Track and Intensity Forecasting of Hurricanes: Impact of Convection-Permitting Resolution and Global Ensemble Kalman Filter Analysis on 2010 Atlantic Season Forecasts

MING XUE AND JORDAN SCHLEIF

Center for Analysis and Prediction of Storms, and School of Meteorology, University of Oklahoma, Norman, Oklahoma

FANYOU KONG, KEVIN W. THOMAS, YUNHENG WANG, AND KEFENG ZHU

Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

(Manuscript received 9 June 2012, in final form 29 April 2013)

ABSTRACT

Twice-daily 48-h tropical cyclone (TC) forecasts were produced for the fall 2010 Atlantic hurricane season using the Advanced Research core of the Weather Research and Forecasting (WRF-ARW) model on a large 4-km grid covering much of the northern Atlantic. WRF forecasts initialized from operational Global Forecast System (GFS) analyses based on the gridpoint statistical interpolation (GSI) three-dimensional variational data assimilation (3DVAR) system and from experimental global ensemble Kalman filter (EnKF) analyses, and corresponding global GFS forecasts were intercompared. For the track, WRF forecasts show improvement over GFS forecasts using either set of initial conditions (ICs). The EnKF-initialized GFS and WRF are also better than the corresponding GSI-initialized forecasts, but the difference is not always statistically significant. At all lead times, the WRF track errors are comparable to or smaller than the National Hurricane Center (NHC) official track forecast error, with those of the EnKF WRF being smallest. For weaker TCs, more improvement comes from the model (resolution) than from the ICs. For hurricane intensity TCs, EnKF ICs produce better track forecasts than GSI ICs, with the best forecast coming from WRF at most lead times. For intensity, EnKF ICs consistently outperform GSI ICs in both models for weaker TCs. For hurricane-strength TCs, EnKF ICs produce forecasts statistically indistinguishable from GSI ICs in either model. For all TCs combined, WRF produces about half the error of the corresponding GFS simulation beyond 24 h, and at 36 and 48 h, the errors are smaller than those from NHC official forecasts. The improvement is even greater for hurricane-strength TCs. Overall, the WRF forecasts initialized with EnKF ICs have the smallest intensity error, and the difference is statistically significant compared to the GFS forecasts.

1. Introduction

The National Hurricane Center (NHC) forecasts of tropical cyclone (TC) track have improved significantly since 1990, thanks to improvements in observing systems, data assimilation (DA) techniques, and numerical weather prediction models (Rappaport et al. 2009). However, despite the improvement in track forecasts, forecasts of TC intensity have not improved much. Intensity forecasts are difficult because small-scale, inner-core processes are very important when predicting changes in TC intensity, and global models typically lack the resolution necessary to resolve the intense vortex circulation in the TC inner-core region. It has been hypothesized that a high-resolution grid would improve TC forecasting, especially that of intensity, because of the improved ability of high-resolution grids to resolve the inner-core structures of TCs and the strong gradients near the vortex center associated with intense hurricanes.

There have been only a limited number of studies focusing on the impact of grid resolution on forecasting TCs, based on real-time forecasts of a number of TCs. Such studies include Davis et al. (2010, hereafter D10) and those documented in the National Oceanic and Atmospheric Administration (NOAA) High-Resolution Hurricane Forecast Test Report (HRHFT; DTC 2009). The real-time forecasts in D10 used a version of the

Corresponding author address: Dr. Ming Xue, Center for Analysis and Prediction of Storms, University of Oklahoma, 120 David Boren Blvd., Norman, OK 73072. E-mail: mxue@ou.edu

DOI: 10.1175/WAF-D-12-00063.1

Advanced Research core of the Weather Research and Forecasting (WRF-ARW) model called the Advanced Research Hurricane WRF (AHW). It included the same 10 Atlantic TCs as were examined in the HRHFT: 6 from 2005 and 4 from 2007. D10's 120-h forecasts used ICs produced by a cycling regional EnKF run on a 36-km grid. Two sets of forecasts were made: one with a single 12-km grid, and one with triple-nested 12-, 4-, and 1.33-km grids. The 12-km grid was fixed while the 4- and 1.33-km grids followed the TCs. D10 found that there was no meaningful difference between storm position errors in the 12-km and nested higher-resolution forecasts. However, TC intensity (in terms of the maximum 10-m wind) was somewhat better forecasted on the nested grids than on the single 12-km grid, and the difference was statistically significant at 72 h and beyond. In particular, the intensity forecast for hurricanes of category 3 strength and stronger benefited the most from the nested grids. The 12-km forecasts tended to exhibit a negative intensity bias for stronger TCs, and the nested-grid forecasts showed a positive intensity bias for weaker TCs.

HRHFT was a study aimed at improving hurricane intensity forecasts. Six modeling groups (from which, only five produced usable results in real time) produced forecasts for 10 tropical cyclones of interest from 2005 and 2007 using five different modeling systems. The NOAA/Atlantic Oceanographic and Meteorological Laboratory (AOML) used the Experimental Hurricane WRF (HWRFX; Gopalakrishnan et al. 2011) model. The Mesoscale and Microscale Meteorology Division (MMM) of the National Center for Atmospheric Research (NCAR) used the AHW (Davis et al. 2008). The Naval Research Laboratory (NRL) ran a TC-optimized version of the Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS-TC; Hendricks et al. 2011). The University of Wisconsin-Madison (UWM) used their own Nonhydrostatic Modeling System (UW-NMS, Tripoli 1992). Finally, the University of Rhode Island (URI) used the Geophysical Fluid Dynamics Laboratory (GFDL) ocean-atmosphere model (Bender et al. 2007). The models were configured with multiple levels of nested grids (listed in the order of the modeling groups above): 27-9-3, 12-3-1.33, 81-27-9-3, and 12–3 km, and $\frac{1}{2^{\circ}}$, $\frac{1}{6^{\circ}}$, $\frac{1}{12^{\circ}}$, and $\frac{1}{18^{\circ}}$ combinations to examine the impacts of spatial resolution. Evaluations of these forecasting results are presented in DTC (2009). The study found that the use of higher resolution did not necessarily lead to an improvement in TC track forecasting. Only the AOML and MMM forecasts exhibited a significant improvement in track error at more than one lead time when using fine grid spacing. On the other hand, intensity forecasts were sometimes better: forecasts except for those of NRL and UWM showed improvements in forecasting intensity when using higher resolutions. In general, the use of a high-resolution grid reduced the magnitude of the negative intensity bias, and in some cases the bias became positive. Given the varied results from the above studies, further investigation into the impact of resolution is clearly needed. Further research into the use of physics schemes, oceanatmosphere coupling, initialization, etc. is also believed to be necessary in order to develop more accurate TC forecasts (DTC 2009).

In the fall of 2010, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma carried out a real-time forecast experiment for TCs of that season. Twice-daily (0000 and 1200 UTC) 48-h experimental TC forecasts were produced in real time for most of the northern Atlantic basin on a single high-resolution grid with 4-km grid spacing, using the WRF-ARW (Skamarock et al. 2008). Two sets of 4-km forecasts were produced: one initialized from the operational gridpoint statistical interpolation (GSI; Kleist et al. 2009) three-dimensional variational data assimilation (3DVAR) analysis of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) and one from the experimental global ensemble Kalman filter (EnKF) ensemble mean analyses produced by the NOAA/Earth System Research Laboratory (ESRL; Hamill et al. 2011b). The two WRF forecasts were compared with the corresponding global GFS model forecasts initialized from the operational GSI and experimental ESRL EnKF analyses, respectively. These forecasts allow us to examine the impact of high resolution and the initial conditions (ICs) on TC forecasts, including the forecasts of track and intensity. One important distinction between the CAPS experiment and the real-time forecasts mentioned earlier is that those forecasts all used relatively small, TCfollowing nested high-resolution grids as opposed to a single large high-resolution grid in the CAPS experiments.

The rest of this paper is organized as follows. Section 2 discusses the forecast models and their configuration, as well as the initial and boundary conditions used. Section 3 presents the intensity and track verification results, and section 4 provides a summary and our conclusions.

2. Forecast models and configurations

Version 3.1 of WRF-ARW was used in the CAPS forecasts. The forecast grid had 1801×901 grid points and is centered at 24°N, 63°W, spanning $70^{\circ} \times 30^{\circ}$ in latitude–longitude over the North Atlantic Ocean. The grid had 4-km spacing in the horizontal, and 51 vertical levels. Figure 1 shows the 4-km forecast domain along



FIG. 1. The seven named TCs examined in this study with their best tracks plotted on the WRF forecast domain.

with the tracks of the seven named storms examined in this study. This large 4-km domain was chosen to cover most of the tropical Atlantic, to track TCs from their genesis through possible recurvature or landfall, and the use of a single grid avoids complication and uncertainty related to the use of multiple nested grids and the associated, often large, lateral boundary condition influence (e.g., Warner et al. 1997). The use of a single grid permits a cleaner comparison between WRF and the global model running at two different resolutions, while at the same time noting that forecast differences can arise both from using finer resolution and from using the WRF model instead of the GFS. For continental stormscale real-time forecasting, 4-km resolution has been successfully used for several years (Xue et al. 2010) and has been shown to produce second-day forecasting guidance similar to corresponding 2-km forecasts (Schwartz et al. 2009) but is much better than the same model run at 20-km resolution (Clark et al. 2009). Often, 4-km resolution is referred to as convection permitting or marginally cloud resolving.

The WRF model employs the Thompson microphysics scheme (Thompson et al. 2006, 2008). Goddard shortwave radiation parameterizations (Chou and Suarez 1999; Chou et al. 1998), the Rapid Radiative Transfer Model (RRTM) for longwave radiation (Mlawer et al. 1997), the Noah land surface model (Ek et al. 2003), the Mellor–Yamada–Janjić boundary layer physics (Janjić 1990), no cumulus scheme, and positivedefinite moisture advection (Skamarock and Weisman 2009).

Two WRF-ARW forecasts were initialized each day at 0000 and 1200 UTC: one using the operational GFS analysis and forecasts to provide the initial and lateral boundary conditions, respectively (referred to as WRF-GSI), and another using the experimental global EnKF ensemble mean analysis produced by the NOAA/ESRL (Whitaker et al. 2008; Hamill et al. 2011b) as the ICs and the corresponding deterministic GFS model forecasts for lateral boundary conditions (referred to as WRF-EnKF). The global EnKF analysis assimilates all observations used by the GSI analysis and, in addition, includes observations of NHC advisory minimum sea level pressure (TCVitals; Hamill et al. 2011a,b). The GSI analysis employs vortex relocation to move the TC from its analyzed position to its actual location (Liu et al. 2000); vortex relocation is not used by the EnKF.

Within the global EnKF DA cycles, the GFS ensemble was run at T254L64 resolution (~47 km at 25°N). The ensemble mean EnKF analyses (Hamill et al. 2011b) used to initialize the GFS deterministic forecasts (GFS-EnKF) were run at T574L64 resolution (~21 km at 25°N). The operational GFS model forecasts based on GSI 3DVAR ICs (Kleist et al. 2009) also had T574L64 resolution (referred to as GFS-GSI). The WRF-GSI and WRF-EnKF forecasts will be evaluated together with the GFS-GSI and GFS-EnKF forecasts.

The primary difference between the 3DVAR and EnKF methods is with their determination of the background error covariance (e.g., Li et al. 2012). The operational GSI 3DVAR for GFS uses static background error covariance derived from historical forecasts using the so-called National Meteorological Center (NMC, now known as NCEP) method (Parrish and Derber 1992). Hence, the covariance is basically unaware of the presence of TCs in the forecast background, and is unable to produce dynamically consistent TC analyses unless comprehensive TC vortex-scale



Average Absolute Track Error

FIG. 2. (a) Absolute errors of TC track forecasts. Error bars represent the two-tailed 90% confidence interval using the block bootstrap distribution method. The number-*n* axis below the forecast-time axis represents the number of forecasts which decreases over time due to dissipation or movement of TCs outside of the forecast domain. (b) The 90% confidence interval of the mean difference between each of the variables plotted in (a). If the interval does not include zero, the difference is statistically significant at the 90% confidence level.

observations are available. The EnKF, however, derives the background error covariance from a forecast ensemble that is specific for the analysis time, and this flowdependent, TC-aware error covariance gives EnKF the ability to produce dynamically consistent TC analyses from a limited number of observations. For example, EnKF is able to correct errors in the wind field by assimilating the TC minimum sea level pressure advisory data (Hamill et al. 2011b).

Both the WRF and GFS forecasts were verified against the best-track data from NHC. The two WRF forecasts were compared to each other and to the GFS-GSI and GFS-EnKF forecasts to see how the grid resolution and ICs impact TC track and intensity forecasts. In addition, the forecasts were split into two groups: those for weaker cyclones initially at tropical depression or tropical storm strength, and those for stronger cyclones initially at hurricane strength. This is to determine how the ICs and the high-resolution grid affected forecasts for TCs of different initial intensity.

Verification statistics for all CAPS TC forecasts from 11 September through 9 October 2010, covering named storms Igor through Otto, were calculated against the NHC best-track data. The TC cases are limited to Igor



FIG. 3. As in Fig. 2, but for TCs classified as tropical depressions and tropical storms.

through Otto because the ESRL EnKF analysis and corresponding GFS forecast data were not available to CAPS outside this period. There were other occasions when the EnKF-initialized global forecasts were not run and therefore are unavailable for use. TC track and intensity forecasts were determined for those TCs with a closed circulation within the WRF forecast domain; Igor, Julia, and Lisa spent portions of their lifespans outside the domain (Fig. 1). In addition, occasionally a closed circulation was not found in the model initial conditions even though the NHC initiated advisories on the system at the time. With no circulation identified in the initial conditions, the corresponding forecasts were not included in the verification statistics. All of these examples lead to 44 TC forecast samples at the time of the initial conditions that are evaluated in this study. As a reference, we also include the official track forecast errors from NHC along with our track verification.

Throughout this paper, statistical significance is determined by using block bootstrap resampling, as in Hamill et al. (2011a). Because different forecasts for a particular TC are correlated, each set of TC forecasts (for Igor, Julia, etc.) was put into a block, and blocks are randomly selected 1000 times with replacement to construct a new set of bootstrap observations. From this sample, the mean is calculated, along with a 90% confidence interval. To determine whether the difference between two means is statistically significant, one set of bootstrap samples is subtracted from the other and a mean difference with a 90% confidence interval are



FIG. 4. As in Fig. 3, but for TCs classified as hurricanes.

computed and plotted in each figure. If the mean difference confidence interval between forecast pairs does not include zero, the difference is statistically significant at the 90% confidence level.

3. Results

a. Absolute track error

Absolute track error (ATE) is defined as the great circle distance between the forecast position and the best-track position of a TC center. In Fig. 2, the ATEs averaged over all TCs are shown for each of the four models from the initial analysis time (0 h) every 6 h until 48 h. At the initial time, the operational GFS analyses are better at determining the initial locations of the TC center positions compared to the EnKF ICs, likely due to the use of the vortex relocation technique (Liu et al. 2000). While the EnKF analyses also benefited from the assimilation of advisory minimum sea level pressure (MSLP) observations (TCVitals; Hamill et al. 2011a; Hamill et al. 2011b), the EnKF does not force the analyzed location to exactly match that of the best track. In the 6-h forecast, the track errors of both WRF forecasts are slightly worse than either of the global forecasts, which may be due to an adjustment caused by interpolating the global analysis to the high-resolution grid. However, in the 12-h forecasts exceeds that of the WRF forecasts. A similar pattern of behavior was

observed in the AOML and MMM runs of HRHFT (DTC 2009), which report that the track error in low-resolution forecasts increases faster than in the high-resolution forecasts, while the other HRHFT experiments and D10 showed little change or degradation of track forecasts using a high-resolution grid.

At 48h, WRF-EnKF forecasts have the smallest ATEs while GFS-GSI forecasts have the largest ATEs, about 50% larger than those of WRF-EnKF. This difference is statistically significant (Fig. 2b). At 6 h, WRF performs worse than GFS. At 12 and 18h, WRF performs better than GFS, but the difference between models is smaller than the difference between the EnKF and GSI ICs when using the same model (Fig. 2b). Between 24 and 42 h, both WRF forecasts outperform the GFS-GSI forecasts, and the benefit of convectionpermitting resolution clearly outweighs the benefit of the advanced EnKF DA approach here. In addition, both WRF-GSI and WRF-EnKF outperform GFS-GSI at the 90% confidence level at 24 and 30 h. Compared to the NHC official track forecasts, the WRF forecast ATEs are comparable or smaller (except for WRF-GSI at 48 h), while the global forecast ATEs are comparable to or higher than the official forecasts. The much larger absolute error differences between forecasts using GSI and EnKF ICs in both WRF and GFS at 48 h (see also Fig. 2a) are due to the poor predictions of Hurricane Lisa when initialized using GSI (not shown). Overall, the improved track forecasts in WRF-EnKF, especially when compared with GFS-GSI, show the benefit of using the high-resolution WRF model with the EnKF ICs.

To see further the track forecast performance for TCs of different intensity, the ATE was calculated separately for TCs below (Fig. 3) and above (Fig. 4) the hurricane strength threshold at the initial time. For weaker TCs (Fig. 3a), WRF-EnKF performs worst before 12 h. At 18 h, WRF-EnKF ATE decreases and becomes the second-best performer next to GFS-EnKF, and the difference is statistically significant compared to GFS-GSI. From 24 to 36 h and at 48 h, the high-resolution WRF produces smaller track errors than GFS. Forecasts from the same model are comparable using either IC: EnKF is slightly better some of the time and worse at other times (Fig. 3b).

For TCs of hurricane strength (Fig. 4a), the advantage of the EnKF IC is much clearer, with the EnKFinitialized forecast errors always smaller than the GSI-initialized forecast errors with the same model. The improvement seen when using EnKF ICs in the GFS (WRF) model is statistically significant (Fig. 4b) at 48 h (12 h). There is also a clear advantage with WRF-EnKF over GFS-EnKF, especially from 12 to 30 h when the difference is statistically significant. For



FIG. 5. Schematic of cross- and along-track errors in relation to actual and forecast hurricane tracks.

hurricane-strength TCs, there is little uncertainty about the initial center locations; the initial track errors in both GSI and EnKF are about 20 km. The initial track errors are larger for weaker TCs, ranging between 50 and 70 km, consistent with the larger difficulty in locating the center of the circulation in weak TCs. The mean track forecast errors for weaker TCs are also larger, exceeding 250 km in GFS-GSI forecasts at 48 h, while those of stronger hurricanes stayed below 200 km. Structural asymmetry and the larger influence of environmental steering flows are thought to be the cause of the larger ATE with weaker TCs.

b. Along- and cross-track error

As illustrated in Fig. 5, along-track error (AITE) is defined as the component of absolute error in the direction of the actual track of the TC. Similarly, cross-track error (CrTE) is defined as the component of absolute error in the direction perpendicular to the actual track.

The average absolute along- and cross-track errors for TC tracks are plotted in Figs. 6 and 7, respectively. Between 12 and 36 h, the GFS forecasts have larger AlTEs than the WRF forecasts, with the difference being largest at 24 h (about 50% larger; Fig. 6a). At the later



FIG. 6. As in Fig. 2, but for absolute along-track errors.

times, the differences become smaller and the GFS actually produced slightly smaller AITEs than WRF. Neither set of ICs exhibits a clear advantage in either model. The CrTEs of the four model forecasts are similar before 30 h (Fig. 7a), with no clear advantage in any one forecast. In general, cross-track errors are larger than along-track errors, especially for longer lead times. This indicates that forecasts have more difficulty determining the direction of a TC than its speed of movement. In terms of track bias, both WRF-GSI and WRF-EnKF forecasts are slightly faster than, and to the left of, the best track (not shown). This was also found by D10 for both of their 12-km and nested grids, for all lead times except 120 h.

c. Absolute wind speed error

The use of high-resolution WRF has the potential to significantly reduce the TC intensity forecast error. Figure 8 shows the absolute errors in the TC maximum 10-m wind speed from the four model forecasts. It is clear that EnKF does a much better job in analyzing the initial TC intensity in terms of the maximum surface wind than do the GSI analyses. The global EnKF analyses include TCVitals estimates of cyclone locations and minimum central pressure (Hamill et al. 2011b), which, through flow-dependent cross covariance, can directly update the wind fields and reduce the surface wind speed error. However, the absolute wind speed error (AWSE)



FIG. 7. As in Fig. 2, but for absolute cross-track error.

grows rapidly in GFS-EnKF from about 6 m s^{-1} at the initial time to above 10 m s^{-1} at 6h. This is due to the inability of the coarse-resolution GFS model to support intense vortices analyzed by the EnKF using TCVitals data (Hamill et al. 2011a). This rapid error growth is also observed in WRF-EnKF, but at a somewhat slower rate. Direct assimilation of TCVitals data on the highresolution WRF grid may help (but is not done here). After the initial adjustment period, the AWSE remains above 10 m s^{-1} in GFS forecasts, with the errors of GFS-EnKF remaining lower than those of GFS-GSI except for the final time. In contrast, the AWSE in WRF forecasts decreases slightly with time until 42 h, with the errors of WRF-EnKF being smaller than those of WRF-GSI except at 6 and 12 h. The reduction in wind speed forecast error with time indicates the ability of the highresolution WRF model to spin up strong TCs that may not have been properly analyzed at the initial time, producing a more dynamically balanced cyclone. By contrast, the GFS forecasts lack the resolution necessary to produce a dynamically balanced cyclone of realistic intensity, resulting in increasing error differences between the GFS and WRF forecasts. By 48 h, the GFS-EnKF error is almost twice that of WRF-EnKF, and the GFS-GSI error is about 50% more than the WRF-GSI error.

Overall, the AWSEs of WRF forecasts are much smaller than those of GFS forecasts, with statistical significance (Fig. 8b), while the EnKF-initialized forecasts are slightly better than corresponding GSI-initialized



FIG. 8. As in Fig. 2, but for absolute wind speed errors.

forecasts, but the differences are not statistically significant. For comparison, three experiments in HRHFT (DTC 2009) exhibited statistically significant improvement in wind speed error at some lead times when using a high-resolution grid. The AOML 3-km experiment exhibited improvement for lead times between 30 and 48 h. The MMM 1.33-km experiment only saw significant improvement at 18-h lead time. The NRL 3-km experiment showed significant improvement at 24 and 42 h. In the UWM experiment, the 3-km experiment showed improvement at 48–84-h lead times. Neither D10 nor the remainder of the HRHFT experiments found a statistically significant improvement in intensity forecasts.

The wind speed forecast errors for strong and weak TCs evolve quite differently. After 6h, the error

generally increases with time for tropical depressions and tropical storms (Fig. 9a), but generally decreases with time for hurricane intensity TCs (Fig. 10a), especially before 36 h. For the weaker TCs, the wind speed error remains at similar levels in the first 6 h, except for WRF-GSI, in which the error decreases noticeably due to spinup. There are also small decreases in wind speed error in WRF-EnKF out to 18 h. These are indications of vortex spinup in the high-resolution model, starting from initial vortices that are somewhat too weak. Similar spinup does not happen in the GFS forecasts due to the lack of resolution, and the error increases monotonically with time (Fig. 9a) except at 42 h. The WRF errors also increase with time after the initial spinup period. For the weak TCs, the wind speed errors are consistently ranked



FIG. 9. As in Fig. 8, but for TCs classified as tropical depressions and tropical storms.

in descending order for GFS-GSI, GFS-EnKF, WRF-GSI, and WRF-EnKF after 12h, again showing the benefits of both high resolution and EnKF DA. At 36–48h, WRF-EnKF forecasts for weak TCs perform the best, especially compared to GFS-GSI where the difference is statistically significant (Fig. 9b).

For hurricane intensity TCs (Fig. 10a), there is a large jump in the wind speed error from an initially low level to more than twice as much at 6h in both EnKFinitialized forecasts. As discussed earlier, this is mainly due to the low resolution at which the EnKF analyses were produced (at T256L64, ~48 km at 25°N), and the lack of full dynamic consistency and balance among state variables in the ICs. In stronger TCs, assimilation of TCVitals observations in the EnKF analysis creates a central pressure that is in a good agreement with the minimum sea level pressure but not necessarily in very good balance with the vortex circulation and temperature field. Therefore, the hurricanes undergo a spindown process before they are spun up again. The spinup is especially clear in the WRF forecasts: the errors decrease from $12-13 \text{ m s}^{-1}$ at 6 h to $6-7 \text{ m s}^{-1}$ at 36 h before they increase again. In GFS forecasts, there is no initial spindown; errors decrease until 36 h before increasing again. The resolution played a greater role than the initial conditions in reducing error in the intensity forecast, as there was little difference between GSI- and EnKF-initialized forecasts of the same model. Overall, the stronger TCs benefit much more from higher resolution than do the weaker TCs for intensity forecasts,



Average Absolute Wind Speed Error -Hurricanes

FIG. 10. As in Fig. 8, but for TCs classified as hurricanes.

and at 48 h, the WRF-EnKF forecast error is slightly smaller for hurricanes than for weaker TCs. This is different from the results of D10, which showed that the wind errors for weak TCs were larger than those for strong TCs on a 1.3-km nested grid.

TC wind speed as forecasted by global models is known to have a large negative bias; that is, the maximum wind speed is forecasted to be much lower than the best-track wind speed. This is due to the inability of global models to resolve small-scale TC structures and properly capture their intensity changes. The wind speed biases of the four sets of forecasts are plotted in Fig. 11. Both GFS forecasts have consistently negative biases of about 10 m s^{-1} or larger at all forecast times, and the GFS-EnKF forecast biases are slightly smaller than those of GFS-GSI except at 48 h, but the differences are not statistically significant. Starting from 6 h, the WRF forecast biases are generally $\frac{2}{3}-\frac{1}{3}$ of the GFS forecasts, and the EnKF-initialized biases are smaller than the GSI-initialized forecasts for the WRF model between 6 and 42 h. At 42 h, the negative bias of WRF-EnKF is about 2 m s^{-1} , well within the best-track wind speed estimation uncertainty (Torn and Snyder 2012). The results of D10 for wind speed bias on a 12-km grid are similar: there is a negative bias through 48-h lead time. However, their 1.33-km nested grid had a positive bias. The AOML experiment in HRHFT yielded underforecasting of intensity for both low- and high-resolution



FIG. 11. As in Fig. 2, but for biases of TC maximum surface wind speed forecasts.

forecasts at all lead times up to 48 h; however, the negative bias was reduced in high-resolution forecasts. Both MMM and UWM systems resulted in underforecasting (overforecasting) of TC intensity in low- (high-) resolution forecasts.

d. Absolute minimum sea level pressure error

MSLP gives another measure of TC intensity; we look at both MSLP and maximum surface wind speed because conclusions based on the two are not always the same, at least quantitatively. The mean absolute MSLP errors are shown in Fig. 12. In general, the MSLP errors tell a similar story as the wind speed error, except that the GFS-EnKF forecast errors are higher than those of GFS-GSI from 18h onward, but they are statistically indistinguishable. Both EnKF-based forecasts went through an error increase before 6h due to spindown. Beyond 6 h, the GFS forecast errors show an overall increasing trend while the WRF forecast errors show an overall decreasing trend. The error differences between GFS and WRF runs are clearly statistically significant after 18 h while those between the GSI and EnKF forecasts of the same model are not (Fig. 12b). We note that most of the WRF MSLP forecast errors are below 8 hPa in day-2 forecasts. At 48 h, the WRF MSLP forecast error uncertainty is smaller than that of global forecasts, and the WRF forecasts cut the MSLP error in half compared to the global forecasts.

The MSLP forecast bias is related to the wind speed forecast bias. The large negative wind speed bias for global models in Fig. 11 is accompanied by a large positive MSLP bias in Fig. 13. The high-resolution WRF model reduces the MSLP forecast bias to smaller than 3 hPa from 30 h onward, compared to over 10 hPa in the



Absolute MSLP error

FIG. 12. As in Fig. 2, but for absolute MSLP errors.

global forecasts over the same period. Figure 13 also shows that the MSLP biases are larger in the EnKFinitialized forecasts in both the GFS and WRF models at the later times (though statistically insignificant; Fig. 13b), even though the WRF-GSI MSLP errors are larger than those of WRF-EnKF at 18h and beyond (Fig. 12). This suggests that there are more MSLP error cancellations in WRF-GSI forecasts than in WRF-EnKF when bias is calculated.

The MSLP forecast bias was also calculated separately for weak and strong TCs. For tropical storms and tropical depressions (Fig. 14), the MSLP bias for both GSI and EnKF ICs is initially positive and under 5 hPa. After staying relatively constant for the first 18 h, GFS-GSI and GFS-EnKF MSLP biases grow to over 10 hPa at 48 h. Meanwhile, WRF-GSI and WRF-EnKF biases decrease to near zero at 42 h, before increasing back to near 5 hPa at 48 h. For hurricanes (Fig. 15), the initial positive bias is larger (near 10 hPa for EnKF ICs, and 13 hPa for GFS ICs) because the resolution of the analyses is too coarse to resolve the actual intensity of the hurricane. For all lead times, the MSLP bias for GFS-GSI stayed between 12 and 15 hPa and the bias for GFS-EnKF increased to about 15–17 hPa from 12h onward. For the high-resolution forecasts, WRF-GSI biases generally decrease with time, becoming negative at 30 h, and decreasing to -5 hPa at 48 h. WRF-EnKF's bias increases at first as the model variables become more dynamically consistent, then decreases to near zero at 48 h. The high-resolution model predicts



FIG. 13. As in Fig. 2, but for biases of TC MSLP forecasts.

hurricanes of more accurate intensity with small biases at longer lead times, though the differences in intensity error and bias between GSI- and EnKF-initialized forecasts from the same model are not always statistically significant.

4. Summary and conclusions

This study examines the impacts of high, convectionpermitting model resolution and EnKF data assimilation on the track and intensity forecasts of 2010 tropical cyclones in the Atlantic basin. The twice-daily, 48-h forecasts used the WRF-ARW model and a single large 4-km grid covering much of the North Atlantic. Two sets of ICs were used for parallel WRF forecasts: one set is the NCEP operational GFS analyses (produced by the operational GSI 3DVAR data assimilation scheme), and another is the experimental ESRL global EnKF ensemble mean analysis. The use of the 4-km highresolution grid was hypothesized to improve hurricane forecasting, particularly intensity forecasting, due to the ability of the high-resolution grid to resolve the innercore structures and processes important to forecasting TC intensity. The single large grid also avoids the complications (e.g., discontinuities and interactions across domain boundaries; the initialization of fine grids using coarse-resolution solutions when the grid moves) and uncertainties (e.g., domain movement and effects on track forecasting) involved with the use of multiple movable nested grids. The use of a single grid also allows the representation of TC environments entirely on the high-resolution grid. Such a practice was also hypothesized to help improve track forecasting. The use of two sets of ICs in the 4-km WRF forecasts and the availability of global GFS forecasts from the same sets of ICs, provide us with an opportunity to evaluate the impacts



FIG. 14. As in Fig. 13, but for TCs classified as tropical depressions and tropical storms.

of the 3DVAR-based operational global analyses and experimental EnKF global analyses on forecasts at both low convection-parameterizing and high convectionpermitting resolutions. Using best-track data from the National Hurricane Center, verification statistics were calculated for both sets of WRF forecasts, and compared with those of GFS model forecasts. The main conclusions are listed below. Overall, the findings are encouraging and demonstrate that the use of a high-resolution model in hurricane forecasting is an avenue worthy of further exploration while the theoretically more advanced EnKF data assimilation method produces better track forecasts at both coarse global-model and high convection-permitting resolutions.

 Significant improvement to track forecasts was observed in the high-resolution WRF forecasts initialized with EnKF, compared to global GFS model forecasts initialized with GSI (Fig. 2). This difference is significant from 12–30 h (every 6 h) and at 48 h. There is a slight improvement in the track forecast in EnKFinitialized GFS over GSI-initialized GFS at all lead times for all TCs combined, but the improvement is not statistically significant. At all lead times 12 h apart, WRF-EnKF forecast track errors are smaller than the NHC official forecasts. Most of the error in track forecasts is found to be due to cross-track error, indicating a greater difficulty in pinpointing the direction of the TC movement than the forward speed (Figs. 6 and 7).

• For TCs below hurricane strength, more improvement in the track forecasts comes from high resolution at 48 h, while the differences due to the ICs are smaller in both the WRF and GFS models (Fig. 3). While EnKF



FIG. 15. As in Fig. 13, but for TCs classified as hurricanes.

ICs improve the track forecasts for weak TCs in both models at 12 and 18 h, GSI ICs produce smaller errors initially and at 36 and 42 h. For hurricane-intensity TCs, the EnKF ICs improve both the GFS and WRF forecasts, and high-resolution WRF track forecasts are better than GFS forecasts except at 48 h, but the differences are generally not statistically significant.

• The high-resolution WRF forecasts produce significantly improved intensity forecasts at 24 h and beyond, both in terms of maximum 10-m wind speed (Fig. 8) and minimum sea level pressure (Fig. 12). For all TCs combined together, the maximum surface wind speed errors of the high-resolution WRF are about one-third smaller than those of GFS forecasts beyond 24 h. The error differences are even larger for TCs of hurricane strength (Fig. 10), with the wind speed errors in the high-resolution forecasts being less than half of the GFS forecast error. An improvement in intensity for weak TCs (Fig. 9) is also observed in the high-resolution forecasts, but the improvement is not as dramatic as that for strong TCs. Examining the impact of ICs on intensity forecasts, EnKF ICs improve the wind speed forecasts in both GFS and WRF models for most of the lead times compared to GSI ICs, but the improvements are not statistically significant and are much smaller than those due to increased model resolution. Intensity improvements attributed to ICs, both in terms of wind speed and MSLP, are found to be mostly associated with weaker TCs.

 Global forecasts have negative wind speed biases, underestimating hurricane intensity, while the highresolution WRF forecasts reduce this bias by up to ²/₃ compared to GFS forecasts (Fig. 11). The GSIinitialized WRF overintensifies hurricane-strength TCs in terms of MSLP error at 30-h lead time and beyond (Fig. 12).

Further studies are possible to explore the impacts of resolution and data assimilation on hurricane forecasting. The spindown followed by spinup observed in this study from dynamically imbalanced TCs represented in the EnKF ICs leads to the question of whether performing EnKF data assimilation directly on the high-resolution grid would lead to a better intensity forecast, especially at shorter lead times. Forecasts of longer ranges than the 48-h forecasts examined here should also be explored, which would require more computational resources. Related questions of interest to the TC forecasting community include if a high-resolution grid improves forecasts of TC genesis and rapid intensification. Other avenues worthy of study include which physics packages are best for highresolution TC forecasting, and examining the potential benefit of ocean-atmosphere coupling.

Similar forecasts are being carried out for the Pacific basin. Large forecast samples will also increase the statistical robustness of the conclusions. Even-higher convection-resolving resolutions as well as high-resolution assimilation of any available TC inner-core observations may be beneficial to further improving the intensity or even track forecasting. These are topics for future research.

Acknowledgments. This research was primarily supported by grants from the Office of Naval Research (ONR) Defense EPSCOR program (N00014-10-1-0133) and ONR funding (N00014-10-1-0775). The global EnKF runs were supported by the NOAA THORPEX and Hurricane Forecast Improvement Program (HFIP) to ESRL. The global EnKF analyses used to initialize WRF forecasts were provided by Jeff Whitaker of ESRL. The WRF forecasts were carried out at the University of Oklahoma Supercomputing Center for Education and Research (OSCER). A mass storage system at the Pittsburgh Supercomputer Center was used for data archival. The first author was also supported by NSF Grants AGS-0802888, OCI-0905040, AGS-0941491, AGS-1046171, and AGS-1046081. Wei Li performed the data archiving, and Xuguang Wang and Keith Brewster provided input on the forecast grid configuration.

REFERENCES

- Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane–ocean prediction system and a summary of its performance. *Mon. Wea. Rev.*, **135**, 45.
- Chou, M.-D., and M. J. Suarez, 1999: A shortwave radiation parameterization for atmospheric studies. NASA Tech Rep. NASA/TM-104606, 40 pp.

- —, —, C.-H. Ho, M. M. H. Yan, and K.-T. Lee, 1998: Parameterizations for cloud overlapping and shortwave single-scattering properties for use in general circulation and cloud ensemble models. *J. Climate*, **11**, 202–214.
- Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-permitting and large convection-parameterizing ensembles. *Wea. Forecasting*, 24, 1121–1140.
- Davis, C., and Coauthors, 2008: Prediction of landfalling hurricanes with the Advanced Hurricane WRF model. *Mon. Wea. Rev.*, 136, 1990–2005.
- —, W. Wang, J. Dudhia, and R. Torn, 2010: Does increased horizontal resolution improve hurricane wind forecasts? *Wea. Forecasting*, 25, 1826–1841.
- DTC, 2009: High-resolution hurricane test final report. Hurricane Forecast Improvement Project, Developmental Testbed Center, 95 pp. [Available online at http://www.dtcenter.org/ plots/hrh_test/HRH_Report_30Sept.pdf.]
- Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, V. Koren, and J. D. Tarpley, 2003: Implementation of the upgraded NOAH land-surface model in the NCEP operational mesoscale Eta Model. J. Geophys. Res., 108, 8851, doi:10.1029/2002JD003296.
- Gopalakrishnan, S. G., F. Marks, X. Zhang, J.-W. Bao, K.-S. Yeh, and R. Atlas, 2011: The experimental HWRF system: A study on the influence of horizontal resolution on the structure and intensity changes in tropical cyclones using an idealized framework. *Mon. Wea. Rev.*, **139**, 1762–1784.
- Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011a: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, 139, 668–688.
- —, —, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011b: Predictions of 2010's tropical cyclones using the GFS and ensemble-based data assimilation methods. *Mon. Wea. Rev.*, 139, 3243–3247.
- Hendricks, E. A., M. S. Peng, X. Ge, and T. Li, 2011: Performance of a dynamic initialization scheme in the Coupled Ocean– Atmosphere Mesoscale Prediction System for Tropical Cyclones (COAMPS-TC). Wea. Forecasting, 26, 650–663.
- Janjić, Z. I., 1990: The step-mountain coordinate: Physical package. Mon. Wea. Rev., 118, 1429–1443.
- Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, 24, 1691–1705.
- Li, Y., X. Wang, and M. Xue, 2012: Assimilation of radar radial velocity data with the WRF ensemble–3DVAR hybrid system for the prediction of Hurricane Ike (2008). *Mon. Wea. Rev.*, 140, 3507–3524.
- Liu, Q., T. Marchok, H.-L. Pan, M. Bender, and S. Lord, 2000: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. NOAA Tech. Procedures Bull. 472, 7 pp.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res., **102** (D14), 16663–16682.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, 24, 395– 419.

- Schwartz, C., and Coauthors, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372.
- Skamarock, W. C., and M. L. Weisman, 2009: The impact of positivedefinite moisture transport on NWP precipitation forecasts. *Mon. Wea. Rev.*, **137**, 488–494.
- —, and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Rep. NCAR/TN-475+STR, 113 pp.
- Thompson, G., P. R. Field, W. D. Hall, and R. M. Rasmussen, 2006: A new bulk microphysical parameterization for WRF (& MM5). WRF User's Workshop, Boulder, CO, NCAR, 5.3. [Available online at http://www.mmm.ucar.edu/wrf/users/ workshops/WS2006/abstracts/Session05/5_3_Thompson.pdf.]
- —, —, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115.

- Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. Wea. Forecasting, 27, 715–729.
- Tripoli, G. J., 1992: A nonhydrostatic mesoscale model designed to simulate scale interaction. *Mon. Wea. Rev.*, **120**, 1342– 1359.
- Warner, T. T., R. A. Peterson, and R. E. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, 78, 2599–2617.
- Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482.
- Xue, M., and Coauthors, 2010: CAPS realtime storm scale ensemble and high resolution forecasts for the NOAA Hazardous Weather Testbed 2010 Spring Experiment. Preprints, 25th Conf. on Severe Local Storms, Denver, CO, Amer. Meteor. Soc., 7B.3. [Available online at https://ams.confex. com/ams/pdfpapers/176056.pdf.]