After a Decade Are Atlantic Tropical Cyclone Gale Force Wind Radii Forecasts Now Skillful?

JOHN A. KNAFF

NOAA/Center for Satellite Applications and Research, Fort Collins, Colorado

CHARLES R. SAMPSON

Naval Research Laboratory, Monterey, California

ABSTRACT

The National Hurricane Center (NHC) has a long history of forecasting the radial extent of gale force or 34-knot (kt; where $1\,kt = 0.51\,m\,s^{-1}$) winds for tropical cyclones in their area of responsibility. These are referred to collectively as gale force wind radii forecasts. These forecasts are generated as part of the 6-hourly advisory messages made available to the public. In 2004, NHC began a routine of postanalysis or "best tracking" of gale force wind radii that continues to this day. At approximately the same time, a statistical wind radii forecast, based solely on climatology and persistence, was implemented so that NHC all-wind radii forecasts could be evaluated for skill. This statistical wind radii baseline forecast is also currently used in several applications as a substitute for or to augment NHC wind radii forecasts. This investigation examines the performance of NHC gale force wind radii forecasts in the North Atlantic over the last decade. Results presented within indicate that NHC's gale force wind radii forecasts have increased in skill relative to the best tracks by several measures, and now significantly outperform statistical wind radii baseline forecasts. These results indicate that it may be time to reinvestigate whether applications that depend on wind radii forecast information can be improved through better use of NHC wind radii forecast information.

1. Introduction

The estimation and forecast of surface winds associated with tropical cyclones (TCs) are important to a variety of public, private, and governmental stakeholders and applications. To provide information about TC risks, the National Hurricane Center (NHC) makes 6-hourly forecasts of TC tracks, intensities, and structures for all active TCs. The initial and forecast TC wind structures are provided in terms of the maximum radial extent of 34, 50, and 64 knots (kt; $1\,kt = 0.51\,m\,s^{-1}$) or gale force, damaging, and hurricane force winds in quadrants surrounding the TC. These are collectively referred to as wind radii. NHC forecasts hurricane force wind radii through 36 hours, and damaging and gale force wind radii through 72 hours, while intensity and track are forecast through 120 hours.

The forecasting of TC structure/wind radii at NHC was last described in the refereed literature by Rappaport et al. (2009). An update to that information is provided here. The 34-kt wind radii forecast process starts with the analysis. Data available for analysis include scatterometry, satellite estimates such as those from AMSU, aircraft data, and sporadic ship, buoy, and land observations. Scatterometry has the ability to provide the best picture of the 34-kt wind field, but the data are intermittent and often only sample a part of the storm. Current forecasting of 34-kt radii tries to capture bulk trends in the size of the storm. Is the wind field growing or shrinking? Are asymmetries developing because of the heading and forward speed, or a change in the synoptic pattern? Guidance available for 34-kt wind radii forecasts includes wind radii climatology/persistence models, and regional hurricane models such as the Hurricane Weather Research and Forecast Model (HWRF); however, the performance of the regional models is often limited by the skill of their intensity forecasts. In addition, global models such as the Global Forecast System (GFS) model and the European Centre

Corresponding author address: John Knaff, NOAA/NESDIS, CIRA, Campus Delivery 1375, Colorado State University, Fort Collins, CO 80523-1375.
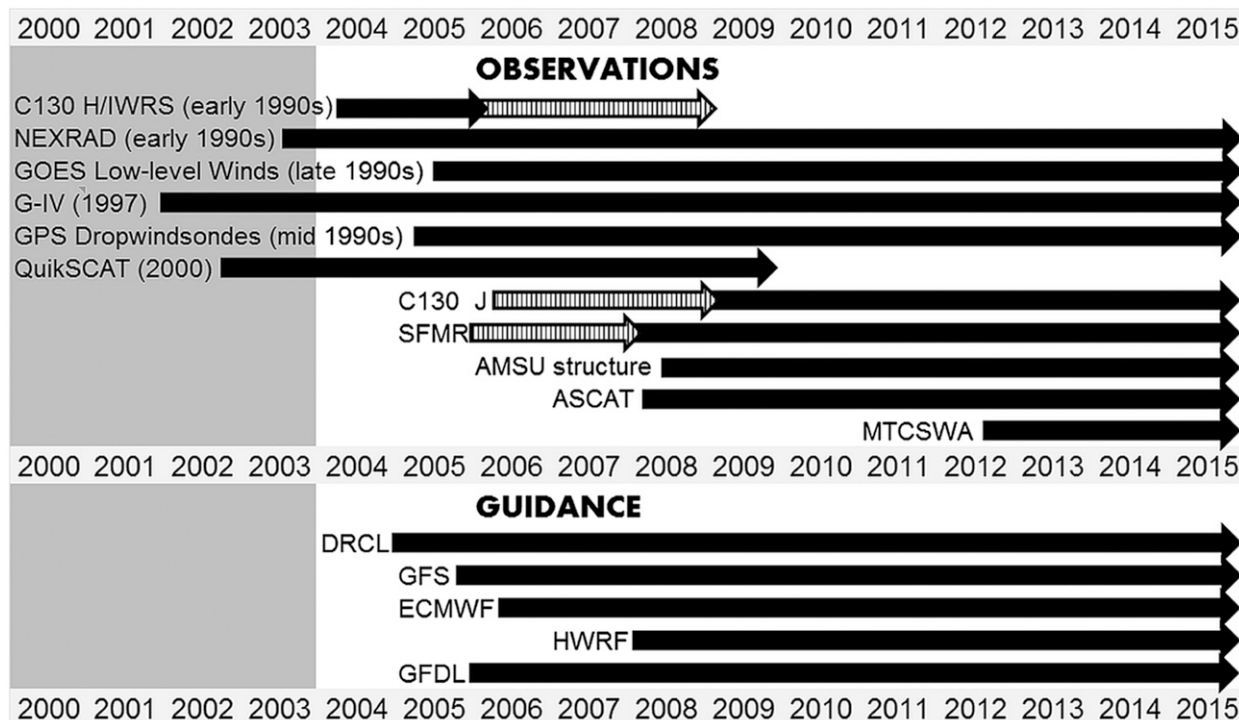E-mail: john.knaff@noaa.gov

FIG. 1. Illustration showing the (top) historical availability of observations to assess wind radii and TC structure and (bottom) forecast guidance available for making wind radii forecasts at NHC. Intermittent availability is indicated by the nonsolid arrows. Best tracking of wind radii began in 2004 and continues to the present (white background). Before that time, wind radii information was not best tracked (gray background). Observations include low-level aircraft reconnaissance (C130 H and C130 J), where IWRS stands for the Improved Weather Reconnaissance System; upper-level aircraft reconnaissance [Gulfstream-IV (G-IV)]; coastal radar (NEXRAD); satellite observations from Geostationary Operational Environmental Satellite (GOES) instruments and AMSU; scatterometry from QuikSCAT (Graf et al. 1998) and Advanced Scatterometer (ASCAT; Gelsthorpe et al. 2000) instruments; Stepped Frequency Microwave Radiometer (SFMR; Uhlhorn et al. 2007) surface wind estimates from aircraft reconnaissance; and the Multisatellite Platform Tropical Cyclone Wind Analysis (MTCSWA; Knaff et al. 2011). Wind radii forecast guidance includes DRCL, the NCEP GFS, the ECMWF, the HWRF, and the GFDL hurricane model.

for Medium-Range Weather Forecasts (ECMWF) model have shown an increased ability to detect trends in overall storm size in recent years as a result of increased resolution. In cases where storms undergo extratropical transition, global models can now be used quantitatively for forecasts of the 34-kt wind field (M. Brennan, NHC, 2015, personal communication).

In 2004, despite the difficulty of estimating wind radii from the available data, NHC began postseason reanalysis (i.e., best tracking) of wind radii, which provides an improved historical record of wind radii observations and allows studies like ours. Prior to that year, wind radii information was only available from the Tropical Cyclone Vitals Database (TC Vitals[1]) used for initializing model guidance (e.g., Tallapragada et al.

2014; Kurihara et al. 1995), a practice that continues today. Like all best-tracking activities, the resulting estimates of wind radii are based on operational practices, available technology, and observations. The errors in those observations are thus a function of the methods, technology, and observations available at the time. Figure 1 shows many of the observational technology changes that have occurred since 2000 as based on operational data archives and information contained in Rappaport et al. (2009). It is important to note that because there are relatively few tools to estimate wind radii, errors associated with the best-track estimates still may be as high as 25%–40% (Knaff and Harper 2010).

Accompanying the 2004 change in the best-tracking procedures was the development and operational implementation of the first purely statistical wind radii climatology and persistence (CLIPER) model [DRCL in the Automated Tropical Cyclone Forecasting System (ATCF; Knaff et al. 2007)]. DRCL has been run operationally at NHC since 2003 and offers a stable baseline forecast to assess skill in other wind radii forecasts. In a seasonal

---

[1] In the Automated Tropical Cyclone Forecast System (Sampson and Schrader 2000), these are the CARQ entries in the aid deck (adeck). TC Vitals have also been referred to as "the bogus" in past literature.

sense, wind radii errors from DRCL can be used to assess seasonal difficulty and normalize wind radii forecast performance for seasonal differences. In addition, the DRCL has been used in the operational Monte Carlo wind speed probability product (DeMaria et al. 2009, 2013) that provides forecasts of the probability of hurricane force, damaging, and gale force winds based on the official forecast, a 5-yr sample of track and intensity errors, track guidance spread, and the climatological errors associated with the variation of wind radii (via DRCL). In addition to the DRCL output, operational TC wind structure guidance has also been provided by global and regional hurricane models via tracker software developed at the Geophysical Fluid Dynamics Laboratory (GFDL) and described in Marchok (2002) and more recently in Tallapragada et al. (2014). Figure 1 also shows when model guidance became available.

Significant TC intensity (DeMaria et al. 2014) and track forecast guidance improvements [Heming and Goerss (2010), and references within] have occurred. However, few attempts have been made to assess and document TC structure forecast skill or improvements. In one such attempt, Knaff et al. (2006) found that NHC 2005 Atlantic gale force wind radii forecast errors were comparable to DRCL beyond 36 hours, that numerical weather prediction–based wind radii guidance was poor, and that if the best-track intensity was known, the DRCL forecast would have improved by 3%–11% over the 72-h forecast period. The latter point suggests that with improved intensity forecasts one would expect both DRCL and the official NHC wind radii forecast to also improve.

With a decade of best-tracked wind radii, this paper will examine the evolution of gale force wind radii prediction in the North Atlantic TC basin, where the observations of TC structure are arguably the most accurate. Official forecasts from NHC will be examined in terms of mean absolute errors, mean error or bias, probability of detection, and probability of false detection versus the performance of an operational baseline model, DRCL, to determine if there has been any improvement in the ability to forecast TC structure. Details of how the verification is conducted and results follow.

2. Data and methods

The verification of the maximum extent of gale force winds R34 is based on postseason/final best-track data, operational forecasts made by NHC (OFCL), and operational forecasts made by DRCL during the period 2004–13. The R34 is estimated and forecasted in Earth-relative quadrants (northeast, southeast, southwest, and northwest) surrounding TCs that have intensities of 34 knots or greater. NHC makes OFCL forecasts of

intensity and track through 120 hours and OFCL forecasts of R34 through 72 hours. The input data for each DRCL forecast are the corresponding OFCL track and intensity forecast and results in DRCL being available for all OFCL forecasts, which ensures a fair baseline forecast comparison can be constructed.

This study will concentrate on R34 verification statistics since R34 is likely best observed/estimated and is available more frequently. The authors are aware of the R34 quality and dependency issues and will attempt to address those issues throughout the manuscript. All the forecast data used in this study are contained in the ATCF (Sampson and Schrader 2000) databases and are freely available from NHC.

To calculate verification statistics, forecast values of R34 in each quadrant and at each forecast lead time are compared to the final best-track values. The occurrence of zero-valued wind radii introduces an added complication of verifying wind radii. The zero-valued wind radii typically occur when storms are near the 34-kt intensity or storm translation speeds are large (i.e., $>8\,\mathrm{m\,s^{-1}}$). For this study, the following verification strategy is adopted. If any of the quadrants in the best-track have nonzero wind radii, all quadrants for that case are verified. That strategy allows the individual quadrant statistics to be combined to form a single measurement of mean absolute error MAE and bias (i.e., the mean error) for each 6-hourly forecast lead time and results in an approximately 20%–25% increase in the number of cases. Since the forecasts of R34 are in units of nautical miles (n mi; 1 n mi = 1.85 km) and of intensity are in units of knots, these units will be used throughout.

To evaluate the ability of forecasts to discriminate the occurrence of R34 and to complement the MAE and bias statistics, the probability of detection POD, probability of false detection POFD, and Peirce skill score PSS (=POD − POFD) are also determined from 2×2 contingency tables.

To keep the verification statistics presented here succinct, this study presents just the combined, all-quadrant statistics. Values of MAE, bias, POD, and POFD are calculated for both OFCL and DRCL forecasts in a homogeneous manner (i.e., they include identical realizations). These basic statistics are then used to calculate the PSS, as well as the percent of forecast improvement relative to DRCL (or skill). Using these results, a trend analysis is performed to determine if OFCL forecasts of R34 have improved over the last decade. The trends will be calculated using the sample sizes of the individual years so that the number of forecasts is accounted for explicitly. Statistical significance in this paper is assessed using a Student's t test assuming one tail and the 95% level. The results of that analysis are presented in the next section.

TABLE 1. Sample sizes (number of verification times), mean best-track intensity (kt), mean of R34 (n mi), and std dev of R34 (n mi) associated with the annual R34 verification subsets.

Year	Sample size			Mean best-track intensity (kt)	R34	
	24 hours	48 hours	72 hours		Mean (n mi)	Std dev (n mi)
2004	301	250	218	68	114	77
2005	428	353	278	59	102	63
2006	194	163	128	54	151	103
2007	111	73	50	55	133	104
2008	292	240	194	60	114	66
2009	81	57	49	57	107	69
2010	292	240	194	60	115	79
2011	258	201	156	54	108	69
2012	333	267	213	53	127	99
2013	105	65	33	43	94	63

3. Results

Since the number of cases, the mean intensity, and the mean and standard deviation of R34 of each year's verification sample are important aspects of the statistical analyses and discussion, they are presented in Table 1. It is noteworthy that the mean intensity varies from 68 knots in 2004 to 43 knots in 2013. The annual variation of seasonal mean intensity is important since more intense TCs tend to have more symmetric features (i.e., fewer missing quadrant values) and generally larger R34. For instance, in our verification sample the mean R34 increases by approximately $1.3\,(n\,mi)\,kt^{-1}$ of intensity with a regression coefficient R of 0.52. Also, since the number of cases will influence the linear trends calculated from MAE and skill, one should note that active years like 2004, 2005, 2008, 2010, and 2012 will receive greater weight than inactive years like 2009 and 2013. The seasonal mean and standard deviation of R34 covary significantly ($R = 0.93$). Furthermore, the seasonal mean intensity and mean R34 are positively correlated ($R = 0.10$), and the annual standard deviation of R34 is negatively correlated ($R = -0.04$), but neither of these relationships is statistically significant.

The first verification statistic examined is the annual time series of MAEs associated with the all-quadrant R34 forecasts. Figure 2 shows those results for both DRCL and OFCL forecasts at lead times of 24, 48, and 72 hours accompanied by the linear trend associated with each time series. These plots show both upward trends in the DRCL errors and downward trends in the OFCL forecasts, both of which are statistically significant at all times. The upward trends in DRCL are thought to be primarily driven by changes in the mean annual intensity, which has been generally falling, as shown in Table 1. In fact, DRCL MAEs decrease with increasing sample intensity with downward trends of -0.25, -0.55, and $-0.79\,(n\,mi)\,kt^{-1}$ explaining 11%,

26%, and 29% of the variance for 24-, 48-, and 72-h forecasts, respectively. The downward trends in OFCL MAEs over time are also statistically significant and are the first suggestion that R34 forecasts have been improving over the last decade.

Figure 3 shows the time series of OFCL and DRCL forecast biases for 24-, 48-, and 72-h forecasts. It appears that the biases of both OFCL and DRCL are correlated from 2004 to 2009, with OFCL biases being generally closer to zero. This is particularly evident for the 72-h forecasts. The relaxation to climatology by DRCL beyond 36 hours is also evident in these biases, where seasons with large R34 in Table 1 (2006, 2007, and 2012) had relatively large negative biases. In and around 2010 and thereafter, OFCL biases are noticeably closer to zero. This suggests that some of the reduction in MAE shown in Fig. 2 is the result of less-biased forecasts of R34. One speculation is that the bias reduction is related to intensity forecasts, which have also been improving (DeMaria et al. 2014).

Another way of investigating improvements in R34 forecasts is to construct skill diagrams, where skill is the percent improvement of MAE with respect to DRCL at each forecast lead time. Skill trends are positive and statistically significant, with annual improvements of 2%, 3%, and 3% for 24-, 48-, and 72-h forecast lead times, respectively. Figure 4 shows the multiyear skill of the OFCL forecasts. Here, we have averaged the two 3-yr periods (2004–06 and 2007–09) and one 4-yr period (2010–13). Multiyear skill averages provide large enough samples to assess statistical significance, adjusted for 30-h serial correlation,[2] and clearly illustrates

---

[2] The effective sample size used for the Student's t test is estimated to be the number of 30-h samples contained in the dataset, which was described as the time between effectively independent samples (Leith 1973).
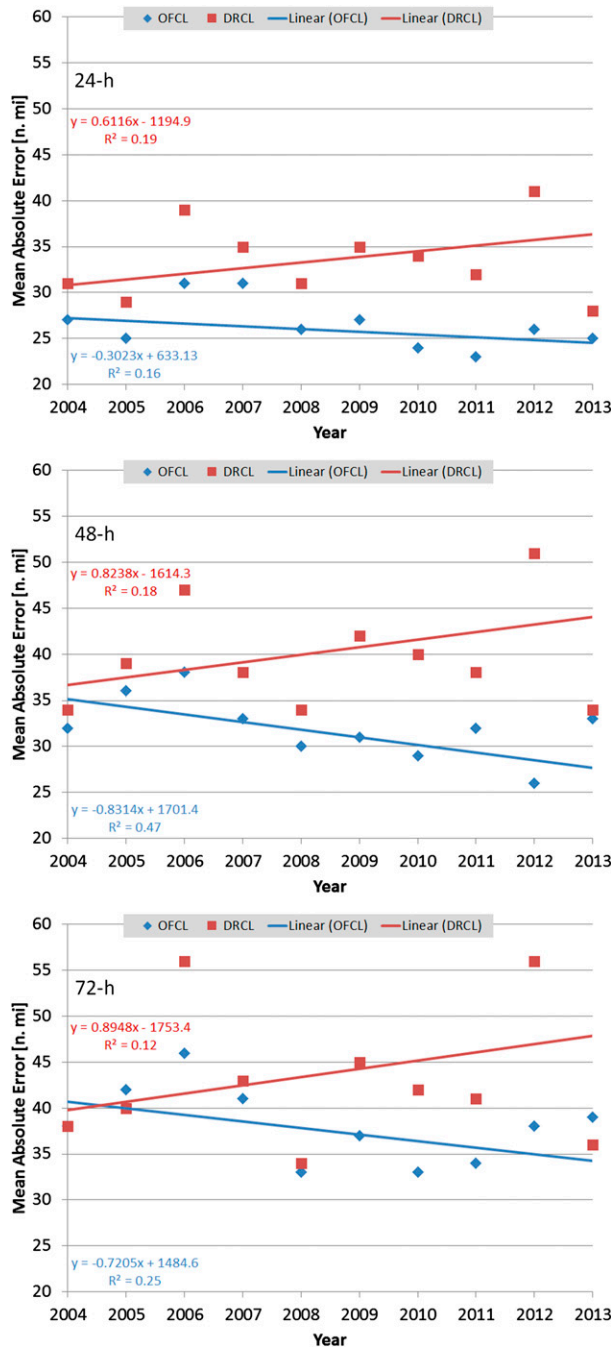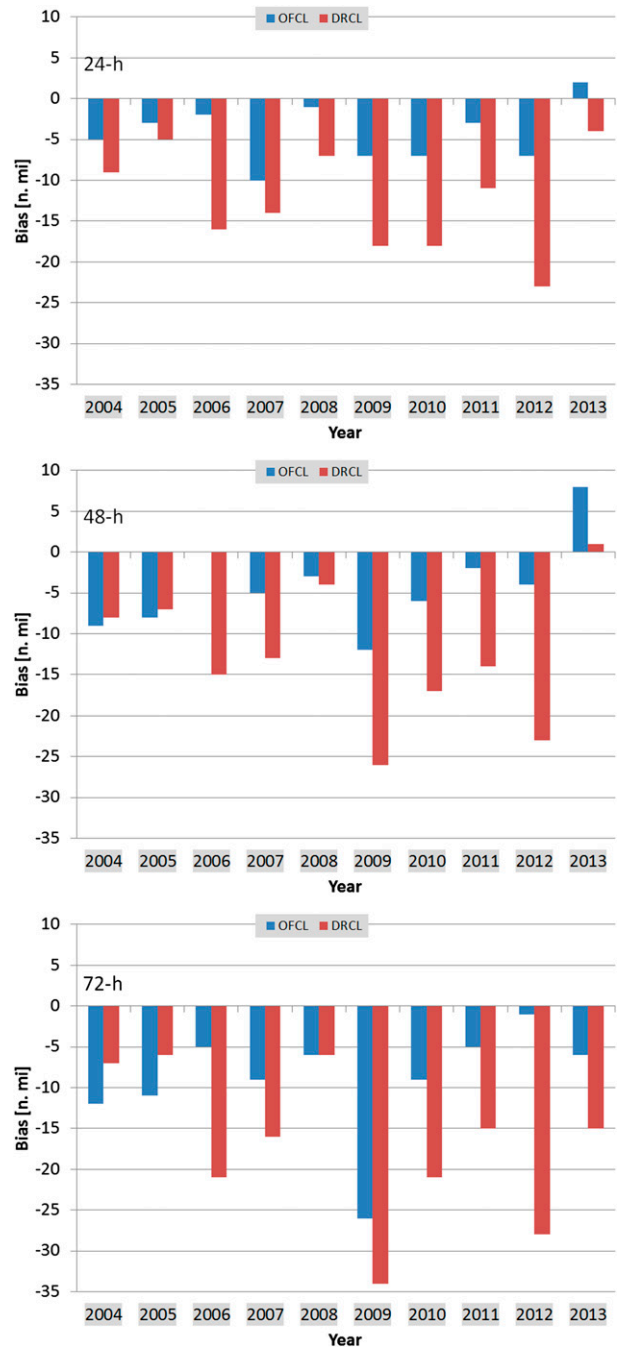
FIG. 2. Time series of annual R34 forecast MAEs associated with OFCL (blue) and DRCL (red) forecasts for lead times of (top) 24, (middle) 48, and (bottom) 72 hours. Along with the annual MAEs, linear trends for each model have been calculated based on the MAEs and the numbers of cases (see Table 1). Trend equations and R^2 statistics are provided for each trend.

FIG. 3. Time series of annual R34 forecast biases associated with OFCL (blue) and DRCL (red) forecasts for lead times of (top) 24, (middle) 48, and (bottom) 72 hours.

that the most significant improvements occurred during the 2010–13 time period. Much like the results presented in Knaff et al. (2006), the skill of the OFCL forecasts is statistically significant (larger markers) through 24

hours in the first two time periods, suggesting little or no skill improvement during 2004–09. However, the 2010–13 OFCL forecast skill is both larger in magnitude at all time periods and the statistical significance extends through the 72-h forecast lead time.

To complete the evaluation, we investigate the probability of detection and false alarms via the PSS. The
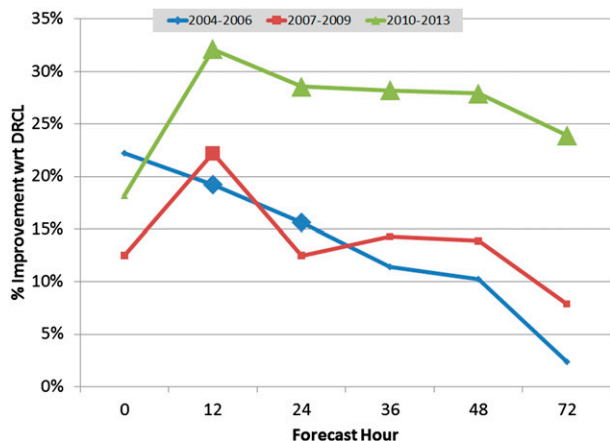
FIG. 4. Percent improvement of MAEs with respect to DRCL forecasts for the periods 2004–06 (blue), 2007–09 (red), and 2010–13 (green). Statistical significance, accounting for 30-h serial correlation, is indicated by the larger line markers.

PSS defines the accuracy of the forecast in predicting the correct category, relative to that of random chance. The PSS ranges from -100% to 100%, and 0% indicates no skill while 100% indicates perfect skill. In this case, the category is the existence of nonzero R34 in the various quadrants. Figure 5 shows the time series of PSS for OFCL and DRCL for the 24-, 48-, and 72-h forecast lead times. Trend lines, again weighted for the number of cases, are provided for each forecast lead time and model. Both OFCL and DRCL have skill at 24, 48, and 72 hours based on this statistic, but the year-to-year variations are quite large. The only significant trends are DRCL (downward) at 24 hours, and both DRCL and OFCL (upward) at 72 hours. The 24-h downward trend of DRCL is likely related to the mean intensity of the seasons, whereas the improved PSSs at 72 hours are likely related to the decreasing MAEs of the OFCL intensity forecast at 72 hours (Cangialosi and Franklin 2014).

One can also compare the PSSs of the OFCL and DRCL by constructing the percent improvement of OFCL PSS relative to DRCL. While this is an uncommon statistical approach, in this case it offers some additional insight into the performance of the OFCL R34 forecasts over the last decade. Figure 6 shows the percent improvement of the OFCL PSSs relative to the DRCL PSSs for the 24-, 48-, and 72-h forecast lead times along with the associated linear trends. It shows that all lead times have statistically significant upward trends. However, the trends suggest that the OFCL forecast has only recently shown improvement over DRCL in its ability to forecast the occurrence of nonzero R34 values. The crossover to a skillful depiction of nonzero R34 values occurs in 2007, 2011, and 2012, for the 24-, 48-, and 72-h forecasts. These statistics offer additional
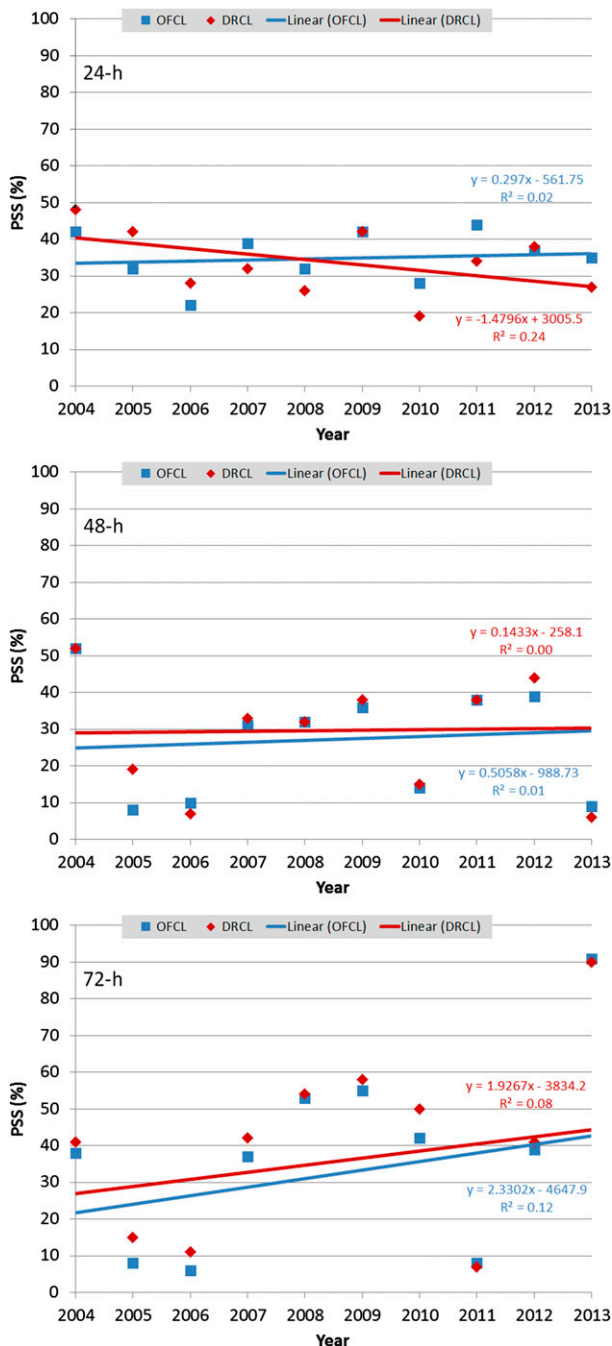

FIG. 5. Time series of annual R34 PSS associated with OFCL (blue) and DRCL (red) forecasts for lead times of (top) 24, (middle) 48, and (bottom) 72 hours. PSS temporal trends for each model have been calculated based on the PSSs and the numbers of cases (see Table 1) and are provided by the blue and red lines. Equations and corresponding R^2 statistics are provided for each trend.

evidence that the OFCL R34 forecasts have been steadily improving relative to the R34 in the best tracks and have recently outperformed the purely statistical forecasts of DRCL. Also, the DRCL R34 forecasts tend
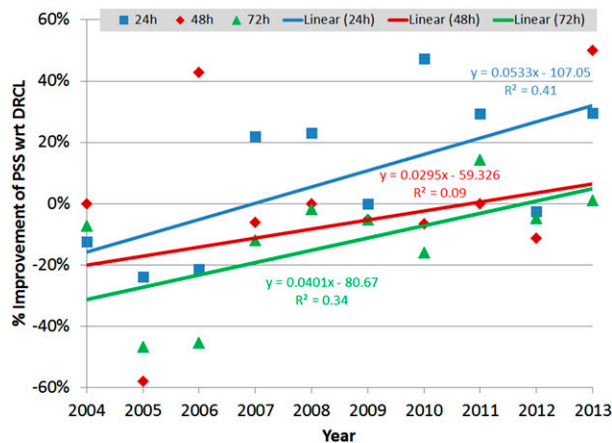
FIG. 6. Time series of annual percentage improvements in the PSSs associated with OFCL and DRCL forecasts for lead times of 24 (blue), 48 (red), and 72 (green) hours are provided as points. Linear trends for each forecast time have been calculated based on the MAEs and the numbers of cases (see Table 1) and are shown as blue, red, and green lines for 24, 48, and 72 hours, respectively. Equations and R^2 statistics are provided for each trend.

to become more symmetric with forecast time, which is not necessarily realistic. NHC wind radii look more realistic in the longer forecast leads, at least anecdotally.

4. Conclusions, discussion, and recommendations

The results presented indicate that NHC has reached a point where its 72-h gale force wind radii forecasts are generally better than DRCL or skillful. We recognize that there is some debate about whether the gale force wind radii in the best tracks can serve as ground truth for forecast evaluation because of concerns over sparse, intermittent, and poor quality observations. In an independent study that made use of only the highest quality best-track data (coincident with aircraft reconnaissance) during 2008–12 to address the shortcomings in best-track wind radii estimation, it was found that NHC/OFCL wind radii average forecast errors increased with forecast time, but were skillful (J. Cangialosi and C. Landsea 2015, personal communication; Cangialosi and Landsea 2015, manuscript submitted to *Wea. Forecasting*, hereafter CL15). We have also conducted experiments that introduce random errors (maximum of 40%) to the best-track ground truth, but doing so does not alter our conclusions. Finally, there is concern that the gale force wind radii best tracks are dependent on the forecasts. We think that is a valid concern. However, this effect likely fades as the forecast lead increases since wind radii persistence has an *e*-folding time of roughly 32 hours (Knaff et al. 2007). In this study, we also explicitly account for 30-h serial

correlation in the best tracks and forecasts. Even with these reduced sample sizes, our conclusions hold. Regardless of these issues, one can, at minimum, state that NHC gale force wind radii forecasts have become more representative of the gale force wind radii in the best tracks to a point where they appear to be more representative than DRCL forecasts.

These results suggest that significant progress is being made in the ability to forecast gale force wind radii. This makes the OFCL forecasts more beneficial to products (and users) that require wind radii information. Since DRCL is available for all OFCL forecast times, it is used as a substitute for OFCL wind radii forecasts in some applications (e.g., DeMaria et al. 2009, 2013), while other applications (e.g., Sampson et al. 2010) extrapolate the existing OFCL forecast wind radii to 120 hours. Neither approach seems optimal given the advances in skill of NHC OFCL gale force wind radii (and possibly of the other wind radii as well) reported here. Rather it appears that these algorithms would be better served by leveraging advances in NHC wind radii forecasts than the statistical proxies currently employed.

We speculate that two drivers may be responsible for improvements in the OFCL gale force wind forecasts. The first is improvements in model forecasts. A cursory examination of gale force wind radii forecasts from several numerical weather prediction systems suggests that those forecasts were also skillful during the 2011–13 time period, mirroring similar findings from CL15 and from J. Cangialosi and C. Landsea (2015, personal communication). This certainly was not the case in 2005 (Knaff et al. 2006). Findings like these have led to a study discussing the development of skillful multimodel gale force wind radii forecasting methods (Sampson and Knaff 2015, manuscript submitted to *Wea. Forecasting*). The second driver of these improvements is related to the improvement in intensity guidance documented in DeMaria et al. (2014). Intensity improvements impact the yes/no detection of gale force winds as well as fundamentally determining their extent and symmetry [e.g., as in Knaff et al. (2006, 2007)].

Finally, the progress indicated by this study's results would simply not be possible without the decade-long history of wind radii contained in the best track. Despite the purported issues with these measurements, they are invaluable for product development and validation. We therefore would encourage other forecast centers to produce postseason analyses of wind radii, noting that a similar recommendation recently came out of the World Meteorological Organization's Eighth International Workshop on Tropical Cyclones hosted by South Korea.

REFERENCES

Cangialosi, J. P., and J. L. Franklin, 2014: 2013 National Hurricane Center forecast verification report. NOAA/NHC, 84 pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2013.pdf.]

DeMaria, M., J. A. Knaff, R. Knabb, C. Lauer, C. R. Sampson, and R. T. DeMaria, 2009: A new method for estimating tropical cyclone wind speed probabilities. *Wea. Forecasting,* **24,** 1573–1591, doi:10.1175/2009WAF2222286.1.

——, and Coauthors, 2013: Improvements to the operational tropical cyclone wind speed probability model. *Wea. Forecasting,* **28,** 586–602, doi:10.1175/WAF-D-12-00116.1.

——, C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.,* **95,** 387–398, doi:10.1175/BAMS-D-12-00240.1.

Gelsthorpe, R. V., E. Schied, and J. J. W. Wilson, 2000: ASCAT—Metop's Advanced Scatterometer. *ESA Bull.,* Vol. 102, European Space Agency, Frascati, Italy, 19–27. [Available online at http://www.esa.int/esapub/bulletin/bullet102/Gelsthorpe102.pdf.]

Graf, J. E., W.-Y. Tsi, and L. Jones, 1998: Overview of QuikSCAT mission—A quick deployment of a high resolution, wide swath scanning scatterometer for ocean wind measurement. *Proc. IEEE Southeastcon '98,* Orlando, FL, IEEE, 314–317, doi:10.1109/SECON.1998.673359.

Heming, J., and J. Goerss, 2010: Track and structure forecasts of tropical cyclones. *Global Perspectives on Tropical Cyclones,* J. C.-L. Chan and J. D. Kepert, Eds., World Scientific Series on Asia-Pacific Weather and Climate, Vol. 4, World Scientific, 287–323, doi:10.1142/9789814293488_0010.

Knaff, J. A., and B. A. Harper, 2010: Tropical cyclone surface wind structure and wind–pressure relationships. *Proc. WMO Int. Workshop on Tropical Cyclones—VII,* La Reunion, France, WMO, KN1.1–KN1.35. [Available online at http://www.wmo.int/pages/prog/arep/wwrp/tmr/otherfileformats/documents/KN1.pdf.]

——, C. Guard, J. Kossin, T. Marchok, C. Sampson, T. Smith, and N. Surgi, 2006: Operational guidance and skill in forecasting structure change. *Proc. WMO Int. Workshop on Tropical Cyclones—VI,* San Juan, Costa Rica, WMO, 160–184. [Available online at http://severe.worldweather.org/iwtc/document/Topic_1_5_John_Knaff.pdf.]

——, C. R. Sampson, M. DeMaria, T. P. Marchok, J. M. Gross, and C. J. McAdie, 2007: Statistical tropical cyclone wind radii prediction using climatology and persistence. *Wea. Forecasting,* **22,** 781–791, doi:10.1175/WAF1026.1.

——, M. DeMaria, D. A. Molenar, C. R. Sampson, and M. G. Seybold, 2011: An automated, objective, multi-satellite-platform tropical cyclone surface wind analysis. *J. Appl. Meteor. Climatol.,* **50,** 2149–2166, doi:10.1175/2011JAMC2673.1.

Kurihara, Y., M. A. Bender, R. T. Tuleya, and R. J. Ross, 1995: Improvements in the GFDL Hurricane Prediction System. *Mon. Wea. Rev.,* **123,** 2791–2801, doi:10.1175/1520-0493(1995)123<2791:IITGHP>2.0.CO;2.

Leith, C. E., 1973: The standard error of time-average estimates of climate means. *J. Appl. Meteor.,* **12,** 1066–1069, doi:10.1175/1520-0450(1973)012<1066:TSEOTA>2.0.CO;2.

Marchok, T. P., 2002: How the NCEP tropical cyclone tracker works. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology,* San Diego, CA, Amer. Meteor. Soc., P1.13. [Available online at https://ams.confex.com/ams/pdfpapers/37628.pdf.]

Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting,* **24,** 395–419, doi:10.1175/2008WAF2222128.1.

Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (version 3.2). *Bull. Amer. Meteor. Soc.,* **81,** 1231–1240, doi:10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2.

——, P. A. Wittmann, and H. L. Tolman, 2010: Consistent tropical cyclone wind and wave forecasts for the U.S. Navy. *Wea. Forecasting,* **25,** 1293–1306, doi:10.1175/2010WAF2222376.1.

Tallapragada, V., and Coauthors, 2014: Hurricane Weather Research and Forecasting (HWRF) Model: 2014 scientific documentation. Developmental Testbed Center, 105 pp. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.6a_ScientificDoc.pdf.]

Uhlhorn, E. W., P. G. Black, J. L. Franklin, M. Goodberlet, J. Carswell, and A. S. Goldstein, 2007: Hurricane surface wind measurements from an operational stepped frequency microwave radiometer. *Mon. Wea. Rev.,* **135,** 3070–3085, doi:10.1175/MWR3454.1.