The Prediction of Northern Hemisphere Tropical Cyclone Extended Life Cycles by the ECMWF Ensemble and Deterministic Prediction Systems. Part I: Tropical Cyclone Stage*

K. I. HODGES AND R. EMERTON

Department of Meteorology, University of Reading, Reading, United Kingdom

(Manuscript received 22 November 2013, in final form 5 June 2015)

ABSTRACT

This study has explored the prediction errors of tropical cyclones (TCs) in the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS) for the Northern Hemisphere summer period for five recent years. Results for the EPS are contrasted with those for the higher-resolution deterministic forecasts. Various metrics of location and intensity errors are considered and contrasted for verification based on IBTrACS and the numerical weather prediction (NWP) analysis (NWPa). Motivated by the aim of exploring extended TC life cycles, location and intensity measures are introduced based on lower-tropospheric vorticity, which is contrasted with traditional verification metrics. Results show that location errors are almost identical when verified against IBTrACS or the NWPa. However, intensity in the form of the mean sea level pressure (MSLP) minima and 10-m wind speed maxima is significantly underpredicted relative to IBTrACS. Using the NWPa for verification results in much better consistency between the different intensity error metrics and indicates that the lower-tropospheric vorticity provides a good indication of vortex strength, with error results showing similar relationships to those based on MSLP and 10-m wind speeds for the different forecast types. The interannual variation in forecast errors are discussed in relation to changes in the forecast and NWPa system and variations in forecast errors between different ocean basins are discussed in terms of the propagation characteristics of the TCs.

1. Introduction

Tropical cyclones (TCs) develop over warm tropical waters from weak incipient disturbances, such as easterly waves (Chen et al. 2008; Kerns and Zipser 2009; Thorncroft and Hodges 2001). In many cases they move poleward where they can undergo extratropical transition (Jones et al. 2003). Previous studies of TCs in numerical weather prediction (NWP) have mainly focused on the operationally defined parts of the life cycle

(Aberson 2008; Majumdar and Finocchio 2010), primarily because evaluation of forecast errors has been against best track data. This paper considers alternative approaches to forecast verification of TCs, based on the NWP analyses (NWPa), as a precursor to studying forecast errors of the extended TC life cycles.

Operational forecasting centers have developed sophisticated global forecasting systems, initially as deterministic systems, which typically produce a single medium-range forecast of 5–10 days. More detailed shorter-range forecasts are often produced by higher-resolution limited area models. The global deterministic forecasts are often used to predict TC paths but are less often used to predict intensities because of insufficient model resolution to capture the processes important for the development of TCs.

Limits of predictability imposed by the chaotic nature of the atmosphere have motivated the development of ensemble prediction systems (EPSs) to provide probability distributions of forecast states. EPSs start a set of forecasts from slightly different, but valid, initial states to sample the initial state errors (Froude et al. 2013).

Denotes Open Access content.

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/MWR-D-13-00385.s1>.

Corresponding author address: Kevin Hodges, Department of Meteorology, University of Reading, Harry Pitt Building, Whiteknights, Reading RG6 6AL, United Kingdom.
E-mail: k.i.hodges@reading.ac.uk

DOI: 10.1175/MWR-D-13-00385.1

Additionally, the forecast model can also be perturbed throughout the forecast to sample the forecast model uncertainty. EPSs have advantages over a deterministic system in that they provide a measure of forecast uncertainty and can provide guidance on the likelihood of extreme events earlier than a single forecast (Buizza and Hollingsworth 2002; Froude et al. 2007b). The ensemble mean forecast, though smoother, can produce lower errors than a single forecast (using the same model configuration), since the averaging process removes the less predictable spatial scales, retaining the more predictable scales. In a perfect ensemble, the spread should be realistic and sufficient to cover all possible outcomes (Leutbecher and Palmer 2008), and ideally the spread should be able to predict the forecast error (Hopson 2014). An underdispersive spread indicates that the ensemble does not fully represent all possible forecast states and may not provide an accurate forecast, as the true atmospheric state may be outside the ensemble range. The spread depends on the perturbation method and the number of ensemble members. EPSs are generally run at lower resolutions than deterministic systems owing to the additional computational requirements; however, both are integral parts of a modern NWP system.

It is important to assess the potential of NWP systems to predict TCs in order to improve them. Forecast verification is complex, requiring an understanding of the uncertainties in the verification data, and encompasses a number of approaches. How forecasts are used informs the verification information required; this can be different for a forecaster wanting information on forecast accuracy of surface properties and a model developer wanting a broader range of information on model performance (WMO 2013). For TCs, the oft-used approach is to use observations directly or to use the postseason reanalyzed “best tracks.” However, in other studies a more common approach is to use both observations and the NWPa (Pinson and Hagedorn 2012). Global NWPa are generally not used for verification of TCs because current spatial resolutions make them incapable of correctly simulating TC wind intensities, though the latest systems can overpredict the depth (see example). Operational verification of TCs relies on observations from satellites, aircraft reconnaissance, and forecaster interpretation, which all have associated uncertainties. In addition, while some operational centers produce information on the pre- and post-TC stages, this information is not always in a form that allows easy and automatic quantitative verification of these life cycle stages to be performed [e.g., the text-based reports from the National Hurricane Center (NHC)], nor is it produced consistently for different ocean basins. The data

most often used for verification are the postseason best track data, such as IBTrACS (Knapp et al. 2010). IBTrACS provides a global dataset of observed TCs produced from the operationally analyzed observations. The information on the pre- and post-TC stages, if available, is not currently incorporated into the best track data, which generally only cover the parts of the life cycle declared as a TC, though some extratropical transition information may be included. Uncertainties in these data are discussed in the appendix. Because of the limitations of the life cycles in best track data and of traditional verification approaches, it is important to consider alternatives, in particular if the extended life cycles of TCs, which include the pre- and post-TC stages, are to be considered. It is also important more generally to provide a broader understanding and quantification of forecast error (Vukicevic et al. 2014; WMO 2013).

This study uses the NWPa to verify TC location and a lower-tropospheric measure of vortex intensity as a means of providing further insight into forecast errors and the relationship between different forecast types. This is motivated by wanting to contrast forecast errors over the extended life cycles of TCs, including the pre- and post-TC stages, which are presented in a following paper. To this end, new metrics are introduced that are suitable for verifying the different stages of the extended life cycles. The focus of this paper is to assess the new proposed metrics and how they relate to verification based on traditional methods and best tracks. The focus of the paper is TC forecast error in the Northern Hemisphere (NH), using the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble and deterministic forecasts.

The aims of this paper are as follows:

- (i) to evaluate and compare different metrics for forecast errors of TCs,
- (ii) to determine and contrast forecast errors of TCs for the ECMWF deterministic and ensemble forecasting systems,
- (iii) to determine the interannual variability of the forecast errors of TCs and their relation to changes in the forecast system,
- (iv) to determine the geographical differences in forecast errors of TCs, and
- (v) to determine the error–spread relationship of the ensemble for the preceding cases.

The paper proceeds with a description of the data and methodology, followed by an example to illustrate the methodology, then by the results and their interpretation, and finally the conclusions.

TABLE 1. Summary of changes to the ECMWF IFS over the period of 2008–12. A full description of model changes can be found at <http://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model>.

Date	Model cycle	Change
8 Sep 2009	35r3	Assimilation system extended to estimate and correct for systematic model errors in the stratosphere. First implementation of weak-constraint 4D-Var.
		Revised stochastic physics.
26 Jan 2010	36r1	Deterministic model horizontal resolution increased from 25 to 16 km.
		EPS horizontal resolution increased from 50 to 32 km for the first 10 days and from 80 to 63 km for days 11–15.
24 Jun 2010	36r2	New method for providing initial-time perturbations for the EPS.
		Differences between members of an EDA are used instead of the evolved singular vectors to create initial spread between EPS forecast members.
9 Nov 2010	36r4	New cloud parameterization scheme.
		Revision of stochastically perturbed physical tendencies.
		Introduction of spectral stochastic backscatter scheme.
		Retuned initial perturbation amplitudes.
15 Nov 2011	37r3	Modification of the entrainment/detrainment of convection.
19 Jun 2012	38r1	Updated background-error covariance statistics using EDA, providing updated error statistics for 4D-Var that are more consistent with the current forecast model and observing systems.
		Modified convective downdraft entrainment, improving surface wind speeds in tropics.
		Improved dealiasing of the pressure gradient term, reducing numerical noise and horizontal diffusion in the forecast.
		EDA perturbations redefined using the ensemble mean instead of control as the reference.
		Variance of the singular vector initial perturbation increased to compensate for reduction in spread caused by change of the EDA perturbations.

2. Data and methods

a. Data

This study uses forecasts, produced 12 hourly, from the ECMWF deterministic and ensemble systems. The periods considered are May–October (MJJASO) for the years 2008–12. This results in fewer changes in the forecast system but provides sufficient TCs to produce robust statistics. A total of 1840 forecasts are considered. The deterministic and ensemble systems are constituents of the ECMWF Integrated Forecast System (IFS), which for this period used model cycles 32r3 to 38r1 (cf. Table 1). The IFS model consists of a semi-Lagrangian spectral dynamical model. The NWPa are produced by assimilating a wide range of observations, using incremental four-dimensional variational data assimilation (4D-Var) (Courtier et al. 1994). For the deterministic system the horizontal resolution changed in January 2010 from T799 (25 km) to T1279 (16 km); there are 91 sigma vertical levels and the forecasts are integrated to 10 days. The EPS horizontal resolution changed in January 2010 from T399 (50 km) to T639 (30 km) for the first 10 days of the forecast and from T255 (80 km) to T319 (60 km) for the last 5 days (Table 1); there are 62 sigma levels in the vertical. A control forecast is run from the optimal initial state at the same resolution as the perturbed ensemble. The perturbations are produced from the leading singular vectors (Ehrendorfer and Tribbia 1997) for the extratropics with extra singular vectors computed specifically

for TCs (Puri et al. 2001). Since June 2010 the singular vectors have been supplemented by perturbations from an ensemble of data assimilations (EDA) (Bonavita et al. 2011). To sample the model uncertainty, two schemes of stochastic physics are used: perturbed parameterization tendencies (Buizza et al. 1999) and kinetic energy backscatter (Shutts et al. 2011). System changes over the study period are summarized in Table 1.

To explore the prediction error, IBTrACS (Knapp et al. 2010) and ECMWF NWPa are both used for verification. Verification against the best tracks is of course very important, in particular for impacts, where low-level wind speeds are the primary interest. However, this is not without difficulties related to uncertainties in the observations that are incorporated in IBTrACS (Knapp et al. 2013; Landsea and Franklin 2013; Torn and Snyder 2012) and their emphasis on the main TC stage. The NWPa will also have uncertainties. When using such data for verification it is important to understand their uncertainties, which are discussed further in the appendix.

b. Tracking

The 6-hourly relative vorticity on the 850-, 700-, and 600-hPa pressure levels is used for the tracking, though the 600-hPa level is not available for the EPS. Data are extracted from the ECMWF archive for the full forecast periods. The NWPa are extracted for the May–November period to provide sufficient data for the verification of forecasts started at the end of October. The

cyclone tracking algorithm (Hodges 1994, 1995, 1999) has been used in numerous studies of TCs (Bengtsson et al. 2007b; Manganello et al. 2012; Strachan et al. 2013) as well as the work of Froude et al. (2007b) and Froude (2010), who performed studies similar to that performed here but focused on extratropical cyclones. It has also been used to study tropical easterly waves (EWs) (Hopsch et al. 2007; Serra et al. 2010). The algorithm identifies more of the TC life cycle than normally considered, including the EW and postextratropical transition stages. Previous studies with the algorithm used only 850-hPa relative vorticity, but Serra et al. (2010) showed that using the vertical average of vorticity between 850 and 600 hPa produced more coherent tracks. This approach is used here to produce tracks of TCs significantly longer than those in IBTrACS (Knapp et al. 2010). While not essential for this part of the study, which only focuses on the more restrictive TC stage, the method described is important for the Part II paper (Hodges and Emerton 2015, manuscript submitted to *Mon. Wea. Rev.*, hereafter Part II), which explores how forecast error depends on different stages of TC life cycle, including the pre- and post-TC stages.

The vertically averaged vorticity is spectrally filtered to reduce high-frequency spatial noise prior to applying the tracking. This entails truncation to T63 and tapering the spectral coefficients (Sardeshmukh and Hoskins 1984). Additionally, the large-scale background is removed for total wavenumbers ≤ 5 . Initial identification is performed by applying a vorticity threshold of $5 \times 10^{-6} \text{ s}^{-1}$ and identifying the gridpoint maxima. These are used to determine the off-grid vorticity maxima locations (Hodges 1995), which provide a better approximation for location and smoother tracks (see the appendix). The maxima are initially organized into sets of tracks using a nearest neighbor scheme. These are refined to produce an optimal set of tracks by minimizing a cost function for track smoothness, subject to adaptive constraints on the time-step displacement and track smoothness (Hodges 1999). Tracks that last longer than 2 days are retained for further analysis. The tracking is applied to both the NWPa and forecasts. Note that the tracking makes no distinction between TCs and other cyclonic systems; how the TCs are identified is described in section 2c.

To provide verification against IBTrACS, the maximum 10-m wind speed and minimum pressure are added to the NWPa and forecast tracks. This is done for pressure by referencing the tracks to the MSLP data (Bengtsson et al. 2009), which finds the nearest pressure minima within a 5° radius (geodesic) of the vorticity center. For the wind speeds a direct search of gridpoint values within 6° of the vorticity center is made.

TABLE 2. Number of TCs identified in the NWPa that match those in IBTrACS by ocean basin, category, and year over the period of May–November 2008–12.

Basin	All years	2008	2009	2010	2011	2012
All	315	69	67	55	62	62
Atlantic	72	14	9	19	15	15
Eastern Pacific	81	19	20	12	14	16
Western Pacific	111	25	27	15	20	24
Northern Indian	32	9	5	7	8	3
TS (<category 1)	183	43	43	31	37	29
Category 1	47	10	8	8	8	13
Category 2	26	5	5	7	5	4
Category 3	27	5	6	3	5	8
Categories 4 and 5	32	6	5	6	7	8
IBTrACS	319	69	69	56	63	62

c. Verification

In studies of TCs in climate model data and re-analyses, objective criteria are often applied to identify TC-like systems (Bengtsson et al. 1995, 2007a). Which systems are identified as TCs and how many are identified can be sensitive to these criteria. If identification is performed during the tracking, typically only the TC stage is obtained, whereas if identification is performed, post-tracking more of the life cycle can be obtained. To avoid these issues, to focus on TCs that are observed, and to obtain the extended life cycles, the NWPa tracks are used to identify the TCs in the forecasts using track matching. The NWPa TCs are also used for the verification to contrast with using IBTrACS and justify their use for the following study looking at different life cycle stages.

The TCs in IBTrACS are identified in the NWPa tracks using track matching (Hodges et al. 2011). Tracks match if they overlap in time by any amount to account for the disparate lifetimes of the NWPa and IBTrACS tracks, and their mean separation is less than 4° (geodesic) and are the track pair with minimum mean separation. A summary of the numbers of identified NWPa TCs compared with those in IBTrACS is given in Table 2. Nearly every track for the chosen years and season in IBTrACS is identified in the NWPa, with those missing being short-lived, weak systems.

To identify the identically same TC in the forecasts, the forecast tracks are matched against the NWPa tracks (Froude et al. 2007a). A forecast track is matched if the spatial separation of its first four points is less than or equal to 4° and is the track with the smallest separation for the four points. Only forecast tracks that have their first point within the first 3 days of the forecast start are considered to exclude matches by chance. The ensemble mean tracks are computed by averaging the ensemble tracks, for location and intensity (however measured), for each TC, as long as a minimum of five members is

present for each lead time step. The ensemble spread is also computed for location and intensity.

In this paper, only the TC stages, as defined by IBTrACS, are considered. This entails restricting the NWPa tracks, for verification, to the IBTrACS time periods. Both nonhomogeneous and homogeneous samples are considered. Nonhomogeneous samples use all the available data, while homogeneous samples use only track information from different forecasts over identical time periods, and which have the same track in each forecast. Results obtained for both nonhomogeneous and homogeneous samples for the combined years are found to be rather similar (see online supplemental material). The imposition of homogeneity has the result of reducing the sample sizes with lead time, though for the multiyear data the sample sizes are still large enough to produce stable statistics throughout the forecast range, though restricted to the deterministic forecast period of 10 days (see online supplemental material). If smaller sample sizes are used—for example, by considering individual years—then some differences do occur between homogeneous and nonhomogeneous results, mainly at the longer lead times, though they are not sufficient to change any conclusions and, in fact, the nonhomogeneous samples appear to produce more stable statistics at longer lead times (see results for 2008 in online supplemental material). For this reason only the nonhomogeneous samples are used here, since they provide greater statistical stability over more of the forecast range.

Statistics are computed for location error, absolute intensity error, intensity bias, and the ensemble spread for location and intensity. Forecast location errors based on using IBTrACS are contrasted with those based on using the NWPa MSLP minima and lower-tropospheric T63 vorticity locations (LVL) from the tracking. Intensity errors depend on the intensity measure. Operationally, TC intensity verification focusses on the low-level wind speeds and occasionally the minimum pressure; these are also considered here based on IBTrACS. However, one of the main focuses of the study is to explore measures of vortex intensity error that can be used to analyze the extended life cycles, hence the use of lower-tropospheric T63 vorticity intensity (LVI), obtained from the tracking. The LVI can be considered as a measure of lower-tropospheric vortex strength. The LVI results are contrasted with the intensity measures based on IBTrACS and the NWPa MSLP and 10-m wind speeds. The IBTrACS wind speeds are converted to 1-min sustained values where necessary using the standard conversion factor (see [appendix](#)).

The 95% confidence intervals (CIs) are also computed from the standard errors for each statistic. When calculating

the CIs, serial correlation of the forecasts should be considered. To address this issue the approach described in [Aberson and DeMaria \(1994\)](#) is used. This is based on the “runs above and below the mean” method ([von Storch and Zwiers 1999](#)). Following [Aberson and DeMaria \(1994\)](#) the “separation time,” t_s , and the “effective sample size,” N_e , are computed. The t_s and N_e are related by $t_s = t(N/N_e)$, where t is the separation time between forecasts (12 h) and N is the actual number of samples. The standard error is computed as $\sigma/\sqrt{N_e}$, where σ is the sample standard deviation. This accounts for serial correlation when computing the CIs and, since N_e is invariably smaller than N , results in wider CIs. The t_s and N_e have been computed for each statistic, forecast type, and field used. In general the location errors have lower values of t_s than the intensity errors or biases. The variation in t_s with forecast type and statistic makes using a subjectively chosen fixed value to infer N_e ([Aberson and DeMaria 1994](#)) problematic, so here the computed values are used to compute the CIs. The actual impact of serial correlation on the CIs is found to be relatively small, possibly because of the large sample sizes used and because the largest values of t_s occur for the earlier lead times where the standard deviations are smallest. Some examples of the variation of t_s are shown in the online supplemental material.

3. Results

a. Example

The methodology is illustrated by applying it to a study of the prediction errors of Hurricane Sandy by the ECMWF system. Sandy caused 147 deaths and losses amounting to \$5 billion (U.S. dollars) ([Blake et al. 2013](#)) when it came ashore in New York in October 2012. Sandy was particularly well sampled by aircraft reconnaissance ([Blake et al. 2013](#)) with 24 missions flown into the storm. [Magnusson et al. \(2014\)](#) highlighted how well the ECMWF deterministic system was able to predict the landfall of this storm and the dependence on the resolution of the forecast model. EPS and deterministic forecast results, together with the NWPa track, are shown in [Fig. 1](#). The NWPa track starts as a weak EW, east of the Caribbean, moving into the Caribbean where it undergoes a rapid change in direction before moving north and intensifying. This accords well with the poststorm report from the National Hurricane Center ([Blake et al. 2013](#)). The time of the abrupt change in direction (22 October) is effectively where the best track starts, highlighting the potential of the method to identify the earlier stages of the TC life cycle.

Considering a forecast initiated close to the start of the best track (1200 UTC 23 October 2012) ([Fig. 1](#)), it is

FIG. 1. Hurricane Sandy forecast results for forecast started 1200 UTC 23 Oct 2012: (a) EPS and deterministic tracks; (b) MSLP intensity histories (hPa); (c) 10-m wind speed intensity histories (m s^{-1}); and (d) LVI histories ($\times 10^{-5} \text{ s}^{-1}$). In (a)–(d) the blue lines are the tracks, intensity, and LVI histories of the individual ensemble members, black lines for the NWPa, green for the ensemble mean, red for the deterministic forecast, and magenta for the control forecast (intensities and LVI only). The orange lines in (b) and (c) are the intensity histories from IBTrACS.

apparent that the majority of the ensemble members have a track where Sandy moves northward and makes landfall, albeit spread over a broad region of coastline (Fig. 1a). The deterministic track follows the NWPa track reasonably well until the storm turns westward to make landfall. The ensemble mean also follows the NWPa track well until the westward turn and is closer to the deterministic track after this point.

For intensity, the full-resolution (T1279) MSLP minima and 10-m wind speeds are considered together with the LVI. These are compared with IBTrACS pressure minima and 1-min sustained wind speeds. For MSLP (Fig. 1b), the first thing to notice is that the NWPa intensity history lies within the spread of the ensemble. However, the IBTrACS intensity history is outside or on the edge of the ensemble, which would

likely result in an underdispersive spread if compared with the ensemble mean error obtained using IBTrACS for verification. The ensemble mean, control, and NWPa all appear to underestimate the depth of the cyclone, leading up to the maximum intensity (greatest depth) relative to IBTrACS. The deterministic forecast appears more similar to IBTrACS, except for the deepening that occurs on 25 October, which is not captured by any of the forecasts. This rapid deepening is associated with the development of Sandy into a category 3 hurricane in the Caribbean (Blake et al. 2013). Both the NWPa and control forecast reach similar depths to IBTrACS, while the ensemble mean is significantly weaker. However, the deterministic forecast appears to attain a deeper minimum that occurs slightly later than observed. Some ensemble members also attain greater depths.

FIG. 2. Location error (geodesic degrees) statistics as a function of forecast lead time computed from all NH cyclones: (a) number of points available, (b) MSLP minima verified against IBTrACS locations, (c) MSLP minima verified against NWPa MSLP minima locations, and (d) LVL verified against NWPa LVL. The transparent shaded regions indicate the 95% confidence intervals.

For 10-m wind speed intensities (Fig. 1c), the NWPa significantly underrepresents the intensity compared with IBTrACS, in particular during the early stage when Sandy rapidly intensifies to category 3. However, they appear more similar during the later stages before Sandy comes ashore on 29 October. None of the forecasts capture the early rapid intensification, with the majority of the ensemble members, the ensemble mean, and control forecasts underpredicting the intensity relative to the NWPa. While the deterministic forecast also underpredicts the wind speed intensity relative to IBTrACS during the early part of the forecast, near landfall the forecast intensity compares quite well to the intensity in the NWPa and IBTrACS.

As IBTrACS does not generally include the pre- and post-TC stages, the use of the LVI is introduced for use in Part II of the study. The LVI results (Fig. 1d) are

qualitatively similar to those for the full-resolution pressure and wind speeds, with the NWPa LVI history embedded within the spread of the ensemble. The LVI for the control and deterministic forecasts agree quite well with the NWPa values except at maximum LVI where they tend to overpredict. Also apparent is the much smoother ensemble mean LVI history compared to the NWPa and the control and deterministic forecasts.

b. Northern Hemisphere forecast error

1) LOCATION

Location errors with lead time, for the combined five years, are shown in Fig. 2, based on IBTrACS (Fig. 2b), the NWPa MSLP minima (Fig. 2c), and the LVL (Fig. 2d). The number of data points, with lead time, that contribute to the statistics (Fig. 2a) is shown to provide

an additional indication of reliability of the statistics. The ensemble members have point numbers an order of magnitude larger than the control, deterministic, and ensemble mean forecasts, with a peak around day 2 before the numbers fall away.

In general, location errors are largest for the individual ensemble members irrespective of which verification dataset is used. This is followed by the control and then the deterministic forecast, with the lowest errors occurring for the ensemble mean. The initial error growth over the first few days is relatively slow for all forecast types but becomes more pronounced at longer lead times. Similar results were found by Froude et al. (2007b) for extratropical cyclones, who suggested that the control and deterministic forecasts have smaller errors than the ensemble members because they are run from an optimal initial state, with no convergence at longer lead times owing to the stochastic physics that only impact the ensemble members. That the ensemble mean has the lowest error is a consequence of the averaging process which emphasizes the more predictable scales. The ensemble mean error only diverges from the control (and deterministic) forecast after day 5. The ensemble spread is almost identical to the ensemble mean error until day 4 when it starts to deviate and indicates that for location the EPS is underdispersive at longer lead times. This is again irrespective of the verification dataset used. The confidence intervals for each of the forecast types are generally quite narrow, reflecting the relatively large sample sizes; this is particularly the case for the ensemble members. For the control, deterministic, and ensemble mean forecasts the width of the confidence intervals increases with lead time and there is some overlap between them, suggesting that they may not be significantly different. However, while the nonoverlap of confidence intervals is often taken as a measure of significance, this can be misleading (Wolfe and Hanley 2002). The errors over the forecast range are relatively insensitive to the verification data used; in particular, using the LVL gives comparable results to verification using NWPa MSLP minima or IBTrACS.

A different perspective is obtained if the early part of the forecasts are considered (0–2 days), shown in the insets of Fig. 2. This indicates that the deterministic forecast has the lowest location error, which lasts until ~day 5 (not shown), but the difference from the ensemble mean error is relatively small. The “jump” in the deterministic forecast errors in the first 6 h for the verification using the NWPa is likely due to adjustment from the initial state. The IBTrACS verification could be different from this because location is often subjectively smoothed (WMO 2013).

TCs can be followed for periods as long as 15 days, which is only limited by their life cycle and the length of the forecast. At long lead times the location errors can be larger than 10° . Other studies using more traditional approaches (Buizza 2010; Magnusson and Erland 2013) have noted error saturation at longer lead times, which is often considered as a limit of predictability and related to the upscale cascade of errors from the initial conditions (Lorenz 1969). This does not appear to be the case for the location errors, which increase monotonically. Forecasts beyond day 10 may not be very useful as a result of the large errors; however, they are perhaps useful in very broad terms as early warning of the potential for a TC to occur.

2) INTENSITY AND LVI

While surface wind speeds are an important metric for impacts, it is instructive to consider other intensity measures (WMO 2013). Here, the forecast 10-m wind speed and MSLP minima are verified against IBTrACS before being contrasted with their verification against the NWPa. Verification using the LVI is also introduced.

Intensity errors based on IBTrACS (Fig. 3) show that errors have a similar behavior for both wind speed and MSLP. The initial error growth is faster than for location for all forecast types. However, the error growth rate reduces at longer lead times with a suggestion of saturation. The “jumps” at day 10 are due to the changes in the ensemble resolution at this point. It is obvious that the forecasts have large errors relative to IBTrACS for both wind speed and MSLP for all forecast types, which is well known for the current prediction of TC intensities in global NWP (WMO 2013). However, apart from these large errors, it is apparent that there is little consistency between the different forecast types using the different intensity measures. For MSLP, the results show little difference between the control and ensemble members. The ensemble mean has lower errors than both the control and ensemble members. In the 0–4-day range the deterministic errors start off lowest then increase to be comparable with the control and ensemble members. For the wind speed it is difficult to see any systematic differences in the errors of the ensemble members, control, and ensemble mean, whereas deterministic errors are systematically lower. These results are reflected in the biases with larger biases for the lower-resolution ensemble forecasts than for the deterministic forecasts. The biases show underprediction, indicated by increasing positive errors for MSLP and negative errors for the wind speed. For both MSLP and wind speed the spreads are systematically and significantly lower than the ensemble mean error; this is primarily because the spread only depends on the ensemble

FIG. 3. Intensity error and bias statistics as a function of forecast lead time, verified against IBTrACS and computed from all NH cyclones: (a) MSLP intensity errors (hPa), (b) MSLP intensity biases (hPa), (c) 10-m wind speed intensity errors (m s^{-1}), and (d) 10-m wind speed intensity biases (m s^{-1}). The transparent shaded regions indicate the 95% confidence intervals.

forecasts. Even taking the bias into account, the slower growth of spread relative to the error suggests the ensemble is underdispersive using the IBTrACS verification. The fact that relative errors and biases show a lack of consistency using the two different intensity measures may stem from the size of the errors, swamping any subtle differences between the forecast types. It may also be due to the different levels of uncertainty in the observations (see [appendix](#)), with winds generally having a higher degree of uncertainty than pressure, with dependence on the TC intensity ([Torn and Snyder 2012](#)).

While it is important to be aware of the underprediction of TC forecast intensity, an alternative insightful approach is to verify intensities against the NWP, as with the locations. The advantage of this approach is that the verifying data is self-consistent with the forecast data in terms of being produced with the same model at similar spatial scales. The uncertainty in this

approach can be explored using NWP products from other operational centers (presented in a following paper). Improving the forecasts relative to the NWP may also lead to improvements when compared against observations, which can be of interest to model developers.

[Figure 4](#) shows the forecast errors obtained using the NWP for verification, full-resolution MSLP, 10-m wind speed, and LVI. The errors show some similarities with the verification against IBTrACS in that the error growth starts fast and then tends to saturate. However, the errors now show more consistency between the forecast types for the different intensity measures, with the control and ensemble members having comparable errors over the forecast range. The ensemble mean now has consistently lower errors than the control and ensemble members, for all measures throughout the forecast range. The deterministic forecast starts out with the lowest errors but these gradually grow with forecast lead

FIG. 4. Intensity error and bias statistics as a function of forecast lead time, verified against the NWPa and computed from all NH cyclones: (a) MSLP intensity errors (hPa), (b) MSLP intensity biases (hPa), (c) 10-m wind speed intensity errors (m s^{-1}), (d) 10-m wind speed intensity biases (m s^{-1}), (e) LVI errors ($\times 10^{-5} \text{ s}^{-1}$), and (f) LVI biases ($\times 10^{-5} \text{ s}^{-1}$). The transparent shaded regions indicate the 95% confidence intervals.

time to become similar to the control errors, though the lead time at which they become greater than the ensemble mean errors varies with the intensity measure; for the LVI it is quite rapid, but for the MSLP and wind speed this does not occur until after day 2. The “jumps” in error at day 10 due to the resolution change of the EPS are still noticeable, though most prominent for the full-resolution metrics. The spread for all intensity measures indicates that the ensemble is underdispersive, even for the LVI where a common resolution is used. For biases, the MSLP and wind speed generally show consistent results, with the ensemble forecasts all indicating underprediction. For these two measures, biases are also consistent for the deterministic forecasts, though smaller than the ensemble biases. For the LVI, although consistent with the other two measures in indicating underprediction for all forecast types, there is less difference between the ensemble and deterministic errors, reflecting the common resolution used. It is also apparent that similar differences between the forecasts occur at day zero as for location, with a “jump” associated with adjustment for the deterministic forecast.

c. Interannual forecast error

To explore interannual differences in forecast errors and whether they may be related to changes in the forecast system over the study period, results are contrasted for each year for both deterministic and ensemble mean forecasts. Results should be considered tentative owing to the lower amount of data available, when considering each year separately. Also, large-scale modes of interannual variability are known to have an impact on the occurrence of TCs and their properties (Frank and Young 2007), which may impact their year-to-year prediction. Results for each year are shown in Fig. 5 for the LVL/LVI. The LVL errors (Figs. 5a,b) show that the ensemble mean forecasts have lower errors than the deterministic system after day 5 for all years, similar to results in section 1. For the ensemble mean there are indications that LVL errors are lower after 2009. However, the differences between the years are relatively small with considerable overlap between the confidence intervals, which are now wider than for the results for the combined years (Fig. 2d). Differences may be related to the changes in the system made around 2009/10 such as improvements to the 4D-Var data assimilation system and an increase in horizontal resolution (Table 1). This is complicated by the fact that there was a La Niña/El Niño in 2008/09, though not a particularly strong event. El Niño–Southern Oscillation (ENSO) has a large impact on TCs, not only in terms of numbers but also their paths, related to the large-scale steering flow (Camargo et al. 2007; Colbert and Soden

2012). Recurring TCs and the ET tend to have a detrimental impact on forecast location errors (Part II) (Jones et al. 2003).

For the LVI (Figs. 5c and 5d) the ensemble mean has generally smaller errors than the deterministic forecasts for all years. For both forecasts there are indications that the LVI errors are lower for the years 2008 and 2009, though the differences are relatively small. Larger differences are seen in the LVI biases (Figs. 5e and 5f), which show a general underprediction in both the deterministic and ensemble mean forecasts, which increases in the later years following the resolution increases in 2010. This is a rather counterintuitive result as increasing resolution might be expected to reduce the error and bias. However, as well as changes in resolution, a new cloud parameterization scheme has been introduced and modifications to the convection scheme made (Table 1). The data assimilation and stochastic physics have also changed (Table 1). These changes can have an impact not only on the forecasted intensity of TCs but also how the TCs are represented by the NWP.

Results for the error–spread relationship of the ensemble are shown in Figs. 5g and 5h for the ensemble mean LVL and LVI, respectively. For LVL there is improvement in the error–spread relationship after 2010 with the error and spread becoming closer, though the spread is still underdispersive. For LVI the error–spread contrast indicates an underdispersive ensemble throughout the whole forecast, but with improvements after 2010 prior to day 2. These improvements in the early lead time error–spread relationship in later years are possibly related to the introduction of EDA perturbations and changes to the stochastic physics in 2010 (cf. Table 1). This was found to improve the EPS spread in the tropics, when evaluations were applied to more traditional fields (Berner et al. 2009).

The LVL/LVI results can be contrasted with those using the full-resolution MSLP locations and 10-m wind speed intensities verified against the NWP (Fig. 6). For the MSLP location (Figs. 6a and 6b), similar results to those for the LVL are obtained, with the same indication of slightly lower errors after the resolution changes in 2010. Perhaps more interesting are the results for the intensity errors and biases. The intensity errors (Figs. 6c and 6d) show similar results to those for the LVI but more clearly showing larger errors for intensity after the resolution changes in 2010. This is further reflected in the intensity biases (Figs. 6e and 6f), which show increasing underprediction of intensity after the resolution changes for both the deterministic and ensemble mean forecasts.

A further interesting contrast is obtained if the full-resolution MSLP locations and 10-m wind speed

FIG. 5. Error statistics as a function of forecast lead time for each year, based on LVL (geodesic degrees) and LVI ($\times 10^{-5} \text{ s}^{-1}$) for the (left) deterministic forecast and (right) ensemble mean forecast: (a) deterministic LVL error, (b) ensemble mean LVL error, (c) deterministic LVI error, (d) ensemble mean LVI error, (e) deterministic LVI bias, (f) ensemble mean LVI bias, (g) ensemble mean LVL error and spread (geodesic degrees), and (h) ensemble mean LVI error (dashed lines) and spread (solid lines). LVL error in units of geodesic degrees and LVI errors and bias in units of $\times 10^{-5} \text{ s}^{-1}$. The transparent shaded regions indicate the 95% confidence intervals.

FIG. 6. Error statistics as a function of forecast lead time, verified against the NWP, for each year for the NH cyclones based on the full-resolution MSLP minima for location (geodesic degrees) and the 10-m wind speed (m s^{-1}) for intensity for the (left) deterministic and (right) ensemble mean forecasts: (a) deterministic location error, (b) ensemble mean location error, (c) deterministic intensity error, (d) ensemble mean intensity error, (e) deterministic intensity bias, and (f) ensemble mean intensity bias. Location error in units of geodesic degrees and intensity errors and bias in units of m s^{-1} . The transparent shaded regions indicate the 95% confidence intervals.

FIG. 7. As in Fig. 6, but for verification against IBTrACS for pressure minima locations and 10-m wind speed.

intensities are verified against IBTrACS (Fig. 7). For location (Figs. 7a and 7b), the results are similar to those of the MSLP and LVL when verified against the NWPa, with similar interannual variations in the location error. The intensity and bias results are also interesting, with lower errors now obtained after the resolution changes in 2010 up to day 5 and generally becoming larger after

this (Figs. 7c and 7d). The biases generally reflect the intensity errors, and although the bias indicates significant underprediction of the wind speed intensity for all years (Figs. 7e and 7f), the biases are smaller for the later years up to day 5.

The results for intensity using IBTrACS verification is intuitively what might be expected with an increase in

FIG. 8. Regions for TC region subsetting.

resolution of the forecast model, whereas the results using NWPa verification might be considered counter-intuitive. Similar results were found by [Pinson and Hagedorn \(2012\)](#) in a study of the prediction of surface wind speed in the extratropics, making use of both NWPa and observations for verification.

d. Forecast error in different ocean basins

The forecast errors for different ocean basins are also considered. TCs are allocated to particular ocean basins based on their maximum intensities, using the NWPa LVL and LVI. Four regions are considered: Atlantic (ATL), eastern Pacific (EP), western Pacific (WP), and northern Indian Ocean (NIO) ([Fig. 8](#)).

The results for LVL ([Figs. 9a–d](#)) show the lowest errors occur in the EP and NIO with the ATL and WP basins showing much larger errors, in particular at lead times beyond day 4. Errors are comparable across the basins at the earliest lead times. In all basins, the relationship between the LVL errors of the different forecast types remains the same and similar to the full NH results. The results become less reliable past day 10, in particular in the NIO, owing to the reduced amount of data resulting from the track truncation to match the observed tracks. Using the full life cycles results in much more stable statistics, with narrower confidence intervals (not shown).

The LVI errors ([Figs. 9e–h](#)) show a similar behavior, with the EP and NIO having the smallest errors and biases, whereas the WP region has the largest errors and biases (underprediction). The ATL region is similar to the WP. Again, the relationship between the errors of the different forecast types in each basin is similar to the full NH results. The WP appears to have the worst error–spread relationship, indicating that the EPS is more underdispersive in this region compared to the other basins.

A similar picture is observed using the full-resolution MSLP locations and 10-m wind speed intensities, verified against the NWPa; these are shown in [Fig. 10](#). For location ([Figs. 10a–d](#)) the results are similar to those for the LVL ([Figs. 9a–d](#)), with the ATL and WP basins having larger errors beyond day 4 and the NIO and EP

having lower errors. The relationship between the forecast types is also similar to the LVL results, except for NIO at the longer lead times, which are very noisy. Similar relationships occur between the different forecast types for the 10-m wind speed in the different ocean basins ([Figs. 10e–h](#)), as seen using the LVI ([Figs. 9e–h](#)). While a little noisier, the intensity errors appear to grow faster and attain larger magnitudes with lead time in the ATL and WP basins, with those in the EP and NIO basins being smaller, though the NIO again has less reliable results at longer lead times. The intensity biases ([Figs. 10i–l](#)) mirror those obtained using the LVI ([Figs. 9i–l](#)), though with the usual differences between the ensemble and deterministic forecasts related to resolution. In general, the biases show underprediction of the intensities when compared with the NWPa, though less so for the deterministic forecasts. However, the biases reflect the errors, with stronger negative biases in the ATL and WP basins than in the EP and NIO basins.

Using IBTrACS for verification ([Fig. 11](#)), the location errors are similar to those when using the NWPa and LVL with the same behavior in the different ocean basins, including the relationship between the different forecast types. More interesting are the intensity results for the wind speeds, which show rather similar results to those for LVI and the full resolution verification against the NWPa for the different ocean basins, albeit with the similar large errors and biases seen before ([Fig. 3](#)). In particular the largest errors are seen in the WP with associated large underprediction biases. The ATL has the next largest errors and biases but now only marginally different from the EP. The lowest errors occur in the NIO with relatively small negative biases, albeit the statistics are quite noisy.

The location errors can be understood in terms of the TC propagation characteristics in the different ocean basins. In the ATL and WP a large proportion of the TCs tend to recurve and move into higher latitudes, gaining speed as they do so, resulting in a higher degree of uncertainty in their location ([Jones et al. 2003](#)), with a larger ensemble spread. For the EP and NIO basins on the other hand, very few TCs manage to move out of the

FIG. 9. Error statistics as a function of forecast lead time in different ocean basins for all years considered, 2008–12. (top) LVL errors (geodesic degrees), (middle) LVI errors ($\times 10^{-5} \text{ s}^{-1}$), and (bottom) LVI biases ($\times 10^{-5} \text{ s}^{-1}$).

FIG. 10. As in Fig. 9, but for verification against the NWPa for full-resolution MSLP location and surface wind speed intensities.

FIG. 11. As in Fig. 9, but for verification against IBTrACS for MSLP location and surface wind speed intensities.

tropics and the propagation speeds are typically fairly slow; in particular, the EP TCs tend to propagate almost zonally. The consequence of this is that the EP and NIO TCs have lower location errors and smaller ensemble spread than the other basins, which may indicate that the steering flow is more predictable in these basins with less extratropical interactions. The differences between the ocean basins could also be related to differences in secondary, less predictable aspects of TC motion such as that associated with β drift (Bin et al. 1998; DeMaria 1985). This imposes a secondary steering effect on the TCs, which could be more important in the ATL and WP, since the β -drift effect increases with latitude and intensity (Bin et al. 1998). However, it may be more subtle than this as TC genesis can occur via a variety of mechanisms (McTaggart-Cowan et al. 2013). For these different development mechanisms, the steering environments may also be different (Bin et al. 1998).

It is tempting to try and relate the intensity and LVI errors to the location and LVL errors, since it appears the intensity and LVI are correlated with the location errors, according to the ocean basins. This could be due to the fact that large errors in location are likely to result in TCs moving in very different environments from those observed and, hence, they will develop differently, resulting in very different intensities. However, even if the location is accurately predicted, this provides no guarantee that intensities will also be correctly predicted. This can be seen using a simple regression analysis to test the hypothesis of a link between intensity and location errors; however, no obvious correlation could be found (not shown). However, the relationship between location and intensity errors is likely to be complex, with the environmental steering flow, β drift, and the Fujiwara effect (Bin et al. 1998) all contributing to the TC steering. The layer-averaged environmental steering and β drift, in particular, are both influenced by TC intensity (Bin et al. 1998), which is likely to make any relationship between TC intensity and steering highly nonlinear and complex. Other environmental factors that affect TC, such as sea surface temperature (SST), wind shear, and humidity, may also differ between the basins and play a role in the variations in intensity and LVI error and their relationship with location.

4. Summary and conclusions

The study has focused on the prediction errors of TCs as simulated by the ECMWF deterministic and ensemble forecast systems. The results can be summarized as follows:

- (i) The LVL and MSLP minima produce location errors similar to using IBTrACS for verification.

- (ii) Intensity in the form of MSLP or 10-m wind speed is significantly underpredicted when verified against IBTrACS, but there is little consistency between the errors of the different forecast types when contrasted using the MSLP and 10-m wind speeds.
- (iii) Verifying against the NWPa results in more consistent relative forecast errors for the different intensity and LVI measures (cf. Figs. 3 and 4).
- (iv) For NWPa verification, generally the ensemble mean has lower errors than the ensemble members, control, and deterministic forecasts for both location and intensity. For location, the deterministic forecasts have marginally lower errors over the first 5 days.
- (v) Location appears to be more predictable than intensity, with slower error growth for location in the initial part of the forecast.
- (vi) In general, the ensemble forecasts are underdispersive—more so for intensity and LVI than location.
- (vii) There is interannual variation in forecast error, possibly associated with changes to the forecast system but also due to the difficulty in predicting changes in TCs from year to year.
- (viii) There are large variations in forecast error and ensemble spread in different ocean basins, which is hypothesized to be related to differences in the large-scale environment, whether TCs recurve, and to secondary effects such as β drift.
- (ix) The use of the LVL is shown to provide comparable results to other verification measures of location providing confidence in its use for the Part II study of the TC life cycles. The LVI is also shown to provide useful insights into forecast errors that are complementary to the intensity measures of MSLP and 10-m wind speed, showing similar relationships between the different forecast types using NWPa verification. This provides confidence in the use of LVI in the study of extended TC life cycles.

There are clearly significant deficiencies in the prediction of the intensities of TC in the ECMWF deterministic and EPS forecasts, when compared with IBTrACS. The current ECMWF IFS is based on a hydrostatic model. Increasing the spatial resolution, allowing physical processes to be explicitly resolved, such as convection, requires a nonhydrostatic model. Improved forecasts could also be achieved by improving the initial state by making use of new space-based and airborne measurements of winds and moisture together with the latest assimilation techniques (Atlas 2010; Pu et al.

2010). However, assimilating space-based observations in the vicinity of TCs is still challenging owing to deficiencies in the first-guess fields where large differences between the first-guess and observations can result in observation rejection (Pu 2009). Observation quality in cloudy areas and rain contamination of sensors also reduces their usefulness. This is especially important for the inner-core regions of TCs, which are still not resolved well in the global forecast models (Pu 2009). This is currently a very active area of research to enable the use of precipitation- and cloud-affected data from satellites in the data assimilation (Bauer et al. 2011)—in particular in the vicinity of TCs using very-high-resolution regional forecasting systems (Zhang et al. 2013), which resolve the inner-core regions better.

Verification against observations or best track data is of paramount importance for forecast users. However, observation uncertainties are still an issue when contrasting between forecast types, and better-quality, frequently sampled observations should be a priority. This could be achieved with better observational platforms and better analysis of observations. Using the NWPa for verification provides more consistent verification, though this is sensitive to observational inhomogeneity and model formulation, with the NWPa themselves underestimating intensity. However, observation use is rigorously quality controlled in the NWPa. This study suggests using both IBTrACS and NWPa can provide useful insights into forecast performance.

The LVL and LVI measures have been contrasted with other error metrics to justify their use in exploring the prediction of TCs over their full life cycle, including pre- and post-TC stages (Part II). The use of LVL has been shown to provide comparable results to using full-resolution NWPa MSLP minima as well as verifying against IBTrACS. The use of the LVI provides an alternative means of evaluation for intensity, which is consistent with using the full-resolution NWPa, MSLP, and wind speed. The use of LVL and LVI can provide further insights into the prediction of TCs by EPSs, thus providing a basis for intercomparing the prediction of TCs by EPSs from different centers. Using a common lower resolution for verification focuses on more predictable scales and provides a fairer comparison between different forecast types and different forecasting systems, which vary greatly in resolution. It is also useful for contrasting the prediction of TCs in different ocean basins and providing insights into changes in the forecast system.

An improved understanding of TC forecast error and its variation with region and environment could be obtained using a frozen forecasting system. This could be based on reanalyses, which are gradually moving to higher resolutions than previously used, and which could

provide initial conditions for forecasts, including EPSs. This would provide a much larger TC sample to work with than used here without the uncertainty introduced by system upgrades. They could also be used to examine the impact of the changing observing system. Such data are starting to be produced as “reforecast” or hindcast data, though not in general at the resolutions of current operational systems (Hamill et al. 2013).

Acknowledgments. The authors would like to acknowledge ECMWF for making the data available via a member state special project. The authors would also like to thank the reviewers for their constructive comments, which have improved the paper.

APPENDIX

Verification Data Uncertainties

The observation-based data used in this study is IBTrACS (Knapp et al. 2010). This is a postseason reanalyzed dataset based on observations from satellites, aircraft reconnaissance, and forecaster interpretation. It is essentially a subjective assessment of TC location and intensity where location is subjectively smoothed and hence differs from the operationally determined location (WMO 2013). Missing data are also an issue in IBTrACS, particularly for intensity.

Understanding observation uncertainty of TCs (Knapp et al. 2013; Landsea and Franklin 2013; Torn and Snyder 2012) is important to have confidence in their use for verification, either directly or as best track data. For example, Bowler (2008) has highlighted the impact of observational error on the verification of deterministic and EPS forecasts. The WMO report on verification methods (WMO 2013) states “biases in observations can lead to wrong conclusions” and “in the absence of perfect observations it is impossible to determine true forecast errors.”

Torn and Snyder (2012) highlighted the dependence of uncertainty on TC intensity, with greater location uncertainty for weaker TCs, and greater intensity uncertainty for strong TCs, when intensities are contrasted between those from the Dvorak technique (Dvorak 1975) and aircraft reconnaissance. Observations of TCs over the oceans mainly depend on the interpretation of satellite imagery and data, supplemented by aircraft reconnaissance. While aircraft reconnaissance is still conducted routinely in the North ATL for TCs with potential to make landfall [see Williamson et al. (2014), 5.5.1.3], and to some extent in the EP, in other ocean basins it is much less common. For example, routine reconnaissance by the Joint Typhoon Warning Center (JTWC) of WP TC ended in 1987 (Knapp et al. 2013),

FIG. A1. (a) Probability density of the mean separation distances between the NWPa tracks and IBTrACS for the same period based on the T63 vorticity locations (LVL) and those based on the gridpoint and off-grid full resolution MSLP minima; solid line is the vorticity maxima, dotted line is the gridpoint MSLP minima, dashed line is the off-grid MSLP minima; bin widths are 0.2° (~ 22 km). (b) Instantaneous intensity differences between the NWPa tracks and IBTrACS for 10-m (1 min) wind speed (black lines) and MSLP (gray lines) (IBTrACS – NWPa) at the T799 resolution (2008–12) and T1279 resolution (2010–12). (c) Scatterplot of MSLP differences (IBTrACS – NWPa) vs latitude. (d) Scatterplot of 10-m wind speed (1 min) differences (IBTrACS – NWPa) vs latitude.

though some limited reconnaissance is still carried out in the vicinity of Taiwan (DOTSTAR) (Wu et al. 2005) and some limited field campaigns have also been conducted such as for The Observing System Research and Predictability Experiment (THORPEX) Pacific Asian Regional Campaign (T-PARC) (Black 2012; Quinting et al. 2014). No reconnaissance is available in the NIO. However, aircraft reconnaissance observations also have uncertainties associated with them, associated with the flight path and wind variations inside TCs (Uhlhorn and Nolan 2012). This can make “the determination of fully two-dimensional surface winds problematic” (Rappaport et al. 2009). Even in the best observed region of the North ATL, only about 30% of TC lifetimes are sampled and less in the EP (Nolan et al. 2014; Rappaport et al. 2009), though this has changed

somewhat in more recent years with increased reconnaissance in the EP and the deployment of multiple aircraft and drones.

Accurately determining the location of TCs using satellite infrared, visible, or water vapor imagery depends on whether an eye has formed (Torn and Snyder 2012). This can be improved with the use of microwave sounder data, such as from the Advanced Microwave Sounding Unit (AMSU), which can penetrate the clouds, but has the drawback that, being on a polar-orbiting platform, does not sample at synoptic times (Kidder et al. 2000).

Determining intensity mainly depends on the Dvorak technique (Dvorak 1975; Velden et al. 2006a), which can have significant uncertainties (Kossin and Velden 2004), with different agencies using different implementations

(Kueh 2012; Velden et al. 2006b), though these can be bias corrected (Knaff et al. 2010; Kossin and Velden 2004). TC intensities are mostly considered in terms of surface wind speed. The WMO recommends a 10-min averaging period for maximum sustained winds at 10 m; however, different agencies use different averaging periods with 1 min being common (Knapp and Kruk 2010). Conversion factors can be used to convert between the different averaging periods with the often used value being 0.88, though recent reevaluation suggests a value of 0.93 should be used over oceans and the 0.88 value being more appropriate for “off-land” areas (Harper et al. 2010). Determining winds from surface instruments and their uncertainty has been examined by Nolan et al. (2014), who found considerable variation in uncertainty depending on the measure used. Recently, alternative intensity metrics have been studied that focus on the “observable and resolvable low-wavenumber intensity within the vortex” (Vukicevic et al. 2014) that have been shown to provide more realistic estimates of forecast intensity errors.

NWPa are often used for forecast verification, though mostly when independent observations are unavailable or are poor quality. Their usefulness derives from the fact that they provide the best estimate of the atmospheric state at scales similar to those of the forecasts. They are less often used for the verification of TCs because of insufficient resolution to reproduce the observed TC wind intensities. The NWPa also have uncertainties, associated with the distribution and quality of the observations and how they are assimilated, and the formulation of the forecast model. The NWPa 10-m winds also have uncertainties as these are not directly computed by the model but are obtained as a diagnostic and are sensitive to how this is done. Additionally, the wind speed is an instantaneous value, representative of the model time step and spatial resolution, which is different from the observational sampling of winds.

This paper introduces the LVL and LVI to verify TCs, motivated by the aim of examining prediction errors of the extended TC life cycles (Part II). To provide confidence in the use of the NWPa LVL for verification, it has been compared with locations from IBTrACS and NWPa full-resolution MSLP minima (grid point and off grid). The mean separation distances, for all three comparisons (Fig. A1a), are generally less than 1° and less than 0.5° at the distribution peak. The NWPa LVL separation from IBTrACS is generally smaller than for the full-resolution MSLP gridpoint minima. The off-grid MSLP minima results in smaller separation distances for some TCs, though differences with using the LVL are small. These results are within the uncertainty estimates for location in best track data reported by Torn and Snyder (2012). NWP systems continue to increase in resolution

and may soon achieve 5 km globally (Magnusson et al. 2014); hence, location can be determined with greater accuracy. This will require improvements in observations to improve their use for verification.

To understand the differences between the intensities of TCs in the NWPa and IBTrACS, Fig. A1b shows the instantaneous differences (IBTrACS – NWPa) for MSLP and 10-m wind speed. This shows that the NWPa underestimate intensity using either measure, more obviously for wind speed. Motivated by the Hurricane Sandy case study (section 3), Figs. A1c and A1d show scatterplots of intensity differences (IBTrACS – NWPa) versus latitude for MSLP and 10-m wind speed. These show that the largest errors occur at low latitudes, south of 25°N , and can in fact be very large, whereas at high latitudes the differences are much smaller.

REFERENCES

- Aberson, S. D., 2008: An alternative tropical cyclone intensity forecast verification technique. *Wea. Forecasting*, **23**, 1304–1310, doi:10.1175/2008WAF2222123.1.
- , and M. DeMaria, 1994: Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Wea. Rev.*, **122**, 2804–2815, doi:10.1175/1520-0493(1994)122<2804:VOANBH>2.0.CO;2.
- Atlas, R., 2010: Review of observing system simulation experiments to evaluate the potential impact of lidar winds on weather prediction. *Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Honolulu, HI, IEEE, 2587–2590, doi:10.1109/IGARSS.2010.5654376.
- Bauer, P., G. Ohning, C. Kummerow, and T. Auligne, 2011: Assimilating satellite observations of clouds and precipitation into NWP models. *Bull. Amer. Meteor. Soc.*, **92**, ES25–ES28, doi:10.1175/2011BAMS3182.1.
- Bengtsson, L., L. M. Botzet, and M. Esch, 1995: Hurricane-type vortices in a general circulation model. *Tellus*, **47A**, 175–196, doi:10.1034/j.1600-0870.1995.t01-1-00003.x.
- , K. I. Hodges, and M. Esch, 2007a: Tropical cyclones in a T159 resolution global climate model: Comparison with observations and reanalyses. *Tellus*, **59A**, 396–416, doi:10.1111/j.1600-0870.2007.00236.x.
- , —, —, N. Keenlyside, L. Kornblueh, J.-J. Luo, and T. Yamagata, 2007b: How may tropical cyclones change in a warmer climate? *Tellus*, **59A**, 539–561, doi:10.1111/j.1600-0870.2007.00251.x.
- , —, and N. Keenlyside, 2009: Will extratropical storms intensify in a warmer climate? *J. Climate*, **22**, 2276–2301, doi:10.1175/2008JCLI2678.1.
- Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, doi:10.1175/2008JAS2677.1.
- Bin, W., R. L. Elsberry, W. Yuqing, and W. Liguang, 1998: Dynamics in tropical cyclone motion: A review. *Chin. J. Atmos. Sci.*, **22**, 416–434.
- Black, P. G., 2012: Tropical cyclone unusual intensity and structure change in the western North Pacific observed by reconnaissance aircraft during TPARC/TCS08 and ITOP/TCS10. *Trop. Cyclone Res. Rev.*, **1**, 75–88, doi:10.6057/2012TCRR01.09.

- Blake, E. S., T. B. Kimberlain, R. J. Berg, J. P. Cangialosi, and J. L. Beven II, 2013: Tropical cyclone report: Hurricane Sandy. National Hurricane Center Tech. Rep. AL182012, 157 pp. [Available online at http://www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf.]
- Bonavita, M., L. Raynaud, and L. Isaksen, 2011: Estimating background-error variances with the ECMWF ensemble of data assimilations system: Some effects of ensemble size and day-to-day variability. *Quart. J. Roy. Meteor. Soc.*, **137**, 423–434, doi:[10.1002/qj.756](https://doi.org/10.1002/qj.756).
- Bowler, N. E., 2008: Accounting for the effect of observation errors on verification of MOGREPS. *Meteor. Appl.*, **15**, 199–205, doi:[10.1002/met.64](https://doi.org/10.1002/met.64).
- Buizza, R., 2010: Horizontal resolution impact on short- and long-range forecast error. *Quart. J. Roy. Meteor. Soc.*, **136**, 1020–1035, doi:[10.1002/qj.613](https://doi.org/10.1002/qj.613).
- , and A. Hollingsworth, 2002: Storm prediction over Europe using the ECMWF ensemble prediction system. *Meteor. Appl.*, **9**, 289–305, doi:[10.1017/S1350482702003031](https://doi.org/10.1017/S1350482702003031).
- , M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, doi:[10.1002/qj.49712556006](https://doi.org/10.1002/qj.49712556006).
- Camargo, S. J., A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, 2007: Cluster analysis of typhoon tracks. Part II: Large-scale circulation and ENSO. *J. Climate*, **20**, 3654–3676, doi:[10.1175/JCLI4203.1](https://doi.org/10.1175/JCLI4203.1).
- Chen, T.-C., S.-Y. Wang, M.-C. Yen, and A. J. Clark, 2008: Are tropical cyclones less effectively formed by easterly waves in the western North Pacific than in the North Atlantic? *Mon. Wea. Rev.*, **136**, 4527–4540, doi:[10.1175/2008MWR2149.1](https://doi.org/10.1175/2008MWR2149.1).
- Colbert, A. J., and B. J. Soden, 2012: Climatological variations in North Atlantic tropical cyclone tracks. *J. Climate*, **25**, 657–673, doi:[10.1175/JCLI-D-11-00034.1](https://doi.org/10.1175/JCLI-D-11-00034.1).
- Courtier, P., J. N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387, doi:[10.1002/qj.49712051912](https://doi.org/10.1002/qj.49712051912).
- DeMaria, M., 1985: Tropical cyclone motion in a nondivergent barotropic model. *Mon. Wea. Rev.*, **113**, 1199–1210, doi:[10.1175/1520-0493\(1985\)113<1199:TCMIAN>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<1199:TCMIAN>2.0.CO;2).
- Dvorak, V. F., 1975: Tropical cyclone intensity analysis and forecasting from satellite imagery. *Mon. Wea. Rev.*, **103**, 420–430, doi:[10.1175/1520-0493\(1975\)103<0420:TCIAAF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1975)103<0420:TCIAAF>2.0.CO;2).
- Ehrendorfer, M., and J. Tribbia, 1997: Optimal prediction of forecast error covariance through singular vectors. *J. Atmos. Sci.*, **54**, 286–313, doi:[10.1175/1520-0469\(1997\)054<0286:OPOFEC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1997)054<0286:OPOFEC>2.0.CO;2).
- Frank, W. M., and G. S. Young, 2007: The interannual variability of tropical cyclones. *Mon. Wea. Rev.*, **135**, 3587–3589, doi:[10.1175/MWR3435.1](https://doi.org/10.1175/MWR3435.1).
- Froude, L. S. R., 2010: TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. *Wea. Forecasting*, **25**, 819–836, doi:[10.1175/2010WAF2222326.1](https://doi.org/10.1175/2010WAF2222326.1).
- , L. Bengtsson, and K. I. Hodges, 2007a: The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. *Mon. Wea. Rev.*, **135**, 315–333, doi:[10.1175/MWR3274.1](https://doi.org/10.1175/MWR3274.1).
- , —, and —, 2007b: The prediction of extratropical storm tracks by the ECMWF and NCEP ensemble prediction system. *Mon. Wea. Rev.*, **135**, 2545–2576, doi:[10.1175/MWR3422.1](https://doi.org/10.1175/MWR3422.1).
- , —, and —, 2013: Atmospheric predictability revisited. *Tellus*, **65**, 19022, doi:[10.3402/tellusa.v65i0.19022](https://doi.org/10.3402/tellusa.v65i0.19022).
- Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau Jr., Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, doi:[10.1175/BAMS-D-12-00014.1](https://doi.org/10.1175/BAMS-D-12-00014.1).
- Harper, B. A., J. D. Kepert, and J. D. Ginger, 2010: Guidelines for converting between various wind averaging periods in tropical cyclone conditions. World Meteorological Organization Tech. Rep., 64 pp. [Available from World Meteorological Organization (WMO), 7 bis, avenue de la Paix, P.O. Box 2300, CH-1211 Geneva 2, Switzerland.]
- Hodges, K. I., 1994: A general method for tracking analysis and its application to meteorological data. *Mon. Wea. Rev.*, **122**, 2573–2586, doi:[10.1175/1520-0493\(1994\)122<2573:AGMFTA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<2573:AGMFTA>2.0.CO;2).
- , 1995: Feature tracking on the unit sphere. *Mon. Wea. Rev.*, **123**, 3458–3465, doi:[10.1175/1520-0493\(1995\)123<3458:FTOTUS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2).
- , 1999: Adaptive constraints for feature tracking. *Mon. Wea. Rev.*, **127**, 1362–1373, doi:[10.1175/1520-0493\(1999\)127<1362:ACFFT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<1362:ACFFT>2.0.CO;2).
- , R. W. Lee, and L. Bengtsson, 2011: A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *J. Climate*, **24**, 4888–4906, doi:[10.1175/2011JCLI4097.1](https://doi.org/10.1175/2011JCLI4097.1).
- Hopsch, S. B., C. Thorncroft, K. I. Hodges, and A. Ayyer, 2007: West African storm tracks and their relationship to Atlantic tropical cyclones. *J. Climate*, **20**, 2468–2483, doi:[10.1175/JCLI4139.1](https://doi.org/10.1175/JCLI4139.1).
- Hopson, T. M., 2014: Assessing the ensemble spread–error relationship. *Mon. Wea. Rev.*, **142**, 1125–1142, doi:[10.1175/MWR-D-12-00111.1](https://doi.org/10.1175/MWR-D-12-00111.1).
- Jones, S. C., and Coauthors, 2003: The extratropical transition of tropical cyclones: Forecast challenges, current understanding, and future directions. *Wea. Forecasting*, **18**, 1052–1092, doi:[10.1175/1520-0434\(2003\)018<1052:TETOTC>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1052:TETOTC>2.0.CO;2).
- Kerns, B., and E. Zipser, 2009: Four years of tropical ERA-40 vorticity maxima tracks. Part II: Differences between developing and nondeveloping disturbances. *Mon. Wea. Rev.*, **137**, 2576–2591, doi:[10.1175/2008MWR2545.1](https://doi.org/10.1175/2008MWR2545.1).
- Kidder, S. Q., M. D. Goldberg, R. M. Zehr, M. DeMaria, J. F. W. Purdom, C. S. Velden, N. C. Grody, and S. J. Kusselson, 2000: Satellite analysis of tropical cyclones using the Advanced Microwave Sounding Unit (AMSU). *Bull. Amer. Meteor. Soc.*, **81**, 1241–1259, doi:[10.1175/1520-0477\(2000\)081<1241:SAOTCU>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<1241:SAOTCU>2.3.CO;2).
- Knaff, J. A., D. P. Brown, J. Courtney, G. M. Gallina, and J. L. Beven, 2010: An evaluation of Dvorak technique–based tropical cyclone intensity estimates. *Wea. Forecasting*, **25**, 1362–1379, doi:[10.1175/2010WAF2222375.1](https://doi.org/10.1175/2010WAF2222375.1).
- Knapp, K. R., and M. C. Kruk, 2010: Quantifying interagency differences in tropical cyclone best-track wind speed estimates. *Mon. Wea. Rev.*, **138**, 1459–1473, doi:[10.1175/2009MWR3123.1](https://doi.org/10.1175/2009MWR3123.1).
- , —, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, doi:[10.1175/2009BAMS2755.1](https://doi.org/10.1175/2009BAMS2755.1).
- , J. A. Knaff, C. R. Sampson, G. M. Riggio, and A. D. Schnapp, 2013: A pressure-based analysis of the historical western

- North Pacific tropical cyclone intensity record. *Mon. Wea. Rev.*, **141**, 2611–2631, doi:[10.1175/MWR-D-12-00323.1](https://doi.org/10.1175/MWR-D-12-00323.1).
- Kossin, J. P., and C. S. Velden, 2004: A pronounced bias in tropical cyclone minimum sea level pressure estimation based on the Dvorak technique. *Mon. Wea. Rev.*, **132**, 165–173, doi:[10.1175/1520-0493\(2004\)132<0165:APBITC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0165:APBITC>2.0.CO;2).
- Kueh, M.-T., 2012: Multiformality of the tropical cyclone wind–pressure relationship in the western North Pacific: Discrepancies among four best-track archives. *Environ. Res. Lett.*, **7**, 024015, doi:[10.1088/1748-9326/7/2/024015](https://doi.org/10.1088/1748-9326/7/2/024015).
- Landsea, C. W., and J. L. Franklin, 2013: How good are the best tracks? Estimating uncertainty in the Atlantic hurricane database. *Mon. Wea. Rev.*, **141**, 3576–3592, doi:[10.1175/MWR-D-12-00254.1](https://doi.org/10.1175/MWR-D-12-00254.1).
- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, doi:[10.1016/j.jcp.2007.02.014](https://doi.org/10.1016/j.jcp.2007.02.014).
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, doi:[10.1111/j.2153-3490.1969.tb00444.x](https://doi.org/10.1111/j.2153-3490.1969.tb00444.x).
- Magnusson, L., and K. Erland, 2013: Factors influencing skill improvements in the ECMWF forecasting system. *Mon. Wea. Rev.*, **141**, 3142–3153, doi:[10.1175/MWR-D-12-00318.1](https://doi.org/10.1175/MWR-D-12-00318.1).
- , J.-R. Bidlot, S. T. K. Lang, A. Thorpe, N. Wedi, and M. Yamaguchi, 2014: Evaluation of medium-range forecasts for Hurricane Sandy. *Mon. Wea. Rev.*, **142**, 1962–1981, doi:[10.1175/MWR-D-13-00228.1](https://doi.org/10.1175/MWR-D-13-00228.1).
- Majumdar, S. J., and P. M. Finocchio, 2010: On the ability of global ensemble prediction systems to predict tropical cyclone track probabilities. *Wea. Forecasting*, **25**, 659–680, doi:[10.1175/2009WAF2222327.1](https://doi.org/10.1175/2009WAF2222327.1).
- Manganello, J. V., and Coauthors, 2012: Tropical cyclone climatology in a 10-km global atmospheric GCM: Toward weather-resolving climate modeling. *J. Climate*, **25**, 3867–3893, doi:[10.1175/JCLI-D-11-00346.1](https://doi.org/10.1175/JCLI-D-11-00346.1).
- McTaggart-Cowan, R., T. J. Galarneau, L. F. Bosart, R. W. Moore, and O. Martius, 2013: A global climatology of baroclinically influenced tropical cyclogenesis. *Mon. Wea. Rev.*, **141**, 1963–1989, doi:[10.1175/MWR-D-12-00186.1](https://doi.org/10.1175/MWR-D-12-00186.1).
- Nolan, D. S., J. A. Zhang, and E. W. Uhlhorn, 2014: On the limits of estimating the maximum wind speeds in hurricanes. *Mon. Wea. Rev.*, **142**, 2814–2837, doi:[10.1175/MWR-D-13-00337.1](https://doi.org/10.1175/MWR-D-13-00337.1).
- Pinson, P., and R. Hagedorn, 2012: Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. *Meteor. Appl.*, **19**, 484–500, doi:[10.1002/met.283](https://doi.org/10.1002/met.283).
- Pu, Z., 2009: Assimilation of satellite data in improving numerical simulations of tropical cyclones: Progress, challenge and development. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, S. K. Park and L. Xu, Eds., Springer-Verlag, 163–176.
- , L. Zhang, and G. D. Emmitt, 2010: Impact of airborne Doppler wind lidar data on numerical simulation of a tropical cyclone. *Geophys. Res. Lett.*, **37**, L05801, doi:[10.1029/2009GL041765](https://doi.org/10.1029/2009GL041765).
- Puri, K., J. Barkmeijer, and T. N. Palmer, 2001: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. J. Roy. Meteor. Soc.*, **127**, 709–731, doi:[10.1002/qj.49712757222](https://doi.org/10.1002/qj.49712757222).
- Quinting, J. F., M. M. Bell, P. A. Harr, and S. C. Jones, 2014: Structural characteristics of T-PARC Typhoon Sinlaku during its extratropical transition. *Mon. Wea. Rev.*, **142**, 1945–1961, doi:[10.1175/MWR-D-13-00306.1](https://doi.org/10.1175/MWR-D-13-00306.1).
- Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419, doi:[10.1175/2008WAF2222128.1](https://doi.org/10.1175/2008WAF2222128.1).
- Sardeshmukh, P. D., and B. J. Hoskins, 1984: Spectral smoothing on the sphere. *Mon. Wea. Rev.*, **112**, 2524–2529, doi:[10.1175/1520-0493\(1984\)112<2524:SSOTS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<2524:SSOTS>2.0.CO;2).
- Serra, Y. L., G. N. Kiladis, and K. I. Hodges, 2010: Tracking and mean structure of easterly waves over the Intra-Americas Sea. *J. Climate*, **23**, 4823–4840, doi:[10.1175/2010JCLI3223.1](https://doi.org/10.1175/2010JCLI3223.1).
- Shutts, G., M. Leutbecher, A. Weisheimer, T. Stockdale, L. Isaksen, and M. Bonavita, 2011: Representing model uncertainty: Stochastic parametrizations at ECMWF. *ECMWF Newsletter*, No. 129, ECMWF, Reading, United Kingdom, 19–24. [Available online at <http://old.ecmwf.int/publications/newsletters/pdf/129.pdf>.]
- Strachan, J., P. L. Vidale, K. I. Hodges, M. Roberts, and M.-E. Demory, 2013: Investigating global tropical cyclone activity with a hierarchy of AGCMs: The role of model resolution. *J. Climate*, **26**, 133–152, doi:[10.1175/JCLI-D-12-00012.1](https://doi.org/10.1175/JCLI-D-12-00012.1).
- Thorncroft, C., and K. I. Hodges, 2001: African easterly wave variability and its relationship to Atlantic tropical cyclone activity. *J. Climate*, **14**, 1166–1179, doi:[10.1175/1520-0442\(2001\)014<1166:AEWVAI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<1166:AEWVAI>2.0.CO;2).
- Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. *Wea. Forecasting*, **27**, 715–729, doi:[10.1175/WAF-D-11-00085.1](https://doi.org/10.1175/WAF-D-11-00085.1).
- Uhlhorn, E. W., and D. S. Nolan, 2012: Observational under-sampling in tropical cyclones and implications for estimated intensity. *Mon. Wea. Rev.*, **140**, 825–840, doi:[10.1175/MWR-D-11-00073.1](https://doi.org/10.1175/MWR-D-11-00073.1).
- Velden, C., and Coauthors, 2006a: The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bull. Amer. Meteor. Soc.*, **87**, 1195–1210, doi:[10.1175/BAMS-87-9-1195](https://doi.org/10.1175/BAMS-87-9-1195).
- , and Coauthors, 2006b: Supplement to: The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bull. Amer. Meteor. Soc.*, **87**, S6–S9, doi:[10.1175/BAMS-87-9-Velden](https://doi.org/10.1175/BAMS-87-9-Velden).
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 494 pp.
- Vukicevic, T., E. Uhlhorn, P. Reasor, and B. Klotz, 2014: A novel multiscale intensity metric for evaluation of tropical cyclone intensity forecasts. *J. Atmos. Sci.*, **71**, 1292–1304, doi:[10.1175/JAS-D-13-0153.1](https://doi.org/10.1175/JAS-D-13-0153.1).
- Williamson, S. P., and Coauthors, 2014: National hurricane operations plan. Office of the Federal Coordinator for Meteorological Services and Supporting Research FCM-P12-2014, 166 pp. [Available online at <http://www.ofcm.noaa.gov/nhop/13/nhop13.htm>.]
- WMO, 2013: Verification methods for tropical cyclone forecasts. WWRP/WGNE Joint Working Group on Forecast Verification Research WWRP 2013-7, 89 pp. [Available online at http://www.wmo.int/pages/prog/arep/wwrp/new/documents/WWRP_2013_7_TC_verification_15_Nov_en.pdf.]
- Wolfe, R., and J. Hanley, 2002: If we're so different, why do we keep overlapping? When 1 plus 1 doesn't make 2. *CMAJ*, **166**, 65–66.
- Wu, C.-C., and Coauthors, 2005: Dropwindsonde Observations for Typhoon Surveillance near the Taiwan Region (DOTSTAR): An overview. *Bull. Amer. Meteor. Soc.*, **86**, 787–790, doi:[10.1175/BAMS-86-6-787](https://doi.org/10.1175/BAMS-86-6-787).
- Zhang, M., M. Zupanski, M.-J. Kim, and J. A. Knaff, 2013: Assimilating AMSU-A radiances in the TC core area with NOAA operational HWRF (2011) and a hybrid data assimilation system: Danielle (2010). *Mon. Wea. Rev.*, **141**, 3889–3907, doi:[10.1175/MWR-D-12-00340.1](https://doi.org/10.1175/MWR-D-12-00340.1).