Ensemble Kalman Filter Assimilation of Simulated HIWRAP Doppler Velocity Data in a Hurricane

JASON A. SIPPEL AND SCOTT A. BRAUN

NASA Goddard Space Flight Center, Greenbelt, Maryland

FUQING ZHANG AND YONGHUI WENG

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

ABSTRACT

This study utilizes ensemble Kalman filter (EnKF) observing system simulation experiments (OSSEs) to analyze the potential impact of assimilating radial velocity observations of hurricanes from the High-altitude Imaging Wind and Rain Airborne Profiler (HIWRAP). HIWRAP is a new Doppler radar mounted on the NASA Global Hawk unmanned airborne system that flies at roughly 19-km altitude and has the benefit of a 25–30-h flight duration, which is 2–3 times that of conventional aircraft. This research is intended as a proof-of-concept study for future assimilation of real HIWRAP data. The most important result from this research is that HIWRAP data can potentially improve hurricane analyses and prediction. For example, by the end of a 12-h assimilation period, the analysis error is much lower than that in deterministic forecasts. As a result, subsequent forecasts initialized with the EnKF analyses also improve. Furthermore, analyses and forecasts clearly benefit more from a 12-h assimilation period than for shorter periods, which highlights a benefit of the Global Hawk's potentially long on-station times.

1. Introduction

It has become increasingly evident over recent years that ensemble Kalman filter (EnKF) assimilation of high-resolution observations from tropical cyclones can improve tropical cyclone analyses and forecasts. To this end, Doppler velocity data (Zhao and Jin 2008; Zhang et al. 2009, hereafter Z09; Zhang et al. 2011, hereafter Z11; Weng and Zhang 2012, hereafter WZ12; Aksoy et al. 2012; Dong and Xue 2013) have been particularly helpful. Though Doppler data from coastal Weather Surveillance Radar-1988 Doppler (WSR-88D) radars can certainly benefit analyses, perhaps more useful are data obtained from reconnaissance aircraft because that data can improve prediction several days before landfall. Among published studies, the results of Z11 best demonstrate the systematic advantage of assimilating airborne Doppler velocity data. After assimilating radial velocity observations from the National

Oceanic and Atmospheric Administration (NOAA) WP-3D tail Doppler radar over a total of 61 missions from 14 tropical cyclones, they found a significant reduction in forecast error relative to operational guidance. Their deterministic track forecasts initialized with EnKF analyses were as good as or better than those from operational models, and error in their intensity forecasts was 20%–40% lower than that in the official NHC forecast.

Even when ensemble data assimilation is used to initialize tropical cyclone vortices, difficulties still remain. A common problem has been that many tropical cyclones are too far out over the ocean to reach with current manned aircraft, which severely limits the in situ observations that can be taken. Alternatively, even when storms are within range, the typical on-station time for radar-bearing aircraft is less than 6 h. As a result, there are often large gaps in aircraft-based data even for storms that are relatively close to the United States. Thus, the typically short duration of aircraft-based radar observations may not be sufficient to analyze some tropical cyclones.

One potential solution to these difficulties has come with the use of high-altitude, long-duration, unmanned

Corresponding author address: Dr. Jason A. Sippel, NASA GSFC, Code 612, Greenbelt, MD 20771.
E-mail: jason.sippel@nasa.gov

aircraft systems (UASs). In 2010, the National Aeronautics and Space Administration (NASA) Genesis and Rapid Intensification Processes (GRIP) experiment used a Global Hawk UAS equipped with a Doppler radar to study Atlantic hurricanes (Braun et al. 2013). The Global Hawk is an ideal platform for observing tropical cyclones because of its long flight duration (up to 30 h), high altitude (>18 km), and ability to carry large payloads. If deployed from the eastern United States, the 20 000-km flight range allows the Global Hawk to be on station over storms for 12–16 h in the central Atlantic and 16–22 h in the western Atlantic (Braun et al. 2013).

For the purposes of this study, we are most interested in radial velocity (Vr) data from the High-Altitude Imaging Wind and Rain Airborne Profiler (HIWRAP; Heymsfield et al. 2008), the Doppler radar flown on the Global Hawk during GRIP (Braun et al. 2013). Since HIWRAP is a new instrument with a different scanning geometry than radars previously used for tropical cyclone analyses, nothing is known about the potential usefulness of its data for assimilation purposes. Because the Global Hawk flies at altitudes much higher than operational aircraft, HIWRAP scans downward into tropical cyclones with incidence angles of 30° and 40° (i.e., elevation angles of −60° and −50°). This means that the radar observes a relatively large component along the vertical axis. Unfortunately, vertical velocity tends to be small in scale, noisy, and weakly correlated with other state variables (Poterjoy and Zhang 2011), which might degrade analyses using HIWRAP data. In an extreme example of a 0° incidence angle (observing only vertical velocity), Vr would not be useful for hurricane initialization. Thus, a significant motivation of this study is to investigate whether the portion of horizontal wind that HIWRAP observes is sufficient to produce an accurate analysis.

There are several reasons why we have taken the simulated-observation approach to these experiments. First and foremost, HIWRAP was flown for the first time during GRIP, and radial velocity data were not available for assimilation tests when the experiments herein were conducted. Second, key benefits of observing system simulation experiments (OSSEs) include the ability to accurately characterize the error of the assimilation system and to easily assess the benefits of assimilating multiple types of observations or using different observation network configurations. Therefore, this study will necessarily make use of simulated HIWRAP observations to provide a proof-of-concept study for future OSSEs and real-data assimilation studies.

The remainder of this study is organized as follows. Section 2 describes the storm to be studied and the
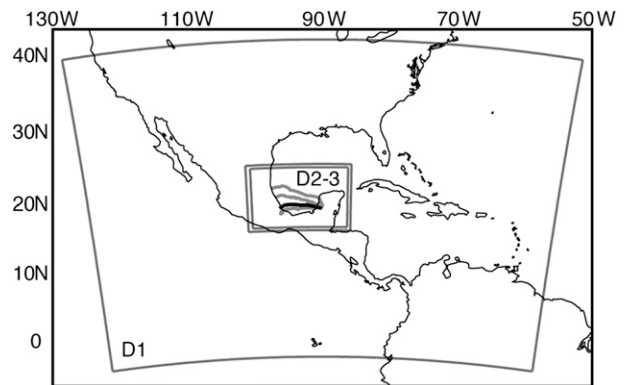

FIG. 1. Domain setup for WRF along with the observed (black) and realization (gray) tracks of Karl from 0000 UTC 16 Sep to 0000 UTC 18 Sep 2010.

methodology for the OSSEs. Section 3 examines EnKF analyses as well as deterministic and ensemble forecasts from those analyses. Section 4 explores various sensitivities, and section 5 follows with the discussion and conclusions.

2. Background and methods

The OSSEs herein will focus on Hurricane Karl, the only GRIP hurricane for which adequate HIWRAP data were gathered. The pre-Karl depression formed from a westward-moving low pressure area in the central Caribbean, and the system attained tropical storm strength north of Honduras early on 15 September 2010. Karl intensified fairly quickly to near-hurricane intensity, making landfall on the Yucatán Peninsula shortly after 1200 UTC 15 September. The system weakened only slightly as it moved across the peninsula, and the inner core remained largely intact as it emerged over the Bay of Campeche early on 16 September (Braun et al. 2013). Thereafter, the storm rapidly intensified to a hurricane by 0000 UTC 17 September and to a major hurricane by 1200 UTC that same day. Karl moved westward across the Bay of Campeche, and it made landfall just before 1800 UTC north of Veracruz, Mexico (Fig. 1).

The period of Karl we have chosen to investigate corresponds with unprecedented sampling of the storm during GRIP (Braun et al. 2013). A U.S. Air Force WC-130 arrived around 1100 UTC 16 September, and thereafter up to five aircraft simultaneously monitored Karl through 0800 UTC 17 September. The Global Hawk arrived around 1800 UTC 16 September and stayed over the storm for roughly 14 h, during which time it overpassed the eye 20 times. From the standpoint of data assimilation, this is an ideal period to study

TABLE 1. Descriptions of the various assimilation experiments along with a listing of subsequent deterministic and ensemble forecasts. The parenthetical note for the forecasts in AOERR indicates the number of forecasts performed for each analysis time (i.e., with different random error vectors).

Expt	Description	Deterministic	Ensemble
NODA	Forecast from GFS initial conditions at 0000 UTC 16 Sep 2010 with no subsequent assimilation	NODA-DET	NODA1-ENS, NODA2-ENS, and NODA3-ENS
CTRL1	Experiment with good first-guess error covariance structure and assimilation from 12 to 24 h	CTRL1-DET	CTRL1-ENS
CTRL2–3	Experiment with an alternate choices for the truth and thus poor first-guess error covariance structure and assimilation from 12 to 24 h	CTRL2-DET and CTRL3-DET	CTRL2-ENS and CTRL3-ENS
VLEG	Experiment with CTRL1 truth realization but alternate choice for flight pattern with variable-length legs and assimilation from 12 to 24 h	VLEG-DET	VLEG-ENS
AOERR	Experiments exactly as in CTRL1, but with different random error applied to raw observations	AOERR-DET (10)	AOERR-ENS (3)
FLTRW	Experiment exactly as in CTRL1, but observations where $w/\text{Vr} > 0.25$ are rejected	No forecast	No forecast
MIX05	Experiment exactly as in CTRL1, but with $\alpha = 0.5$ instead of 0.8	No forecast	No forecast
P3GEO	Experiment with CTRL1 truth realization, but with the P3 tail Doppler radar geometry instead of the HIWRAP geometry	No forecast	No forecast

because of the abundance of data from multiple sources. Our eventual goal is to use extensive OSSEs and real-data assimilation to compare the benefits of assimilating different data types, including Vr from both HIWRAP and the WP-3D tail Doppler radar.

a. Initial no data assimilation and truth simulations

All simulations herein use the Weather Research and Forecasting model, version 3 (WRF V3.1.1). The domain setup, which is shown in Fig. 1, uses two-way nests to achieve 3-km grid spacing over the Bay of Campeche and surrounding terrain. Domain 2 is kept to the minimum possible size to reduce the computation burden associated with the large number of runs required for this study. All model domains have 35 vertical layers, and the model top is set at 10 hPa. Model physics choices include the Kain–Fritsch cumulus scheme (Kain and Fritsch 1990, 1993) on the 27- and 9-km grids, and the Yonsei State University scheme (Noh et al. 2003) for planetary boundary layer (PBL) processes. To include at least some representation of model error, simulations to be regarded as realizations of the "truth" use the Goddard microphysics with graupel (Tao and Simpson 1993), whereas the forecast model simulations use the WRF single-moment 6-class microphysics with graupel (Hong et al. 2004). Radiative processes are calculated using the Rapid Radiative Transfer Model longwave (Mlawer et al. 1997) and Dudhia shortwave (Dudhia 1989) schemes.

Deterministic and ensemble experiments with no data assimilation (NODA) were created with WRF in order to establish benchmarks of forecast error for the OSSEs. The deterministic forecast (hereafter NODA-DET; see Table 1 for a full list of experiment nomenclature) was created by integrating the Global Forecast System (GFS) analysis from 0000 UTC 16 September forward for 48 h. Lateral boundary conditions used for this simulation came from the 6-hourly GFS forecast files. Meanwhile, a set of 31 initial and boundary condition perturbations to the GFS analysis and forecast was used to create an ensemble of forecast realizations. The random large-scale differences from GFS were created by implanting noise derived from National Centers for Environmental Prediction (NCEP) background error covariance statistics into the WRF variational data assimilation system (e.g., Barker et al. 2004). This approach is commonly used to create ensembles in WRF, both for data assimilation and ensemble sensitivity studies (e.g., Z09; Sippel and Zhang 2008). Model integration for all NODA forecasts commences at 0000 UTC 16 September, as Karl was exiting the Yucatán Peninsula, and lasts for 48 h.

We assess assimilation from several different truth simulations here to demonstrate the repeatability of the results, particularly when the first-guess ensemble delivers poor error covariance. The initial and boundary conditions for each truth simulation were chosen from among the aforementioned 31 realizations, a selection that was based mostly upon storm tracks in the initial ensemble. The best scenario (hereafter TRUTH1) is one in which the truth tracks near the center of the envelope (Fig. 2a) and yields an initial background error
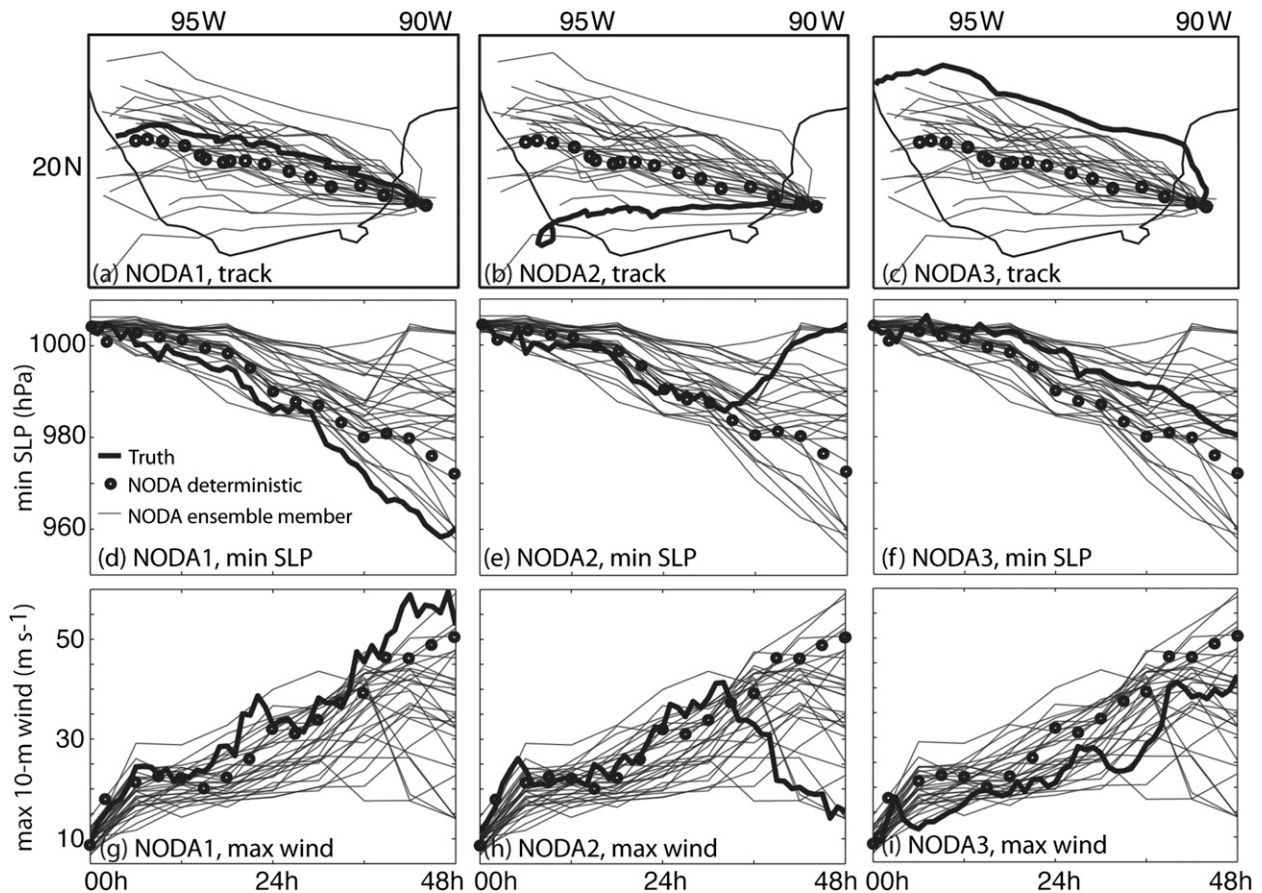
FIG. 2. (left) TRUTH1, (middle) TRUTH2, and (right) TRUTH3 realizations along with their respective NODA ensemble and deterministic forecasts of (a)–(c) track, (d)–(f) minimum SLP, (g)–(i) maximum 10-m wind speed beginning at 0000 UTC 16 Sep. NODA-DET data are shown every 3 h.

covariance field that produces generally accurate analysis increments. Undesirable scenarios (hereafter TRUTH2 and TRUTH3) occur when the truth moves along the edge of the ensemble envelope (Figs. 2b,c) so that the initial background error covariance field can yield highly inaccurate increments.

The method for selecting the truth simulations necessitates three slightly different NODA ensembles (hereafter NODA1-ENS, NODA2-ENS, and NODA3-ENS). Each ensemble comprises the remaining 30 realizations not used to create its respective truth simulation. For example, after the selection of the TRUTH1 initial and boundary conditions, NODA1-ENS was created with the remaining 30 members. Though the NODA ensembles differ by one member, the effect is barely noticeable in Fig. 2, and the ensemble mean for each is almost identical.

In addition to having different tracks, the three truth cyclones also vary in structure and intensity. The TRUTH1 storm evolves most similarly to Hurricane Karl in the sense that it rapidly intensifies from roughly 10 to $60 \, \mathrm{m \, s^{-1}}$ over the 48-h forecast period (Fig. 2g). The storm is fairly asymmetric during the first 36 h (Figs. 3a,c), but it becomes more symmetric as it approaches and exceeds the category-3 threshold (Figs. 2g and 3e). Figure 3a shows that the cyclone has an organized inner core (i.e., at least a partial ring of convection around the center and an associated near-center wind maximum) by 24 h, and by 48 h it has a symmetric eyewall and rain-free eye (Fig. 3e). Meanwhile, the storm in TRUTH2 intensifies at a slightly slower pace during the first 18 h, but it briefly exceeds the TRUTH1 intensity after a period of accelerated intensification from 18 to 30 h (Fig. 2h). This intensification abruptly halts at 30 h, presumably due to interaction with land, though it does not make landfall until after 36 h (Figs. 2e,h and 4c). Although the TRUTH2 cyclone is much more asymmetric than that in TRUTH1, it does manage to establish an inner core with some convection wrapping around the center (Figs. 4a,c). Finally, the TRUTH3 cyclone evolves somewhat differently than the first two. It intensifies at a slower rate (Figs. 2f,i) and takes longer to establish an
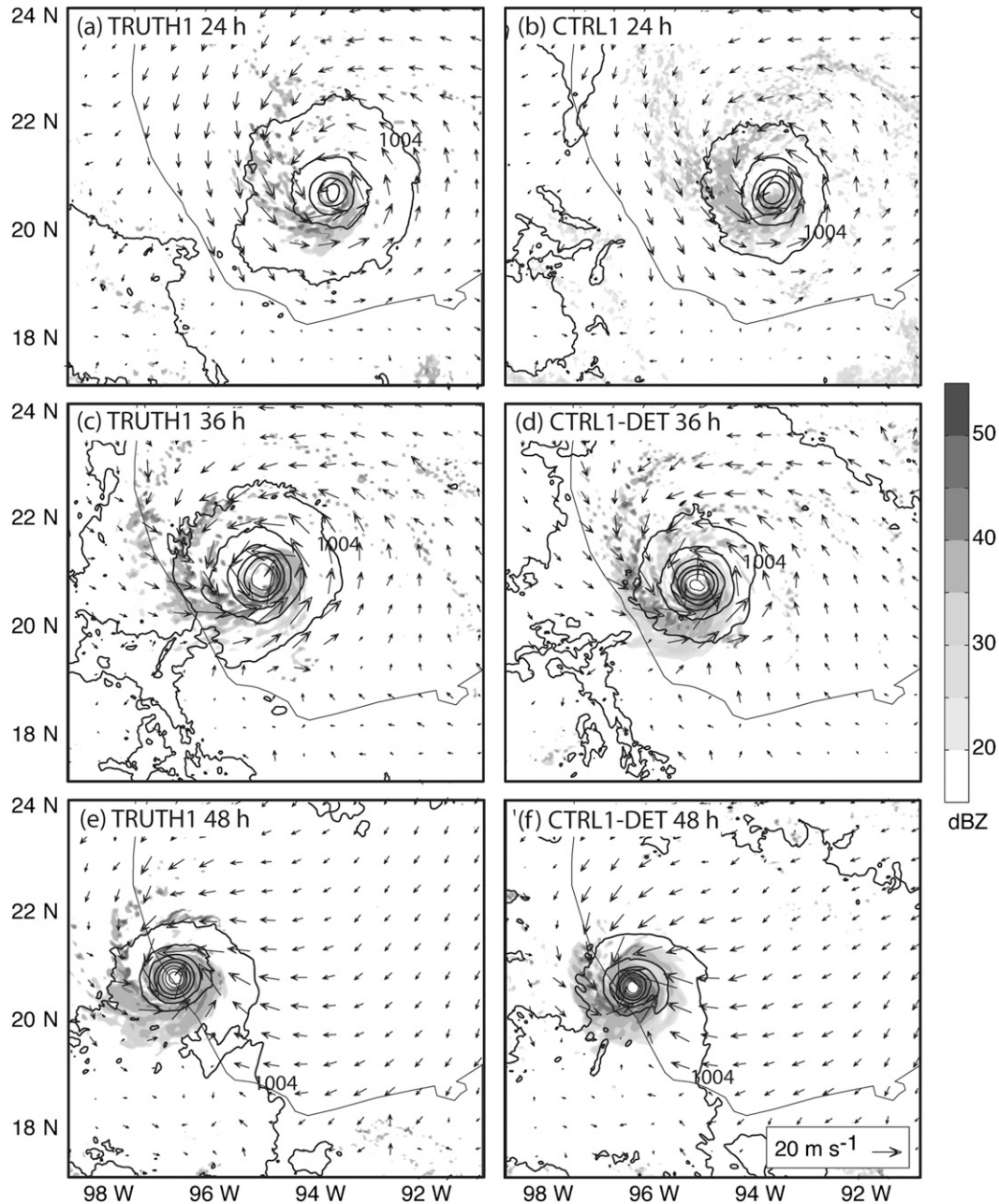
FIG. 3. (left) TRUTH1 realization and (right) final CTRL1 analysis and CTRL1-DET forecast of SLP (contoured every 8 hPa up to 988 hPa and every 4 hPa thereafter), 10-m winds [see (f) inset for reference vector], and reflectivity at 1-km altitude (shaded every 5 dBZ) at (a),(b) 24; (c),(d) 36; and (e),(f) 48 h from a base time of 0000 UTC 16 Sep.

organized inner core (Fig. 5a). It also has a broader circulation than the first two truth realizations (Figs. 3–5c), which could help explain its slower intensification rate. The storm strengthens more quickly from 30 to 48 h, which coincides with the establishment of an inner core (Fig. 5c), though it never exceeds category-1 intensity. Although the system is strongly asymmetric through

36 h, it rapidly organizes thereafter and becomes more symmetric than the stronger TRUTH1 storm by the time of its landfall (Figs. 3e and 5e).

The NODA deterministic forecast, which is shown in Figs. 2 and 6, is associated with varying degrees of error depending on the version of the truth being considered. NODA-DET's track error for TRUTH1 is small, though
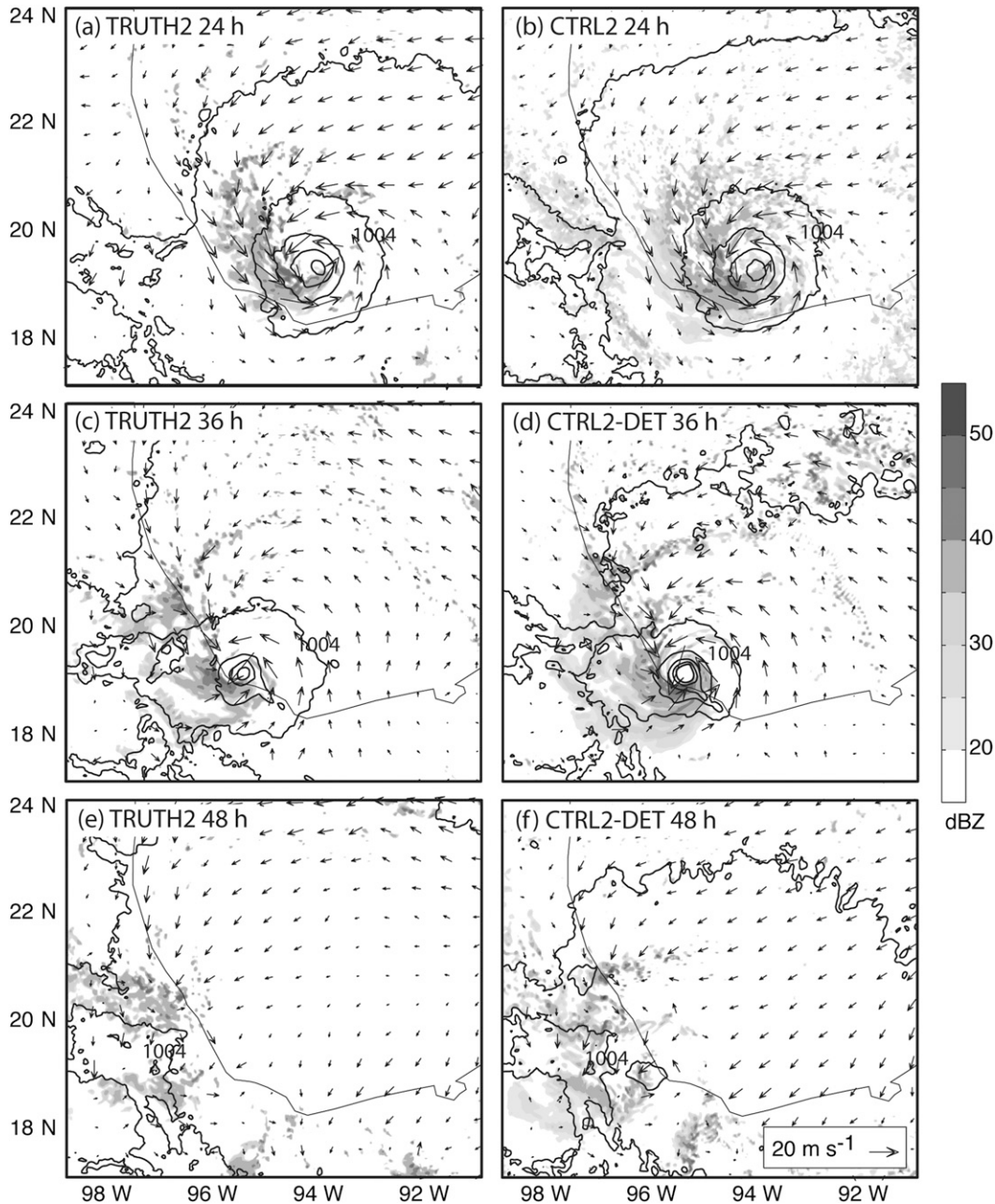
FIG. 4. As in Fig. 3, but for (left) TRUTH2 and (right) the CTRL2 analysis and CTRL2-DET forecast.

there is a consistent south-of-track bias (Fig. 2a). The forecast intensity is generally less than that in TRUTH1 (Figs. 2d,g), and the intensity error magnitude occasionally exceeds that of operational forecasts with similar lead times (Cangialosi and Franklin 2011). This is particularly true toward the end of the forecast, when NODA-DET underestimates the maximum winds by more than $10 \, \mathrm{m \, s^{-1}}$. The structure of the NODA-DET storm (Fig. 6) is similar to that in TRUTH1 (Fig. 3), though the NODA-DET inner core does not have as

much convection wrapping around the center. Meanwhile, the track error for both TRUTH2 and TRUTH3 is quite large (Figs. 2b,c). Though the NODA-DET intensity forecast for TRUTH2 is initially accurate (Figs. 2e,h), it becomes poor after about 30 h because of the large track error and concomitant difference in landfall time. Finally, the evolution of the TRUTH3 intensity (Figs. 2c,f) and structure (Fig. 5) are also different from the NODA-DET simulation. The NODA storm becomes organized much more quickly, and it exhibits
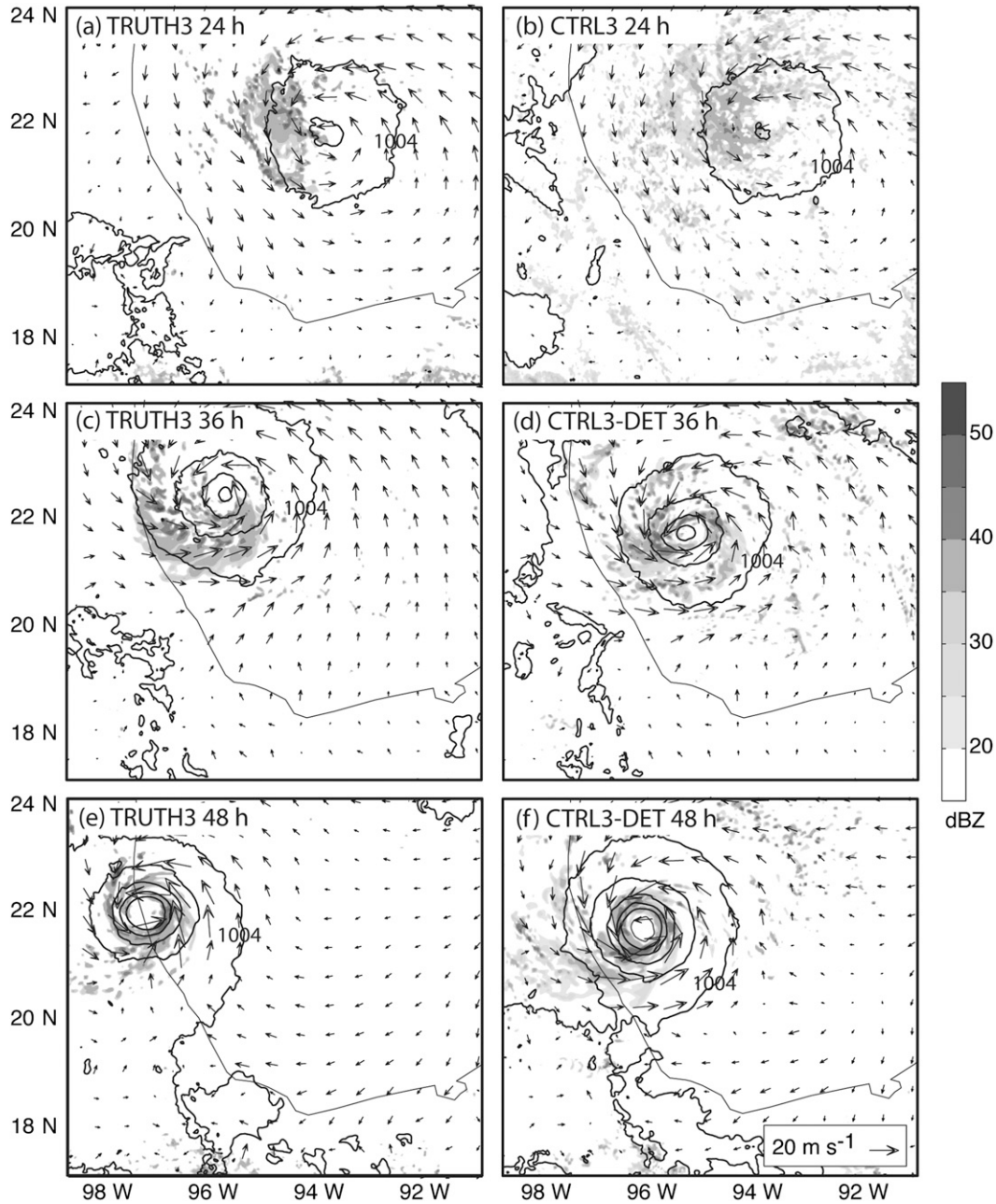
FIG. 5. As in Fig. 3, but for (left) TRUTH3 and (right) the CTRL3 analysis and CTRL3-DET forecast.

a consistently positive intensity bias that is equal to or larger than operational forecast error.

In addition to the aforementioned intentional error of the NODA-ENS tracks, there is also a varying degree of error associated with the ensemble intensity forecasts. Though NODA-DET exhibits a fairly good representation of TRUTH1, the NODA1-ENS intensity forecast for TRUTH1 is poor (Figs. 2d,g). This is particularly true after 30 h, when the truth storm traverses the edge of the ensemble intensity envelope. The TRUTH2 intensity is

initially nearer the center of the NODA2-ENS envelope (Figs. 2e,h), but the ensemble intensity forecast becomes poorer with time. Just before 24 h, the TRUTH2 storm intensifies rather quickly so that it is more intense than most NODA2-ENS members, but after 30 h it becomes weaker than most members because of its much earlier landfall. Conversely, the ensemble intensity forecast for TRUTH3 starts off poorly, but by 48 h the TRUTH3 storm lies roughly in the middle of the NODA3-ENS intensity envelope (Figs. 2f,i).
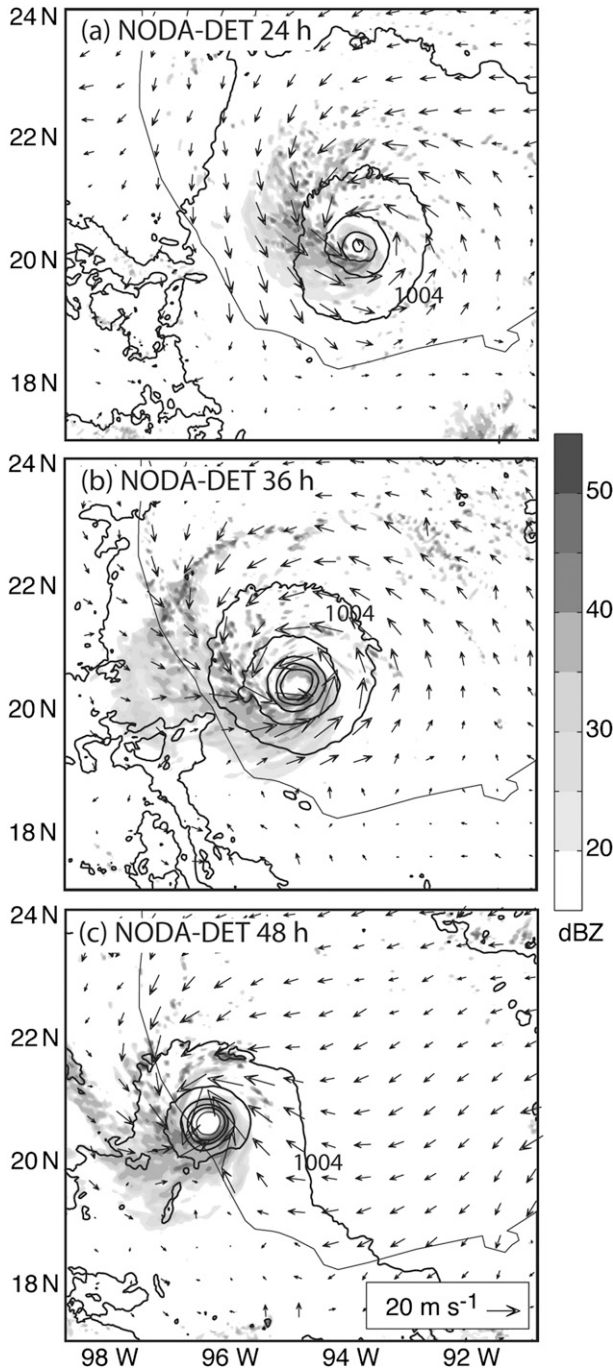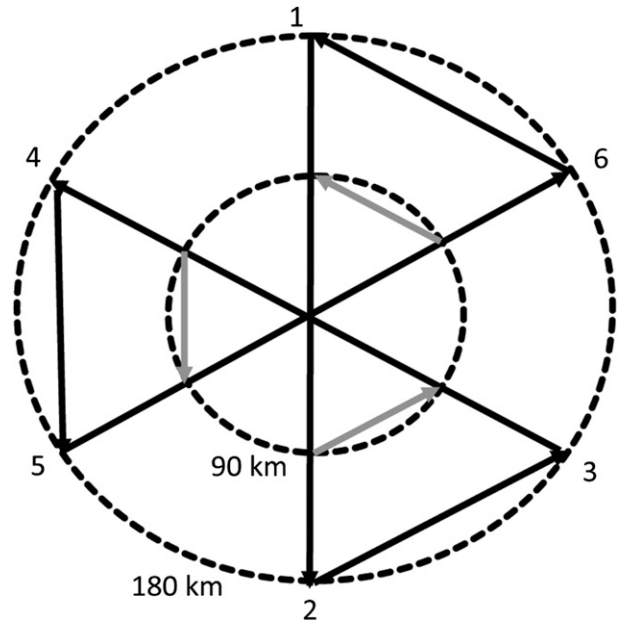
FIG. 6. As in Fig. 3, but for NODA-DET.

FIG. 7. A comparison of the flight pattern used in VLEG with those used in the CTRL (black lines only) experiments.

most of the precipitating region in the simulated storms. Figure 8a more clearly demonstrates this method by superposing a 1-h flight segment upon reflectivity at 1 km from the 19-h model output time in TRUTH1. Observations were taken in instantaneous conical scans from 12 to 24 h,[1] an on-station time that is slightly less than that for Karl (i.e., 14 h), but is a much longer period than is possible with operational aircraft. In addition, for each assimilation cycle, observations for an entire hour were gathered from a single time rather than over the 1-h period. This method bypasses the real-data assimilation problems associated with storm movement as observations are collected.

The simulated radar geometry for most experiments is similar to that found with HIWRAP, but some simplifications have been made in gathering the observations. We assumed a 19-km radar altitude and constant radar elevation angle of $-50°$ (Fig. 8b), which is consistent with the Global Hawk altitude and the HIWRAP Ku-band geometry (G. Heymsfeld 2011, personal communication). Though not demonstrated in Fig. 8b, the height of the simulated radar beam accounts for Earth's curvature and beam refraction using Eq. (3.12) of

b. Simulated observations

Simulated Vr observations for most experiments were gathered from the truth realizations along a butterfly pattern similar to that shown with the solid black lines in Fig. 7. The Global Hawk ground-relative speed of $170\,\mathrm{m\,s^{-1}}$ (330 kt) was assumed, and center-crossing transects were made sufficiently long (360 km) to observe

---

[1] This period for assimilation was chosen to overlap with the actual Global Hawk flight over Karl in order to facilitate comparison with forthcoming real-data experiments; however, assimilation here begins approximately 6 h earlier than in reality to allow for longer postassimilation forecasts before landfall.
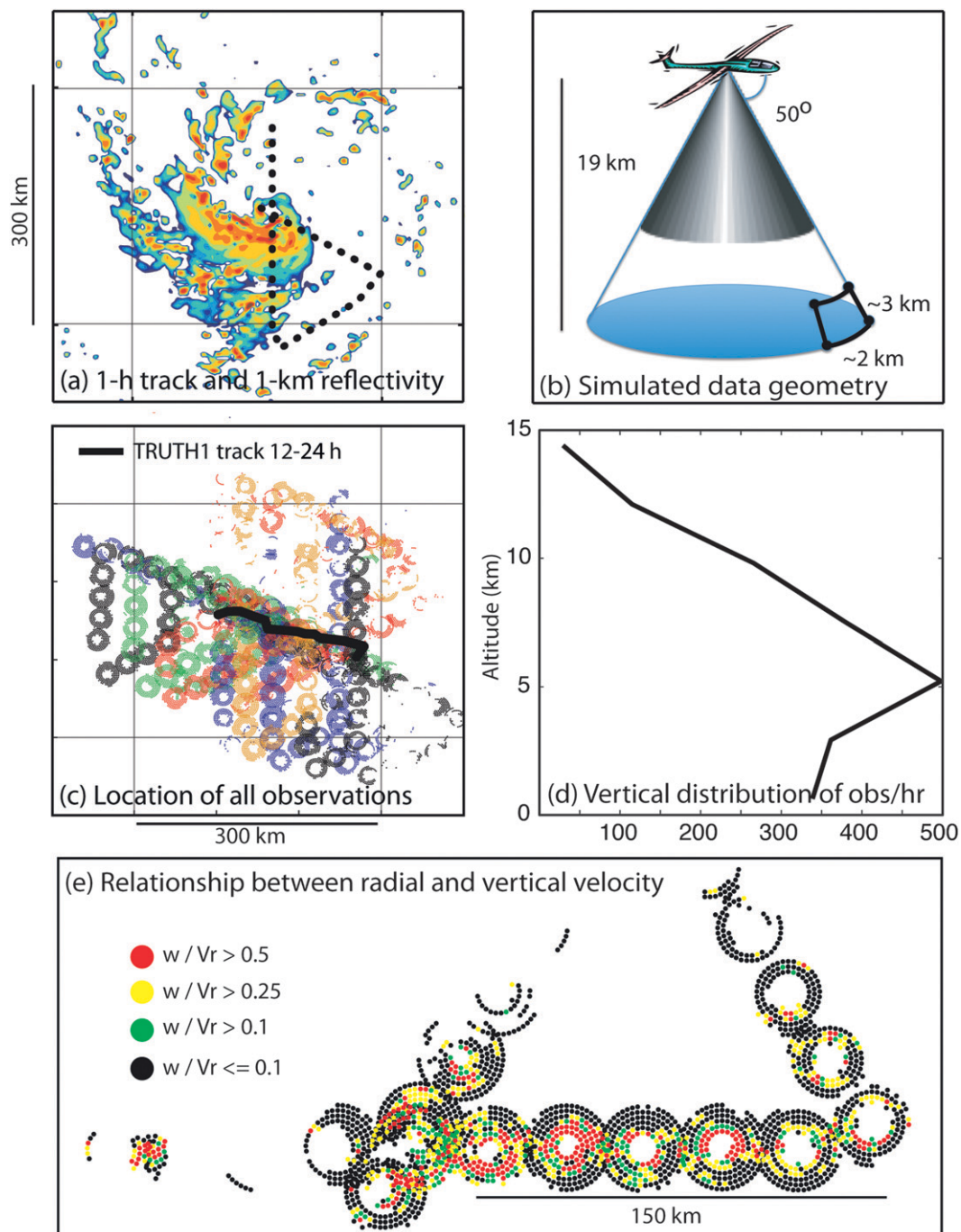
FIG. 8. Schematics showing the method for gathering data and data distribution. (a) A sample 1-h flight track at 19 h superposed upon 1-km reflectivity. (b) An illustration of the 3D distribution of a single simulated scan. (c) The track of the cyclone from 12 to 24 h superposed upon the location of all data points with data points color coded by the hour they were assimilated. (d) The average vertical distribution of observations. (e) A zoomed in view of the track in (a) showing individual radial velocity observations that are strongly (red), moderately (yellow), and weakly (green) affected by vertical velocity, as well as those that are generally unaffected (black). Length scales are provided for reference.

Reinhart (1997). Beamwidth was not accounted for. Instead, simulated radial velocities were calculated by linearly interpolating u, v, and w wind components in the truth simulations. As shown in Fig. 8b, Vr observations on the conical surface are spaced approximately 3 km radially and 2 km azimuthally. Observations were only gathered for points below 16-km altitude where attenuated reflectivity exceeded 10 dBZ, and they were stored along with the calculated range, elevation angle, azimuth, and radar location (i.e., the observation format is exactly as in Z11 and WZ12). This setup allows for roughly 1500–2000 observations per hour in conical scans that are spaced approximately 28 km apart (e.g., Figs. 8c,e). Though this number is orders of magnitude less than the number of raw observations available from HIWRAP, it is consistent with the number of observations assimilated in recent studies that have assimilated Vr (e.g., Z11 and WZ12). Finally, no consideration was given to particle fall speed, and a normally distributed, unbiased error with $3 \, m \, s^{-1}$ standard deviation was added to the simulated observations.

c. EnKF setup

This study uses the same EnKF setup as in Z09 and WZ12. The NODA ensembles provided the initial ensemble covariance and boundary conditions, and all experiments used a spinup time of 12 h before beginning assimilation. Forecast covariance for subsequent hourly analyses was obtained by integrating EnKF perturbations from the previous analysis forward. To control filter divergence resulting from sampling and model error, the covariance relaxation of Zhang et al. [2004, their Eq. (5)] was used to inflate the covariance at updated grid points via a weighted average of the prior and posterior perturbations. Here, except for when otherwise noted, all experiments use a weight α of 0.8, which means 80% of the prior perturbations were used.

Covariance was localized with the Z09 successive covariance localization (SCL) procedure, which uses the Gaspari and Cohn (1999) fifth-order correlation function with a varying radius of influence (ROI). In particular, the EnKF assimilated ⅑ of observations onto all domains with a horizontal ROI of 1215 km. Meanwhile, an additional ⅔ of observations were ingested into domains 2 and 3 with a horizontal ROI of 405 km, and on domain 3 the EnKF assimilated the remaining ⅚ of observations with a 135-km horizontal ROI. Furthermore, all domains utilized a 35-point vertical ROI (i.e., distance in terms of model vertical layers). The ROIs and percentage of observations assimilated into each domain are the same as in WZ12 and Z11 (Z09 used slightly different fractions but the same ROIs).

The EnKF observation operator is also the same as that utilized in the aforementioned studies. Using the previously calculated azimuth, elevation, and range data, innovations were computed for each Vr observation with the same linear interpolation and corrections for beam height that were used to create the observations. Note that the operator is thus perfect and not a source of error here.

Finally, using the above information and an assumption of $3 \, m \, s^{-1}$ observation error, the EnKF provided hourly updates to the three wind components (u, v, and w), temperature T, water vapor mixing ratio, cloud water mixing ratio, rainwater mixing ratio, perturbation geopotential, and perturbation column dry air mass. Our past experience has shown that updating ice-related variables does not improve the EnKF analysis, so they are not updated here. For more detailed information on the EnKF analysis procedure, please see Z09.

d. Description of experiments

In general, the experiments herein can be classified as those relating to the three control runs (CTRL1–3) and sensitivity experiments (see Table 1). In the CTRL experiments, Vr observations from the three truths were assimilated from 12 to 24 h (i.e., 13 cycles). A number of sensitivity experiments that are derived from TRUTH1 and discussed in section 4 assess the sensitivity to observation error (AOERR), choice of flight pattern leg length (VLEG), removal of observations with strong vertical velocity (FLTRW), covariance mixing coefficient (MIX05), and radar geometry (P3GEO). In addition, deterministic and ensemble forecasts corresponding to the CTRL, VLEG, and AOERR analyses will be used to assess forecast improvement gained from data assimilation.

3. CTRL analyses and forecasts

This section presents the results of assimilating all hourly Vr observations from 12 to 24 h in the three truth scenarios. In addition, ensemble and deterministic forecasts initialized from the EnKF analyses are also assessed in a subjective sense.

a. Analysis comparison with NODA forecasts

An important point to recognize from Fig. 9, which shows the 12–24-h evolution of intensity and track in the NODA and CTRL experiments, is that the NODA-DET forecast of the point intensity metrics during this period is at times very good. This is particularly true for TRUTH2 (Figs. 9e,h), where the difference between the NODA-DET and actual intensity is much less than typical best-track analysis error for a category 1 with
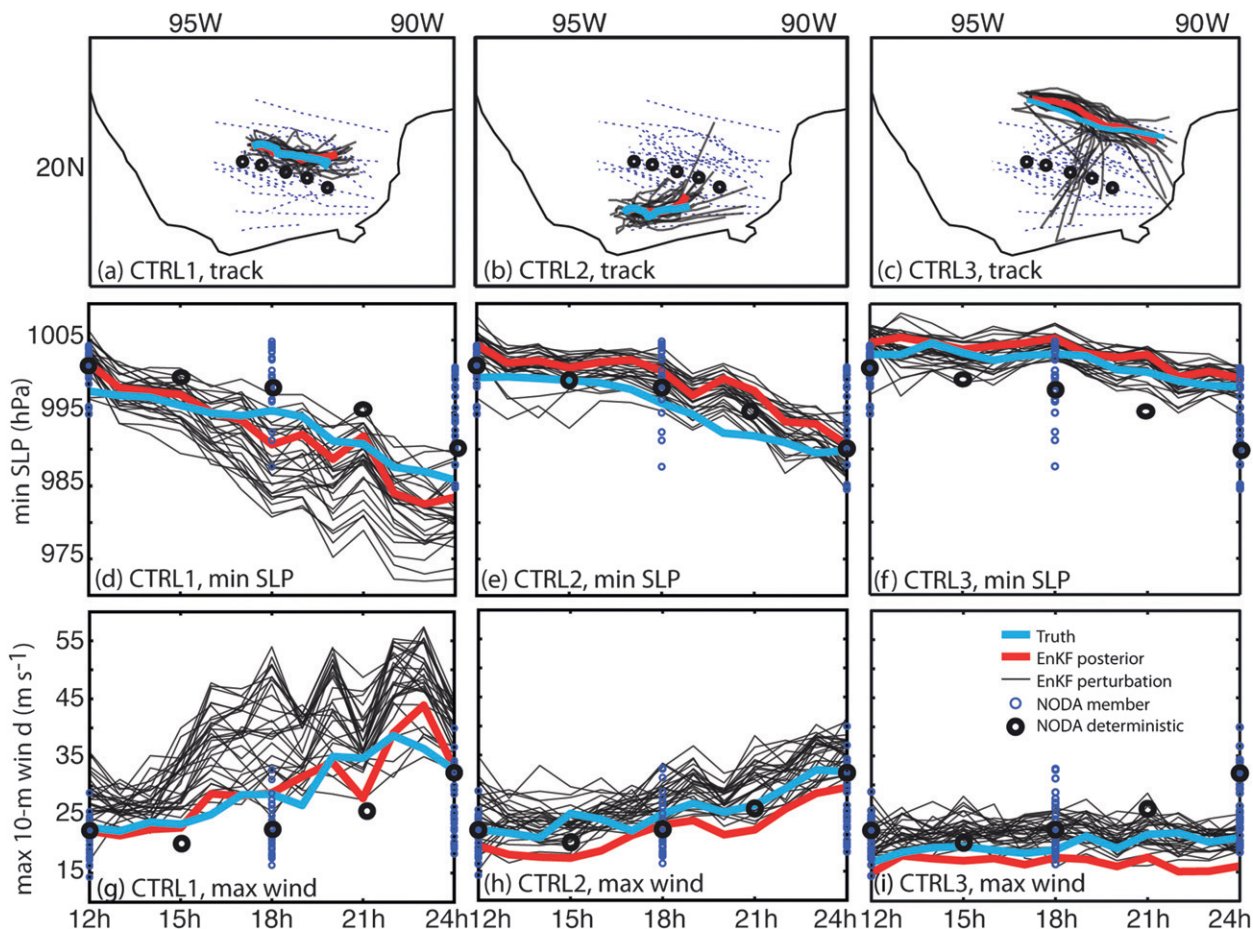
FIG. 9. Truth evolution, NODA deterministic forecasts, and EnKF analyses of (a)–(c) track, (d)–(f) minimum SLP (hPa), and (g)–(i) maximum 10-m wind (m s^{-1}) for the (left) TRUTH1, (middle) TRUTH2, and (right) TRUTH3 scenarios during the assimilation period. The thick cyan line denotes the truth, circles represent NODA forecasts from Fig. 2 (large black, deterministic; small blue, ensemble), the thick red line shows the EnKF posterior, and the thin black lines denote EnKF perturbations. The NODA-DET data are shown every 3 h.

aircraft reconnaissance [roughly 5 m s^{-1} for maximum winds and 4 hPa for minimum sea level pressure (SLP); Landsea and Franklin (2013)]. Meanwhile, compared to TRUTH1, NODA-DET is on average too weak by roughly 4 hPa in terms of minimum SLP (Fig. 9d) and 4 m s^{-1} in terms of maximum 10-m winds (Fig. 9g). This again is approximately commensurate with operational forecast error, though the NODA-DET wind error peaks around 9 m s^{-1} at 21 h. Finally, intensity forecast error for TRUTH3 is initially very low, though by the end of the assimilation period it is roughly double that of operational forecasts.

In general, Fig. 9 shows that the EnKF posterior mean analyses are better than the NODA-DET forecasts in terms of point intensity and location metrics. With roughly 2–3 m s^{-1} and 1–2 hPa average error in maximum wind and minimum SLP, respectively, CTRL1 (Figs. 9d,g) and CTRL3 (Figs. 9f,i) better represent the point intensity metrics than does NODA-DET. Meanwhile, the

CTRL1 analysis clearly corrects the south-of-track bias from NODA-DET (Fig. 9a), resulting in a track that is very close to TRUTH1. The correction for the CTRL2 and CTRL3 tracks (Figs. 9b,c) is even larger, with significant improvements to both position and trajectory. The only exception to the improvement in point metrics is that the CTRL2 intensity analysis is slightly worse than NODA-DET (Figs. 9e,h), but this inferior performance is not particularly surprising considering the very small NODA intensity errors for TRUTH2. Average analysis errors for CTRL2 are about 3.5 m s^{-1} for maximum wind and 3.5 hPa for minimum SLP, which is only slightly worse than in CTRL1 and CTRL3.

The EnKF posterior perturbations are shown in Fig. 9, though the perturbation distributions of minimum SLP and maximum wind are not necessarily expected to be centered upon the truth or posterior mean for Eulerian assimilation. If the location of the storm center were the same in all members, then the mean of the minima or
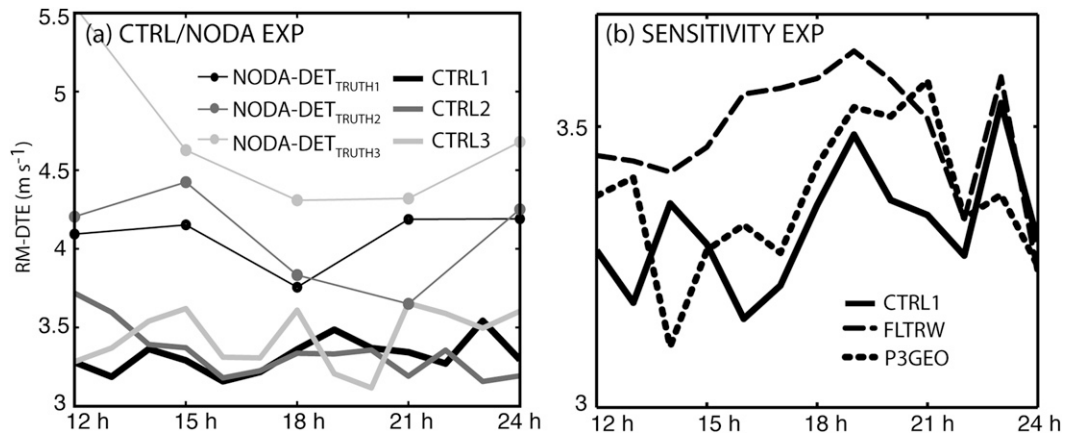
FIG. 10. Time evolution of area-average RM-DTE within 180 km of the truth center in (a) CTRL1–3 and the corresponding NODA-DET experiments and (b) CTRL1 and sensitivity experiments FLTRW and P3GEO.

maxima intensity metrics in each member would roughly equal the posterior mean value and be distributed around the truth (assuming the assimilation is performing well). As it is, there is always some degree of position spread (Figs. 9a–c) so that the wind and pressure fields near the center have very large ensemble variance. Since posterior perturbations add up to nearly zero, at any given point in the inner core it is likely that a few members will have wind speeds (surface pressure) greater (lower) than the maximum (minimum) in the posterior mean. At other near-center grid points, particularly for the noisier wind fields, it is possible or even likely that other members will have a higher intensity than the maximum in the posterior mean. When this is considered over a number grid points, it is inevitable that many members have maximum winds greater than the posterior mean, while somewhat fewer members should have a lower minimum pressure than the mean. This pattern is clearly seen in Fig. 9, and it is similar to that seen in Fig. 4 of Z09. This issue will be explored more in section 5.

While the point metrics of analysis performance are in some sense useful because of their operational utility, the root-mean of difference total energy (RM-DTE) gives a more complete picture of analysis error. In this commonly used metric (e.g., Mitchell et al. 2002; Zhang et al. 2006; Meng and Zhang 2007; Sippel and Zhang 2008, 2010), $DTE = 0.5(u'u' + v'v' + kT'T')$, where $k = C_p/T_r$, $C_p = 1004.9 \, \text{J kg}^{-1} \text{K}^{-1}$, and $T_r = 270 \, \text{K}$. Here, DTE is computed over the horizontal domain for each vertical level by assessing the differences (primes) between the NODA-DET forecast or various analyses and the truth, and the root mean is then taken over each vertical column. Differences in RM-DTE error are associated with different errors in location, structure, and intensity of the cyclone as well as errors in the background.

With this in mind, Figs. 10 and 11 illustrate net benefits of assimilation in terms of RM-DTE. In particular, Fig. 10 plots the time evolution of area-average RM-DTE within 180 km of the truth center (i.e., the same radius as the flight pattern) in the EnKF analyses and NODA-DET. To isolate RM-DTE associated with structure or intensity error from that caused by position error, the domain in NODA-DET was horizontally translated for this calculation so that the center matched the center in the truth experiments. Meanwhile, Fig. 11 shows the analysis improvement considering *both* structural and position error. In particular, Fig. 11 (left column) shows how well the analyses perform relative to NODA-DET by plotting the *difference* in RM-DTE between the analyses and the NODA forecast at 24 h, the end of the assimilation period. Finally, Fig. 11 (right column) shows the relative reduction of RM-DTE error between the EnKF analyses and the NODA-DET forecasts [i.e., $(\text{RM-DTE}_{\text{CTRL}} - \text{RM-DTE}_{\text{NODA-DET}})(\text{RM-DTE}_{\text{NODA-DTE}})^{-1}$].

The results from Figs. 10 and 11 demonstrate that Vr assimilation substantially reduces net RM-DTE error on multiple scales. On average, Fig. 10 shows that errors in storm structure or intensity in all analyses stay relatively constant with time, roughly between 3.25 and 3.75 m s^{-1}. In contrast, RM-DTE error for NODA-DET typically ranges from 3.75 to 4.5 m s^{-1}. Thus, in terms of storm structure within the radius of the butterfly flight pattern, the analyses reduce error by roughly 20%. Meanwhile, Fig. 11 gives a better sense of overall error reduction by the end of the assimilation period. All three analyses reduce RM-DTE error surrounding the cyclone centers, and peak error reduction occurs between the 24-h truth position (end of the solid line) and the NODA-DET forecast position (black or white ×). This is the region where forecast flow directly opposes truth flow. CTRL2
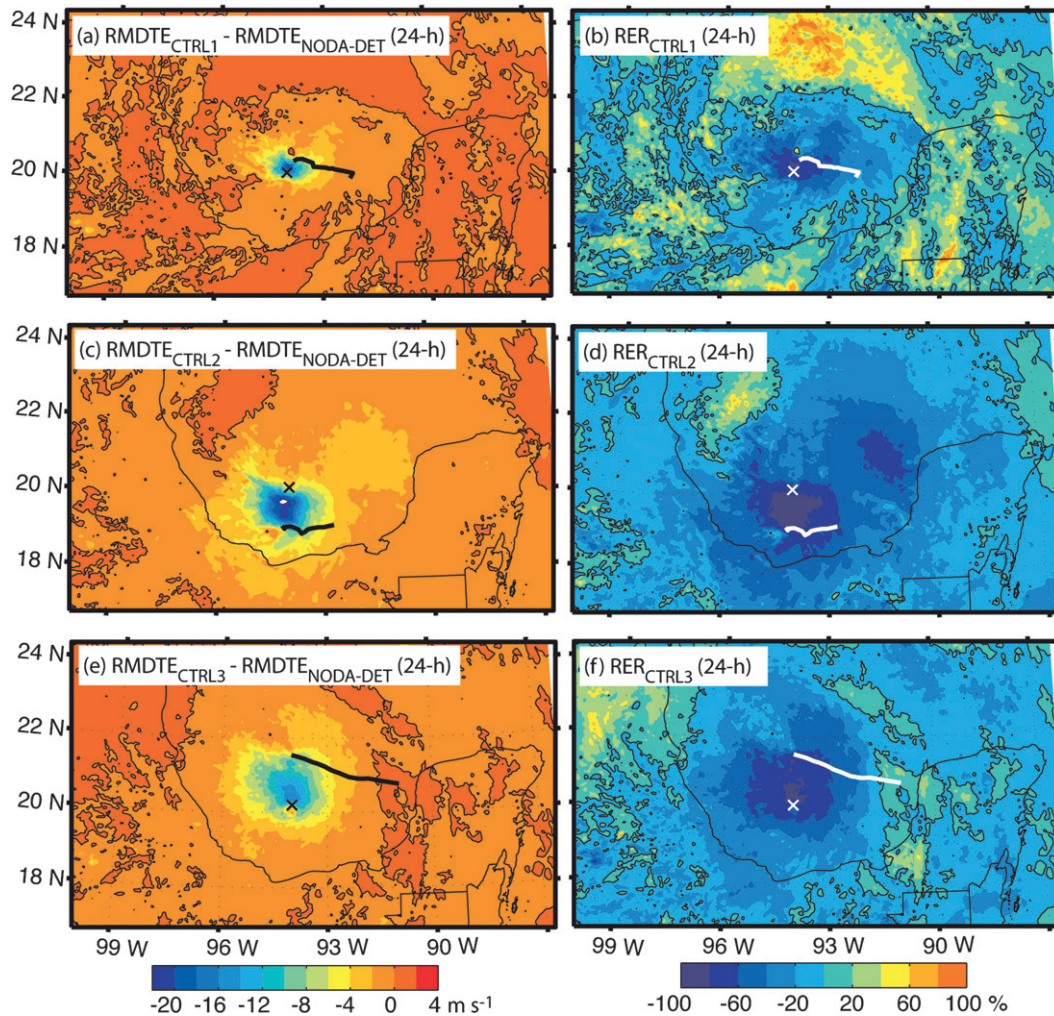
FIG. 11. (left) Difference RM-DTE and (right) relative error reduction between the deterministic NODA forecasts and EnKF analyses after 12 h of assimilation (i.e., at 24 h). The thick black or white lines show the TRUTH tracks during the 12-h assimilation period. A thin black line denotes the zero-difference line, and a black or white × denotes the 24-h forecast position from NODA-DET.

and CTRL3 have a greater area of very large error reduction, which can mostly be explained by the large NODA-DET position errors for TRUTH2 and TRUTH3. Peak error reduction is greatest in CTRL2 due to both the position error and the relatively strong wind field in TRUTH2. In addition, error reduction in CTRL2 and CTRL3 encompasses nearly all of domain 3, which suggests that EnKF assimilation of HIWRAP data is capable of correcting problems with the background analysis that extend well beyond the tropical cyclone.

Though there are some regions of large percent error increase in CTRL1 (Fig. 11b), this result must be interpreted carefully. Examination of Fig. 11a reveals that the change in RM-DTE in areas far from the TRUTH1 center is small ($<2\,\mathrm{m\,s^{-1}}$), and the large

percent error increase is associated with relatively small initial errors in NODA-DET (i.e., a small denominator). Furthermore, in an average sense there is very little change in RM-DTE farther than 250 km from the TRUTH1 center in Fig. 11a, despite the widespread positive values in percent error. While the ideal condition would be for error in these regions to decrease, we view this issue to be relatively minor considering the small error changes involved. Furthermore, assimilation of data external to the cyclone, which is beyond the scope of this study, would likely reduce such errors. It is also possible that a refined method for covariance localization would improve results at larger radii, though the SCL method has consistently proven successful in the past.
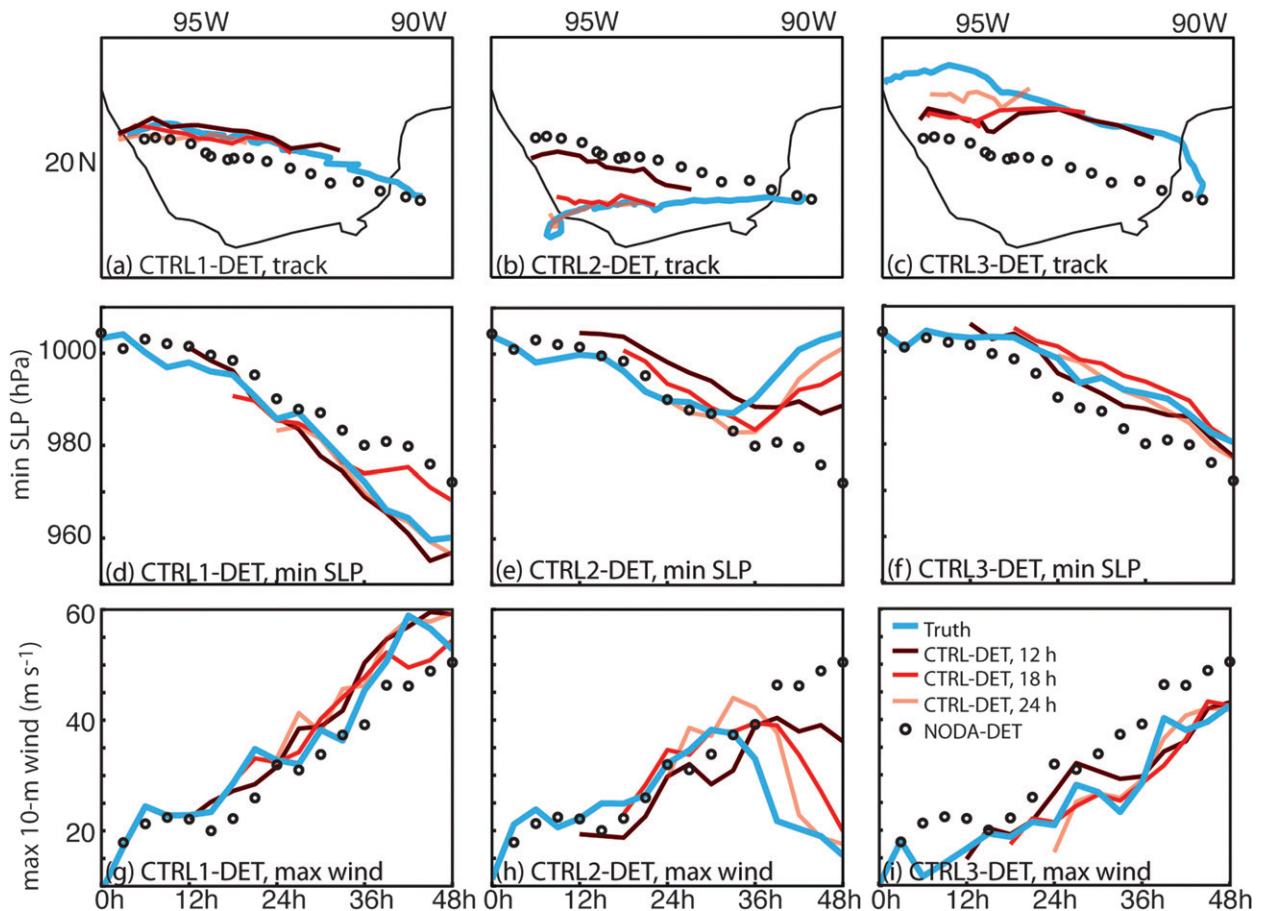
FIG. 12. A comparison of (a)–(c) track, (d)–(f) minimum SLP, and (g)–(i) maximum 10-m winds from the truth (cyan) and deterministic forecasts based on the 0000 UTC 16 Sep GFS (circles) and subsequent EnKF analyses (varying shades of red for analysis times are indicated in the legend) for the (left) TRUTH1, (middle) TRUTH2, and (right) TRUTH3 scenarios. The NODA-DET data are shown every 3 h.

b. Deterministic and ensemble forecasts initialized from analyses

Here we compare NODA forecast error for the truth realizations to that in deterministic and ensemble forecasts initialized from the EnKF analyses. The EnKF-initialized forecasts began at 12, 18, and 24 h (i.e., after 1, 7, and 13 cycles). For reasons that will become apparent in section 5, this subsection focuses on a qualitative assessment of the error.

It is clear that assimilation has a benefit for the CTRL1 forecasts despite the relatively small corrections to the intensity metrics (e.g., Figs. 9d,g) in the analysis. For example, CTRL1 intensity forecasts for TRUTH1 from all three start times perform better than NODA-DET (Figs. 12–13d,g). In fact, two of the three CTRL1-DET SLP forecasts are nearly perfect through 48 h, whereas the NODA-DET error in this period is 15–20 hPa. These same CTRL1-DET forecasts also better represent the

maximum wind speed achieved by the TRUTH1 storm ($\sim 60\,\mathrm{m\,s^{-1}}$), albeit with a 3–6-h error in timing. Though the CTRL1 intensity forecast ensemble means are not as good as the corresponding CTRL1 deterministic forecasts, they have roughly 50% of the error of the NODA1-ENS mean by 48 h, and their 95% confidence intervals better overlap the truth (Figs. 13d,g). In addition, the forecast precipitation structure from CTRL1-DET initialized at 24 h is more accurate than that from NODA-DET. For example, CTRL1-DET has a symmetric ring of precipitation around the cyclone center at 48 h (Fig. 3f), which is similar to that seen in TRUTH1 at the same time (Fig. 3e). Meanwhile, NODA-DET at 48 h forecasts a substantial gap in the northwestern eyewall (Fig. 6b), which is consistent with its somewhat weaker intensity. The CTRL1-DET tracks also improve upon the small south-of-track bias seen in NODA-DET (Fig. 12a), and they follow very close to TRUTH1. Though CTRL1-ENS forecasts do not improve upon the NODA1-ENS tracks
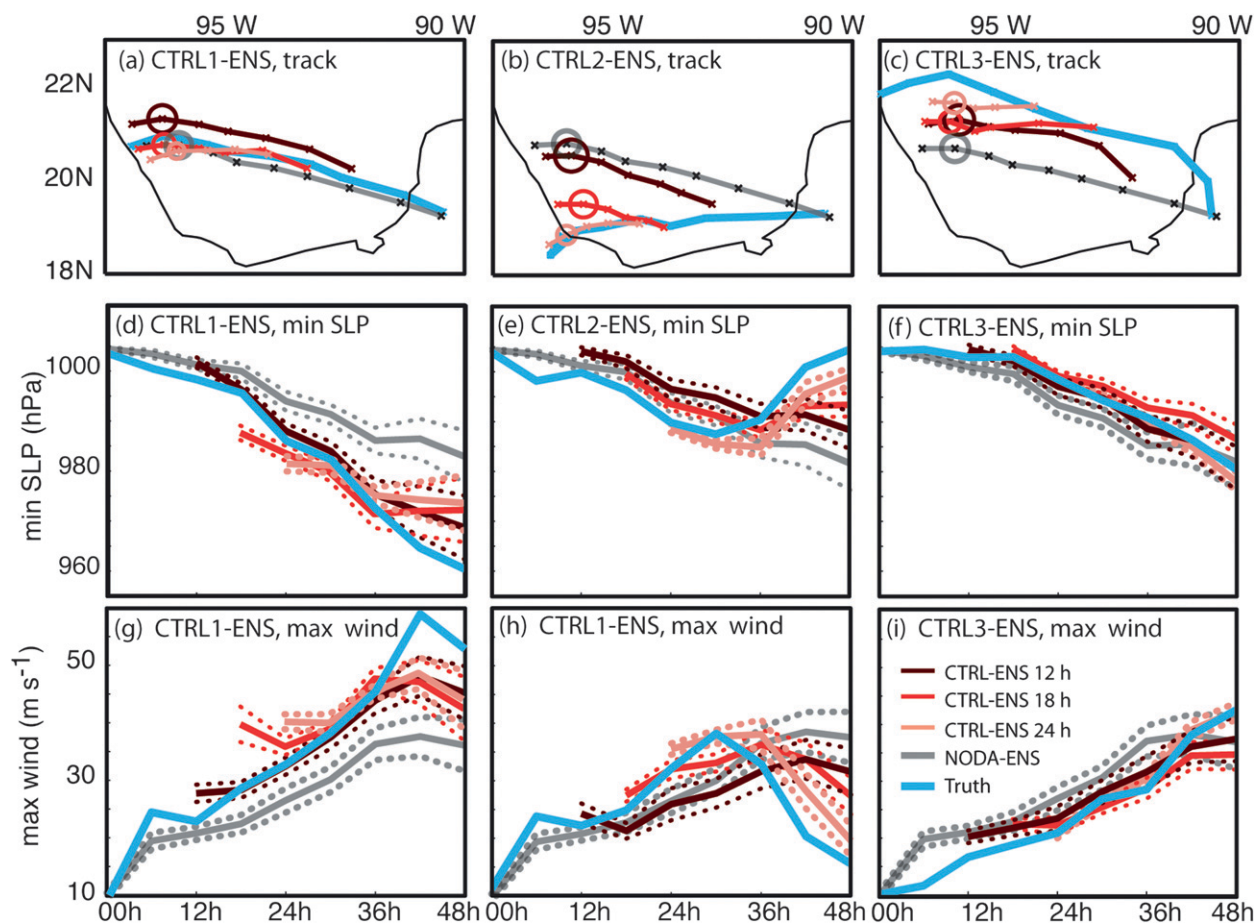
FIG. 13. A comparison of (a)–(c) track, (d)–(f) minimum SLP, and (g)–(i) maximum 10-m winds from the truth (cyan) with the NODA ensembles (gray) and ensemble forecasts initialized with subsequent EnKF analyses (varying shades of red for analysis times are indicated in legend) for the (left) TRUTH1, (middle) TRUTH2, and (right) TRUTH3 scenarios. Solid lines denote the ensemble mean. Circles in (a)–(c) indicate the 95% confidence intervals on the ensemble mean position at 42 h (spread in the zonal and meridional directions is very similar; 48-h position is not shown because of difficulties tracking storms in a number of members over high terrain after landfall), while dotted lines in (d)–(i) indicate 95% confidence intervals on the ensemble mean intensity.

(Fig. 13a), this result must be taken in context since 48-h track error from the NODA1-ENS mean is only about 45 km, which is much lower than current average operational error for the same lead time. On the other hand, the mean intensity error from the NODA1-ENS is roughly 3 times larger than average operational error. Thus, it is much easier for the CTRL1 ensembles to outperform NODA1-ENS in terms of intensity than it is in terms of track. These results demonstrate that even when forecast error is not particularly large, there is room for improvement using the EnKF.

Forecast improvements for TRUTH2 are larger than those for TRUTH1. As was previously mentioned, the NODA forecast intensity error for TRUTH2 is large after 30 h, mainly because of its large track error and early landfall. CTRL2 assimilation clearly improves the subsequent deterministic and ensemble track forecasts,

especially those initialized at 18 and 24 h (Figs. 12–13c). As a result, the intensity forecasts for TRUTH2 also improve markedly, and the latter two CTRL2-initialized forecasts capture well the truth evolution (Figs. 12–13e,h). The forecasts initialized at 24 h are particularly successful, and their cyclones move almost exactly along the TRUTH2 track. Finally, Figs. 4d,f reveals that the precipitation distribution from the CTRL2-DET forecast initialized at 24 h is also similar to the truth before and after landfall.

The results from CTRL2 demonstrate a case when a long assimilation period is especially desirable. The effects of the poor first guess are obvious in the first CTRL2-initialized forecasts, which are worse than the subsequent forecasts. In CTRL2-DET, the analyzed vortex is broad and disorganized because the analysis is not completely eliminating the incorrectly placed vortex

from the prior forecast (not shown). The correction to the steering flow is also insufficient, and the result is a storm that tracks similarly to that in NODA-DET but intensifies more slowly. The EnKF required multiple cycles to pull the CTRL2 analysis more closely to TRUTH2 and thus lead to a more accurate forecast. Figures 12 and 13 clearly show the large improvement in the forecast initialized at 18 h (7 cycles) and some continued improvement in the forecast initialized at 24 h (13 cycles).

Forecasts initialized from CTRL3 also tend to improve with time, though the tendency is not as strong as with the CTRL2 forecasts. Deterministic and ensemble track forecasts initialized from all three start times are better than those of NODA (Figs. 12–13c), and the forecasts from 24 h are arguably the best. It is notable that the final track forecasts from CTRL3 are poorer than those from CTRL2, and sensitivity tests (not shown) reveal that this is as a result of a failure of the EnKF to adequately remove analysis error of the background (i.e., steering) flow. Recall, however, that the CTRL3 analysis is still more accurate than NODA-DET over most of domain 3 (Fig. 9e), and assimilation of data from outside the storm could further correct the background. Despite the difficulties with track, the CTRL3 intensity forecasts are all very good. All three CTRL3-DET forecasts follow very close to the TRUTH3 minimum SLP and maximum winds and accurately depict a weaker storm than in NODA-DET. The CTRL3-ENS intensity forecasts are also generally better than NODA3-ENS, and the best forecast is that initialized at 24 h. The means of both maximum winds and minimum SLP from the final CTRL3-ENS forecasts are nearly identical to the TRUTH3 intensity metrics, and they also have quite narrow confidence intervals (Figs. 13f,i). Thus, not only is this forecast highly accurate, but it also comes with high certainty. Finally, the precipitation distribution from the CTRL3-DET forecast in Fig. 5 also improves upon that from NODA-DET in Fig. 6, particularly by 48 h.

4. Sensitivity experiments

This section examines the sensitivity of the results to random observation error, variations in flight track and radar geometry, removing observations with a large w component, and the covariance mixing coefficient. These experiments give important insights into the CTRL1 behavior and into how changes to observation and analysis strategies can impact results.

a. Alternate observation error experiments

To understand the intrinsic variability of the above forecasts, the alternate observation error (AOERR)

experiments in Fig. 14 examine the forecast sensitivity to random observation error. A total of 10 different analyses were created at 12, 18, and 24 h by using the CTRL1 prior forecasts and assimilating TRUTH1 observations with different[2] random error vectors. The error vector was only perturbed for the final analysis for each set of forecasts because of computational time and storage constraints (i.e., for each forecast time this method only requires 10 additional analyses without cycling). The observations were assimilated in exactly the same order in all experiments so that each experiment used the same set of covariance ROIs for any given observation. From the resultant analyses and perturbations, sets of 10 deterministic and three ensemble[3] forecasts were performed for each time.

Perhaps the most important result from the deterministic AOERR experiments in Figs. 14a–f is that large forecast spread can develop as a result of only random error in observations. This spread is particularly evident for forecasts from the 12- and 18-h analysis times, wherein the SLP envelope at 48 h is at least 20 hPa (Figs. 14a,b), and the maximum 10-m wind envelope is about 15 m s^{-1} (Figs. 14d,e). The spread decreases for the set of forecasts initialized at 24 h (Figs. 14c,f), and it demonstrates why quantitative assessment of error in the previous section is inappropriate. For example, upon assessment of only the CTRL1-DET intensity forecasts initialized at 12 and 18 h, one might get the impression that 6 h of cycling has done no good (i.e., the CTRL1-DET forecast from 18 h is worse than that from 12 h). However, it turns out that the CTRL1-DET forecast from 12 h is anomalously good compared to the other AOERR forecasts, whereas the CTRL1-DET forecast initialized at 18 h is anomalously bad. The large sensitivity to very small analysis differences is very similar to the results of Zhang and Sippel (2009) and Sippel and Zhang (2010), who noted that initial condition error much smaller than that seen in current operational analyses can quickly grow in the presence of moist convection.

Another important result is that on average the AOERR intensity forecasts perform better with a longer assimilation period. This improvement is especially evident in the AOERR minimum SLP forecasts, where the envelope of forecasts from 12 h barely encompasses TRUTH1, but those initialized at 24 h are clustered very close to the truth. The evolution of the mean in the AOERR ensemble forecasts (Figs. 14g–l) also suggests

---

[2] The error vectors carry the same mean and standard deviation, but individual elements within the vectors are different.

[3] Only 3 of the 10 sets were used to initialize the ensembles because of storage and computational limitations.
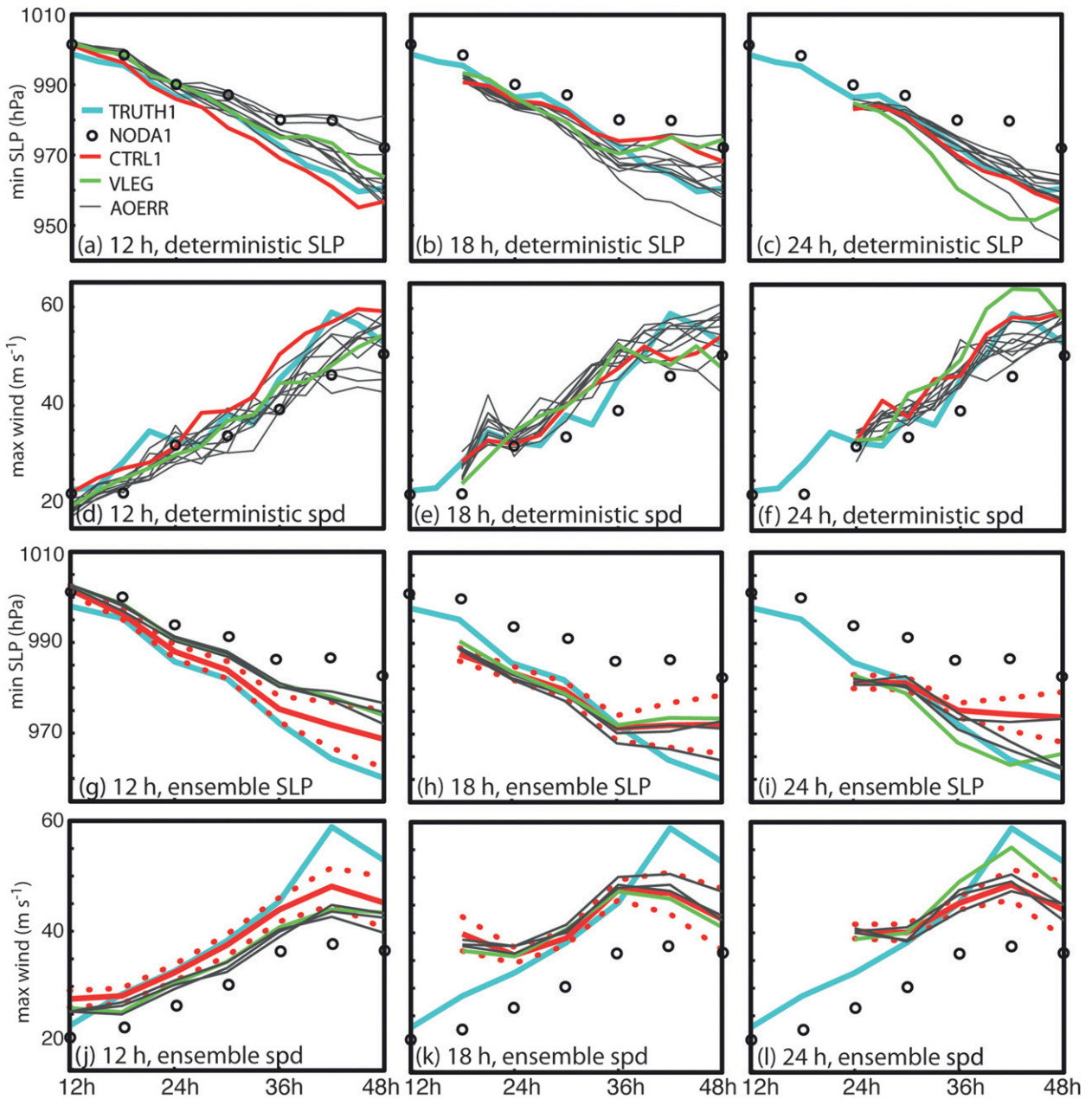
FIG. 14. A comparison of sensitivity (a)–(f) deterministic and (g)–(l) ensemble intensity forecasts with NODA1 (circles), TRUTH1 (cyan), and CTRL1 (red). Thin black and light green lines respectively indicate (a)–(f) deterministic and (g)–(l) ensemble-mean forecasts obtained from the AOERR and VLEG analyses. Dotted red lines indicate the 95% confidence interval on the ensemble mean.

a benefit from assimilation cycling. The ensemble-mean intensity in all three AOERR ensembles initialized at 12 h is less (and more erroneous) than that in CTRL1-ENS. This result again indicates that the relatively good CTRL1 forecasts from that time were somewhat by chance. However, by 24 h, the ensemble-mean minimum SLP from two of the three AOERR experiments follows TRUTH1 more closely than does the mean of CTRL1-ENS.

b. Variable-length leg experiments

An experiment (VLEG) was conducted to investigate whether focusing observations on the inner core at the expense of the outer circulation would have a practical benefit on analyses and forecasts. To do this, we ran an alternate flight pattern through TRUTH1 wherein each complete butterfly pattern with 360-km legs was followed by two complete patterns with 180-km legs (Fig. 7).

Taking more frequent observations in and just outside of the inner core does not provide an obvious net benefit in the VLEG analysis. While time-average RM-DTE in VLEG is lower than in CTRL1 within about 150 km of the track (Fig. 15a), it is higher outside that radius, presumably due to less frequent sampling of the outer regions of the storm. The evolution of maximum 10-m winds is somewhat less noisy in the VLEG analysis and perturbations (Fig. 15b) than in CTRL1 (Fig. 9g), but the VLEG analysis of winds also exhibits somewhat of a weak bias compared to CTRL1.

Likewise, ensemble and deterministic forecasts initialized from the VLEG analysis are not meaningfully different from the CTRL1 forecasts. This is most easily seen by examining the VLEG forecasts in the context of the AOERR forecasts (Fig. 14). Considering the decrease in analysis RM-DTE near the storm center (Fig. 15a), this lack of improvement in the VLEG forecasts might be a result of the degraded analysis of the outer circulation. Though this topic certainly warrants further investigation for operational flight planning, such an analysis is well beyond the scope of this study.

c. Experiment with observations of strong w removed

As was mentioned in the introduction, one potential concern for assimilating Vr with the HIWRAP geometry is the relatively large constant elevation angle, which might result in vertical velocity strongly influencing the analysis. This situation could lead to significant analysis errors due to the short length scale for w variations and concomitant noisy or weak covariance between w and other model state variables (e.g., Poterjoy and Zhang 2011). To show the extent to which w impacts observed Vr, Fig. 8e shows data points from the 19-h output file of TRUTH1 wherein the ratio between w and Vr exceeds 50%, 25%, and 10%. For each of these thresholds, the percentage of Vr observations impacted is 36%, 19%, and 11%, respectively.

We explored this potential problem by running an assimilation experiment wherein observations were rejected when the above ratio exceeded 25%. Curiously, the results in Figs. 10 and 15c,d suggest that observations with a strong w component are actually *beneficial* to the CTRL1 analysis. Area-average RM-DTE near the center in FLTRW is consistently higher than in CTRL1 (Fig. 10), and time-average FLTRW RM-DTE along the track is also higher (Fig. 15c). These results suggest that HIWRAP's geometry may not be a hindrance for accurate EnKF analyses, though it is possible that using a different model or physics configuration would give a slightly different result.

d. Experiment with lower covariance mixing coefficient

Here we examine the sensitivity of the EnKF posterior perturbations of minimum SLP and maximum winds to the choice of the mixing coefficient for covariance inflation. As was noted earlier, the point intensity metrics are not distributed around the EnKF posterior mean in Fig. 9. This result is related to analysis position spread, which is affected by the choice of mixing coefficient. We found that changing α to 0.5 in experiment MIX05, so that analysis covariance uses only 50% of the prior perturbations instead of 80%, significantly modifies the point intensity distributions relative to the truth and EnKF mean. For example, Fig. 15f shows that the EnKF posterior mean in MIX05 is much closer to the center of the perturbation distribution. While MIX05 does provide a slightly better fit to maximum 10-m winds than does CTRL1 (Figs. 9g and 15f), it is not uniformly better in terms of time-average RM-DTE (Fig. 15e). Finally, limited testing suggests that MIX05 also does not produce better forecasts (not shown).

e. Experiment with P3-like radar geometry

A final sensitivity experiment, P3GEO, uses alternate radar geometry similar to that of the NOAA WP-3D tail Doppler radar. For this experiment, we again employ instantaneous radar scans to gather observations along conical surfaces that look forward from the aircraft position at 3-km altitude. The angle between the aircraft bearing and the conical surfaces is 70°, which is consistent with the WP-3D radar geometry (see Fig. 2 of WZ12). This experiment is conducted only for the TRUTH1 simulation, and for the sake of simplicity the center latitude and longitude of all radar scans is exactly the same as in TRUTH1. That is, we assume the same ground-relative plane speed and distance between scans so that this experiment considers only the impact of radar geometry and altitude. The number of observations gathered per hour is roughly the same as in CTRL1. All other aspects of P3GEO are the same as in CTRL1.

The results of P3GEO, which are shown in Figs. 10 and 15, suggest that the differences in radar geometry between the NOAA WP-3D tail Doppler radar and HIWRAP may not play a critical role in determining analysis quality. Errors in maximum 10-m winds (Fig. 15h) are commensurate with those in CTRL1 (Fig. 9g), and average RM-DTE around the storm center is similar in both experiments (Fig. 10b). The distribution of difference RM-DTE (Fig. 15g) shows that CTRL1 generally performs better to the north of the center, and
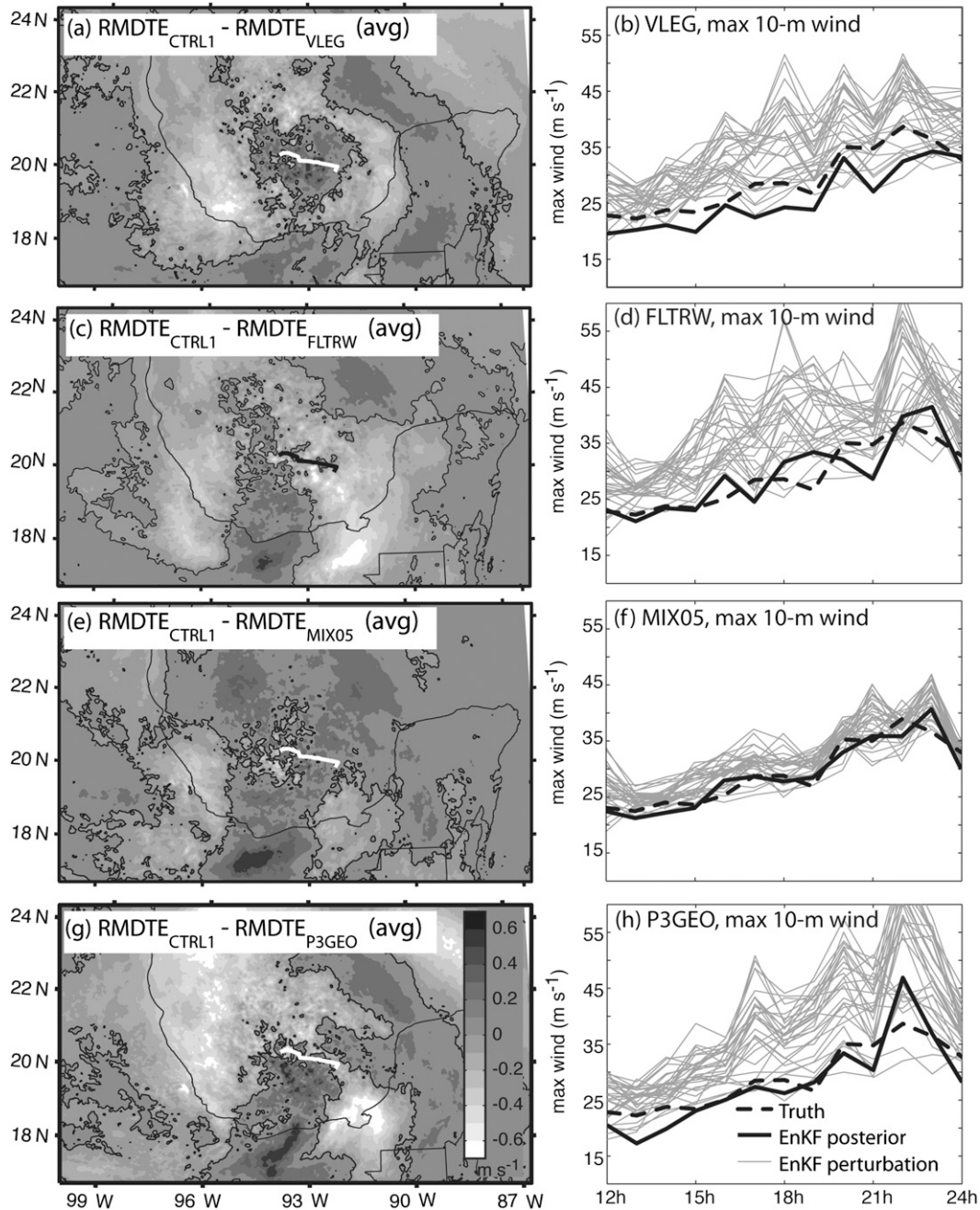
FIG. 15. (left) Time-average difference RM-DTE and (right) evolution of maximum-wind in EnKF analyses for (a),(b) VLEG; (c),(d) FLTRW; (e),(f) MIX05; and (g),(h) P3GEO. RM-DTE differences are calculated between the given analysis and CTRL1 and averaged over the entire analysis period. A thin black line denotes the zero-difference line in (a),(c),(e),(g). The thick white lines in (a),(e),(g) and black line in (c) show the TRUTH1 track during the 12-h assimilation period. For (b),(d),(f),(h), a dashed line represents TRUTH1, a solid line indicates the EnKF posterior, and thin gray lines show the EnKF perturbations.

P3GEO is better to the south, though neither experiment has an obvious advantage. Finally, limited testing (not shown) reveals similar forecast errors in P3GEO to those in CTRL1.

5. Discussion and conclusions

This study has explored the use of an EnKF to assimilate radial velocity observations of a hurricane from

a new instrument and platform. Using simulated data experiments, the analyses ingest hypothetical observations from the HIWRAP Doppler radar, which flies on board NASA's Global Hawk unmanned aerial system. The experiments focus on Hurricane Karl, which rapidly intensified to a major hurricane over the Bay of Campeche in 2010.

The study is designed to assess how the EnKF performs for several different situations in which the first-guess error covariance is either relatively accurate or poor. To accomplish this goal, three different realizations of the truth were selected to generate and assimilate observations into essentially the same initial ensemble (i.e., the covariance was the same but the truth changed). Forecasts without assimilation (NODA) represent the intensity and track of the realizations with varying degrees of accuracy, but at least one aspect of forecast error (i.e., either track or intensity) significantly exceeds operational error in each case.

The most important result from this research is that HIWRAP data can be used for hurricane analyses and prediction. Assimilation carries out over a 12-h period, and by the end of that period the analysis error is much lower than in the deterministic NODA forecasts. The improvement comes from corrections to storm location, structure, and maximum intensity that vary between the realizations. As a result of the improved analyses, a majority of deterministic and ensemble mean forecast error is eliminated, which is in general agreement with recent research that has shown the substantial benefit of assimilating radial velocity data for hurricane analysis and forecasting, both in OSSEs (Aksoy et al. 2012) and real-data (e.g., Z09; Z11; WZ12) experiments.

While the analyses highlight the potential of using an EnKF with HIWRAP data, the differences between the results are also insightful. For example, forecasts initialized from a situation where the first-guess error covariance is relatively accurate (i.e., CTRL1) demonstrate an immediately clear benefit of assimilation. Alternative situations with relatively poor initial error covariance (i.e., CTRL2–3) need more analysis cycles to improve ensemble and deterministic EnKF-initialized forecasts. Thus, when the first guess is bad, the benefit of the Global Hawk's potentially long on-station times becomes more evident.

Even when the first guess is good, the results here suggest that a longer assimilation period is still beneficial. This fact is made clear from forecasts initialized with the AOERR analyses, wherein the error vector of observations assimilated into CTRL1 was perturbed for the final cycle prior to the forecasts. This method effectively made sets of 10 slightly different EnKF analyses for each forecast initialization time. The sets of AOERR intensity forecasts initialized after 13 cycles more closely follow the truth than the forecasts initialized at the earlier times, which demonstrates the benefit of a longer assimilation period.

The AOERR simulations also give insight into the potential for large sensitivity to small initial condition differences. Though the only difference between the deterministic AOERR simulations is in the observation error vector applied before assimilation, storms differ by up to 20 hPa in SLP and $15\,\mathrm{m\,s^{-1}}$ in maximum 10-m wind in as few as 30 h. This result is similar to Zhang and Sippel (2009), who showed that tiny initial analysis errors in tropical cyclones can rapidly amplify in the presence of moist convection.

Other sensitivity experiments give more practical insight into assimilation of HIWRAP data. For example, focusing the flight pattern more on the inner core and less on the outer circulation of the TRUTH1 storm is not beneficial to analyses or forecasts. Though assimilating observations from such a pattern can somewhat reduce analysis error within and just outside the inner-core region, it comes at the expense of analysis quality at larger radii. Subsequent forecasts also do not improve, likely as a result of the degraded outer circulation. Similarly, a simple experiment showed that analyses accuracy may not differ when using a scanning geometry similar to that of the NOAA WP-3D tail Doppler radar. Though further analysis over multiple cases is needed to quantify the general differences as a result of flight track and scanning geometry, results such as this can be very useful for future operational or research flight planning.

While these results are promising, the extent to which assimilating HIWRAP Vr can improve real-time forecasts remains unclear. First, it is important to note that these results are subject to a very small sample of realizations in an OSSE context with a simplified scanning strategy. Analysis errors can vary from storm to storm, though we have attempted to account for that in the experiment design. In general, this OSSE has limited model error, so analysis error with real Vr data would likely be larger than that obtained here. Storm movement through the hourly assimilation cycles and an imperfect observation operator are other sources of error in the real-data case that have not been considered here. Another caveat is that assimilation is limited to simulated Vr data from within the cyclones, and data that are typically available external to the storm has been ignored. We acknowledge that many different data types (e.g., environmental dropsondes, space-based remote sensing, and surface in situ data from ships and buoys) contribute to skillful forecasts, and assimilation of such data here would likely have

reduced analysis and forecast error relative to the NODA experiments regardless of Vr availability. It is also possible for vortex relocation techniques, which have not been used herein, to improve forecasts even when data assimilation is not utilized.

Given the above caveats, we have begun work to more completely assess the benefit of assimilating HIWRAP Vr data as well as other datasets based from the Global Hawks. An initial goal is to use more extensive OSSE and real-data studies to compare the assimilation of Vr from HIWRAP and from the NOAA WP-3D tail Doppler radar. We also wish to assess real-data assimilation of HIWRAP Vr from multiple storms as the data become available; only a thorough examination such as that in Z11 will more precisely demonstrate the general utility of assimilating this data. Finally, a thorough examination of the usefulness of the Global Hawk UAS for forecasting should include assimilation of other data such as radiances from the Scanning High-resolution Interferometer Sounder (S-HIS; Revercomb et al. 2003), dropsondes profiles from the Advanced Vertical Atmospheric Profiling System (AVAPS; Hock and Franklin 1999), wind retrievals from the Tropospheric Wind Lidar Technology Experiment (TWILITE; Gentry et al. 2007), surface wind speed retrievals from the Hurricane Imaging Radiometer (HIRAD; Amarin et al. 2010), and brightness temperatures from the High-Altitude Monolithic Microwave Integrated Circuit (MMIC) Sounding Radiometer (HAMSR; Brown et al. 2007). These instruments will all fly on the Global Hawks as part of the NASA Hurricane and Severe Storm Sentinel (HS3) investigation being conducted during the Atlantic hurricane seasons of 2012–14.

REFERENCES

Aksoy, A., 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 airborne Doppler radar observations in an OSSE. *Mon. Wea. Rev.,* **140,** 1843–1862.

Amarin, R. A., W. L. Jones, J. W. Johnson, C. Ruf, T. L. Miller, and E. Uhlhorn, 2010: The Hurricane Imaging Radiometer wide swath simulation and wind speed retrievals. Preprints, *Int. Geosciences and Remote Sensing Symp.,* Orlando, FL, Institute of Electrical and Electronics Engineers, 4290–4293.

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.

Braun, S. A., and Coauthors, 2013: NASA's Genesis and Rapid Intensification Processes (GRIP) field experiment. *Bull. Amer. Meteor. Soc.,* **94,** 345–363.

Brown, S., B. Lambrigtsen, A. Tanner, J. Oswald, D. Dawson, and R. Denning, 2007: Observations of tropical cyclones with a 60, 118 and 183 GHz microwave sounder. Preprints, *Int. Geoscience and Remote Sensing Symp.,* Barcelona, Spain, Institute of Electrical and Electronics Engineers, 3317–3320.

Cangialosi, J. P., and J. L. Franklin, 2011: 2011 National Hurricane Center forecast verification report. NHC Rep., 76 pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2011.pdf.]

Dong, J., and M. Xue, 2013: Assimilation of radial velocity and reflectivity data from coastal WSR-88D radars using an ensemble Kalman filter for the analysis and forecast of landfalling Hurricane Ike (2008). *Quart. J. Roy. Meteor. Soc.,* **139** (671), 467–487.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.,* **46,** 3077–3107.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723–757.

Gentry, B., and Coauthors, 2007: Development of an airborne molecular direct detection Doppler lidar for tropospheric wind profiling. *Proc. Lidar Remote Sensing for Environmental Monitoring,* San Diego, CA, International Society for Optics and Photonics, U63–U68.

Heymsfield, G., J. Carswell, L. Li, D. Schaubert, J. Creticos, M. Vega, and W. Welch, 2008: Status of the High-Altitude Imaging Wind and Rain Airborne Profiler (HIWRAP). *Proc. NASA Earth Science Technology Conf.,* College Park, MD, National Aeronautics and Space Administration, 5 pp. [Available online at http://esto.nasa.gov/conferences/estc2008/papers/Heymsfield_Gerald_B5P2.pdf.]

Hock, T., and J. L. Franklin, 1999: The NCAR GPS dropwindsonde. *Bull. Amer. Meteor. Soc.,* **80,** 407–420.

Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice-microphysical processes for the bulk parameterization of cloud and precipitation. *Mon. Wea. Rev.,* **132,** 103–120.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.,* **47,** 2784–2802.

——, and ——, 1993: Convective parameterization for mesoscale models: The Kain-Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.,* No. 24, Amer. Meteor. Soc., 165–170.

Landsea, C. W., and J. L. Franklin, 2013: How "good" are the best tracks? Estimating uncertainty in the Atlantic hurricane database. *Mon. Wea. Rev.,* in press.

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.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 2791–2808.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102** (D14), 16 663–16 682.

Noh, Y., W.-G. Cheon, and S.-Y. Hong, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.,* **107,** 401–427.

Poterjoy, J., and F. Zhang, 2011: Dynamics and structures of forecast error covariance in the core of a developing hurricane. *J. Atmos. Sci.,* **68,** 1586–1606.

Reinhart, R. E., 1997: *Radar for Meteorologists.* Rinehart Publications, 428 pp.

Revercomb, H. E., and Coauthors, 2003: Applications of high spectral resolution FTIR observations demonstrated by radiometrically accurate ground-based AERI and Scanning HIS aircraft instruments. *Proc. Multispectral and Hyperspectral Remote Sensing Instruments and Applications,* Vol. 4897, Hangzhou, China, International Society for Optics and Photonics, 11–23.

Sippel, J. A., and F. Zhang, 2008: A probabilistic analysis of the dynamics and predictability of tropical cyclogenesis. *J. Atmos. Sci.,* **65,** 3440–3459.

——, and ——, 2010: Factors affecting the predictability of Hurricane Humberto (2007). *J. Atmos. Sci.,* **67,** 1759–1778.

Tao, W.-K., and J. Simpson, 1993: The Goddard cumulus ensemble model. Part I: Model description. *Terr. Atmos. Oceanic Sci.,* **4,** 35–72.

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.

Zhang, F., and J. A. Sippel, 2009: Effects of moist convection on hurricane predictability. *J. Atmos. Sci.,* **66,** 1944–1961.

——, 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.

——, Z. Meng, and A. Aaksoy, 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.

——, Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Convection-permitting hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter: Humberto (2007). *Mon. Wea. Rev.,* **137,** 2105–2125.

——, ——, 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.

Zhao, Q., and Y. Jin, 2008: High-resolution radar data assimilation for Hurricane Isabel (2003) at landfall. *Bull. Amer. Meteor. Soc.,* **89,** 1355–1372.