imposes restrictions on the ensemble size, thus limiting the maximum degrees of freedom in the state estimate. This sampling deficiency causes the prior covariance matrix to be rank deficient and overestimated between variables at distant grid points. In practice, covariance localization approaches are applied to overcome these shortcomings by reducing the magnitude of off-diagonal terms in the ensemble-estimated covariance matrix. The localization is typically achieved using a simple element-wise multiplication of the covariance matrix with an empirical correlation function (Hamill and Whitaker 2001). In addition to sampling errors, other unresolved sources of error such as those imposed by the imperfect forecast model and nonlinearity may also lead to an underestimation of the true forecast variance, thus causing the filter to fit observations less over time in a process called “filter divergence.” Various means of inflating the forecast covariance have been used to achieve adequate filter performance for repeated assimilation cycles (Anderson and Anderson 1999; Mitchell and Houdekamer 2000; Zhang et al. 2004; Whitaker and Hamill 2012). Many of these sampling deficiencies are demonstrated in the context of tropical cyclone data assimilation by Aksoy et al. (2012) and Sippel et al. (2013) using simulated radar observations.

This study uses real data to examine the accumulative effects of sampling errors in a cycling EnKF data assimilation system. A set of independent data assimilation cycles for Hurricane Katrina (2005) are run from 1430 to 2000 UTC 25 August 2005, a period in which Katrina approached the Florida coast while intensifying from a tropical storm to a category 1 hurricane. The data assimilation experiments are configured identically, except that ensemble size and the degree of covariance relaxation are varied between cases. The purpose of this study is to examine how sampling errors can influence the evolution of storm structure during cycling. The model grid spacing, choice of ensemble sizes, and detail of our analysis limits this investigation to one case study. Nevertheless, our findings may be applicable to a large variety of cases in which inner-core observations are assimilated at short cycles to “spin up” a tropical cyclone with proper dynamic and thermodynamic structure.

The organization of the manuscript is as follows. Section 2 contains the details regarding model and experiment setup for this study. Sections 3 and 4 describe the analysis and forecast results from our set of cycling data assimilation cases, respectively. Section 5 provides the summary and conclusions.

2. Methodology

The EnKF data assimilation system described in (Meng and Zhang 2008a,b; Weng and Zhang 2012) is used in this study for the Weather Research and Forecasting Model (WRF) (Skamarock et al. 2008). It follows the square root algorithm described by Whitaker and Hamill (2002) to update the perturbations around the posterior mean state. Covariance localization is achieved using an element-wise multiplication of the covariance matrix with a Gaspari and Cohn (1999) fifth-order correlation function, and the covariance is inflated after each analysis using the “covariance relaxation to the prior” method proposed in Zhang et al. (2004). This system has been used in real time since 2008 to assimilate routinely collected radial velocity observations from National Oceanic and Atmospheric Administration (NOAA) P3 airborne Doppler radar flight missions (Aberson et al. 2006), and provide forecasts for tropical cyclones in the Atlantic hurricane basin (Zhang et al. 2011).

The current study uses the same observations, model configuration and general EnKF setup as Weng and Zhang (2012) and Poterjoy and Zhang (2011). The Advanced Research WRF version 3.1 (Skamarock et al. 2008) is used with a coarse domain (D1) of 202×181 horizontal grid points at 40.5-km grid spacing, and two two-way nested inner domains with 13.5- and 4.5-km grid spacing (D2 and D3, respectively). The two inner domains automatically follow the storm using the WRF vortex-following algorithm, and represent convection explicitly. All domains use 35 vertical levels, most of which are concentrated in the lowest 8 km, with the model top at 10 mb. Details regarding the method of ensemble generation, choice of physical parameterization schemes, and collection and quality-control procedures for the airborne radar data are provided in Weng and Zhang (2012).

The ensembles are initialized at 0000 UTC 25 August 2005 using the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) Final Analysis (FNL) as the mean, and GFS forecast data for lateral boundary conditions. The initial ensemble perturbations are created using the WRF variational data assimilation system (Barker et al. 2004; Huang et al. 2009) with the cv3 covariance option (Parrish and Derber 1992). The ensemble members are integrated for 14.5 h to

evolve flow-dependent forecast error statistics, which are then centered on the 60-member ensemble mean before assimilating the first batch of airborne radar observations. Data from the remaining flight legs are assimilated at 1530, 1630, 1730, 1900, and 2000 UTC thereafter. Each set of observations is collected within 30 min of the analysis time, which is chosen based on the time of the flight leg. The observation operator projects the background state vector onto the horizontal component of the radar radial winds during data assimilation. In doing so, the vertical component of wind is ignored to avoid the estimation of particle fall speeds for each observation. The assimilated observations span altitudes of 10–15 000 m, with 95% falling below 9200 m, and cover large portions of the inner core;¹ see Fig. 12 of Weng and Zhang (2012).

Each data assimilation experiment uses successive covariance localization (Zhang et al. 2009), which is an empirical means of adjusting the localization radius of influence during sequential data assimilation. The localization implements a 1215-km radius of influence in all three domains to assimilate the first $\frac{1}{9}$ of the observations; the radius of influence is then decreased to 405 km to assimilate another $\frac{2}{9}$ of the observations in D2 and D3, and decreased further to 135 km to assimilate the remaining observations in D3 only. The vertical localization radius of influence is set to 35 vertical levels, which is large enough to have only marginal effects on model levels close to flight level, where most of the verification is performed. Using this configuration, all observations are assimilated serially in the same order during each data assimilation cycle.

To ensure that differences between each data assimilation configuration are due to sampling errors alone, each ensemble shares the same localization configuration and starts from the same prior mean at 1430 UTC 25 August. Our choice of localization radii is based on past studies that use the same ensemble data assimilation system for assimilating airborne radar observations with 30–60 members (Zhang et al. 2009; Weng and Zhang 2012; Zhang et al. 2011). Under a configuration that is typical for hurricane applications, the sensitivity to ensemble size N is examined by performing a set of data assimilation experiments with N set to 30, 60, 120, and 300. While we acknowledge that the optimal choice of localization varies with N , the current study does not address this dependence. The goal of these experiments is to explore how the posterior mean and covariance

change as a result of increasing N in a controlled manner.

Following Zhang et al. (2004), the covariance is inflated after each cycle by relaxing the posterior perturbations (\mathbf{x}_n^{ta}) back to the prior perturbations (\mathbf{x}_n^{tf}):

$$\mathbf{x}_n^{ta} \leftarrow (1 - \alpha)\mathbf{x}_n^{ta} + \alpha\mathbf{x}_n^{tf}. \quad (1)$$

The α in Eq. (1) is called the relaxation coefficient and ranges from 0 to 1, where $\alpha = 0$ implies no relaxation. The experiments that investigate the sensitivity of the data assimilation to N use a constant α of 0.8, a value that has been used frequently in past studies (Meng and Zhang 2008a,b; Torn and Hakim 2008). The 60-member case is then repeated for a second set of experiments using an α of 0.6, 0.4, 0.2, and 0.0. This allows for an examination of how α impacts the data assimilation when N is set to a value that has been demonstrated to be both effective and affordable for the current application (Weng and Zhang 2012; Zhang et al. 2011).

3. Cycling data assimilation results

a. Changes in storm structure during cycling

In situ observations of flight-level (825–840 mb) wind speed, temperature T , pressure P , and relative humidity were collected by instruments on board the NOAA P3 aircraft as it passed through Katrina during the six flight missions. Because these observations are not assimilated during the experiments, they are used in this section to verify the performance of the prior and posterior mean states at each analysis time. Figure 1 shows the prior and posterior mean flight-level wind speed and T plotted against the verifying observations at the first and last update times, along with the intermediate time of 1630 UTC 25 August. Provided that the path of the aircraft through the center of Katrina does not follow a straight line (see Fig. 1b of Weng and Zhang 2012), only the inner ~ 120 km of the vortex is observed during each flight. The verification shows that all cases produce reasonably accurate analyses of flight-level wind speeds in the core, despite the range in ensemble sizes. Nevertheless, the 30-member case produces less accurate results for variables that are not assimilated during this experiment. For example, the 30-member posterior mean T is much lower than observations along the flight track. The larger-ensemble cases verify much closer to the T observations, with the exception of inner-core values during the last cycle where temperatures are stronger than observed; the warm temperature bias in these cases may come from errors in the data assimilation that are not considered in this study (e.g., model

¹ Following Weatherford and Gray (1988), the inner core refers to the portion of the storm within 1° (~ 100 km) of storm center, whereas outer core is defined as the portion between 1° and 2.5° (~ 100 – 250 km).

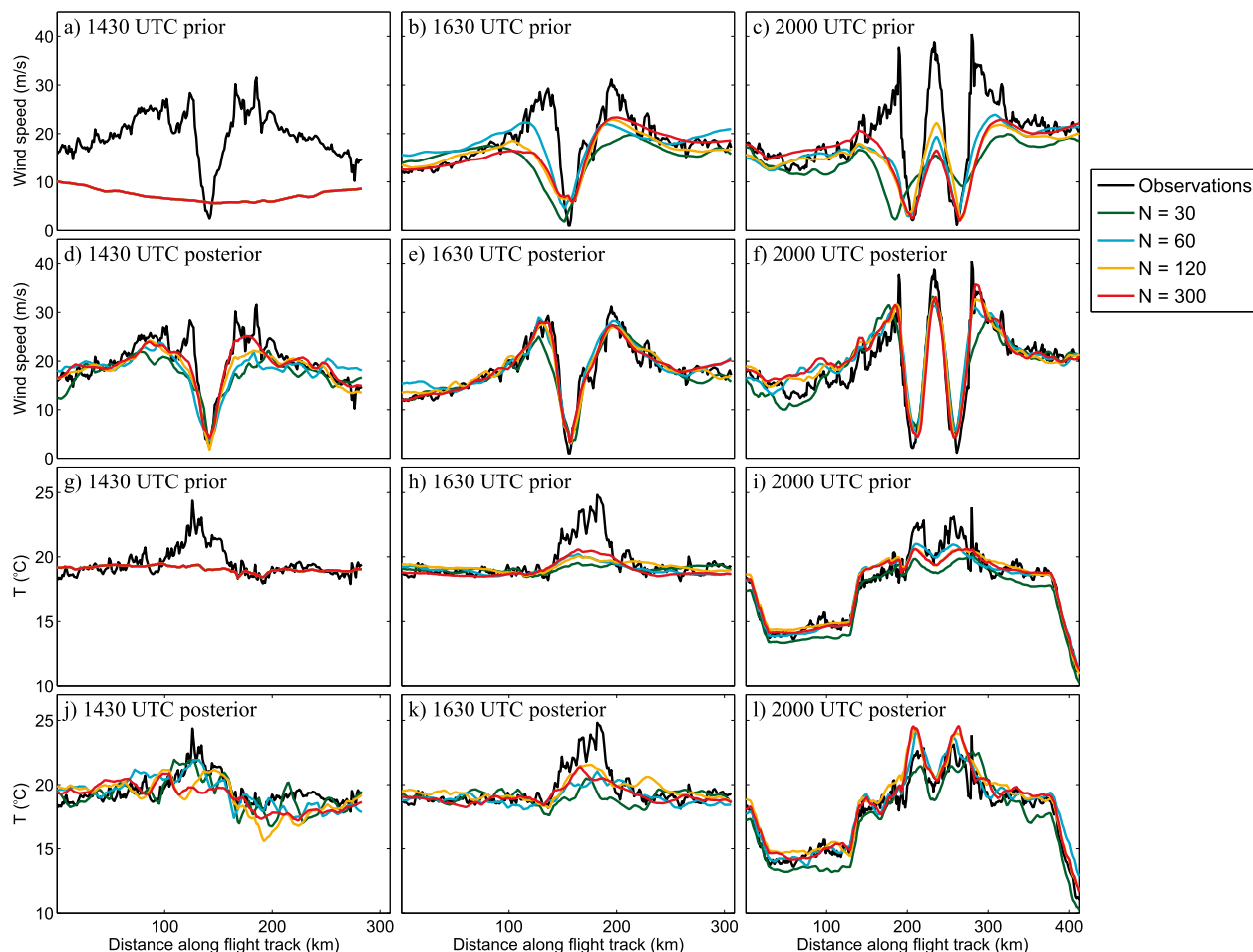


FIG. 1. (a)–(c) Prior and (d)–(f) posterior flight-level winds and (g)–(i) prior and (j)–(l) posterior T verified against aircraft observations at (a), (d), (g), (j) 1430; (b), (e), (h), (k) 1630; and (c), (f), (i), (l) 2000 UTC for ensemble sizes of 30 (green), 60 (light blue), 120 (orange), and 300 (red).

error or correlated observation errors). Though not shown, the verification of posterior mean pressure and humidity is consistent with the flight-level T in Figs. 1g–l in that the 30-member case contains errors that are systematically greater than the higher-ensemble cases during the cycling. Given that temperature, pressure, and moisture are not assimilated in these experiments, this result suggests that the 30-member ensemble may not represent accurate cross correlations between the wind and thermodynamic variables.

While the flight-level verification provides insight into how ensemble size can affect the accuracy of the data assimilation in the inner core, the main objective of this section is to describe the changes in storm structure that emerge from our ensemble size experiments. Figure 2 shows the difference in posterior tangential wind speeds V_θ between the $N = 300$ ensemble and $N = 30, 60$, and 120 ensembles. Model data in these plots come from a subspace of D1 that covers the entire Gulf of Mexico

and Florida, and the 300-member vortex center is used as the reference center for calculating V_θ . The unpaired Student's t test shows that differences of about 1 m s^{-1} or greater are statistically significant at the 90%–95% confidence level. Portions of the flight-level 300-member posterior mean V_θ field are 11 m s^{-1} stronger than in the 30-member ensemble, but the differences decrease substantially when additional members are added. The posterior mean wind differences at each time represent the accumulative effects of errors caused by sample size, which are assumed to be smallest for the 300-member ensemble. The largest values are found in the dashed boxes plotted in Fig. 2, which indicate a $900 \times 900 \text{ km}^2$ region that fits inside D3. Data from the 4.5-km domain will be used for the remaining portions of this manuscript to compare the posterior ensembles in more detail within the dashed box.

The posterior azimuthal mean tangential wind V_θ and P are plotted in Figs. 3a–f at three analysis times to

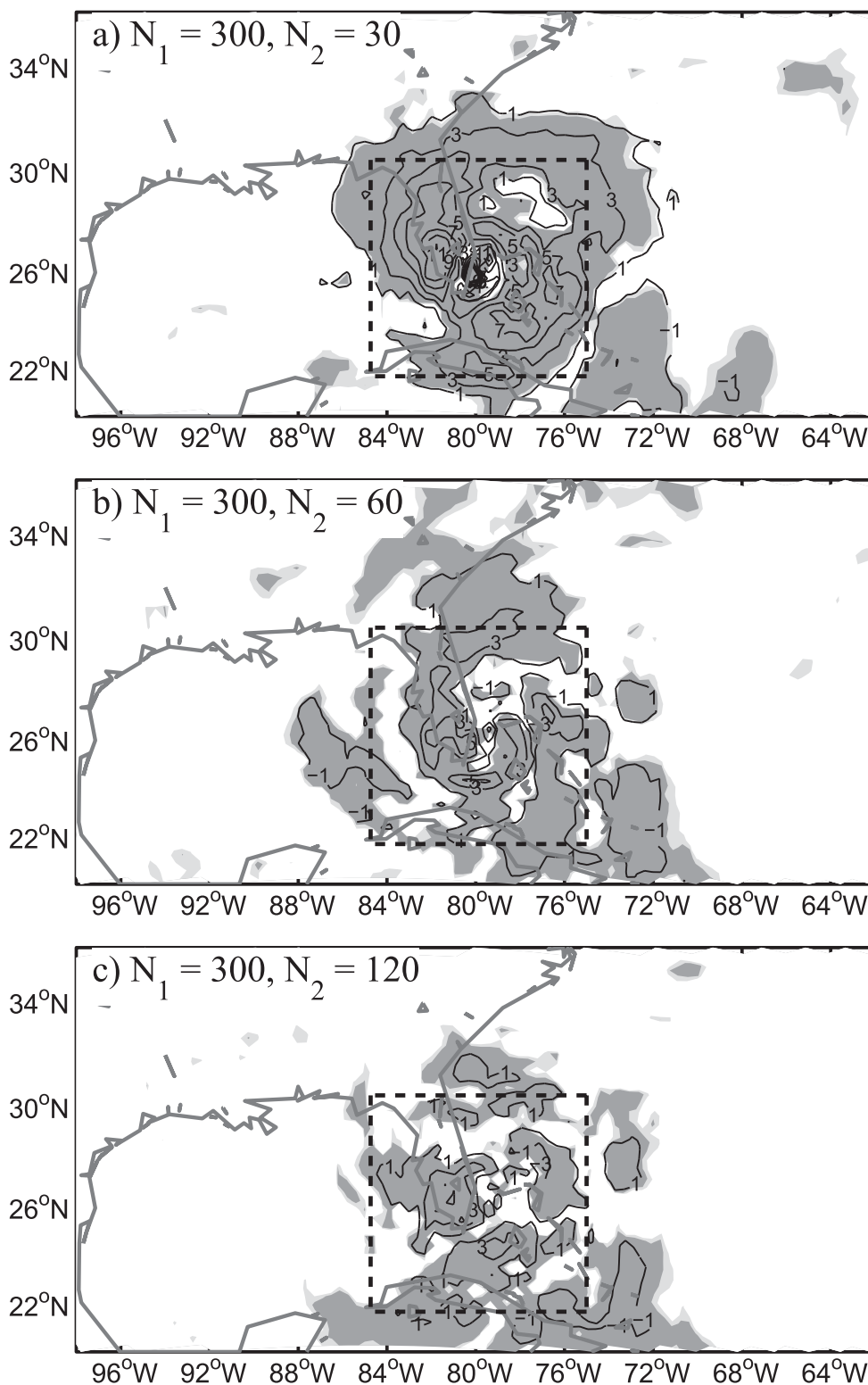


FIG. 2. Differences between the 300-member and (a) 30-, (b) 60-, and (c) 120-member posterior tangential winds contoured every 2 m s⁻¹ at 1500 m for D1 on 2000 UTC 25 Aug. The light shading indicates where the pair of posterior means is statistically unique at the 90% confidence level using the unpaired Student's *t* test. Dark shading indicates a significant difference at the 95% confidence level. The dashed box indicates the subspace in which data from D3 are compared in the subsequent figures.

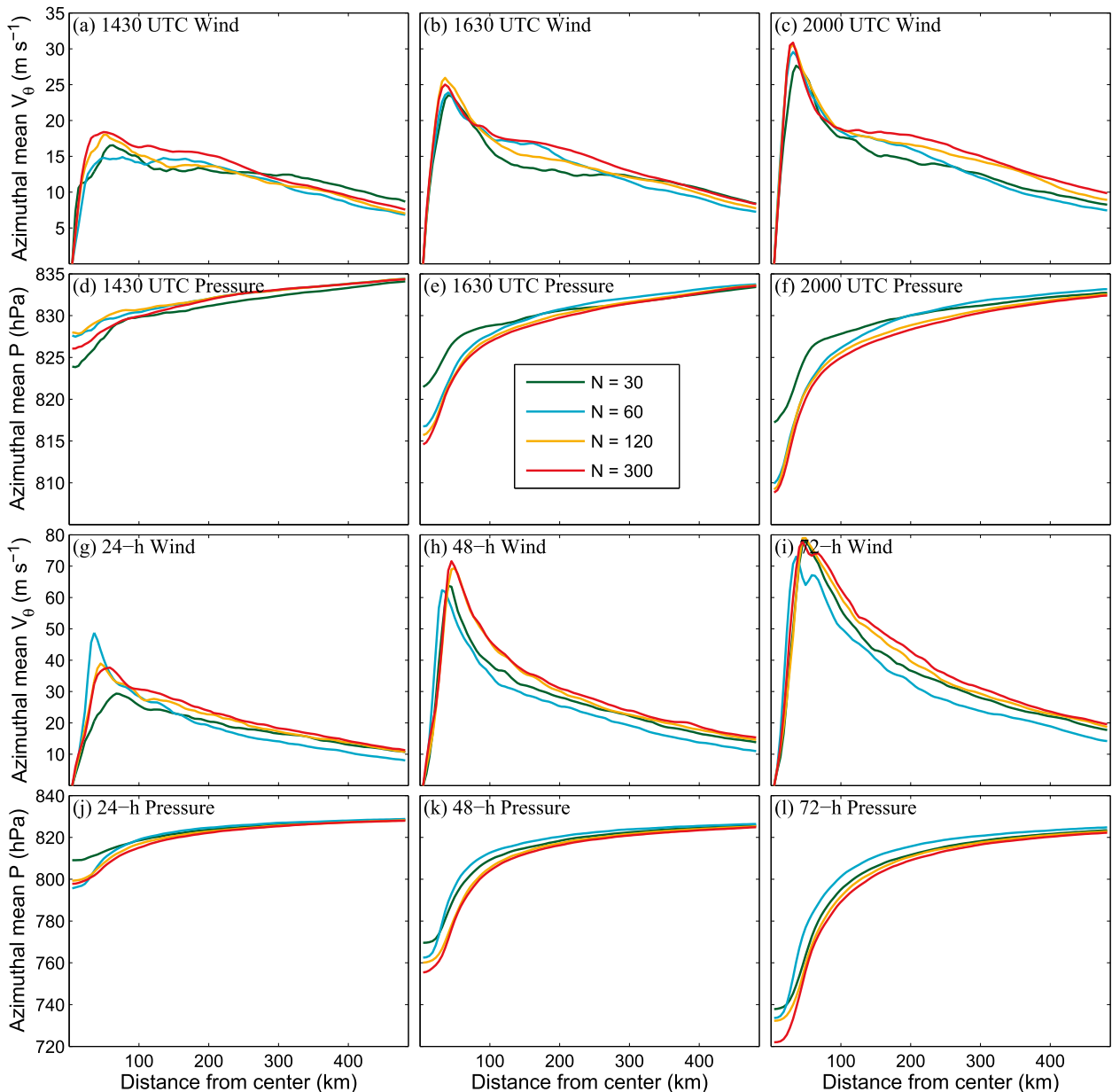


FIG. 3. Axisymmetric posterior (a)–(c) V_θ and (d)–(f) P at 1500 m for (left to right) the first, second, and sixth update times. (g)–(i) As in (a)–(f), but for the deterministic forecasts of V_θ and P for (left to right) 24-, 48-, and 72-h lead times. Values are plotted for 30 (green), 60 (light blue), 120 (orange), and 300 (red) members.

compare the evolution of the axisymmetric storm structure as the observations are assimilated. The vortices produced by the larger ensembles converge toward a similar solution inside 100 km by the last cycling time, owing to the greater concentration of observations in the inner core. Nevertheless, the V_θ and P fields outside the inner core diverge early in the experiments, where analysis increments depend on covariance and cross covariance between the outer-core state vector and winds

close to the storm center. The ensemble covariance will be discussed in more detail in the next section.

The profiles of azimuthal mean P in Fig. 3d show that the 30-member ensemble produces pressure increments that are 2–4 mb lower than all other experiments at the initial time. Each data assimilation experiment uses the same prior mean during the first cycle; therefore, any distinctions in storm structure must come entirely from sampling errors in the ensemble-estimated background

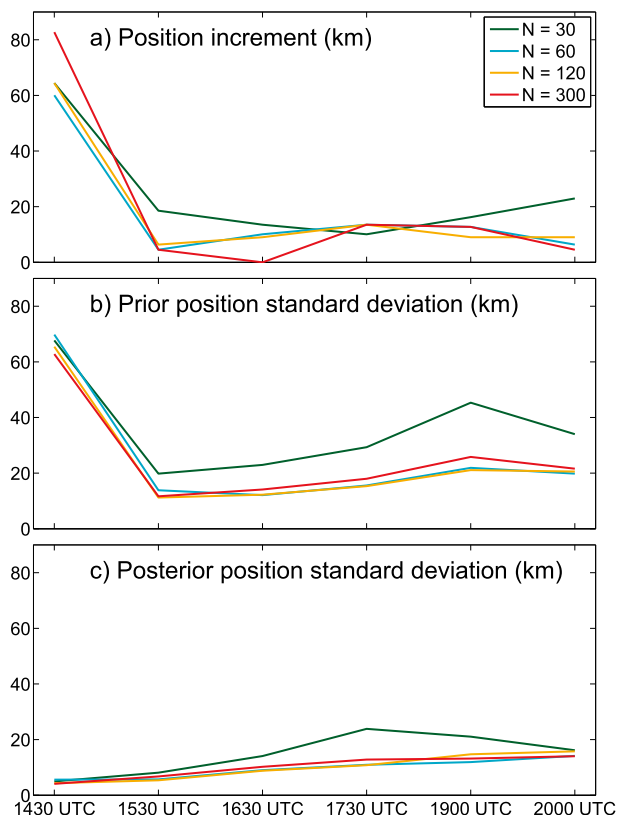


FIG. 4. (a) Displacement between the prior and posterior mean vortex centers, (b) prior ensemble vortex position spread, and (c) posterior ensemble vortex position spread for 30 (green), 60 (light blue), 120 (orange), and 300 (red) members.

error covariance. While the 30-member case yields the deepest central pressure after the first update, it produces anomalously higher central pressures at the remaining assimilation times because of poor ensemble track forecasts between cycles. Figure 4a shows that the displacement between the prior and posterior mean vortex is typically much larger in the 30-member case than in other experiments, which suggests that the effects of storm position errors on the analyses are largest in the 30-member experiment. The 30-member ensemble also produces a prior spread in vortex positions that is nearly 2 times larger than any of the other ensembles during cycling (Fig. 4b). The large position uncertainty in the 30-member case decreases the influence of vortex size and intensity on the ensemble error covariance estimate, which may lead to a degraded azimuthal mean storm structure during data assimilation (Chen and Snyder 2007; Poterjoy and Zhang 2011). Both of these factors contribute to the 30-member case producing the largest flight-level errors (Fig. 1).

Amplitudes of azimuthal wavenumber (n) 0, 1, and 2 posterior variables are plotted for radial profiles of the

inner and outer core at 2000 UTC 25 August in Fig. 5. The profiles are calculated at the approximate altitude of the flight path (1500 m), which is found to be representative of major sampling differences in storm structure. While this height may not be the most representative of the variables examined in this manuscript, its proximity to the independent observations in Fig. 1 make it a practical choice for this comparison. As is shown in Figs. 3a–f, the 60-, 120- and 300-member posterior means converge toward a similar axisymmetric primary circulation and pressure field in the inner 100 km of the vortex, but diverge outside of this region by the last cycle (first column in Fig. 5). The $n = 0$ profiles show that the larger ensembles produce a warmer warm core, more subsidence, and lower pressure in the eye and higher water vapor mixing ratios q_v in the outer core.

Cycling with 30 members produces the largest asymmetries in posterior V_θ and P (Figs. 5b,c,n,o), which is at least partially caused by the large position adjustment during the data assimilation (Fig. 4a). The secondary circulation in the smallest ensemble also differs from the larger cases. Two maxima of V_r for $n = 0-2$ are found in the posterior means: one near the radius of maximum winds at 27 km and another between 100 and 200 km. Sampling errors in the 30-member ensemble lead to larger amplitudes of asymmetric V_r at both of these maxima (green lines in Figs. 5e,f). The $N = 60, 120$, and 300 cases also contain $\sim 1 \text{ m s}^{-1}$ spikes in the w fields for $n = 1$ and 2 near the radius of maximum winds, which are absent in the 30-member posterior mean. The agreement between the larger ensembles on the representation of V_r and w in the inner core suggests that the larger ensembles are approaching a similar solution to Katrina's secondary circulation.

While the posterior means tend toward similar inner-core results when 60 or more members are used, noticeable differences exist between the 60- and 120 to 300-member cases away from the inner core. For instance, the 60-member posterior contains values of axisymmetric V_θ , V_r , and q_v that are lower than the 120- and 300-member cases at radii larger than 100 km. These profiles are accompanied by larger amplitudes in the asymmetric components. Though the asymmetries are not as large as those produced for the 30-member ensemble, these results may be a reflection of the slightly inferior storm position in the priors and posteriors depicted by the smaller ensembles.

b. How the evolving ensemble members affect EnKF updates during cycling

In this section, we describe how the spatial extent of analysis increments changes over the course of the cycling data assimilation experiments. Figure 6 shows the

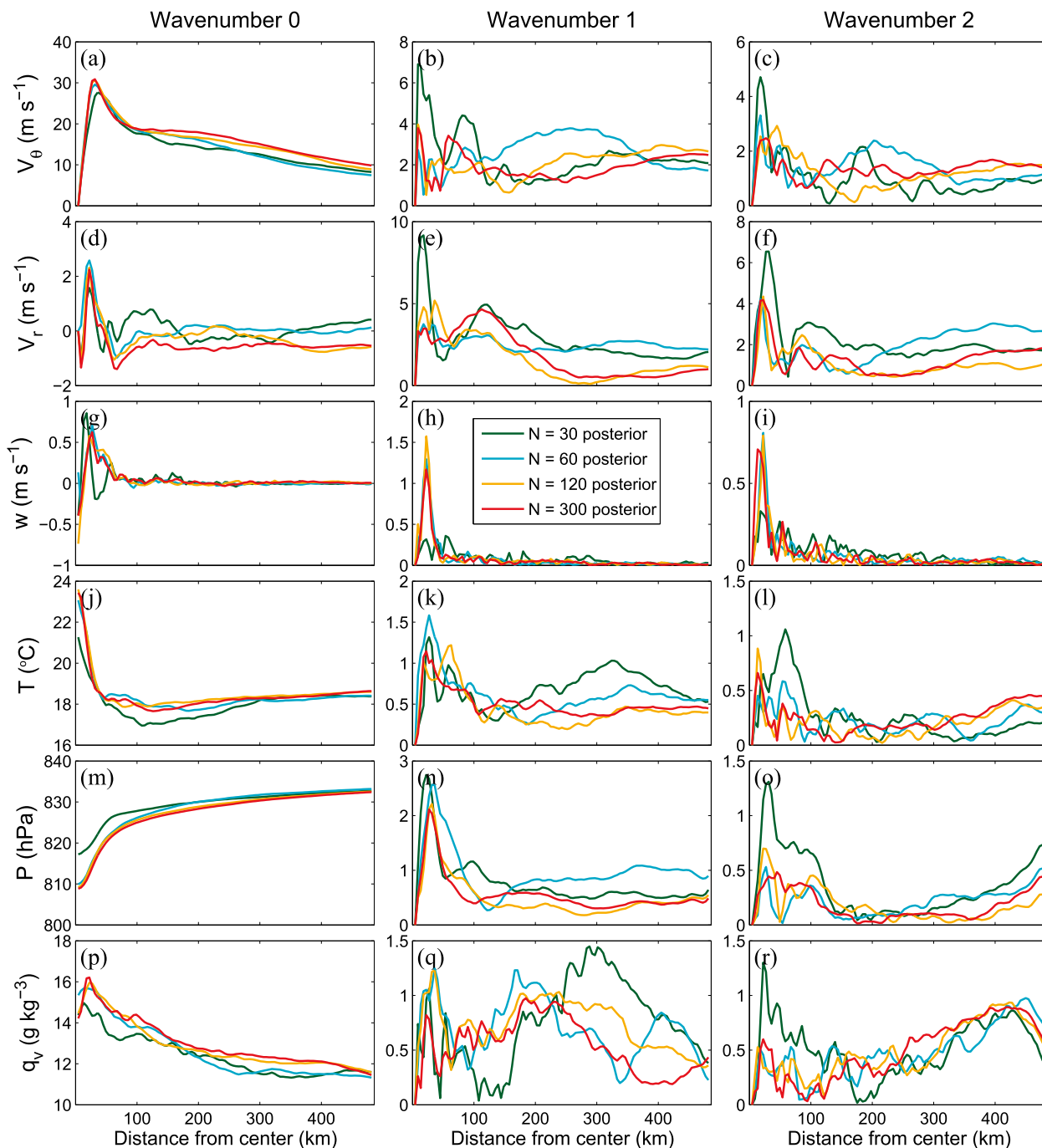


FIG. 5. Wavenumber (left) 0, (middle) 1, and (right) 2 amplitudes of (a)–(c) V_θ , (d)–(f) V_r , (g)–(i) w , (j)–(l) T , (m)–(o) P , and (p)–(r) q_v for the posterior ensemble means. Values are plotted for 30 (green), 60 (light blue), 120 (orange), and 300 (red) members at 2000 UTC.

posterior mean 1500-m V_θ and analysis increments at the three times used in Fig. 3. The increments are calculated by subtracting the prior mean from the posterior mean at each analysis time, while using the posterior mean vortex center as the origin for calculating the tangential winds. At the first data assimilation cycle, analysis

increments act to adjust the horizontal winds within a 400-km radius from the storm center. The large positive increments within 300 km of the vortex center (white contours in Fig. 6), along with negative increments outside this region at some analysis times (black contours along periphery of Figs. 6b–d), act to increase the gradient of V_θ

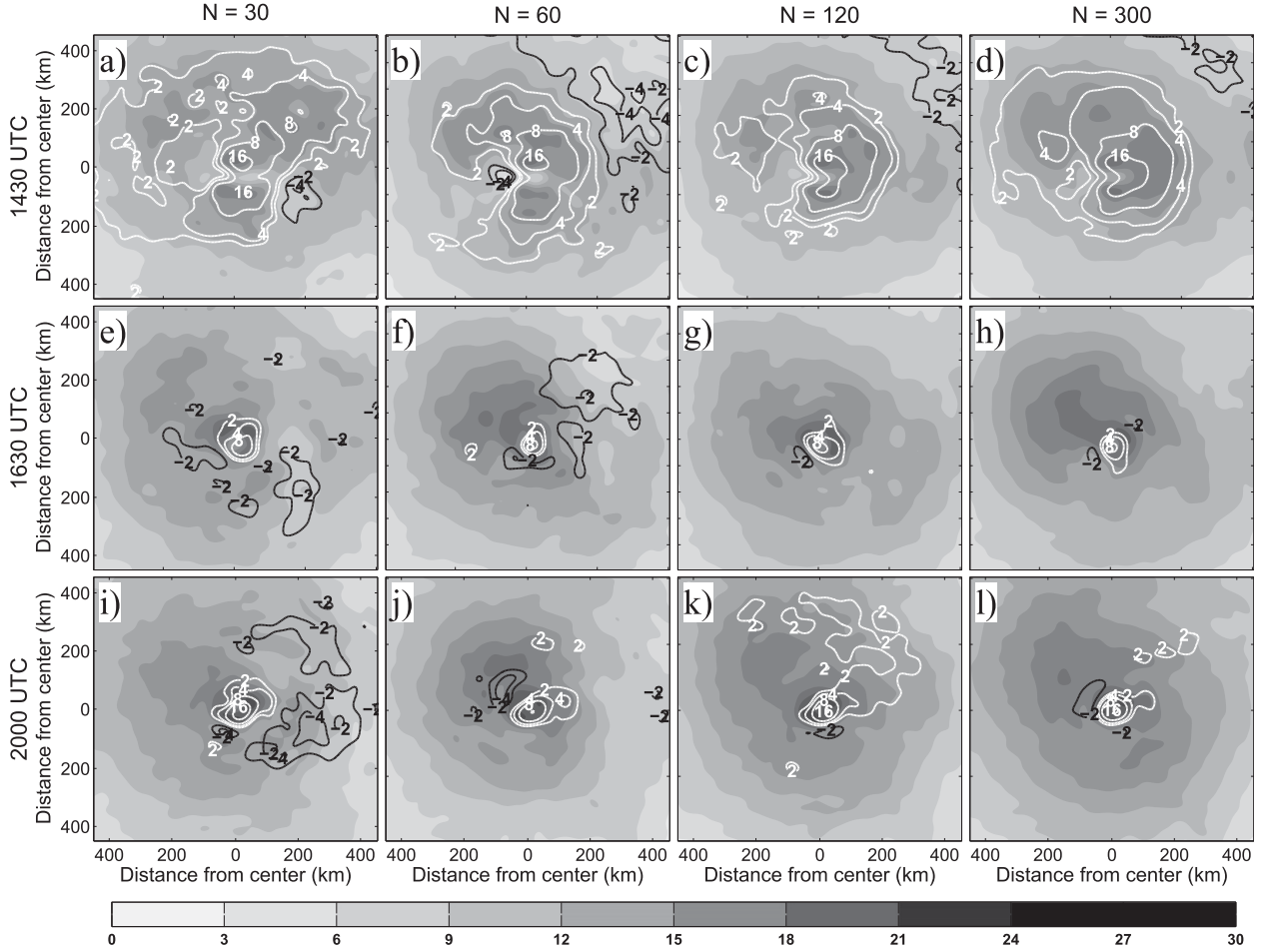


FIG. 6. Posterior mean V_θ shaded every 3 m s^{-1} and analysis increments contoured for 2, 4, 8, and 16 m s^{-1} with positive (negative) values in white (black). Data are plotted for an altitude of 1500-m at (top) 1430, (middle) 1630, and (bottom) 2000 UTC for number of members: (a),(e),(i) 30; (b),(f),(j) 60; (c),(g),(k) 120; and (d),(h),(l) 300.

around the storm, thus producing a strong vortex around the region covered by the radar observations.

The initial ensemble forecast at 1430 UTC contains the weakest storms and largest position spread compared to the cycles that follow, which leads to relatively small values of ensemble variance near the storm center at this time; the variance in tangential winds for the 60-member ensemble is provided in Fig. 7 for reference. After the first data assimilation cycle, the hurricanes in the ensemble begin to adjust and move apart between each successive cycle. This process causes large gradients of wind to be concentrated near the mean storm center (near A in Fig. 7). As wind variance in the inner core increases between successive data assimilation cycles, the variance in the outer core (near B in Fig. 7) remains relatively small. The ensemble members therefore evolve in a manner that leads to larger variance near more dynamically active regions of the domain (i.e., in the inner core of the simulated vortex).

We will present a hypothetical single-observation example to demonstrate how the evolving covariance matrix affects the spread of information from inner-core observations during cycling. The Kalman update equation is given by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\mathbf{y} - H\bar{\mathbf{x}}^f), \quad (2)$$

where $\bar{\mathbf{x}}^f$ and $\bar{\mathbf{x}}^a$ are the forecast and analysis mean state vectors, respectively; \mathbf{y} is the observation state vector; and H is an operator that maps the model state into observation space. Bold sans serif and bold roman fonts are used to represent matrix and vector quantities, respectively, and the overbar denotes an ensemble mean. The \mathbf{K} is the Kalman gain matrix:

$$\mathbf{K} = \mathbf{P}^f H^T (H \mathbf{P}^f H^T + \mathbf{R})^{-1}, \quad (3)$$

which uses the ensemble-estimated prior covariance \mathbf{P}^f and observation covariance matrix \mathbf{R} to spread

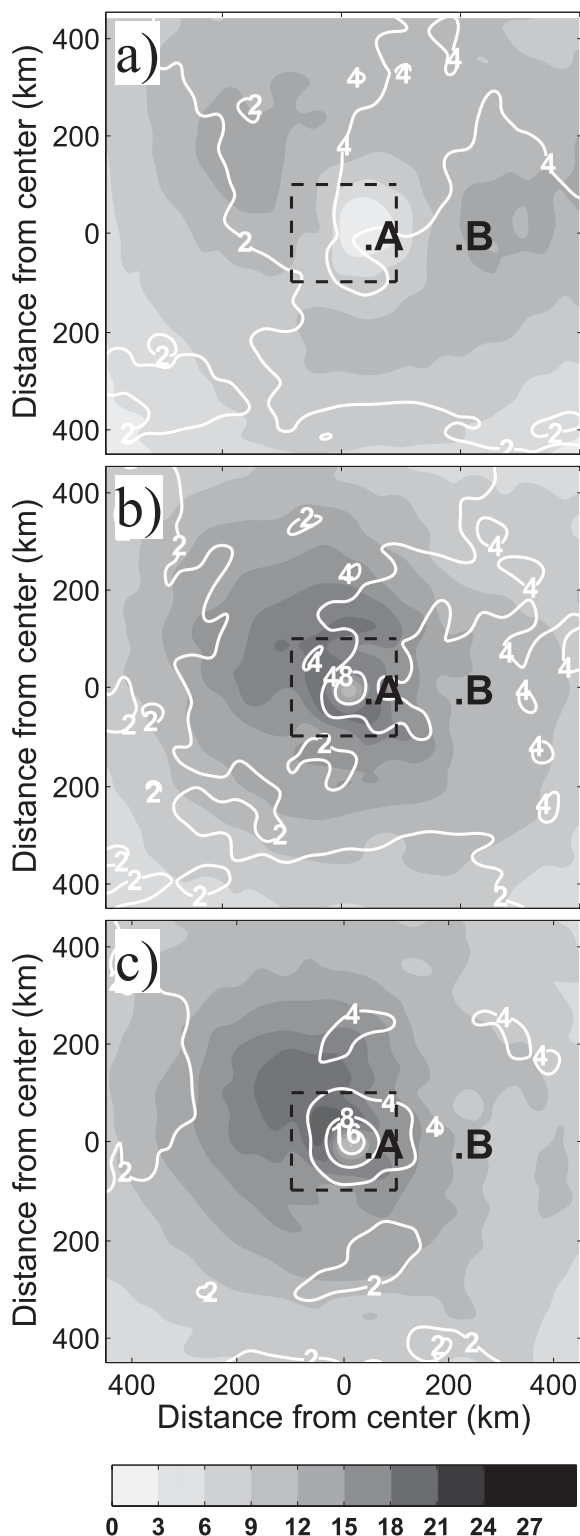


FIG. 7. Prior ensemble mean and standard deviations for the 60-member experiment at (a) 1430, (b) 1630, and (c) 2000 UTC. The gray shading is the ensemble mean V_θ in 3 m s^{-1} increments and the white contours are the standard deviations contoured for 2, 4, 8, and 16 m s^{-1} . As referenced in the text, points A and B indicate the location of a hypothetical observation and state variable to be updated, respectively.

information from observations to the model state vector.

Suppose we wish to assimilate a tangential wind observation y that is located at the inner-core location A in Fig. 7. Let $H\mathbf{x}^f = \bar{x}_A$, where \bar{x}_A is the prior mean projected onto the tangential wind at point A, and $\mathbf{R} = \sigma_y^2$, so that the Kalman gain can be written as

$$\mathbf{K} = \frac{\text{Cov}(x_A, \mathbf{x}^f)}{\sigma_A^2 + \sigma_y^2} = \text{Cor}(x_A, \mathbf{x}^f) \frac{\sigma_A \sigma_{\mathbf{x}^f}}{\sigma_A^2 + \sigma_y^2}. \quad (4)$$

The \mathbf{x}^f in Eq. (4) is a random variable that is represented by the prior ensemble, and the nonbold x is a scalar with position indicated by the capital subscript letter. If y is used to update the prior mean wind at point B (denoted \bar{x}_B) outside the inner core in Fig. 7, then the analysis increment for \bar{x}_B is given by

$$\delta \bar{x}_B = \text{Cor}(x_A, x_B) \frac{\sigma_A \sigma_B}{\sigma_A^2 + \sigma_y^2} (y - \bar{x}_A). \quad (5)$$

The prior variance increases in the inner core after the first cycle (cf. Fig. 7). Though not shown, this result occurs for cases that use $\alpha \geq 0.4$. The relaxation also allows the ensemble to maintain strong correlations between inner- and outer-core winds throughout the cycling. The outer-core analysis increment, $\delta \bar{x}_B$ in Eq. (5), is at a maximum with respect to σ_A when $\sigma_A = \sigma_y$, and decreases to zero as σ_A gets progressively larger. Likewise, $\delta \bar{x}_B$ is maximized at the first analysis time, because σ_A is close to the $\sigma_y = 3 \text{ m s}^{-1}$ value assigned for the airborne Doppler winds. Since σ_B and $\text{Cor}(x_A, x_B)$ change very little during the successive cycles, both the decrease in innovation and increase in σ_A cause $\delta \bar{x}_B$ to be much smaller at later cycles when relaxation is used. More generally, the increase in ensemble variance in the inner core is one factor that contributes to the changing magnitude of outer-core wind increments in Fig. 6.

It follows that the EnKF representation of the outer core weighs greatly on the first data assimilation cycle. This result also holds true for cases that do not use covariance relaxation, in which case ensemble correlations between variables at distant grid points are reduced substantially during the cycling. Each data assimilation experiment uses the same localization and prior mean at 1430 UTC 25 August, so differences in increments at this time must come solely from sampling errors in the ensemble forecast covariance. As a result, the sampling errors from the first analysis time lead to differences in posterior wind and pressure for the experiments at later times. Further evidence is shown in Figs. 3a–c, where the 300-member posterior contains the strongest azimuthal mean outer-core winds at every update time.

c. Verification of ensemble spread in the inner core

The performance of the EnKF depends largely on how well the forecast ensemble approximates the true prior error distribution. The assimilated radar observations (each observation denoted by y_i) are used here to compare ensemble standard deviations with the innovations d_i at each time to measure the effectiveness of the ensembles in estimating the true forecast errors. Any displacement in storm position from the truth is likely to result in biases in the model state projected to point observations. To limit this problem, error statistics are estimated by averaging the value of each metric over annular regions around the storm center at each time. The bins are created at 2-km radii out to 108 km, using the 300-member posterior as the storm center, with each bin representing a different error region that spans all vertical levels. Since the sign of airborne radar radial winds depends on the location of the aircraft and quadrant of the storm being sampled, all y_i are forced to be positive for the verification. This task is accomplished by removing all y_i with magnitudes less than the observation error ($\sigma_o = 3 \text{ m s}^{-1}$), and reversing the sign of the remaining negative y_i and corresponding prior perturbations $H(x_{n,i}^f)$. The modified set of observations are positive, but the ensemble members may be positive or negative to account for values of $H(x_{n,i}^f)$ that have the opposite sign of y_i . Detecting a statistical bias in the ensemble would be nearly impossible without a post-processing procedure of this type, since a positive bias in one quadrant of the storm would otherwise translate into a negative bias in another quadrant for a quasi-axisymmetric wind field, thus giving a total bias near zero.

If the ensemble variance is calibrated properly, the expected value of the squared innovations must be equal to the sum of the forecast error variance in model space and observation error variance (Dee 1995). We perform this comparison by first calculating the mean squared innovations, $\langle (d_i - \bar{d})^2 \rangle$, and mean observation space ensemble variance, $\langle \sigma_{H(x)}^2 \rangle$, for each bin of observations. As in Aksoy et al. (2009), we use the ratio of error terms:

$$R = \frac{\sqrt{\langle \sigma_{H(x)}^2 \rangle}}{\sqrt{\langle (d_i - \bar{d})^2 \rangle - \sigma_o^2}}, \quad (6)$$

to characterize the ensemble spread in each bin. The coefficient R in Eq. (6) is between 0 and 1 when the ensemble spread is too small, and greater than 1 when the ensemble spread is too large. Since the goal of covariance relaxation is to control the ensemble variance between cycles, Fig. 8 shows the metric R and the

ensemble biases for each of the ensemble priors that are available from the 60-member covariance relaxation experiments. Ensemble members at the first update time come from 14.5-h model integrations, causing the spread in vortex position to be nearly as large as the region covered by the radar data (Fig. 4b). With many members forecasting a storm location that is far from the actual center, the variance in winds near the observations are largely underestimated by the ensemble at this time (Fig. 8a), and contain a negative bias of about 15 m s^{-1} around the radius of maximum winds. Experiments that use small amounts of relaxation ($\alpha \leq 0.2$) continue to have too little variance during the cycling, thus demonstrating why inflation is necessary for practical implementations of the EnKF. Nevertheless, too much relaxation can lead to a prior error estimation that is too large, as demonstrated by the last four cycles of the $\alpha = 0.8$ experiment (Figs. 8c–f).

A larger α can lead to more bias near the radius of maximum winds, as shown in Figs. 8j–l. This result follows from the fact that relaxation increases the position spread between cycles, thus reducing the amplitude of inner-core winds in the prior mean. The choice of $\alpha = 0.8$ worked well in previous studies in which a 6- or 12-h time window was used between assimilation cycles (Meng and Zhang 2008a,b; Torn and Hakim 2008), but the statistics presented in Fig. 8 suggest that an α of 0.4 or 0.6 is more appropriate for assimilating inner-core hurricane observations with short lead times between cycles. The amount of relaxation should also depend on ensemble size, with more relaxation required for smaller ensembles. In this context, however, the errors associated with sample size become small compared to other error sources (e.g., model error) as N becomes larger than 60. Though not shown, we find similar biases and R values in the inner core for cases that use an increasing number of members (and $\alpha = 0.8$), because the optimal α depends less on ensemble size as random sampling error decreases. Overall, the choice of relaxation coefficient has little impact on storm structure in the 60-member experiment for $\alpha > 0.2$, but a much larger sensitivity is observed in relaxation experiments using 30 members.

As was discussed in the previous section, the spatial distribution of ensemble variance can limit the amount of information that is passed from observations in dynamically active regions to distant model grid points. We will now explore the consequences of underestimating the background error variance in the vicinity of the inner-core observations. Figure 8a shows that the prior error estimation is between 2 and 3 times too small at 1430 UTC, before the EnKF with covariance relaxation recovers the correct spread by the third assimilation cycle. Assuming that the ensemble provides a more

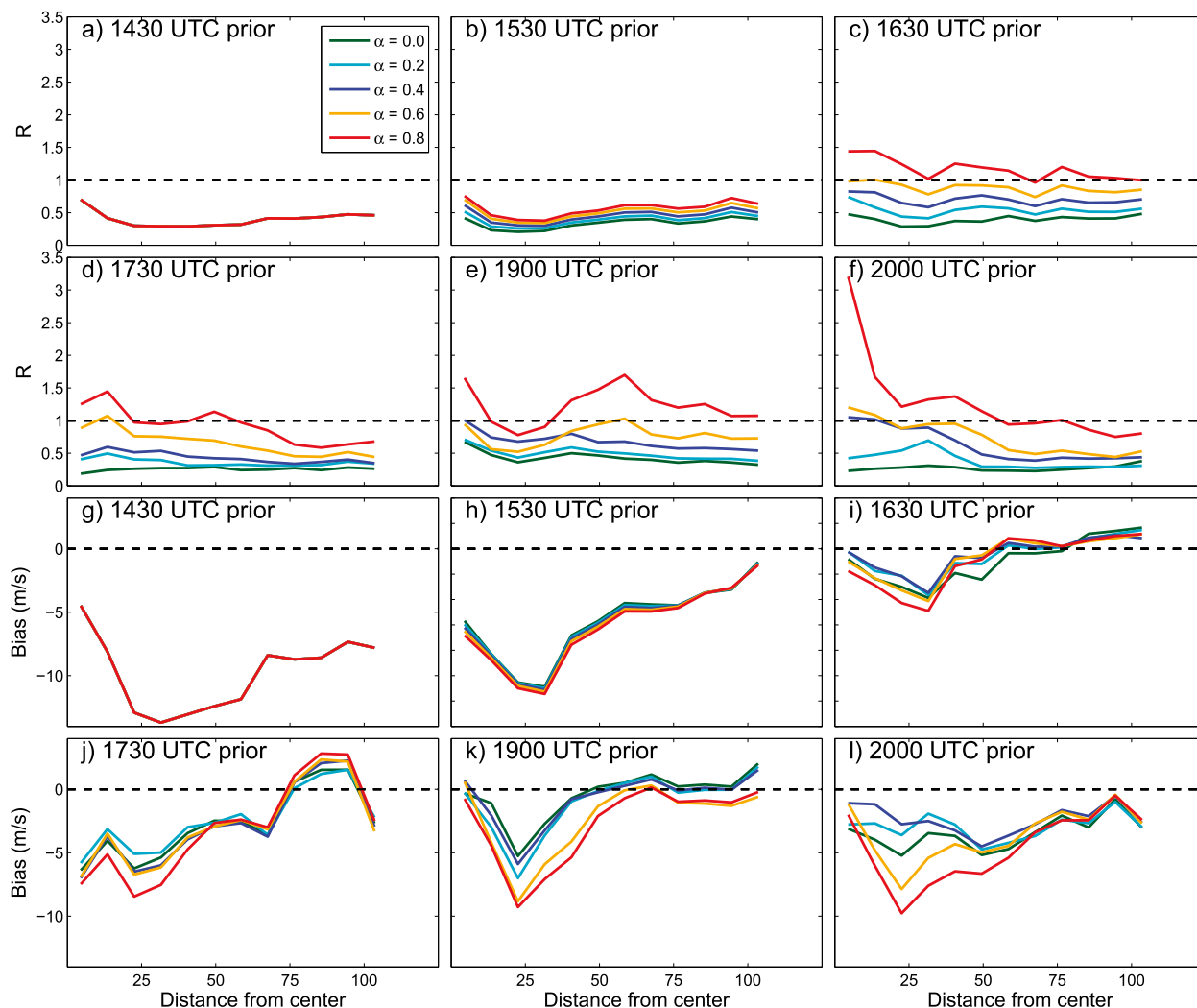


FIG. 8. Verification of prior ensemble spread in the relaxation experiments using 60 ensemble members and the assimilated radar data at each update time. The error ratio (R) and bias are plotted as a function of radius for (a)–(f) and (g)–(l), respectively, for $\alpha = 0.0$ – 0.8 .

accurate error estimation outside the inner core than inside,² the actual analysis increments outside the inner core may be much larger than what would have been made given a properly calibrated prior variance. We examine the impact of variance deficiency in the inner core by recalculating the 1430 UTC posterior mean using an inflated prior variance. The variance inflation follows the function

$$\mathbf{x}_n^f(r) \leftarrow \mathbf{x}_n^f(r)[1 + 0.0075(r+9)^2 e^{-[(r+9)/25]}], \quad (7)$$

² This is a reasonable assumption because a 14.5-h ensemble forecast is more likely to capture the uncertainty in the large-scale wind features than the much smaller vortex-scale uncertainty associated with the location of the inner core.

where r is the distance of a model grid point to the storm center. This function (plotted in Fig. 9a) provides a reasonable fit to the ratio of root-mean-squared ensemble variance to root-mean-squared innovation and observation variance (inverse of the coefficient R in Fig. 8a) at 1430 UTC, and decays exponentially to zero outside 100 km. Figure 9b shows the prior and posterior azimuthal mean tangential winds at 1500 m for the 60-member case with inflation and without inflation (denoted control). The inflation produces azimuthal mean winds in the 60-member posterior that are over 2 m s^{-1} weaker than the control experiment for large parts of the outer core, and reduces some of the outer-core sensitivity to ensemble size (not shown). Taking note of the fact that changes to the inner-core winds are relatively small between these two analyses, we can conclude that the suboptimal background variance

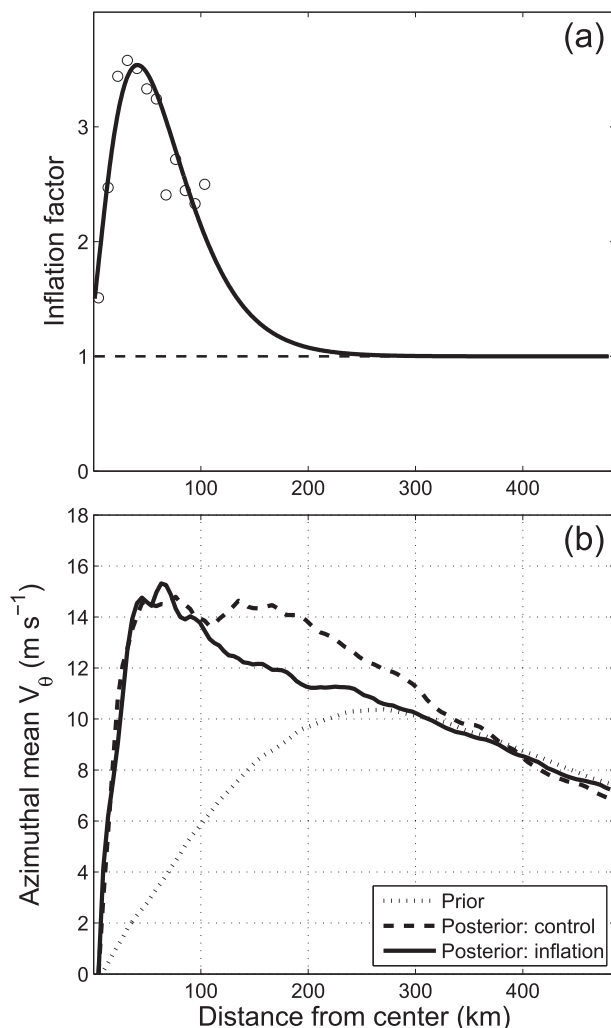


FIG. 9. (a) The function used to tune the prior error variance along with R^{-1} data (circles) from Fig. 8a. (b) Axisymmetric mean 1500-m V_θ for the first analysis time for the prior (dotted line), control 60-member posterior (dashed line), and 60-member posterior with inflation (solid line).

near observations has the most profound impact on distant analysis increments. This example partially explains why the outer-core storm structure demonstrates the most sensitivity to sampling errors in our simulations.

One limitation of the inflation function applied during this experiment is that the variance outside the inner core is assumed to be accurate. The inflation also adjusts the perturbations to fit a variance profile that is estimated from horizontally and vertically averaged innovations, which may not be representative of the true forecast error in this region. Despite these assumptions, the experiment still suggests that the Katrina analyses can benefit from variance inflation near the observed storm center, or possibly a smaller localization radius for

the first data assimilation cycle to avoid spurious corrections outside the inner core. Adaptive inflation techniques that provide spatially and temporally varying covariance inflation (e.g., Anderson 2009; Miyoshi 2011) may also be useful for providing a radially varying inflation factor in this scenario. Updates to the larger scales should come from additional observations that capture the environmental conditions (e.g., dropsonde data or satellite winds) or after the ensemble variance becomes more consistent with the magnitude of innovations. This result has broad implications for ensemble data assimilation at the storm scale, because ensembles can easily underestimate the prior spread in regions of high wind and thermodynamic gradients (e.g., data assimilated in a supercell environment may lead to incorrect analysis increments at larger scales if the ensemble fails to capture the correct model uncertainty in the vicinity of the mesocyclone).

d. Inner-core updates

While the assimilation of radar observations produces only minor adjustments to the large-scale circulation around the hurricane vortex after the first cycle, the EnKF continues to make significant corrections to the inner core. This subsection compares results within the dashed box indicated in Fig. 7 that covers the inner 100 km of the tropical cyclone vortex. The 60-member ensemble is compared with the 300-member ensemble at the last update time because 60 members is found to be the smallest sample that produces reasonably accurate storm structure during the experiments (cf. Fig. 5). Since tropical cyclones can be thought of as quasi-axisymmetric weather systems, prior and posterior ensemble members are decomposed into storm-relative azimuthal wavenumber components to show the contribution of the axisymmetric and asymmetric components during the data assimilation experiments. The wavenumber separation is carried out in a storm-relative reference frame for each member. This procedure involves: locating the vortex center for each member via the Geophysical Fluid Dynamics Laboratory (GFDL) vortex tracker algorithm (Marchok 2010), interpolating all variables to cylindrical coordinates with respect to vortex location, transforming variables into azimuthal wavenumber space, separating the powers for each wavenumber before transforming the variables back to physical space, and interpolating the filtered variables to a Cartesian grid with the approximate center of the posterior mean at the origin. The reference center for each member is fixed with height and storm-motion vectors are not subtracted from the wind field; therefore, the nonzero wavenumbers include both internally

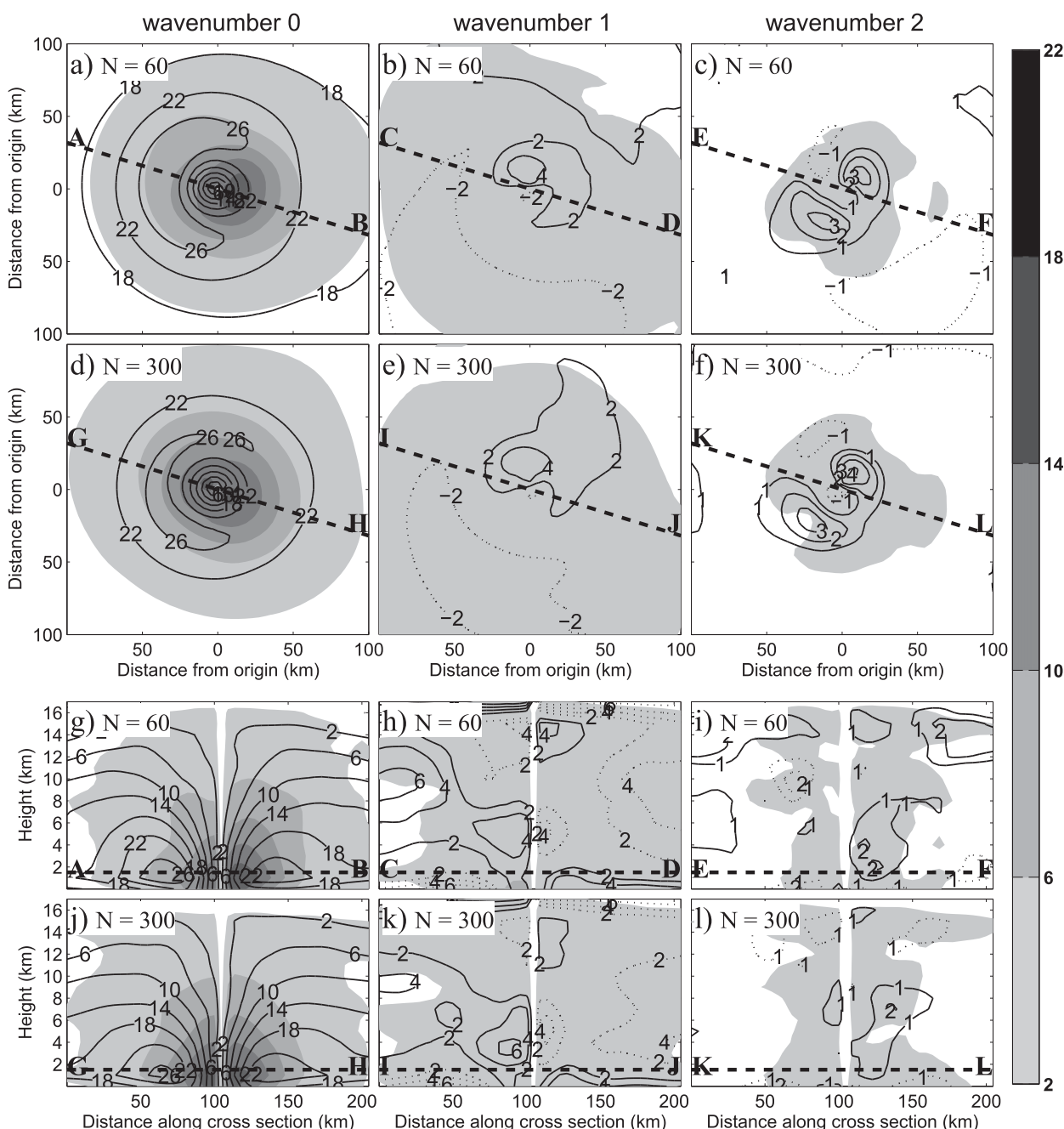


FIG. 10. The [(a)–(c) and (g)–(i)] 60- and [(d)–(f) and (j)–(l)] 300-member ensemble V_θ fields compared in the inner core. Posterior means (solid for positive and dashed for negative) and standard deviations (shaded) at 2000 UTC for (left) $n = 0$, (middle) $n = 1$, and (right) $n = 2$. (a)–(f) Horizontal cross sections through an altitude of 1500 m. (g)–(l) Vertical cross sections through the dashed lines in (a)–(f). The sample size is indicated on the top of each panel.

generated vortex asymmetries as well as those that are induced by the environment. The wavenumber separation is performed with respect to the vortex center of each member, but the ensemble spread in vortex position is maintained when variables are transformed to the ground-relative Cartesian grid.

Figure 10 shows the contribution of the $n = 0, 1$, and 2 wind components to the 60- and 300-member posterior means and standard deviations of V_θ for the last data assimilation cycles. These statistics represent the mean and standard deviations of the filtered ensemble data and should not be confused with a filtered mean and

filtered standard deviations, which would require the use of a common center for all members. Given that Katrina reached category-1 hurricane intensity at this time, the $n = 0$ portion of the ensemble vortices contributes the most to the mean and variance of V_θ , followed by the $n = 1$ and 2 components. Each member is decomposed with respect to its own vortex center in these calculations, so the $n = 0$ portions of the mean and variance in Figs. 10a,d,g,j come from ensembles of axisymmetric vortices with the same position spread as the unfiltered ensembles (Fig. 4a). The contribution from wavenumbers higher than 2 is relatively insignificant compared to the first three components, so they are omitted from the figure. The $n = 1$ and 2 components produce a total of 7 m s^{-1} to the total 1500-m posterior winds inside 50 km (sum of winds in the second and third columns of Fig. 10). Since the filtering is performed in a storm-relative reference frame for each member, the $n = 1$ and 2 winds in the figure must come from collocated asymmetries in the inner core of the ensemble at this time. As in Fig. 5, which shows consistent results between the 60- and 300-member posterior mean winds in the inner core, the ensemble perturbations for these two cases (Fig. 10) share qualitatively similar features. The decomposition of the posterior mean and standard deviations for inner-core radial winds (not shown) also indicate an agreement between the two ensembles, suggesting that the horizontal winds in this region are not sensitive to ensemble size for this case study. Recalling that the smallest radius of influence that is used to localize the ensemble covariance via successive covariance localization is 135 km for these experiments, the close match between the 300- and 60-member ensembles in the inner core shows little sampling sensitivity for ensembles greater than 60 members. This result holds for any reasonable choice of localization [e.g., additional experiments using a fixed localization radius of 405 km for the 60-member ensemble (not shown) provided similar results in the inner core].

Noticeable differences exist between the two ensembles at regions away from the observations (e.g., the $n = 2$ component contributes more to winds in the upper levels of the 60-member posterior mean than in the 300-member posterior mean; Fig. 10i). Figure 2 also shows that sampling differences are largest in the outer core, where no observations are available during the cycling. The smaller ensembles are more likely to produce spurious correlations at great horizontal and vertical distances from the observations (e.g., Hamill and Whitaker 2001), given that the same localization is used for all experiments. Nevertheless, the lack of observations in these regions limits our ability to verify which experiments produce the most accurate results outside the inner core.

Azimuthal wavenumber decomposition for w is shown for the 300- and 60-member posterior ensemble statistics in Fig. 11, using the same approach that was applied in Fig. 10 for V_θ after the last assimilation cycle. Unlike the horizontal winds, the power in w is distributed across a larger spectrum of wavenumbers, so the components are separated into three different bins: $n = 0$, $n = 1, 2$, and 3, and $n > 3$. The $n = 0$ component contributes a maximum of 0.6 m s^{-1} to the posterior mean and 1 m s^{-1} to the total standard deviations of w , showing a weak region of mean ascent around the eyewall in both experiments. The pair of ensembles contains qualitatively similar contributions from low ($n = 1, 2$, and 3) and high ($n > 3$) wavenumbers, each of which favor larger updrafts in the southeast quadrant of the vortex. The largest downdrafts in the eye come from low-wavenumber asymmetries in the members, while the largest updrafts are found in both low and high wavenumbers.

As configured, the EnKF posterior mean in both experiments fails to capture strong updrafts and downdrafts in the eyewall and eye, due to a lack of correlations between the horizontal and vertical winds in the core. This follows from the fact that vortex position uncertainty causes the locations of strongest vertical motion in the ensemble to be displaced in a ground-relative coordinate system, which causes a decorrelation between the storm-relative asymmetric updrafts/downdrafts and the horizontal wind field. As a result, most members contain discrete regions of strong w ($>10 \text{ m s}^{-1}$) in the eyewall that go unchanged after each EnKF analysis. These features are reflected in the relatively large standard deviations of w , which are maximized southeast of the vortex center in the upper levels of the eyewall (Figs. 11b,c,e,f). Prior standard deviations in vortex position at 2000 UTC 25 August decrease from 14.4 km to less than 2 km between the $\alpha = 0.8$ and $\alpha = 0.2$ experiments. Though not shown, the cases that use an α between 0.0 and 0.4 contain more substantial updrafts and downdrafts in the posterior mean; the contribution of each bin of wavenumbers to the total w field in Fig. 11 is about twice as large for these cases. Nevertheless, the standard deviations of w in the eyewall become very small ($<1 \text{ m s}^{-1}$) as α is reduced to 0.0, which is almost certainly an underestimation of the true vertical motion errors, given the assumptions made by the EnKF in producing these analyses. This result supports the recommendation made in section 3c for a lower (but non-zero) relaxation coefficient for ensembles of 60 or more members.

Figure 12 shows the $n = 0, 1$, and 2 contributions of the ensemble T statistics at the last update time. Results are truncated at $n = 2$ because of the relatively insignificant

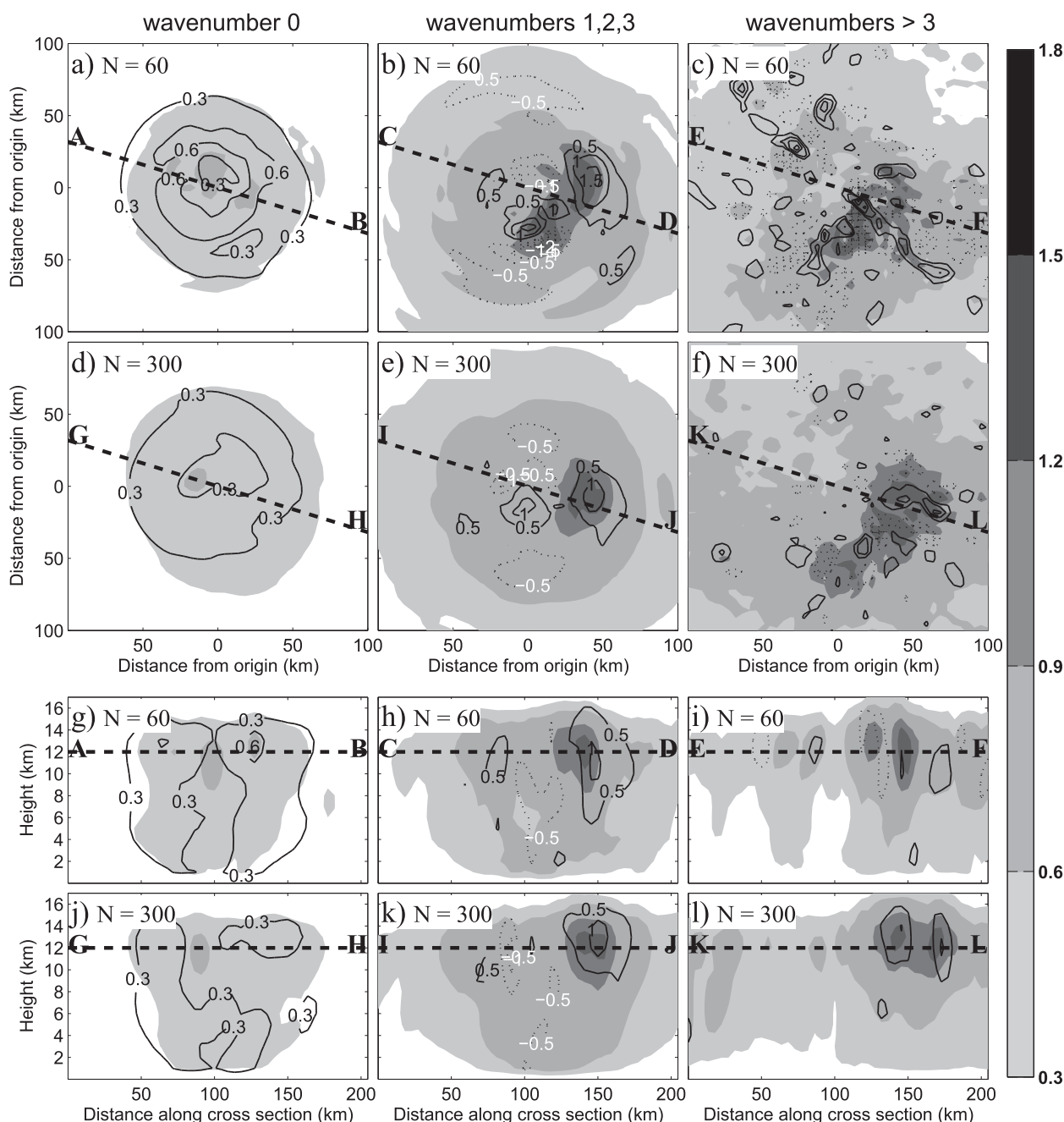
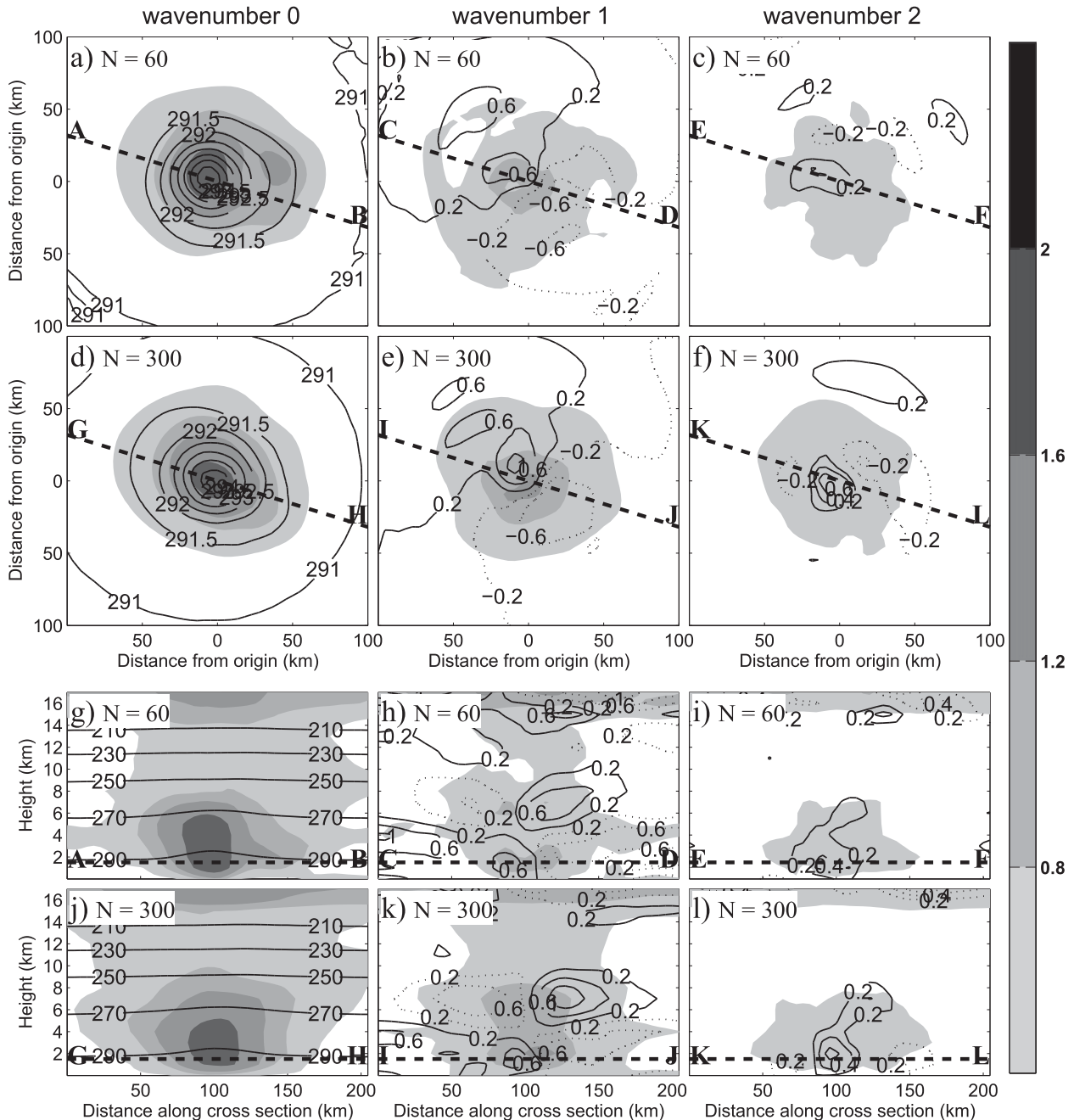


FIG. 11. As in Fig. 10, but for the w field for (left) $n = 0$; (middle) $n = 1, 2$, and 3 ; and (right) $n > 3$. The contours for posterior mean are plotted every 0.3 m s^{-1} for $n = 0$; and 0.5 m s^{-1} for $n > 0$. The horizontal cross sections are through an altitude of 12 000 m.

contribution of the higher wavenumbers to the posterior. The simulated tropical cyclones reach category 1 hurricane intensity by the end of cycling, which leads to thermodynamic structures that are largely dominated by the $n = 0$ component. For the asymmetric portions of T , the $n = 1$ component (second column in Fig. 12) yields slightly cooler (warmer) temperatures ($\pm 1^\circ\text{C}$) in the southeast (northwest) half of the vortex near the surface,

with a reversal of sign at the midlevels. The $n = 2$ component (third column in Fig. 12) contributes an additional $0.1^\circ\text{--}0.3^\circ\text{C}$ to the low- and midlevel temperature near the storm center. The two ensembles produce reasonably similar results for the T decomposition; that is, the $n = 0$ contribution to the mean and standard deviations are within 0.2°C , and the $n = 1$ and 2 values are in phase.

FIG. 12. As in Fig. 10, but for the T field.

The persistent asymmetric midlevel T anomalies coexist with similar $n = 1-2$ moisture anomalies (not shown) for mid- and upper model levels in the southeast portion of the vortices for both ensembles. The location of these thermodynamic anomalies match the posterior w field considerably well, suggesting that the asymmetries are induced by diabatic heating from convective updrafts. Evidence of a strong relationship between updrafts in the eyewall and asymmetric T is given in

Fig. 13, which shows modest correlations (>0.5) between w at a point in the eyewall and the $n = 1-3$ components of T for both ensembles. Correlations between the w and the $n = 0$ portion of T are left out of the figure since they are negligible in this region. Each member of the $N = 300$ ensemble w is filtered in the third column of Fig. 13 to show that the low-wavenumber ($n = 1-3$) w field contributes the most to the correlations. While correlations between w and q_v are not shown here, they

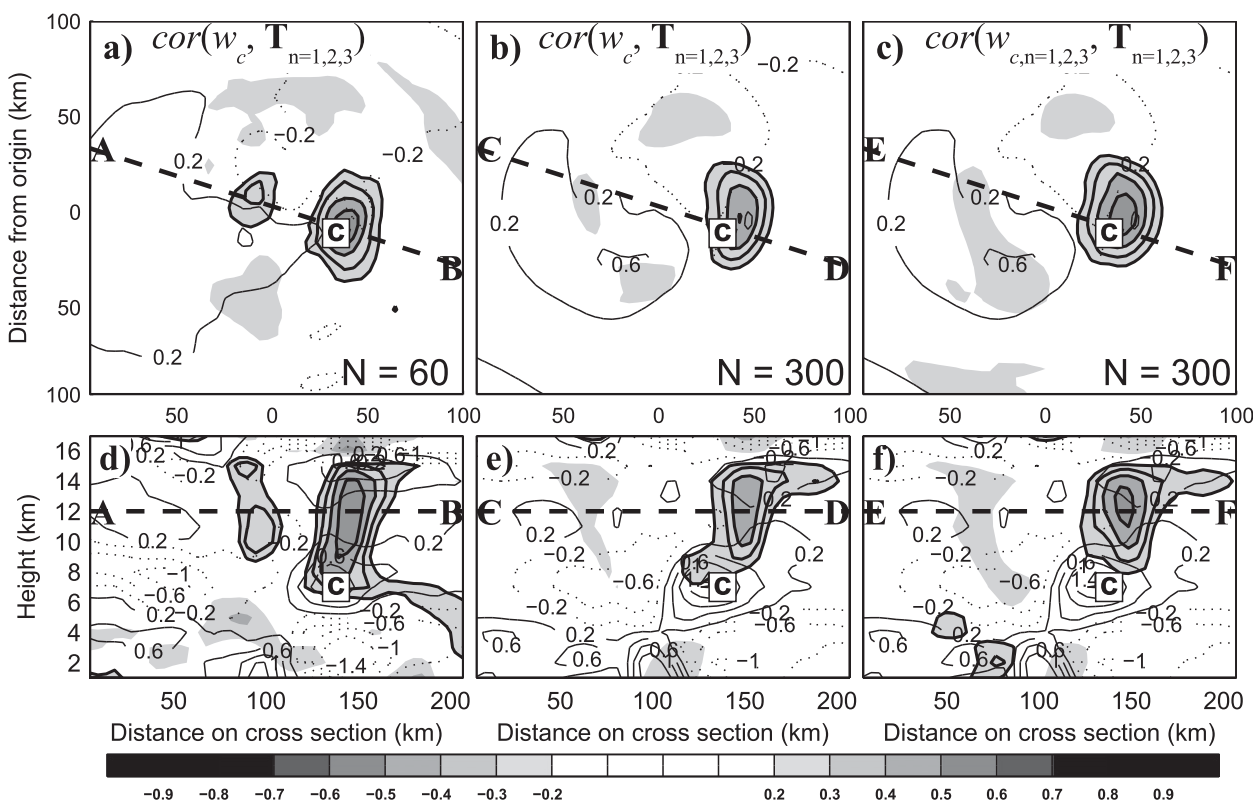


FIG. 13. The $n = 1-3$ contributions to the posterior ensemble mean T contoured (light solid lines for positive, dashed for negative) at 2000 UTC along with correlations (shading) between w at point “c” and the T field. The posterior statistics for the (a),(d) 60- and (b),(e) 300-member ensembles. (c),(f) The w field of each member of the 300-member ensemble filtered to estimate correlations between low-wave-number w and low-wavenumber T asymmetries. (a)–(c) Horizontal cross sections through an altitude of 12 km and (d)–(f) vertical cross sections through the dashed lines in (a)–(c).

follow a similar structure and magnitude in the two ensembles, because warm anomalies in the eyewall must contain higher vapor mixing ratios in order to exceed saturation in these regions. Though neither w nor T are assimilated in these experiments, examples like this one show that a 60-member ensemble can maintain qualitatively similar correlations within the ensemble during cycling.

4. Deterministic forecast results

a. Intensity and track forecasts

Figure 14 compares deterministic forecasts with the National Hurricane Center best track data for all experiments. Each forecast is initialized from the 2000 UTC posterior on 25 August 2005 and run for 130 h to 0600 UTC 31 August. As the simulated storms track westward from Florida, a strengthening ridge over the northern Gulf of Mexico produces strong northeasterly mid- to upper-level flow. This deep-layer steering flow causes the simulated storms (and real storm; Knabb et al. 2005) to track southwestward

from the initialization time before turning northward toward the Gulf Coast. The bias in minimum surface pressure for these simulations was noted in Green and Zhang (2013) to be caused by the choice of surface flux parameterization scheme in WRF (“isftcfx” namelist option). The default surface flux option (used in this study) assumes a monotonically increasing value for the surface drag coefficient at higher wind speeds, which was shown to produce a pressure–wind relationship that is inferior to schemes that cap the surface drag coefficient at a set wind speed. This possible source of model error is expected to have little impact on the EnKF analyses described in the previous section, because the storm intensity during the data assimilation period is too low for the drag-induced pressure bias to occur.

With the exception of a weaker intensity forecast in the 30-member experiment, all cases provided similar predictions for the track and intensity of Katrina, showing little forecast sensitivity to ensemble size and covariance relaxation under the given circumstances. Despite these similarities, notable changes in inner- and outer-core storm structure appear after assimilating the

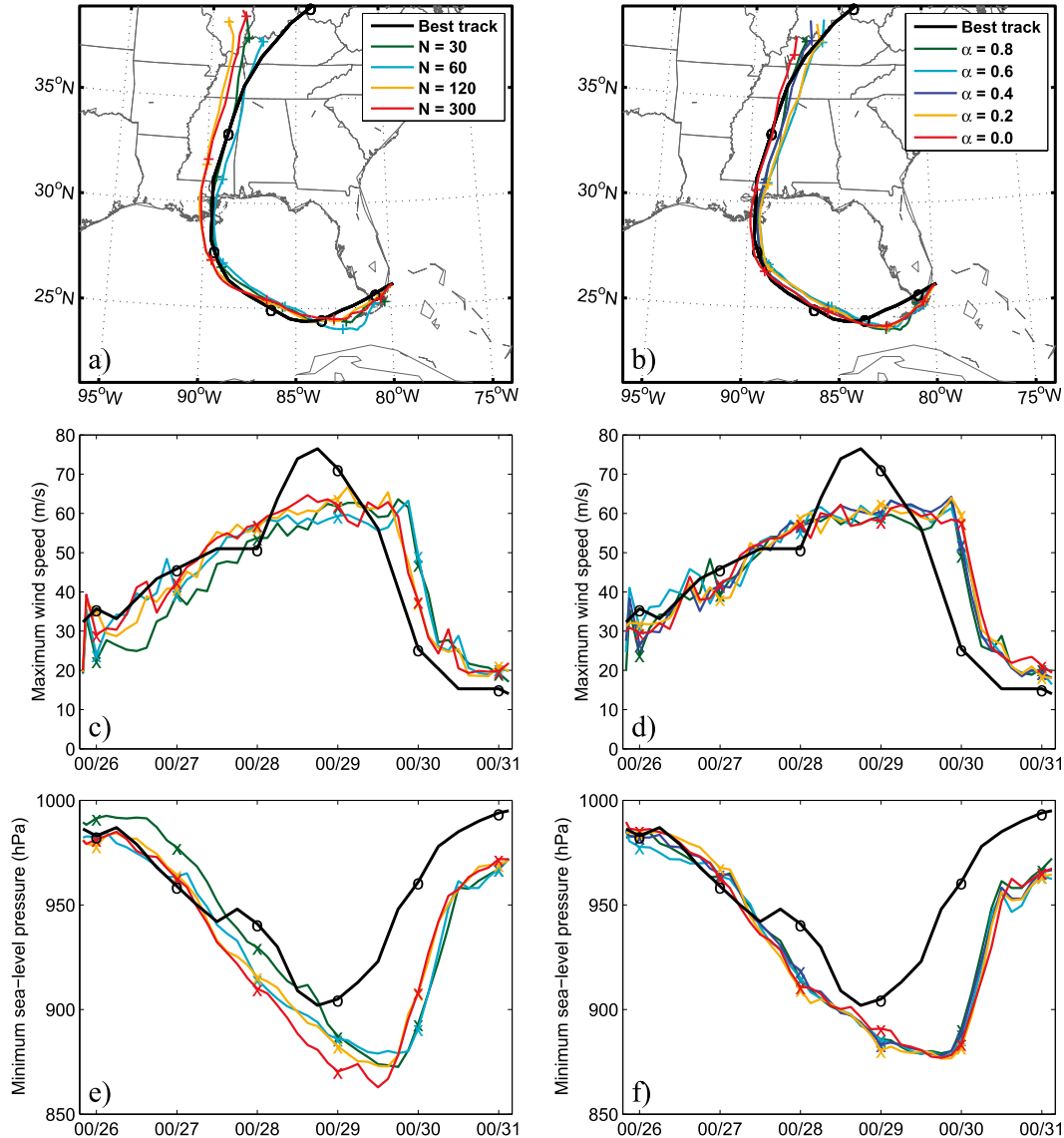


FIG. 14. Deterministic forecasts of (a),(b) track; (c),(d) maximum 10-m winds; and (e),(f) minimum surface pressure for (a),(c),(e) the ensemble size experiment and (b),(d),(f) covariance relaxation experiments for $\alpha = 0.0$ – 0.8 . The ○ and × symbols indicate best track and forecast values, respectively, at 0000 UTC on each day.

radar observations, especially for cases that use different ensemble sizes.

b. Storm structure after initialization

Figures 3g–l show azimuthal mean V_θ and P for 24-, 48-, and 72-h forecasts that were initialized from the 2000 UTC 25 August posterior means. The storm evolution in each of these forecasts suggests that the initial wind and pressure differences in Figs. 3c,f are significant for the development of the tropical cyclone vortex, despite the small variability in track and maximum surface winds (Fig. 14). Sampling errors lead to 5 m s^{-1} or greater azimuthal mean wind differences between the

300- and 60-member posteriors outside of 100 km, with values that exceed 10 m s^{-1} by 72 h in the deterministic forecasts. The 120- and 300-member ensembles produce comparable forecasts of storm structure because of similar posteriors at 2000 UTC, though an explanation for why the 30- and 300-member forecasts provide similar outer-core solutions is nontrivial. While the 30-member posterior mean vortex contains an outer core that closely resembles the vortex in the 60-member case at the end of cycling, the vortex initialized by the 30-member mean undergoes an adjustment toward a solution that more closely matches the cases that uses 120 or more members. The proximity of Katrina to land at the

initialization time makes it difficult to determine exactly how the vortex in the 30-member experiment adjusts toward a similar outer-core wind and pressure field as in the 300-member experiment.

Though not shown, NOAA P3 missions on 27 and 28 August indicate a positive bias in the outer-core winds for the deterministic forecasts initialized from cases that use larger ensembles. This verification is consistent with our results from sections 3b and 3c, where the inner-core variance deficiency at the initial update time is shown to produce a positive outer-core wind bias that persists to the last data assimilation cycle. Sampling errors in the 60-member ensemble happen to offset the error induced by using an underdispersed ensemble, thus causing weaker outer-core winds by the last update cycle (cf. Fig. 3c). However, we acknowledge that other sources of error may have also caused the outer-core wind bias (e.g., the wind field adjusting to the low pressure bias described in section 4).

5. Summary and conclusions

A set of cycling data assimilation experiments are performed for a developing tropical cyclone case study to examine the sensitivity of an EnKF to sampling errors in a real-data application. Airborne Doppler radar observations that were collected from NOAA P3 flight missions for Hurricane Katrina (2005) are assimilated in six short assimilation cycles using ensemble sizes that range from 30 to 300 members and varying degrees of covariance relaxation. The experiments are carried out using WRF, nested down to a cloud-permitting (4.5 km) model grid spacing. Results are examined for the inner- and outer-core region of the vortex, where the data assimilation has the largest effect.

Deterministic forecasts from the EnKF posterior means at the last assimilation cycle produce similar track and intensity predictions for the developing hurricane. Nevertheless, sampling differences in the ensemble-size experiments cause variations in the outer-core pressure and wind fields that persist for at least 3 days in simulations. While covariance relaxation has almost no impact outside the inner core during data assimilation, sampling errors in experiments that use less than 60 members can lead to significant changes in the outer wind and pressure field. These outer-core differences emerge at the first analysis time and carry through to the remaining cycles. Ensemble variance increases rapidly in the inner core during the assimilation cycles, owing to corrections in storm location and intensity by the EnKF. The disproportionately large increase in variance in the inner core, compared to other locations in the domain, contributes to the lack of significant outer-core increments after the first

analysis time. A verification of the inner-core ensemble variance shows that the true forecast error may be underestimated by a factor of 2–3 at the first assimilation time, which causes the initial set of outer-core increments to be too large. Experiments using an inflation factor in the inner core show significant impacts for the outer-core analysis, but additional research is needed to understand the full consequences of the variance deficiency. In general, additional steps should be taken to either inflate the variance near the vortex center or reduce the localization radius when the ensemble contains a significant amount of vortex position spread. Furthermore, the lack of observations in the hurricane outer core may have increased the sensitivity of the outer wind and pressure fields to ensemble size, since the final structure of the posterior vortex relies greatly on how covariance between distant grid points are represented at the first update time.

For the inner core, two experiments are distinguished as outliers early in this study; the 30-member ensemble fails to produce a strong axisymmetric vortex and the ensemble variance in the $\alpha = 0$ (no relaxation) case collapses by the end of cycling. The kinematic and thermodynamic structure of the inner core is consistent among experiments that use 60 or more members and a modest amount of relaxation (≥ 0.4). While the vortex structure in the 60-member cases shows little sensitivity to the choice of α , smaller ensembles are expected to produce a much greater sensitivity to relaxation, owing to the larger sampling error (e.g., Aksoy 2013). The examples show that a 60-member ensemble can contain qualitatively similar wind and temperature asymmetries as a 300-member ensemble after several assimilation cycles. In conclusion, 60 members appears to be a sufficient ensemble size for capturing many of the important features of the tropical cyclone inner core when provided with a high-resolution set of wind observations. Nevertheless, this result is only valid in the vicinity of the radar observations. Though not shown, we also find that the storm structure in the 30-member case can be improved by decreasing the localization radius; however, the solution never approaches the same storm size and intensity that are found with ensembles of 60 or more members.

While it is desirable to reproduce these experiments for additional tropical cyclones, we are limited by the computational expense that is required to assimilate observations using the resolution and ensemble sizes used in this study. The main objective is to explore possible sampling differences that may result from using a range of ensemble sizes and covariance relaxation coefficients for a high-resolution cycling data assimilation case. This study uses the default setup of an ensemble data assimilation system that has been applied over the

past five Atlantic hurricane seasons (Zhang et al. 2011) under NOAA's Hurricane Forecast Improvement Program (HFIP; Gall et al. 2013). Our choice of localization radii and relaxation coefficients may not necessarily be optimal for the set of ensembles, but the configuration has been shown to be reasonable enough to allow for an examination of sampling errors in both the inner- and outer-core areas of a hurricane vortex. Applying an ensemble-size-dependent localization radius of influence and tuning the relaxation coefficient are two ways of coping with some of the model and sampling errors we observed in our experiments. Nevertheless, finding the optimal balance between ensemble size, localization, and relaxation is beyond the scope of the current study.

Acknowledgments. This work was supported in part by the NOAA Hurricane Forecast Improvement Project (HFIP), Office of Naval Research Grant N000140910526 and the National Science Foundation Grant ATM-0840651. We are thankful for the comments provided by Altug Aksoy and an anonymous reviewer for improving the manuscript. The computing was performed at the Texas Advanced Computing Center.

REFERENCES

- Aberson, S. D., B. L. Black, R. W. Burpee, J. J. Cione, C. W. Landsea, and F. D. Marks, 2006: Thirty years of tropical cyclone research with the NOAA P-3 aircraft. *Bull. Amer. Meteor. Soc.*, **87**, 1039–1055, doi:10.1175/BAMS-87-8-1039.
- Aksoy, A., 2013: Storm-relative observations in tropical cyclone data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **141**, 506–522, doi:10.1175/MWR-D-12-00094.1.
- , D. C. Dowell, and C. Snyder, 2009: A multicaser comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824, doi:10.1175/2008MWR2691.1.
- , S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for high-resolution data: The impact of airport Doppler radar observations in an OSSE. *Mon. Wea. Rev.*, **140**, 1843–1862, doi:10.1175/MWR-D-11-00212.1.
- Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, doi:10.1111/j.1600-0870.2008.00361.x.
- , and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758, doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.
- Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914, doi:10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2.
- Chen, Y., and C. Snyder, 2007: Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.*, **135**, 1828–1845, doi:10.1175/MWR3351.1.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145, doi:10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2.
- Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and thermodynamic retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005, doi:10.1175/1520-0493(2004)132<1982:WATRIT>2.0.CO;2.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, doi:10.1029/94JC00572.
- Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.*, **135**, 1846–1868, doi:10.1175/MWR3391.1.
- Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329–343, doi:10.1175/BAMS-D-12-00071.1.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi:10.1002/qj.49712555417.
- Green, B. W., and F. Zhang, 2013: Impacts of air–sea flux parameterizations on the intensity and structure of tropical cyclones. *Mon. Wea. Rev.*, **141**, 2308–2324, doi:10.1175/MWR-D-12-00274.1.
- Hamill, T. H., and J. S. Whitaker, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790, doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.
- Houze, R. A., S. S. Chen, B. F. Smull, W.-C. Lee, and M. M. Bell, 2007: Hurricane intensity and eyewall replacement. *Science*, **315**, 1235–1239, doi:10.1126/science.1135650.
- Huang, X.-Y., and Coauthors, 2009: Four-dimensional variational data assimilation for WRF: Formulation and preliminary results. *Mon. Wea. Rev.*, **137**, 299–314, doi:10.1175/2008MWR2577.1.
- Kalman, R. E., and R. S. Bucy, 1960: New results in linear filtering and prediction theory. *J. Basic Eng.*, **83**, 95–108, doi:10.1115/1.3658902.
- Knabb, R. D., J. R. Rhome, and D. P. Brown, 2005: Tropical cyclone report: Hurricane Katrina, 23–30 August 2005. NOAA/NHC, 43 pp. [Available online at www.nhc.noaa.gov/pdf/TCR-AL122005_Katrina.pdf.]
- Marchok, T., 2010: Use of the GFDL vortex tracker. WRF tutorial for hurricanes, NOAA/GFDL, 40 pp. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/presentations/tutorial02222010/tracker_slides.pdf.]
- Meng, Z., and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect-model experiments. *Mon. Wea. Rev.*, **135**, 1403–1423, doi:10.1175/MWR3352.1.
- , and —, 2008a: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVar in a real-data case study. *Mon. Wea. Rev.*, **136**, 522–540, doi:10.1175/2007MWR2106.1.
- , and —, 2008b: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a month-long experiment. *Mon. Wea. Rev.*, **136**, 3671–3682, doi:10.1175/2008MWR2270.1.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–433, doi:10.1175/1520-0493(2000)128<0416:AAEKF>2.0.CO;2.
- Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble

- transform Kalman filter. *Mon. Wea. Rev.*, **139**, 1519–1535, doi:10.1175/2010MWR3570.1.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763, doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.
- Poterjoy, J., and F. Zhang, 2011: Dynamics and structure of forecast error covariance in the core of a developing hurricane. *J. Atmos. Sci.*, **68**, 1586–1606, doi:10.1175/2011JAS3681.1.
- Sippel, J. A., S. A. Braun, F. Zhang, and Y. Weng, 2013: Ensemble Kalman filter assimilation of simulated HIWRAP Doppler velocity data in a hurricane. *Mon. Wea. Rev.*, **141**, 2683–2704, doi:10.1175/MWR-D-12-00157.1.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 125 pp. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677, doi:10.1175//2555.1.
- Talagrand, O., 1997: Assimilation of observations, an introduction. *J. Meteor. Soc. Japan*, **75**, 191–209.
- Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133**, 1789–1807, doi:10.1175/MWR2898.1.
- Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev.*, **136**, 3947–3963, doi:10.1175/2008MWR2443.1.
- , —, and C. Snyder, 2006: Boundary conditions for a limited-area ensemble Kalman filter. *Mon. Wea. Rev.*, **134**, 2490–2502, doi:10.1175/MWR3187.1.
- Weatherford, C. L., and W. M. Gray, 1988: Typhoon structure as revealed by aircraft reconnaissance. Part I: Data analysis and climatology. *Mon. Wea. Rev.*, **116**, 1032–1043, doi:10.1175/1520-0493(1988)116<1032:TSARBA>2.0.CO;2.
- Weng, Y., and F. Zhang, 2012: Assimilating airborne Doppler radar observations with an ensemble Kalman filter for convection-permitting hurricane initialization and prediction: Katrina (2005). *Mon. Wea. Rev.*, **140**, 841–859, doi:10.1175/2011MWR3602.1.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.
- , and —, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, doi:10.1175/MWR-D-11-00276.1.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.
- , Z. Meng, and A. Aksoy, 2006: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect-model experiments. *Mon. Wea. Rev.*, **134**, 722–736, doi:10.1175/MWR3101.1.
- , Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125, doi:10.1175/2009MWR2645.1.
- , —, J. F. Gamache, and F. D. Marks, 2011: Performance of convection-permitting hurricane initialization and prediction during 2008–2010 with ensemble data assimilation of inner-core airborne Doppler radar observations. *Geophys. Res. Lett.*, **38**, L15810, doi:10.1029/2011GL048469.