Evaluation of the Advanced Hurricane WRF Data Assimilation System for the 2009 Atlantic Hurricane Season

STEVEN M. CAVALLO
National Center for Atmospheric Research,* Boulder, Colorado

RYAN D. TORN
University at Albany, State University of New York, Albany, New York

CHRIS SNYDER, CHRISTOPHER DAVIS, WEI WANG, AND JAMES DONE
National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 14 May 2012, in final form 24 September 2012)

ABSTRACT

Real-time analyses and forecasts using an ensemble Kalman filter (EnKF) and the Advanced Hurricane Weather Research and Forecasting Model (AHW) are evaluated from the 2009 North Atlantic hurricane season. This data assimilation system involved cycling observations that included conventional in situ data, tropical cyclone (TC) position, and minimum SLP and synoptic dropsondes each 6 h using a 96-member ensemble on a 36-km domain for three months. Similar to past studies, observation assimilation systematically reduces the TC position and minimum SLP errors, except for strong TCs, which are characterized by large biases due to grid resolution. At 48 different initialization times, an AHW forecast on 12-, 4-, and 1.33-km grids is produced with initial conditions drawn from a single analysis member. Whereas TC track analyses and forecasts exhibit a pronounced northward bias, intensity forecast errors are similar to (lower than) the NWS Hurricane Weather Research Model (HWRF) and GFDL forecasts for forecast lead times \#60 h (\#60 h), with the largest track errors associated with the weakest systems, such as Tropical Storm (TS) Erika. Several shortcomings of the data assimilation system are addressed through postseason sensitivity tests, including using the maximum 800-hPa circulation to identify the TC position during assimilation and turning off the quality control for the TC minimum SLP observation when the initial intensity is far too weak. In addition, the improved forecast of TS Erika relative to HWRF is shown to be related to having initial conditions that are more representative of a sheared TC and not using a cumulus parameterization.

1. Introduction

Even though there has been a steady decrease in tropical cyclone (TC) track forecast errors since 1985, there has not been a comparable decrease in TC intensity forecast errors over the same period (e.g., Rappaport et al. 2009). One reason for the differences is that TC motion is mostly controlled by large-scale environmental winds (e.g., Wang and Wu 2004), which have been steadily getting better during that period due to improved models, improved data assimilation systems, improved moisture fields through dropsonde observations (e.g., Kamineni et al. 2006; Qu and Heming 2002), and additional targeted observations in the environment surrounding the storm. By contrast, TC intensity is thought to depend on the larger-scale kinematic and thermodynamic environment, the lower boundary condition (e.g., ocean heat content, land), and inner-core processes. Simulating the latter influence likely requires running a numerical weather prediction (NWP) model at grid scales of <4 km so that the finescale structures of the storm are resolved and a cumulus parameterization is no longer necessary (Done et al. 2004; Chen et al. 2007; Davis et al. 2008). Recent
studies suggest that increasing horizontal resolution from $O(10 \text{ km})$ to $O(4 \text{ km})$ alone does not lead to systematic improvements in intensity forecasts (e.g., Davis et al. 2010; Fierro et al. 2009; Gopalakrishnan et al. 2010), which suggests that these models suffer from errors due to inadequate initial conditions and deficiencies in model physics related to TCs.

Most of the reliable dynamical TC intensity models use initial conditions that are relatively coarse compared to the scale of a TC, thus several different techniques have been developed to add more realistic TCs to these initial conditions and relocate the storm to the observed position (e.g., Liu et al. 2000; Hsiao et al. 2010). Essentially, there are two different types of methods for constructing a TC-like vortex. One class of methods replaces the representation of the TC in the initial conditions with a vortex whose properties depend on a number of metrics related to the TC, including size, position, and intensity (e.g., Kurihara et al. 1993). The second class is similar to the first in that a vortex is constructed based on the TC properties, then profiles of that vortex are sampled and assimilated into the initial conditions in a manner similar to actual observations (e.g., Chou and Wu 2008).

Recently, ensemble-based data assimilation techniques, such as the ensemble Kalman filter (EnKF) have gained considerable attention for TC initialization. These techniques can make corrections to the TC (e.g., in terms of its location or amplitude) that are dynamically consistent in their multivariate relationships and in their spatial structure, even given limited observations. Moreover, the ensemble of analyses produced by such a system can be used to initialize ensemble forecasts out to arbitrary lead times. To date, this technique has shown great promise in improving TC track forecasts in National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS; Hamill et al. 2011) compared to the operational data assimilation system at the time and Advanced Research (ARW) Weather Research and Forecasting Model (WRF) forecasts relative to initializing with a global model analysis (Torn and Hakim 2009; Torn 2010). Moreover, assimilating Doppler radar data from both land-based and reconnaissance aircraft with an EnKF has led to improved intensity forecasts for a limited number of cases (Zhang et al. 2009, 2011). Torn (2010) also found that cycling with an EnKF system was particularly effective for weaker TCs, which was attributed to having initial conditions that captured the vertical structure of sheared TCs.

This study describes the results of running the data assimilation similar to the one described in Torn (2010) to initialize real-time 1.33-km Advanced Hurricane Weather Research and Forecasting Model (AHW) TC forecasts during the 2009 season. This activity was part of the National Oceanic and Atmospheric Administration (NOAA) Hurricane Forecast Improvement Project (HFIP), whose 10-yr goal is to reduce the errors in TC track and intensity forecasts by 50%. The application of this combined data assimilation and forecasting system over an entire hurricane season provides a large sample for assessing its potential benefit to TC forecasting, and is consistent with the HFIP’s overarching goals. This paper describes the first year where AHW forecasts are initialized from the same model as a combined real-time data assimilation and forecast modeling system, as there can otherwise be a significant adjustment period when initializing forecasts from a different model’s initial conditions (e.g., Chen and Snyder 2007; Davis et al. 2008). In addition to evaluating the performance of the combined data assimilation and forecasting system, this paper also addresses some of its observed deficiencies.

This manuscript will proceed as follows. The complete initialization and forecasting system will be described in section 2. Section 3 will provide the performance statistics of the initialization and ensemble forecasting system, while single-member high-resolution forecasts will be discussed in section 4. Initial condition and model sensitivities will also be examined and discussed for select cases in section 4. The manuscript will conclude in section 5 with a summary and future directions.

2. System overview

a. Data assimilation

Real-time ensemble analyses are generated each 6 h from 9 August 2009 to 16 November 2009 by cycling a 96-member EnKF system every 6 h over a limited-area domain with 36-km horizontal grid spacing (shown in Fig. 1) and 36 vertical levels. This period included all Atlantic basin TCs during the season, except Tropical Depression One, which existed from 28–29 May. This study uses many of the same data assimilation settings as Torn (2010) and are summarized below; the interested reader is directed to that manuscript for greater detail.

Observations are assimilated using the implementation of the ensemble adjustment Kalman filter in the Data Assimilation Research Testbed (DART; Anderson et al. 2009). These include Automated Surface Observing System (ASOS) stations, ships, buoys, rawindsondes, cloud motion vectors (Velden et al. 2005), aircraft data from the Aircraft Communications Addressing and Reporting System (ACARS), real-time estimates of TC position and minimum sea level pressure (SLP) produced by the National Hurricane Center (NHC; i.e.,
working best track or TCVitals), and dropsondes from aircraft that sampled the TC and its environment. A summary of the observation source and types are given in Table 1. Dropsondes from aircraft that sample the TC itself are excluded within a 200-km radius of the TC position due to the lack of resolution in the data assimilation system. Observation errors are taken from NCEP statistics, except for TC position and minimum SLP, whereas assumed to be 0.1° and 3 hPa, respectively.

Sampling errors owing to a finite ensemble are addressed using adaptive covariance localization and inflation strategies outlined in Torn (2010). The covariances are localized using Eq. (4.10) of Gaspari and Cohn (1999) where the localization value becomes zero 2000 km in the horizontal and 6 km in the vertical from the observation’s three-dimensional location. In regions where there are more than 1600 observations within the localization ellipsoid determined by the above horizontal and vertical length scales, the horizontal and vertical covariance length scales are proportionally reduced using the method outlined in Torn (2010). Adaptive covariance inflation is performed using the Anderson et al. (2009) technique, where the inflation factor is damped by 10% at each assimilation time and the inflation standard deviation is fixed at 0.6.

The ensemble members are advanced in time using version 3.1 of the ARW-WRF (Skamarock et al. 2008) with the following physics choices: Rapid Radiative Transfer Model (RRTM) longwave radiation (Mlawer et al. 1997), the National Aeronautics and Space Administration (NASA) Goddard shortwave radiation (Chou and Suarez 1994), WRF single-moment 5-class (WSM5) microphysics (Hong et al. 2004), Kain–Fritsch cumulus convection (Kain and Fritsch 1990), Yonsei University (YSU) planetary boundary layer (Hong and Pan 1996), thermal diffusion surface physics (Dudhia 1996), and Monin–Obukhov surface layer physics (Paulson 1970; Dyer and Hicks 1970; Webb 1970).

Torn (2010) noted that 36-km resolution systematically underestimated the initial intensity of category 3 and above TCs because of a lack of horizontal resolution. To partially overcome this limitation, a noncycled nesting procedure is employed that is meant to give the benefit of increased resolution within the TC without having to deal with the technical challenges associated with moving nested domains, such as each member having a different nest location and interpolation from the coarse grid. After the analysis is completed on the 36-km grid, a 12-km resolution fixed, two-way interactive nest (1000 km horizontal) centered on the TC position is generated for each system that NHC designated as a tropical depression or greater in real time by interpolating from the 36-km grid. Each ensemble member is then integrated forward to the next assimilation time using this two-way nesting approach. Once the forecast is finished, the nested domains are discarded so that the new analysis is only done on the 36-km domain. Because the forecast on the 36-km grid is simply averaged from the 12-km grid (at least where the 12-km exists), the prior forecast, even on the 36-km grid, will enjoy the benefits of increased resolution for intensity and thus should reduce the low-intensity bias, as will be demonstrated in section 3. This approach was motivated by technical limitations of the model itself; namely, that a new fixed nest has to be initialized solely from the 36-km domain, thus we could not have used an analysis produced on the 12-km nest(s). It should be noted that the data assimilation system can handle updating state variables on multiple nested domains, thus future work

![Fig. 1. Atlantic basin 36-km domain and tropical cyclone tracks based on National Hurricane Center best-track locations of all tropical cyclones during the 2009 Atlantic hurricane season. The lines indicate the track for each TC during its life cycle. Latitude and longitude lines are shown for each 10°.](image-url)

**Table 1. Summary of observations assimilated.**

<table>
<thead>
<tr>
<th>Platform(s)</th>
<th>Type(s)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASOS, ship, buoy</td>
<td>Surface altimeter</td>
<td>NCEP prepbufr</td>
</tr>
<tr>
<td>Rawindsonde</td>
<td>$u$ and $v$ wind components</td>
<td>NCEP prepbufr</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water vapor</td>
<td></td>
</tr>
<tr>
<td>Dropsonde</td>
<td>$u$ and $v$ wind components</td>
<td>NHC</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water vapor</td>
<td></td>
</tr>
<tr>
<td>ACARS</td>
<td>$u$ and $v$ wind components</td>
<td>MADISb</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td></td>
</tr>
<tr>
<td>Satellite wind</td>
<td>$u$ and $v$ wind components</td>
<td>CIMSSc</td>
</tr>
<tr>
<td>TC</td>
<td>Latitude, longitude, Min sea level pressure</td>
<td>NHC advisory data</td>
</tr>
</tbody>
</table>

---

*a NCEP standard binary universal form for the representation (BUFR) of meteorological data files that have been preprocessed for quality control.

*b Meteorological Assimilation Data Ingest System.

c Cooperative Institute for Meteorological Satellite Studies.
will involve developing strategies to assimilate data on the 12-km data and retain the nested analyses.

Prior to the beginning of the season, the potential benefit of employing this noncycled nest strategy was evaluated by running the data assimilation system from 12 August–14 September 2008 in the same configuration as described here. This period was characterized by five different TCs (Fay, Gustav, Hanna, Ike, and Josephine) with varied locations and intensities. In one experiment, denoted “no nest,” the assimilation system uses no nested domains, similar to what was used in Torn (2010), while the second experiment, denoted noncycled nests, is identical to the no-nest experiment, except that it uses the nesting strategy outlined above. The result of this experiment is reported in the next section.

Ensemble initial and boundary conditions are generated using the fixed-covariance perturbation technique described in Torn et al. (2006). This technique generates perturbations about an ensemble mean by sampling from the NCEP error covariances contained in the WRF data assimilation system (Barker et al. 2004). The ensemble mean initial condition on 9 August 2009 was the 24-h GFS forecast valid at that time, while the ensemble-mean lateral boundary conditions are taken from the GFS forecast initialized at the same time.

b. AHW Forecasts

At each time NHC designated a system as a tropical depression (TD) or greater intensity, a higher-resolution forecast was generated in real time using the AHW (Davis et al. 2008), which is a derivative of ARW; Table 2 provides the 48 forecast initialization times. There are several differences between the ARW and AHW, including the addition of a one-dimensional mixed layer ocean model based upon Pollard et al. (1973) that cools the mixed layer based on wind stress and the vertical temperature profile obtained from the Hybrid Coordinate Ocean Model (HYCOM) model (Bleck 2002). In addition, this formulation of the model uses the Donelan et al. (2004) surface drag coefficients, and enthalpy exchange coefficients as described by Dudhia et al. (2008), which is meant to provide better treatment of surface fluxes within TCs. These forecasts use a two-way nested configuration (Michalakes et al. 2005) with a fixed domain having a grid spacing of 12 km that varies by TC, and two movable nests with grid spacings of 4 and 1.33 km. The movable nests are centered on the location of minimum 850-hPa geopotential height, with a search radius centered on the location of the previous vortex position. The nest is repositioned every 15 simulation minutes with the search radius limited to the maximum distance that could be obtained for a vortex with a speed of 40 m s$^{-1}$. All other physics options are the same as the data assimilation system described above, except that the cumulus parameterization is only used on the 12-km domain. We perform the forecasts at high resolution to both test the behavior of AHW at resolutions beyond those typically used operationally and to produce the most skillful predictions possible, especially of intensity. Previous results (e.g., Done et al. 2004; Davis et al. 2008) show that higher resolution does yield benefits relative to coarse-resolution forecasts that use parameterized convection. Furthermore, employing higher-resolution nesting allows the TC inner core to be resolved and moist convection in the core to be explicitly represented.

Table 2. List of nested analysis times and AHW forecast initialization times.

<table>
<thead>
<tr>
<th>TC name</th>
<th>Start–end (No. of analyses)</th>
<th>AHW forecast initialization times (No. of forecasts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ana</td>
<td>1200 UTC 11 Aug–1200 UTC 17 Aug (25)</td>
<td>0000 UTC 12–15 Aug (4)</td>
</tr>
<tr>
<td>Bill</td>
<td>0600 UTC 15 Aug–0600 UTC 24 Aug (37)</td>
<td>0000 UTC 16–23 Aug (8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1200 UTC 15–22 Aug (8)</td>
</tr>
<tr>
<td>Claudette</td>
<td>1200 UTC 16 Aug–1800 UTC 17 Aug (6)</td>
<td>(0)</td>
</tr>
<tr>
<td>Danny</td>
<td>1200 UTC 26 Aug–0600 UTC 29 Aug (12)</td>
<td>1200 UTC 26–28 Aug (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0000 UTC 27–29 Aug (3)</td>
</tr>
<tr>
<td>Erika</td>
<td>0000 UTC 1 Sep–1800 UTC 3 Sep (12)</td>
<td>0000 UTC 1–3 Sep (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1200 UTC 1–3 Sep (3)</td>
</tr>
<tr>
<td>Fred</td>
<td>1800 UTC 7 Sep–1200 UTC 12 Sep (20)</td>
<td>0000 UTC 8–12 Sep (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1200 UTC 8–12 Sep (5)</td>
</tr>
<tr>
<td>TD-Eight</td>
<td>1200 UTC 25 Sep–1800 UTC 26 Sep (6)</td>
<td>0000 UTC 26 Sep (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1200 UTC 26 Sep (1)</td>
</tr>
<tr>
<td>Grace</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>Henri</td>
<td>0000 UTC 7 Oct–1800 UTC 8 Oct (8)</td>
<td>0000 UTC 7 Oct (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1200 UTC 7 Oct (1)</td>
</tr>
<tr>
<td>Ida</td>
<td>1200 UTC 4 Nov–1200 UTC 10 Nov (15)</td>
<td>0000 UTC 8 Oct (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1200 UTC 8 Oct (1)</td>
</tr>
</tbody>
</table>
Forecasts are initialized using a single member of the analysis ensemble described above, where the member is chosen based on the following cost function that measures how close each ensemble member’s position is to the ensemble mean position and the difference between each ensemble member’s minimum SLP and the real-time estimate from NHC:

\[
J(n) = \left( \frac{\text{Lat}_n - \text{Lat}}{\sigma_{\text{Lat}}} \right)^2 + \left( \frac{\text{Lon}_n - \text{Lon}}{\sigma_{\text{Lon}}} \right)^2 \\
+ 2 \left( \frac{\text{MSLP}_n - \text{MSLP}_{\text{NHC}}}{\sigma_{\text{MSLP}}} \right)^2,
\]

where Lat\(_n\), Lon\(_n\), and MSLP\(_n\) are ensemble member \(n\)’s estimate of the TC latitude, longitude, and minimum SLP; MSLP\(_{\text{NHC}}\) is the real-time NHC estimate of the minimum SLP; \(\sigma\) is a normalizing constant (0.15° for latitude and longitude, 4 hPa for minimum SLP); and an overbar indicates the ensemble mean. It is worth noting that when using the ensemble mean instead of a single member’s analysis as the initial condition, there is inherent spatial smoothing imposed on the TC structure. The implications of this are currently being investigated, and will be reported on in a future publication.

3. Performance during the 2009 Atlantic hurricane season

Figure 1 shows the NHC best-track positions of all TCs during the 2009 Atlantic hurricane season. This season consisted of 11 tropical depressions, 9 tropical storms, and 3 hurricanes. Two hurricanes, Bill and Fred, reached category 3 or higher on the Saffir–Simpson scale (Saffir 1973; Simpson 1974), with Bill reaching category-4 strength. In addition, Bill was the longest-lived TC during this period (8.75 days), with several of the TCs being characterized by a lifetime of 3 days or less (7 of 10 TCs).

All forecast metrics are evaluated against NHC postseason best-track data only for those times where the best track indicates a system of TD strength or greater (i.e., remnant lows and extratropical cyclones are excluded). Where appropriate, statistical significance is established using bootstrap resampling, whereby the distribution is resampled 10 000 times and the 90% confidence bounds established,\(^1\) similar to what is done in Davis et al. (2010).

Similar to previous studies, the performance of the data assimilation system is assessed by evaluating the RMS difference between best-track data and the ensemble-mean TC vital statistics (position, minimum SLP, and maximum wind speed) in both the analysis and background (6-h forecast). As a means of comparison, the same verification is performed on the real-time version of other operational models where full cycling is not performed [the Geophysical Fluid Dynamics Laboratory (GFDL) and the National Weather Service (NWS) Hurricane Weather Research Model (HWRF)].

In addition to evaluating the RMS error in the ensemble mean, it is also important to determine whether the variance within the ensemble is appropriate given the ensemble-mean error. This is determined by computing what we term as the “total spread,” which is the square root of the sum of the analysis or background forecast ensemble variance and the best-track error variance, with the latter determined from Torn and Snyder (2012). In a well-calibrated ensemble system, the RMS error should match this quantity (e.g., Houtekamer et al. 2005). It is worth noting that the assimilation system assimilates the real-time estimates of TC position and minimum SLP, not the postseason best-track values; therefore, although it is highly likely that the real-time and postseason best-track information is correlated, we do not consider this correlation when comparing the analysis error with the analysis total spread.

Prior to describing the results from the 2009 season, we present the comparison of the 2008 period with and without cycling nests within the assimilation system. Figure 2a, which shows the analysis and 6-h forecast errors for these two experiments, indicates there is a 22% reduction in the RMS error in 6-h TC minimum SLP forecasts with the noncycled nests, with a bias of nearly zero, compared to the 4-hPa bias without nests. By comparison, the maximum wind speed shows a more modest decrease (1 m s\(^{-1}\)) in both the RMS error and bias (Fig. 2b). The reduction in the RMS error also leads to a better match between the RMS error and the total spread, though the ensemble is still characterized by a lack of variance. Overall, these results indicate a clear benefit from this nesting strategy, which acts to reduce the bias in the background forecast, even without assimilating on this nest.

We now return to the analysis of the 2009 system. The assimilation of observations leads to an improvement in the TC vital metrics, though the degree of improvement depends on the TC intensity (Fig. 3). Over all TCs, there is a statistically significant reduction in the

\(^1\) While serial correlation among forecasts can be accounted for through block bootstrap approaches (Wilks 1997), the irregular temporal spacing of our forecast samples complicate the implementation of these traditional approaches. Therefore, we do not attempt blocked bootstrapping here.
ensemble-mean RMS position error for the analysis compared to 6-h forecasts; however, the analysis position error is nearly twice the value of the 6-h HWRF or GFDL position errors (Fig. 3a). Much of the increased error can be attributed to the large number of times where the TC is categorized as a TD or TS (93/141), which are often characterized by greater uncertainty in position (e.g., Torn and Snyder 2012). Recall that this data assimilation system does not use any kind of vortex repositioning or construction techniques, thus unlike HWRF or GFDL, there is no guarantee that the ensemble-mean TC position will match the real-time estimate of the position. Furthermore, this data assimilation system assumes that the TC vital observations have errors, as was shown in Torn and Snyder (2012), thus the analysis values should not fit the real-time best track values, whereas the HWRF and GFDL systems assume these quantities have zero error. It is worth noting that both the HWRF and GFDL initial position errors are smaller than the uncertainty in TC position for this year (Torn and Snyder 2012). Some of the position error could be related to large-scale wind biases [see Torn (2010)], which can lead to systematic position biases that are difficult to overcome since position observations do not have a strong correlation with the surrounding wind field (not shown).

Next, consider the ensemble total spread, which is also shown in Fig. 3. For all TCs less than category 3, the total spread is smaller than the RMS difference, indicating that the ensemble is spread deficient both in the analysis and 6-h forecast, though the ensemble has too much variance for the major TCs, but this could be related to the small number of cases. By contrast, Torn (2010) obtained a good match between the ensemble-mean analysis RMS difference and spread in TC position for 10 TCs from 2005 and 2007. The reason for this difference is unclear, though it could be related to this particular TC season, and to larger initial position biases for a few of the weaker TCs; this issue is addressed in section 4.

In contrast to TC position, which showed improvement in the ensemble-mean analysis relative to the ensemble-mean background forecast at all intensities, observation assimilation only appears to benefit the ensemble-mean TC minimum SLP for TD/TS in a statistically significant manner (Fig. 3b). Moreover, the background RMS difference is smaller than the total spread for all TCs less than category-3 intensity, indicating that the ensemble actually has too much variance in this quantity, while the opposite is true for major hurricanes. Over all times, the analysis RMS difference in 6-h forecasts are comparable to the operational HWRF and about 20% larger than GFDL. The largest analysis RMS differences are associated with greater than category-3 TCs (many of which are associated with Bill), much of which can be attributed to a weak bias. This large bias is likely due to the lack of horizontal resolution in the data assimilation system, despite the use of noncycled two-way nested domains in the region surrounding TCs within the data assimilation system.

---

2 The nonzero position errors at 0 h for the HWRF and GFDL forecasts are a reflection of the difference between real-time and postseason reanalysis of positions by NHC.
Similar to minimum SLP, ensemble-mean prior and posterior maximum wind speed errors are proportional to intensity, with larger biases for more intense TCs (Fig. 3c). Moreover, the total spread is less than the analysis RMS difference at all intensity categories and the analysis RMS error and total spread is equal to or larger than the background forecast value, which suggests that observation assimilation does not provide any systematic improvement to this quantity. This result is not particularly surprising given that the system does not consistently assimilate wind information in the vicinity of the TC that might improve the analysis maximum wind speed. As discussed by Torn (2010), the higher ensemble spread in the analysis can occur because maximum wind speed is defined as the maximum wind speed at a single grid point near the TC, thus spurious covariances can result in a higher maximum wind speed in some members, while others do not change. Finally, the analysis RMS differences over all storms are 50% and 35% greater than HWRF and GFDL forecasts, respectively, though it should be noted that the 0-h HWRF and GFDL values are smaller than the uncertainty in the maximum wind speed (Torn and Snyder 2012).

b. Forecast verification

The remainder of this section turns to the verification of the higher-resolution (1.33 km) deterministic AHW forecasts (denoted as AHW1) against NHC best-track data and compares it against a homogeneous sample of the real-time operational HWRF (Gopalakrishnan et al. 2010), GFDL (Bender et al. 2007), NCEP GFS (Environmental Modeling Center 2003), Statistical Hurricane Intensity Forecast (SHIFOR; Knaff et al. 2003) Model, and Logistic Growth Equation Model (LGEM; DeMaria 2009) forecasts. Only those times where the TC is designated at least TD in the post-season best track are considered in the verification.

Over the entire season, AHW1 TC position forecasts show mixed errors with respect to other operational guidance (Fig. 4a). Prior to 72 h, the AHW1 forecasts have significantly larger track errors compared to other operational guidance; however, after 96 h, AHW1 track errors are similar to or smaller than HWRF, GFDL, and GFS, though the number of cases at those lead times is small. Prior to 60 h, AHW1 track forecasts are systematically north of the actual position, while the other operational guidance is generally south of the actual TC (Fig. 4b). Much of the northward position bias in AHW1 can be attributed to Erika and Ida (not shown). In the former case, Erika was consistently initialized 100–200 km to the north of the NHC advisory position, even though the real-time TC position was assimilated each 6 h. The assimilation system consistently shifted the position up to 100 km to the south of the background forecast at each analysis time; however, the actual storm continued to reform to the south of its previously observed position, making it difficult for the analysis position to “catch up” with the actual storm. A method of improving the initial track errors associated with Erika will be discussed in section 4b. For Ida, much of the position bias is likely attributable to the large-scale wind biases in AHW1 forecasts for the western Atlantic. As shown in Torn (2010), this region is characterized by easterly and southerly wind biases that move the TC track closer to Central America (cf. his Fig. 6). The exact cause of these biases is unknown, though it is likely related to some model physics deficiency, such as the lack of shallow convection (e.g., Torn and Davis 2012).
In contrast to TC track, the intensity forecasts, measured by either minimum SLP or maximum wind speed, show either similar or reduced error with respect to the other models. With the exception of 0 h, the mean absolute error in AHW1 minimum SLP forecasts is comparable to GFDL and HWRF, while beyond 60 h, AHW1 has smaller errors that are statistically significant (Fig. 4c). For AHW1, the forecast RMS difference increases with forecast hour, then levels off at 9 hPa at 48 h, while the bias indicates AHW1 has TCs that are too weak during early lead times, and too strong later on. GFDL forecasts are characterized by a similar bias, though the forecast RMS difference increases at a faster rate. By contrast, HWRF shows a tendency to overintensify many TCs (elaborated on later), thus the forecast RMS differences are larger than those from other models.

For maximum wind speed, AHW1 starts off with larger errors relative to other models, but is characterized by a modest increase thereafter. By comparison, other operational models start off with lower errors at the initial time, but are characterized by larger error
growth rates (Fig. 4d). The AHW1 error distribution likely results from the initial conditions systematically underestimating the intensity at 0 h because of the lack of resolution in the assimilation system; however, once the model is able to spin up the TC, the bias decreases to near zero by 24 h at which time the mean-absolute error is statistically indistinguishable from the other models. The large biases at 96 h and beyond are due to the model maintaining stronger winds during the extratropical transition of Bill, and delaying the dissipation of Fred. There is very little growth of intensity error over time in AHW1 forecasts. This is similar to the AHW forecasts in Davis et al. (2008), and is partially due to the fact that the domain for which the metric is computed is following the storm, and that initial conditions lead to poor short-term forecasts that subsequently recover over time. Although not reflected in these error statistics because we did not verify any intensity forecasts when the system was not at least a TD, AHW1 correctly predicted the dissipation of the highly sheared TCs, including Ana, Erika, and TD-Eight, while HWRF and to a lesser extent GFDL forecasted these systems to reach category 2 and above. Potential reasons for why AHW1 was able to correctly predict the evolution of these systems better than HWRF will be explored in the next section.

4. Examination of specific cases

The summary verification presented in the previous section revealed a number of issues related to the data assimilation system. Here, we focus on a few times where the forecast performance is relatively poor due to inadequacies in the data assimilation system and test potential solutions to these issues. Throughout this section, the procedures used in the sensitivity experiments will be the same as the real-time configuration, except with the proposed revision. For computational considerations, each of these tests is carried out over a subset of times, rather than the entire 2009 season. This section proceeds by looking at initial intensity errors associated with Bill, which had the largest initial intensity errors, then turns to the initial position errors associated with Erika, which had the largest initial position errors of all TCs during this season, and the reasons why AHW correctly predicted its dissipation.

a. Hurricane Bill

Bill was the longest-lived and most intense TC of the 2009 Atlantic season, forming from an African easterly wave near the Cape Verde Islands on 15 August and remained a tropical cyclone until 24 August (Berg and Avila 2011). At peak intensity on 19 August, the maximum wind was 115 kt (~60 m s⁻¹), and the minimum sea level pressure of 943 hPa occurred on 21 August. Bill moved south and west of Bermuda, recurved to the north and east shortly before reaching the eastern coast of the United States, and subsequently passed directly over Newfoundland, Canada, before soon thereafter becoming extratropical. Figure 5 shows a series of AHW intensity forecasts for Hurricane Bill initialized at 12-h intervals. For initialization times during the first 3 days of Bill’s life (prior to 1200 UTC 18 August), AHW forecasts were able to capture the strong intensification of the system at about the correct time and rate, as measured by the minimum SLP. By contrast, forecasts initialized on or after 0000 UTC 19 August are characterized by minimum SLP values at least 20 hPa greater than observations, which corresponds to analysis times when Bill’s minimum SLP is no higher than 965 hPa (Fig. 5b). It appears that the initial intensity error could prevent the higher-resolution AHW forecast from correctly predicting the minimum SLP until Bill begins to weaken on 21 August.

Closer examination of the data assimilation system suggests that part of the reason for the large mismatch between the model minimum SLP and the real-time estimated value was the quality control system, which is meant to prevent bad observations from getting assimilated. In particular, the minimum SLP observations for Bill were rejected by the assimilation system from 1200 UTC 18 August to 1200 UTC 21 August because the difference between the real-time estimate of the minimum SLP and the ensemble-mean value exceeded 4 times the innovation standard deviation (i.e., the square root of the sum of the ensemble variance and the observation-error variance). This period corresponds to the approximate time when the ensemble-mean analysis intensity errors increase from roughly 10 to 30 hPa and the minimum SLP reaches 943 hPa (Fig. 6). Given the lack of intensity data being incorporated into the assimilation system during these assimilation times, it is possible that forcing the assimilation system to assimilate the minimum SLP observation could reduce the initial intensity error and improve the subsequent intensity forecast.

The above hypothesis is tested by restarting the assimilation system with the 0000 UTC 16 August analysis ensemble and cycling with observations through 0600 UTC 21 August where the same set of observations and data assimilation configuration is used, except that the minimum SLP observation is not subject to the quality control algorithm. Figure 6 shows that forcing the assimilation system to assimilate this observation reduces the ensemble-mean minimum SLP analysis error by nearly 50% during the times of interest, thus it appears that including this observation improves the initial intensity.
Even though the assimilation system produces a better initial intensity for Bill, there is little subsequent benefit to the higher-resolution AHW forecast. Figure 7 shows the 0000 UTC 20 August AHW track and intensity forecast initialized from the analysis that assimilated the minimum SLP observation (denoted experimental) and compares it to the real-time forecast that did not include it (denoted real time). Other initialization times are characterized by similar behavior and are not shown. In the forecast where the minimum SLP observation is assimilated, the TC moves farther west than the control during the first 72 h of the forecast, then recurves and moves farther to the east during longer lead times (Fig. 7a). Although the experimental 0-h minimum SLP is 953 hPa (20 hPa closer to the best track value), the model rapidly adjusts by filling to ~965 hPa during the first 15 min (Fig. 7b) and over the next 18 h, the control and experimental forecast that assimilate the minimum SLP converge toward the same value, which is roughly 20 hPa higher than best track. For maximum wind speed, the forecast that included the minimum SLP observation has a 30 kt (~15 m s\(^{-1}\)) higher maximum wind speed relative to the control by 6 h, which is within 5 kt of the best-track value; however, this improvement is short lived and the two forecasts have similar maximum winds by 12 h (Fig. 7). This result indicates that forcing the assimilation system to use the minimum SLP observation leads to a marginal improvement in Bill’s intensity forecast, which is similar to what Torn (2010) obtained with a lower-resolution version of the model.

The reason for the lack of improvement is unknown at this time, but one possibility is that the TC intensity on time scales of 12–24 h is strongly influenced by the initial TC environment and/or the TC structure, which are not properly corrected by a single observation of the minimum SLP. Comparison of forecasts between stronger and weaker intensification do not reveal large

---

**Fig. 5.** AHW forecasts of minimum sea level pressure for Bill initialized (a) 0000 UTC 15 Aug–1200 UTC 18 Aug and (b) 0000 UTC 19 Aug–1200 UTC 21 Aug. The NHC best-track minimum sea level pressure is shown in black for comparison.
differences in the background shear (not shown). However, G-IV dropsondes launched near Bill during these times indicate the presence of dry air layers and relatively high aerosol concentrations (J. Dunion 2012, personal communication). Note that 1200 UTC 18 August is the first time G-IV dropsondes near Bill are assimilated into the AHW-DART system; these observations reduce relative humidity values in the background forecasts, which could possibly limit further intensification in the model. The radiative effects of aerosol are not considered in this version of the model and this is an area of ongoing research.

b. Tropical Storm Erika

Erika was a weak, highly sheared system that existed from 1–3 September east of the Windward Islands, which formed out of a tropical wave that exited the western coast of Africa on 25 August (Berg and Avila 2011). Although it was short lived, Erika is notable because of large differences between the best-track position and that analyzed in AHW, and because operational models predicted this system would reach category-3 status, while in reality, the storm had degenerated into a remnant low (Fig. 8). In addition, all of these models were characterized by large short-term track errors. The remainder of this section explores potential ways to reduce the initial position errors associated with this TC and the role of both initial condition and model physics on forecast errors.

1) INITIAL POSITION

Even though the AHW assimilation system assimilated the TC latitude and longitude at each initialization time, the analyses were characterized by large position
errors, even several data assimilation cycles after genesis, which in theory could hurt the position forecast. Figure 8b shows that at the first analysis time after the NHC declared Erika a tropical storm, there was a 200-km position error. Over the next day and a half, the position errors are not significantly reduced, even though the position observations were assimilated at each time.

One potential reason for the lack of improvement from assimilating the TC position could be related to the algorithm used to find the TC center, which is based on the cyclone minimum SLP. This operator works well for well-defined TCs that have large horizontal gradients in the mass field; however, weak TCs like Erika are often characterized by relatively flat SLP gradients within a broad vortex and thus an ill-defined SLP minimum. As a consequence, there may not be a systematic covariance between TC position observations and the rest of the state, which could result in smaller corrections to the TC position.

As an alternative, one could specify the center of the TC using the maximum in circulation, or area-averaged vorticity, at a particular vertical level, which should not suffer from an ambiguous center location for weak systems and thus lead to larger covariances with the state and systematic position corrections. This idea is demonstrated in Fig. 9, which shows the analysis increment in 850-hPa geopotential height due to assimilating the TC latitude determined using observation operators that find the center using either minimum SLP or the 800-hPa circulation maximum, where the TC position is determined by computing the 800-hPa area-averaged vorticity for a circle of radius 150 km at each grid point and identifying the location with the maximum value.

For both operators, the increment pattern shows a dipole, with higher (lower) heights on the north (south) side of the TC, which has the net effect of shifting the TC to the south, closer to the real-time best-track position. While the increment pattern is similar with both operators, the increment associated with the circulation maximum operator is about 20% larger than the minimum SLP operator, particularly on the north side, which would produce a larger position increment.

The applicability of the circulation maximum in the position observation operator is tested by cycling the data assimilation system from 1200 UTC 1 September to 0000 UTC 4 September. All other settings are the same, thus any differences are solely attributable to the position operator. Figure 10 shows that the initial position errors of Erika are improved by as much as 31% when using the circulation operator, and that the observations draw the model position toward the actual position at an earlier time.

Although the new operator only produces a 93-km improvement in the initial TC position, the 1200 UTC 2 September AHW forecast initialized from the analysis that used the circulation operator is much closer to the NHC best track (Fig. 11a). In particular, the 42-h track error (at the time of final best-track position) is reduced by 56% compared to the control, while both the minimum SLP and maximum wind speed are improved. One reason for the large improvement in the track forecast given the relatively small change in initial position could be related to the large-scale flow pattern through which Erika moved. Figure 12 shows a comparison of the 0-h 500- (Fig. 12a) and 850-hPa wind (Fig. 12b). At this time, Erika is located within the axis of confluence within a diffluent wind field at lower levels, where a relatively strong circulation appears evident. Thus, vortices located slightly north of the NHC best-track position would have a greater likelihood of moving northwest, while vortices farther south would likely continue on a more westward track. Nevertheless, this result suggests
that using the maximum in circulation is beneficial to the assimilation system, particularly since NHC often uses satellite-derived cloud motions (i.e., wind) to determine the position (C. Landsea 2011, personal communication).

2) INTENSITY

The remainder of this section is devoted to understanding why the AHW forecasts did not intensify Erika, while other operational models, such as HWRF, produced a 100-kt (~51.4 m s\(^{-1}\)) TC. Here, we run additional AHW forecasts to understand the role of initial condition and model physics on Erika’s track and intensity at various lead times.

Even though both the AHW and HWRF initial conditions are characterized by similar intensities, measured through minimum SLP (0-hPa difference) and maximum wind speed [5 kt (~2.6 m s\(^{-1}\))] there are significant differences in TC structure. Figure 13 shows a west–east cross section of the meridional wind through Erika from both the AHW and HWRF initial conditions on 0000 UTC 2 September. In the AHW initial conditions, the winds tilt eastward with height and the wind speed decreases to less than 2 m s\(^{-1}\) around 7–8-km height (Fig. 13a). By contrast, the HWRF has a wind structure with little tilt, where the wind speed is greater than 2 m s\(^{-1}\) up to 15 km (Fig. 13b). At this time, Erika was near a region of 15 m s\(^{-1}\) (~30 kt) of 200–850-hPa westerly shear; thus, it is more likely that the AHW initial condition, which has a structure similar to other sheared TC (e.g., Jones 1995; DeMaria 1996; Frank and Ritchie 2001; Braun and Wu 2007), is more appropriate than the HWRF structure, which is more representative of a deep TC characterized by low shear.

The role of HWRF and AHW initial condition differences in producing the different forecasts is tested by running AHW with the 0000 UTC 2 September operational HWRF\(^3\) initial condition (denoted AHW-HWRF) and comparing it with the control AHW forecast that uses the AHW-DART analysis. All model settings, including the domain, are the same as the real-time AHW forecast, thus, any difference is solely attributable to initial conditions.

Figure 14 indicates that Erika’s track and intensity forecast are quite sensitive to the initial conditions. In addition to the initial position difference, the AHW forecast initialized from the AHW data assimilation system takes a more west-northwest track and moves faster, while the forecast from HWRF initial conditions has a slower and more northwesterly track, though both tracks have large errors relative to the actual TC. The intensity forecast differences are particularly striking, even at relatively short lead times. While the forecast from the AHW initial conditions does not have a maximum wind speed greater than 50 kt (~26 m s\(^{-1}\))
beyond 6 h, the AHW-HWRF forecast maximum wind speed increases by 50 kt (\(\sim 26 \text{ m s}^{-1}\)) during the first 12 h and stays above 80 kt (\(\sim 41 \text{ m s}^{-1}\)) through 80 h. This result suggests that this AHW forecast of Erika was very sensitive to the vortex structure; initializing the forecast with a sheared TC led to a much more accurate forecast, compared to using initial conditions with a vertically coherent vortex.

3) PHYSICS SENSITIVITY

In addition to differences in initial conditions, HWRF and AHW are characterized by different physical parameterizations, which could also explain some of the differences between the two forecasts of Erika. One of the important differences between these two models is that AHW uses explicit convection on the 4- and 1.33-km domains, while the operational HWRF during this season uses a cumulus parameterization on 27- and 9-km domains. Previous work suggests that explicit treatment of convection at horizontal grid spacings of 4 km or less improves the representation of the structure of organized convective systems compared to using a parameterized convection scheme (e.g., Done et al. 2004). This may be particularly true for Erika given the significant vertical tilt due to vertical shear. Most cumulus parameterizations assume that convection occurs in a column and not slantwise, which can help develop a more vertical warm core. To test this possibility, we generated another Erika forecast on the 36- and 12-km domains from 0000 UTC 2 September with identical initial conditions as in real time, and where both domains use the Kain–Fritsch cumulus parameterization.
Although differences exist between the AHW forecast that uses a cumulus parameterization and the forecast that uses explicit convection, these differences are much smaller than what is seen using different initial conditions. Figure 15 shows that while the forecast with explicit convection shows gradual weakening with time, the forecast with parameterized convection has a maximum wind speed that fluctuates around 50 kt (26 m s\(^{-1}\)) and a minimum pressure around 1000 hPa throughout. Moreover, the forecast with parameterized convection is slower and to the right of the control forecast with explicit convection, especially after 36 h. Nevertheless, these results seem to support the hypothesis that explicit convection can lead to more accurate forecast of sheared TCs, but more cases are needed.

5. Summary and conclusions

This manuscript describes the performance of real-time Advanced Hurricane WRF (AHW) forecasts during most of the 2009 Atlantic season. The initial conditions for these forecasts were obtained from a cycling EnKF system that assimilates conventional in situ observations plus TC position and intensity information each 6 h using a 96-member ensemble. At each 0000 or 1200 UTC initialization time where NHC was tracking a system of at least TD strength, a single member of this analysis ensemble is used to initialize a 120-h forecast of the TC using nested domains with 12-, 4-, and 1.33-km horizontal resolution.

In an average sense, the assimilation of observations systematically reduces the error in TC position; however, for intensity the improvement is limited to category 2 or weaker TCs. The latter result is obtained despite using noncycled two-way nested domains with 12-km resolution in the vicinity of TCs in the ensemble data assimilation system. One potential reason for the lack of improvement is that the quality control system...
rejected the TC minimum SLP observation during times with the largest errors. Bypassing this step for the TC minimum SLP observation during Hurricane Bill led to a 50% reduction in the ensemble-mean analysis minimum SLP error relative to the control experiment. Even with the large reduction in the 0-h intensity error, the benefit does not carry forward into the high-resolution forecast; by 24 h, the intensity is similar to the control.

**Fig. 14.** AHW TC (a) track, (b) minimum sea level pressure, and (c) maximum wind forecasts using AHW (blue) and HWRF (red) initial conditions on 0000 UTC 2 Sep 2009. NHC best-track data is shown in black.

**Fig. 15.** Comparison of AHW TC (a) track, (b) minimum sea level pressure, and (c) maximum wind forecasts with cumulus parameterization (blue) and without (red) initialized 0000 UTC 2 Sep 2009. NHC best-track data is shown in black.
where the TC minimum SLP observation is rejected. This result is similar to what has been found in previous studies (e.g., Torn 2010; Hamill et al. 2011), and suggests that assimilating the minimum SLP alone may not be sufficient to correct the initial conditions in a way that will improve the subsequent TC intensity when there is a large difference between the model and observed intensity. This situation may require higher density observations near the TC, such as dropsondes and/or Doppler velocity data (e.g., Zhang et al. 2009) or information about the radial wind profile (e.g., Wu et al. 2010).

Higher-resolution AHW track forecasts are characterized by larger track errors relative to other operational models, such as HWRF, GFDL, and GFS, during the first 60 h, with smaller values at longer lead times. The largest short-range track errors were associated with Erika, which was characterized by 100–200-km initial position errors, even though the TC position was assimilated each analysis time. The lack of improvement from assimilating TC position can be partially attributed to using the minimum in SLP as the TC position, which can lead to ambiguous center locations in weak TCs and smaller covariances between this observation and other model state variables. Rerunning the data assimilation system during Erika with an observation operator that uses 800-hPa circulation to determine TC position lead to a 50% reduction in 0-h position errors and lower track errors in the subsequent forecast. Although this improvement was seen for one particular case, there is reason to believe that a circulation-based position operator may work better within weak TCs compared to the mass-based operator used in previous studies (e.g., Chen and Snyder 2007; Torn and Hakim 2009; Wu et al. 2010). It is worth noting that this circulation operator has been implemented during subsequent hurricane seasons.

During this period, AHW intensity forecasts were characterized by similar or smaller errors relative to other benchmark intensity models, such as HWRF, GFDL, and LGEM. In particular, AHW had much lower errors for the high-shear tropical storms, such as Erika. Much of this skill can be attributed to the initial conditions containing a vertically tilted asymmetric vortex, rather than a vertically aligned axisymmetric vortex, which is assumed by initialization schemes that use vortex specification. Integrating AHW forward with initial conditions that contain a more upright vortex leads to intensity forecasts that are similar to operational models that predicted a category-3 TC, when the storm had dissipated. This result suggests that EnKF data assimilation systems, which have the benefit of flow-dependent error statistics, may be particularly useful for weaker TCs (e.g., Torn 2010; Hamill et al. 2011), which are often characterized by nonclassical structures.

Even though there are a number of positive aspects to this approach, there are also significant drawbacks that need to be addressed. For category-3 and higher TCs, there still exists a significant weak bias, which could be addressed by either cycling with vortex-following 12-km nests, or running with even greater horizontal resolution. Moreover, the assimilation system could also incorporate more observations near the core of the TC, including flight-level reconnaissance data or Doppler radar (e.g., Zhang et al. 2009). The nonzero track biases suggest there are a number of large-scale environmental biases in the model. Given that the same numerical model was used here for both the data assimilation and forecasts, this can be evaluated in the spirit of Phillips et al. (2004) for mesoscale models following Klinker and Sardeshmukh (1992). For example, atmospheric aerosols and corresponding radiative interactions, including dust flow out of the Saharan desert in Africa, are not included in the current configuration of AHW. While Saharan dust may alter the tropospheric temperature distribution by warming (cooling) the lower (upper) troposphere (e.g., Wong et al. 2009), the impact on TC is less clear, and is likely sensitive to the exact concentration and distribution of dust (e.g., Zhang et al. 2007; Rosenfeld et al. 2011). All of these topics are the subject of ongoing research and will be reported on in future work. Finally, this system produced an ensemble of analyses for an entire season, which could be used for TC predictability and sensitivity studies.

Acknowledgments. Satellite wind data was obtained from Chris Velden and Dave Stettner (CIMSS/SSEC). ACARS data for this study were made available to the Earth System Research Laboratory/Global Systems Division by the following commercial airlines: American, Delta, Federal Express, Northwest, United, and United Parcel Service. Dropsonde data were processed using software provided by NOAA/Hurricane Research Division of AOML. This research was supported by the Hurricane Forecast Improvement Project. Three anonymous reviewers provided insightful comments that improved the manuscript.

REFERENCES
Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane–ocean prediction


