On the Predictability and Error Sources of Tropical Cyclone Intensity Forecasts

KERRY EMANUEL

Lorenz Center, Massachusetts Institute of Technology, Cambridge, Massachusetts

FUQING ZHANG

Department of Meteorology, and Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 29 March 2016, in final form 21 June 2016)

ABSTRACT

The skill of tropical cyclone intensity forecasts has improved slowly since such forecasts became routine, even though track forecast skill has increased markedly over the same period. In deciding whether or how best to improve intensity forecasts, it is useful to estimate fundamental predictability limits as well as sources of intensity error. Toward that end, the authors estimate rates of error growth in a ''perfect model'' framework in which the same model is used to explore the sensitivities of tropical cyclone intensity to perturbations in the initial storm intensity and large-scale environment. These are compared to estimates made in previous studies and to intensity error growth in real-time forecasts made using the same model, in which model error also plays an important role. The authors find that error growth over approximately the first few days in the perfect model framework is dominated by errors in initial intensity, after which errors in forecasting the track and large-scale kinematic environment become more pronounced. Errors owing solely to misgauging initial intensity are particularly large for storms about to undergo rapid intensification and are systematically larger when initial intensity is underestimated compared to overestimating initial intensity by the same amount. There remains an appreciable gap between actual and realistically achievable forecast skill, which this study suggests can best be closed by improved models, better observations, and superior data assimilation techniques.

1. Introduction

Much attention has been directed in recent years toward improving tropical cyclone intensity forecasts (Gall et al. 2013), especially since there has been comparatively little improvement in measures of intensity skill in the last few decades (DeMaria et al. 2014). Within the community of scientists working on tropical cyclones, efforts have been directed toward improving dedicated tropical cyclone models (e.g., Gopalakrishnan et al. 2011), real-time in situ and remote sensing observations of storms (e.g., Ruf et al. 2016), assimilation of those observations into models (Zhang and Weng 2015; Weng and Zhang 2016; Zhang et al. 2016), and statistical forecast models, which are still competitive with deterministic models (Kaplan et al. 2015). Numerical modeling of tropical cyclones is especially challenging owing to the very high resolution required to resolve the critical eyewall region (Rotunno et al. 2009), to the complex physics of boundary layers and air–sea interaction at high winds speeds (Nolan et al. 2009; Green and Zhang 2014; Andreas et al. 2015; Green and Zhang 2015), and to the importance of correctly modeling the response of the upper ocean to the storms (Moon et al. 2007; Yablonsky and Ginis 2009). Assimilating observations into models of tropical cyclones also presents great difficulties, partly because methods based on an assumption of Gaussian error distribution are bound to run into problems, since even small errors in the location of strong vortices entail large, non-Gaussian errors in winds at fixed points in space (Ravela et al. 2007). To all

Denotes Open Access content.

Corresponding author address: Kerry Emanuel, Rm. 54-1814, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139.
E-mail: emanuel@mit.edu

DOI: 10.1175/JAS-D-16-0100.1

this, one can add the operational difficulties and expense of observing tropical cyclones. Even today, there are appreciable discrepancies among real-time intensity estimates by different techniques based on satellite- and aircraft-based observations (Landsea and Franklin 2013).

In attempting to improve any type of numerical weather forecast, it is useful to be aware of fundamental and practical predictability limits imposed by the inherent chaotic nature of atmospheric behavior (Lorenz 1969; Zhang et al. 2007; Rotunno and Snyder 2008; Palmer et al. 2014) coupled with limitations on our ability to accurately observe our environment (Sun and Zhang 2016). One may reasonably expect models and data assimilation techniques to improve continuously over time, while the hope that observations will also improve must be tempered by finite economic and technological resources. For example, even if tropical cyclone models and data assimilation techniques improve greatly and if local storm-related error growth were to prove small, forecasts of tropical cyclones will still be circumscribed by the finite predictability of their large-scale environments. It is therefore of both practical and inherent intellectual interest to estimate predictability horizons in the limit that models and data assimilation systems become essentially perfect. On the practical side, this would provide a useful target for long-term efforts to improve tropical cyclone forecasts. A recent case study by Judt et al. (2016) showed that tropical cyclone surface wind is highly scale dependent; the vortex scale circulation is more resistant to upscale error growth and thus longer-term intensity predictions are theoretically possible through extending the predictability of the large-scale environment.

It is the aim of this paper to provide useful quantitative estimates of tropical cyclone intensity predictability in the limit of perfect models, given plausible lower bounds on initial-condition and large-scale environmental errors. We do so by comparing large numbers of simulated forecasts using the same model but with perturbations to the initial and/or environmental conditions; the magnitudes of those perturbations are designed to be plausible lower bounds given today's knowledge and technology. These are then compared to estimates of real-time forecast skill in the North Atlantic region from 2009 to 2015.

Our work builds on previous research on the fundamental predictability of tropical cyclone intensity, beginning with that of Moskaitis (2009), who used the Coupled Hurricane Intensity Prediction System (CHIPS; Emanuel et al. 2004) to show that North Atlantic tropical cyclone evolution is particularly sensitive to initial and environmental conditions when the ambient shear is close to climatological values; higher or lower shear leads to greater predictability. (In the high-shear cases, storms can be reliably forecast to dissipate.) Zhang and Tao (2013) and Tao and Zhang (2015) showed that tropical cyclones are particularly sensitive to shear in their formative stages, with small differences in shear or initial intensity strongly affecting the timing of subsequent intensification.

TABLE 1. CHIPS ensemble definitions.

Ensemble member	Definition
Control	Analyzed initial intensity; GFS shear
1	Add $3\,\mathrm{m\,s^{-1}}$ (5.8 kt) to initial intensity; GFS shear
2	Subtract $3\,\mathrm{m\,s^{-1}}$ (5.8 kt) from initial intensity; GFS shear
3	Add perturbation to initial rate of intensification
4	Subtract perturbation from initial rate of intensification
5	Add $3\,\mathrm{m\,s^{-1}}$ (5.8 kt) to initial intensity, subtract $5\,\mathrm{m\,s^{-1}}$ (9.7 kt) from shear
6	Subtract $3\,\mathrm{m\,s^{-1}}$ (5.8 kt) from initial intensity, add $5\,\mathrm{m\,s^{-1}}$ (9.7 kt) to shear

2. Error growth in real-time intensity forecasts

We begin with an analysis of real-time tropical cyclone intensity forecasts made for the North Atlantic region during the period 2009–15. The starting year was chosen as the first full year that the CHIPS model made intensity forecasts out to 5 days. We compare those forecasts to intensity predictions by two deterministic modeling systems: that of NOAA's Geophysical Fluid Dynamics Laboratory (GFDL; Kurihara et al. 1998) and a seven-member CHIPS ensemble model run routinely at the Massachusetts Institute of Technology. The definitions of the seven ensemble members are listed in Table 1. The CHIPS intensity forecasts use the NHC forecast tracks and environmental wind shear analyzed and forecast by NOAA's Global Forecasting System[1] (GFS). For the present purpose, we use the arithmetic mean of the seven members. The GFDL and CHIPS models have been relatively stable over the period 2009–15.

As a measure of forecast skill we use the root-mean-square difference between the forecast and analyzed maximum wind speed [knots; $1\,\mathrm{kt} = 0.51\,\mathrm{m\,s^{-1}}$], with the analyzed intensity taken from NHC's historical ("best track") data (Landsea and Franklin 2013). One issue

[1] See http://www.emc.ncep.noaa.gov/index.php?branch=GFS.

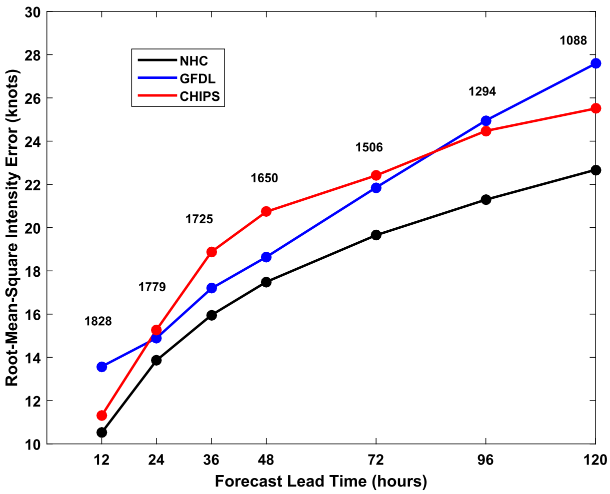

FIG. 1. Root-mean-square forecast intensity errors for North Atlantic tropical cyclones during the period 2009–15. NHC forecasts in black, GFDL in blue, and CHIPS in red. Numbers of forecasts included for each lead time are shown in black above the curves.

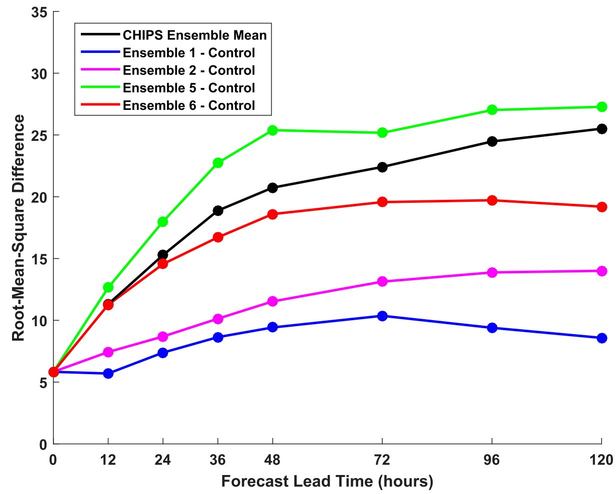

FIG. 2. Root-mean-square differences (kt) between various CHIPS ensemble members and the CHIPS control forecast, indicated by colored curves. The CHIPS ensemble-mean forecast error from Fig. 1 is replotted in black for comparison.

that must be dealt with is the termination of tropical cyclone tracks in the historical data and in the forecasts. In the historical data, tracks may be terminated for a variety of reasons, including dissipation of the storm. Generally, the last data points in historical tracks have wind speeds near 20 kt but there are exceptions. Tropical cyclones simulated by the GFDL and CHIPS models are terminated when their wind speeds drop below about 20 kt. One approach is to calculate intensity error at a given lead time only when, at that lead time, the storm exists in reality, in the official forecasts, and in both model forecasts. While this avoids the termination problem, it fails to account for poor prediction of termination, which is an important practical forecast issue that is not included in official NHC intensity forecast error verification statistics (Cangialosi and Franklin 2013). Here we take a somewhat different approach, calculating intensity error at a given lead time when any one of the real storm, the official forecast, and the models have nonzero intensity. But in this case, assigning a wind speed of zero to a dissipated tropical cyclone will overpenalize a forecast whose timing of dissipation is only mildly in error. To mitigate this problem, we assign an intensity of 20 kt to a storm that has dissipated, when calculating the intensity error. We choose 20 kt as being close to the minimum intensity that is usually carried by models and historical data, before the system is regarded as having dissipated, and this is also close to the mean strength of the background wind in hurricane-prone regions of the tropics. We henceforth refer to this strategy as the "20-kt rule."

Figure 1 shows the root-mean-square intensity error as a function of lead time for the NHC, GFDL, and CHIPS ensemble-mean forecasts from 2009 to 2015. These are homogeneous comparisons, made only when all three forecasts are present, aside from the termination issue discussed above. Bear in mind that intensity estimates of actual Atlantic storms have uncertainty in the range of 8–14 kt, depending on whether the storm has been surveyed by reconnaissance aircraft (Landsea and Franklin 2013).

The errors are similar among the three forecasts, with the CHIPS error small at 12 h but growing more rapidly than the NHC and GFDL forecasts. The error curves are mildly concave, suggesting that error saturation may be achieved at a lead time of 1–2 weeks (which is beyond the lifespan of many tropical cyclones).

The errors shown in Fig. 1 can be regarded as arising from error in the initial intensity (and, in the case of the models, of the initial vortex structure), track, environmental thermodynamic (including upper ocean) and kinematic conditions, and in the case of GFDL and CHIPS, model error (owing to inaccurate physics and/or insufficient grid resolution).

Figure 1 also shows that the CHIPS intensity forecasts are competitive with other guidance, which suggests that the model should be useful for quantifying error sources other than model error.

The divergences between intensity forecasts of ensemble members 1, 2, 5, and 6 (see Table 1) and the control forecast are shown in Fig. 2. (We omit ensemble members 3 and 4, which vary in the initial rate of intensification, as they do not differ greatly from perturbing the intensity itself.) For comparison, we replot the CHIPS ensemble-mean intensity forecast error from Fig. 1. The growth in the root-mean-square differences arise, in this case, only from different initial intensities and, in the case of ensemble members 5 and 6, from shears that differ from the forecast shears by a constant amount.

Errors arising solely from adding $3\,\mathrm{m\,s^{-1}}$ (about 6 kt) to the initial intensity (ensemble member 1) hardly grow at all during the 120-h period, while errors arising from subtracting the same amount from the initial intensity (ensemble member 2) grow a bit faster. This asymmetry may be understood through the regime diagram of tropical cyclone intensity presented by Tang and Emanuel (2010, see their Fig. 5): if a storm is near the stable manifold (top of their Fig. 5), positive perturbations will decay back toward the stable manifold regardless of their magnitude. But a sufficiently large negative perturbation may lead to storm decay and consequent large errors. Weaker storms are also more liable to dissipate entirely, which despite our 20-kt rule, may increase error magnitude.

Adding or subtracting a $5\,\mathrm{m\,s^{-1}}$ perturbation to the magnitude of the environmental wind shear has a larger effect on error growth. Curiously, the dependence on the sign of the initial intensity error is reversed from the case with initial intensity error only: In this case, adding $3\,\mathrm{m\,s^{-1}}$ to the initial intensity while subtracting $5\,\mathrm{m\,s^{-1}}$ from the shear results in larger error growth than the converse. This is perhaps because shear, being a positive definite quantity, will have more nearly a lognormal than a Gaussian probability distribution, so that adding and subtracting the same quantity from the forecast shear samples the probability distribution asymmetrically. (Note that in subtracting $5\,\mathrm{m\,s^{-1}}$ from the forecast shear, the result is not permitted to be negative.)

Even with substantial perturbations to the initial storm intensity and forecast shear, the ensemble divergence is not much larger than the CHIPS ensemble-mean forecast error, which is also driven by model and track error as well as errors in the forecast of the storm's oceanic and atmospheric thermodynamic environment. This suggests that these other sources of error are important in the actual forecast errors.

To further explore the growth of intensity error in a perfect model framework, we now examine the divergence of pairs of CHIPS simulations in which somewhat more realistic perturbations are added to the environmental shear and storm track.

3. Error growth in a tropical cyclone risk assessment model

For the purpose of assessing tropical cyclone risk, the CHIPS model has been adapted to run on a large number of synthesized tropical cyclone tracks, as described in detail in Emanuel et al. (2006) and Emanuel et al. (2008). Here we provide a brief summary of the technique.

The method begins with a global reanalysis or climate model dataset. Monthly mean sea surface temperature and atmospheric temperature and humidity at all pressure levels are calculated together with monthly mean winds at 850 and 250 hPa. The departures of daily winds from their monthly means are also derived at these two levels and used to calculate monthly mean variances and covariances among the four wind components (two at each level). The monthly mean winds and their variances and covariances are used to generate synthetic time series of winds at both levels that have the correct monthly means, variances, and covariances and have power spectra that fall off as frequency cubed, similar to observed flows at synoptic and planetary scales.

To generate tropical cyclone tracks, the ocean basin in question is seeded randomly in space and time with weak protovortices and the synthesized winds are used in a beta-and-advection model to produce the tracks. Then the CHIPS model is run along each track, using the monthly mean thermodynamic environment (linearly interpolated in space and time) and climatological monthly mean upper-ocean thermal conditions. The same winds used to generate the tracks provide the environmental shear that is an important component of the intensity model.

The large majority of storms thus generated fail to intensify and are discarded. Only those cyclones that reach an intensity of at least 40 kt are retained. The technique can therefore be regarded as working on the principle of natural selection.

Applications of this technique to the current climate have been rigorously compared to historical tropical cyclone climatologies (Emanuel et al. 2006). The spatial distribution, seasonal cycle, and response to ENSO are all well simulated by this technique.

For the purpose at hand, we begin by generating a control set of 3100 events in the North Atlantic region, downscaled from NCAR–NCEP reanalyses (Kalnay et al. 1996). In the standard application of the technique to risk assessment, the initial vortices have intensities ranging from 9 to 18 kt. But since we are trying to mimic a forecast system here, we apply initial intensities varying randomly between 10 and 110 kt. We then go back and regenerate the tropical cyclones using the same sequence of random numbers (governing the position and date of the seeds and the initial intensity of the storms) but with small perturbations introduced as summarized in Table 2.

In the first instance, we add 3 kt to the initial intensity of the vortex. This is somewhat smaller than current

TABLE 2. Description of perturbation experiments.

Perturbation expt	Description	No. of overlapping cases
Initial intensity only	Added 3 kt to initial intensity.	2711
Shear only	Shear decorrelates from control over 25 days. Track and initial intensity identical to control.	2916
Shear + initial intensity	Added 3 kt to initial intensity and shear decorrelates from control over 25 days. Track identical to control.	2656
Track	Winds (affecting shear and steering) decorrelate from control over 25 days. Tracks respond to changing steering flow.	2204
Track + initial intensity	As in Track, but added 3 kt to initial intensity.	2051

initial intensity error estimates (Landsea and Franklin 2013) and is provided here as a very optimistic lower bound. Everything else is identical to the control, including the track, thermodynamic, and kinematic environments.

We generate a second set of perturbed events by allowing the winds to slowly decorrelate from the control experiment. This is done by adding an additional random phase to the randomly varying component of the environmental winds (generated using the wind variances and covariances), which grows from zero at the initial time to $\pi/2$ after 25 days. This time scale was chosen as a mildly optimistic estimate of the predictability horizon for large-scale atmospheric flows, given recent estimates of 16–23 days for global predictability (Buizza and Leutbecher 2015).

In this second perturbation experiment, the track is fixed to be identical to the control but the decorrelating winds affect the wind shear used in the CHIPS intensity model. Everything else, including the initial intensity, is also identical to the control. Thus this second experiment focuses on the particular effect of errors in the shear forecast.

The third perturbation experiment is identical to the second but also adds 3 kt to the initial storm intensity.

In the fourth perturbation experiment, the decorrelating winds affect the track of the storm, so that the large-scale environment experienced by the storm (including its thermodynamic environment) drifts away from the control as the storm positions separate. (This is in addition to the Eulerian decorrelation of the wind shear.)

The final perturbation experiment is identical to the previous experiment but 3 kt is also added to the initial storm intensity. This last experiment is meant to encompass most of what affects real forecast error, aside from model error. As such, it may provide a realistic target for tropical cyclone intensity forecast improvement.

When running the perturbation experiments, some of the initial seed vortices that survived in the control experiment do not survive in the perturbation experiment and vice versa. We therefore must identify the overlapping sets of survivors before comparing the control and perturbation tracks. The number of such overlapping events is indicated for each perturbation experiment in Table 2. But bear in mind that the tracks eventually terminate, so the number of events diminishes with lead time. (The errors described here are only calculated, at each lead time, when at least one track exists. We apply to the dissipated events the same 20-kt rule as used to evaluate real-time forecasts.) The evolution of the number of overlapping events as a function of lead time is shown in Fig. 3.

Several real sources of intensity forecast error, besides model error, are not considered here. In each experiment, the initial positions of the control and perturbation vortices are identical, so initial location error is not considered. Errors in forecasting the thermodynamic environment, including upper-ocean temperature, are not considered here except insofar as perturbed tracks take storms into different thermodynamic environments. In our experience, potential intensity in the tropics is slowly varying, although the same cannot be said for midtropospheric humidity, which plays a role in the intensity model. While upper-ocean temperature is

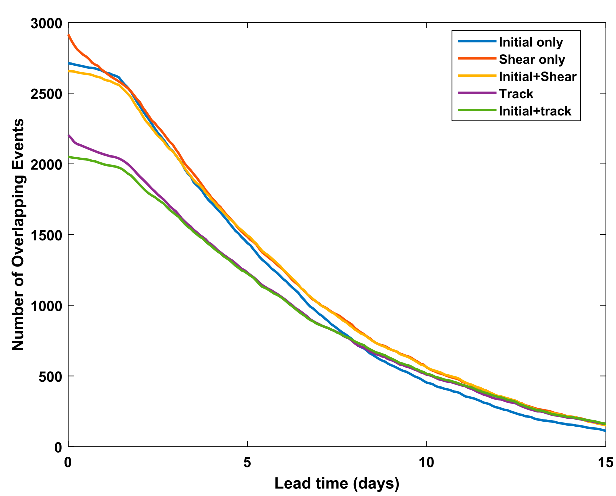

FIG. 3. Number of overlapping events for each perturbation experiment as a function of lead time.

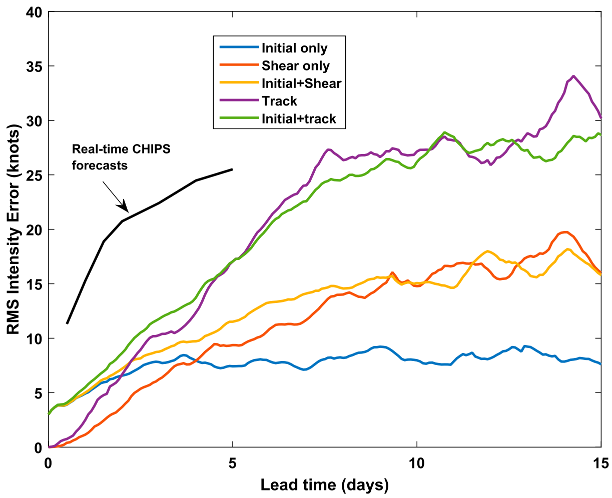

FIG. 4. Root-mean-square intensity differences between the control experiment and five perturbation experiments described in Table 2. For comparison, the black curve displays the CHIPS ensemble-mean real-time forecast errors (same as red curve in Fig. 1).

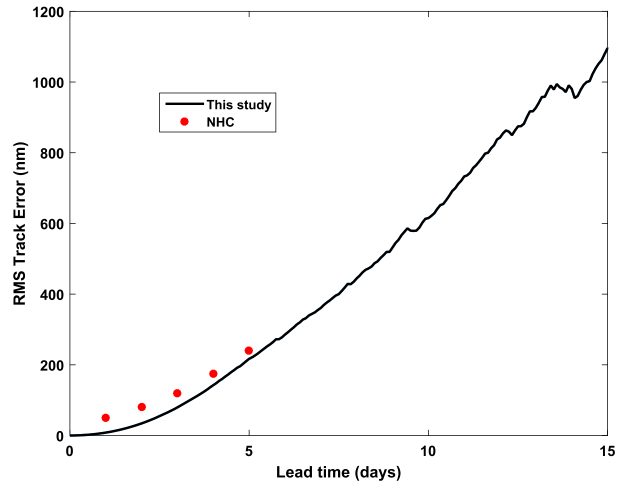

FIG. 5. Track error owing to decorrelating environmental steering flow in the present study compared to actual Atlantic track forecast error over the period 2009–15 according to NHC.

not likely to vary much over the typical lifetime of tropical cyclones, in reality, limited observations of the subsurface ocean can lead to errors in upper-ocean temperature analyses, leading through atmosphere–ocean interaction to errors in tropical cyclone intensity (e.g., Lin et al. 2005). The advent of sea surface altimetry and robotic floats should, however, lead to improved real-time analyses of the upper ocean.

The evolutions with time of the intensity error of each of the perturbation experiments described in Table 2 are shown in Fig. 4. Intensity errors stemming solely from initial intensity error double in not much more than a day and reach saturation values of around 8 kt after about 4 days. (Not surprisingly, these value are comparable to the differences between real-time CHIPS ensemble members 1 and 2 and the control forecasts, shown in Fig. 2.) These saturation values are not very different from current estimates of the accuracy of estimating real tropical cyclone intensity (Landsea and Franklin 2013), suggesting that initial intensity errors are not currently the most important source of error growth, although their large magnitude insures that they dominate forecast errors at early lead times.

Errors owing to incorrectly forecasted shear grow faster, reaching saturation values of 15–20 kt only after 10–15 days and overtaking errors arising from initial intensity error alone after about 4 days.

As the perturbed environmental winds deviate over time from those of the control, the storms move into large-scale kinematic and thermodynamic environments that differ from the control, leading to intensity error, as shown by the magenta curve in Fig. 4. This overtakes the errors that arise from initial intensity error alone (blue curve) after about 2 days. When storms are initialized with intensity error and also subject to track error (green curve), the resulting error grows faster than that owing to initial intensity error alone.

Figure 5 shows that the track error itself is not unrealistic when compared to Atlantic tropical cyclone track errors tallied by NHC over the period 2009–15.[2]

We regard the green curve in Fig. 4 as an estimate of lowest achievable intensity error given a perfect model, 3-kt initial intensity error, and large-scale wind errors that we might expect from global forecasting systems after a decade or so if they continue to improve at the current rate. This also assumes perfect knowledge of environmental thermal conditions, including the temperature of the upper ocean, and of environmental humidity above the boundary layer. As such, the green curve in Fig. 4 is a fairly optimistic estimate of the achievable predictability horizon for tropical cyclone intensity.

Also reproduced in Fig. 4 are the real-time CHIPS ensemble-mean forecast errors (identical to the red curve in Fig. 1). Insofar as the CHIPS ensemble-mean forecast error is not too much worse than other currently available guidance (see Fig. 1), the gap between the black and green curves in Fig. 4 represents the difference between current forecast skill and what is theoretically achievable with a perfect model, with very good observations of initial tropical cyclone intensity and structure and of upper-ocean properties and the means of

[2] See http://www.nhc.noaa.gov/verification/figs/ALtkerrtrd_noTD.jpg.

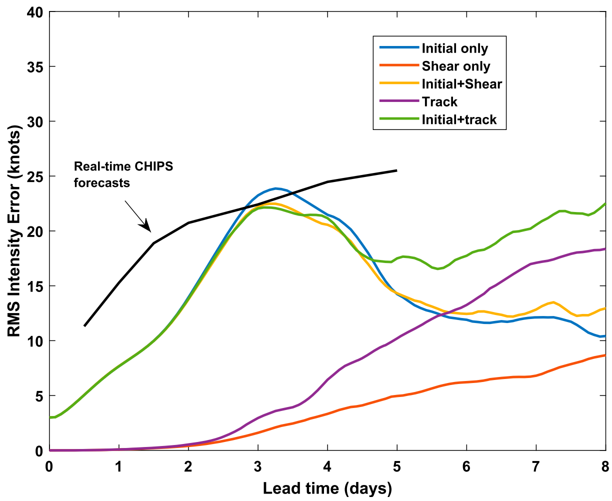

FIG. 6. As in Fig. 3, but for initial intensities between 9 and 18 kt. Note that, in contrast to Fig. 3, results are displayed out to a lead time of only 8 days.

assimilating those observations into the forecast model, and with realistically achievable forecast skill of large-scale environmental conditions.

We perform one variant on the set of perturbation experiments described above, taking the initial protovortices to have intensities ranging between 9 and 18 kt (rather than between 10 and 110 kt). Otherwise, the experiments are identical to those listed in Table 2. The numbers of initially overlapping cases for this set of experiments, corresponding to the rightmost column in Table 2 are, respectively, 505, 2600, 511, 2173, and 476. The smaller number of overlapping events stems from the greater susceptibility of the initially very weak vortices to environmental influences.

The error growth in the case of weak initial intensities is shown in Fig. 6. In this case, we show results only out to lead times of 8 days because the number of overlapping events that survive beyond 8 days, in the case of perturbed initial intensity, is too small to obtain robust statistics.

Error growth owing to initial intensity error is much more rapid in this case and dominates the total perfect model forecast error out to about 5 days. We interpret this as stemming from the greater potential for rapid intensification of initially weak storms compared to storms that are already not far from their potential intensities and to the great sensitivity of the timing of rapid intensification to shear when the initial intensity is small (Zhang and Tao 2013; Tao and Zhang 2015). This suggests that future work might focus on predictability of rapidly intensifying events and/or on events that achieve major hurricane intensity.

4. Summary

Errors in tropical cyclone intensity forecasts arise from myriad sources. This study addresses the relative importance of these sources in a "perfect model" framework, by comparing control and perturbation experiments in which various error sources (other than model error) can be examined individually and in combination. We used the CHIPS model, which while not as skillful as more comprehensive models, can be run many thousands of times and whose real-time forecast skill is not much worse than other guidance (Fig. 1).

Here we focused on forecast intensity errors that arise from errors in initial intensity, forecast environmental wind shear, and storm track. These errors were chosen to be small, with initial intensity error of 3 kt and environmental-flow errors saturating after 25 days, to provide an optimistic upper bound on achievable forecast skill. We did not examine errors caused by initial mislocation of the initial storm or by errors in the forecast atmospheric and oceanic thermodynamic environment, except insofar as track errors cause storms to move into different large-scale environments.

Our results indicate that errors arising from mischaracterization of the initial intensity of the tropical cyclone dominate intensity errors for the first 2–4 days, especially when the initial intensity is small. (We did not examine the effects of initial structural error, as the specification of the radius of maximum wind is effectively the only free structural parameter in CHIPS, but this source of error may be important in real storms.) The high sensitivity to initial intensity error when the initial intensity is small may be owing to the striking sensitivity of the timing of rapid intensification of weak storms to shear documented by Zhang and Tao (2013) and Tao and Zhang (2015).

Beyond 2–4 days, intensity error is increasingly dominated by track error, which in these experiments is owing to errors in environmental winds, which also affect the shear experienced by the storm. We note here that this error source should be less problematic in probabilistic, ensemble-based forecasts for fixed points in space, simply because those ensemble members that affect the point in question will necessarily have approximately the correct steering flow. (Put another way, the intensity error of those ensemble members that do not affect the point in question will anyway be irrelevant to wind probabilities at that point.) These points are discussed in more detail by Zhang and Weng (2015).

Aside from a single set of experiments in which all the initial cyclones are weak (see Fig. 6), we did not explore the sensitivities of forecast error to initial intensity or size, or to translation speed, which the analysis of Zhang et al. (2014) suggests can be important. This is left to future work.

The large gap between today's intensity forecast skill and what is theoretically achievable with a perfect model under reasonably small initial-intensity and large-scale forecast environmental errors (the gap between the black and green curves in Fig. 4) suggests that much improvement is possible at leads times up to about a week. Such improvement will rely on better models and much better characterization of the initial storm through better observations and improved means of assimilating them into the forecast model. Our results also suggest that large ensembles consisting of a diversity of models and initial vortex states and, for larger lead times, different environmental-flow evolutions, would be valuable for quantifying intensity forecast uncertainty.

Acknowledgments. The first author gratefully acknowledges support from ONR through Grant N000141410062. This study was initiated during the second author's sabbatical visit at MIT during fall 2015 sponsored by the Houghton Lecturer Fund, as well as NSF Grant AGS 1305798 and ONR Grant N000140910526. We thank Chris Landsea and two anonymous reviewers for very helpful comments.

REFERENCES

Andreas, E. L, L. Mahrt, and D. Vickers, 2015: An improved bulk air–sea surface flux algorithm, including spray-mediated transfer. *Quart. J. Roy. Meteor. Soc.*, **141**, 642–654, doi:10.1002/qj.2424.

Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141**, 3366–3382, doi:10.1002/qj.2619.

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

DeMaria, M., 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.

Emanuel, K., C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci.*, **61**, 843–858, doi:10.1175/1520-0469(2004)061<0843:ECOTCI>2.0.CO;2.

——, S. Ravela, E. Vivant, and C. Risi, 2006: A statistical deterministic approach to hurricane risk assessment. *Bull. Amer. Meteor. Soc.*, **87**, 299–314, doi:10.1175/BAMS-87-3-299.

——, R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc.*, **89**, 347–367, doi:10.1175/BAMS-89-3-347.

Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329–343, doi:10.1175/BAMS-D-12-00071.1.

Gopalakrishnan, S. G., F. Marks, X. Zhang, J.-W. Bao, K.-S. Yeh, and R. Atlas, 2011: The Experimental HWRF System: A study on the influence of horizontal resolution on the structure and intensity changes in tropical cyclones using an idealized framework. *Mon. Wea. Rev.*, **139**, 1762–1784, doi:10.1175/2010MWR3535.1.

Green, B. W., and F. Zhang, 2014: Sensitivity of tropical cyclone simulations to parametric uncertainties in air–sea fluxes and implications for parameter estimation. *Mon. Wea. Rev.*, **142**, 2290–2308, doi:10.1175/MWR-D-13-00208.1.

——, and ——, 2015: Numerical simulations of Hurricane Katrina (2005) in the turbulent gray zone. *J. Adv. Model. Earth Syst.*, **7**, 142–161, doi:10.1002/2014MS000399.

Judt, F., S. S. Chen, and J. Berner, 2016: Predictability of tropical cyclone intensity: Scale-dependent forecast error growth in high-resolution stochastic kinetic-energy backscatter ensembles. *Quart. J. Roy. Meteor. Soc.*, **142**, 43–57, doi:10.1002/qj.2626.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kaplan, J., and Coauthors, 2015: Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Wea. Forecasting*, **30**, 1374–1396, doi:10.1175/WAF-D-15-0032.1.

Kurihara, Y., R. E. Tuleya, and M. A. Bender, 1998: The GFDL Hurricane Prediction System and its performance in the 1995 hurricane season. *Mon. Wea. Rev.*, **126**, 1306–1322, doi:10.1175/1520-0493(1998)126<1306:TGHPSA>2.0.CO;2.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, doi:10.1175/MWR-D-12-00254.1.

Lin, I.-I., C.-C. Wu, K. Emanuel, I.-H. Lee, C.-R. Wu, and I.-F. Pun, 2005: The interaction of Supertyphoon Maemi (2003) with a warm ocean eddy. *Mon. Wea. Rev.*, **133**, 2635–2649, doi:10.1175/MWR3005.1.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, doi:10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2.

Moon, I.-J., I. Ginis, T. Hara, and B. Thomas, 2007: A physics-based parameterization of air–sea momentum flux at high wind speeds and its impact on hurricane intensity predictions. *Mon. Wea. Rev.*, **135**, 2869–2878, doi:10.1175/MWR3432.1.

Moskaitis, J. R., 2009: Toward improved tropical cyclone intensity forecasts: Probabilistic prediction, predictability, and the role of verification. Ph.D. dissertation, Massachusetts Institute of Technology, 214 pp. [Available online at http://dspace.mit.edu/handle/1721.1/47846.]

Nolan, D. S., J. A. Zhang, and D. P. Stern, 2009: Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of in situ observations and high-resolution simulations of Hurricane Isabel (2003). Part I: Initialization, maximum winds, and the outer-core boundary layer. *Mon. Wea. Rev.*, **137**, 3651–3674, doi:10.1175/2009MWR2785.1.

Palmer, T. N., A. Döring, and G. Seregin, 2014: The real butterfly effect. *Nonlinearity*, **27**, R123–R141, doi:10.1088/0951-7715/27/9/R123.

Ravela, S., K. Emanuel, and D. McLaughlin, 2007: Data assimilation by field alignment. *Physica D*, **230**, 127–145, doi:10.1016/j.physd.2006.09.035.

Rotunno, R., and C. Snyder, 2008: A generalization of Lorenz's model for the predictability of flows with many scales of motion. *J. Atmos. Sci.*, **65**, 1063–1075, doi:10.1175/2007JAS2449.1.

——, Y. Chen, W. Wang, C. Davis, J. Dudhia, and C. L. Holland, 2009: Large-eddy simulation of an idealized tropical cyclone. *Bull. Amer. Meteor. Soc.*, **90**, 1783–1788, doi:10.1175/2009BAMS2884.1.

Ruf, C. S., and Coauthors, 2016: New ocean winds satellite mission to probe hurricanes and tropical convection. *Bull. Amer. Meteor. Soc.*, **97**, 385–395, doi:10.1175/BAMS-D-14-00218.1.

Sun, Y. Q., and F. Zhang, 2016: Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect. *J. Atmos. Sci.*, **73**, 1419–1438, doi:10.1175/JAS-D-15-0142.1.

Tang, B., and K. Emanuel, 2010: Midlevel ventilation's constraint on tropical cyclone intensity. *J. Atmos. Sci.*, **67**, 1817–1830, doi:10.1175/2010JAS3318.1.

Tao, D., and F. Zhang, 2015: Effects of vertical wind shear on the predictability of tropical cyclones: Practical versus intrinsic limit. *J. Adv. Model. Earth Syst.*, **7**, 1534–1553, doi:10.1002/2015MS000474.

Weng, Y., and F. Zhang, 2016: Advances in convection-permitting tropical cyclone analysis and prediction through EnKF assimilation of reconnaissance aircraft observations. *J. Meteor. Soc. Japan*, doi:10.2151/jmsj.2016-018, in press.

Yablonsky, R. M., and I. Ginis, 2009: Limitation of one-dimensional ocean models for coupled hurricane–ocean model forecasts. *Mon. Wea. Rev.*, **137**, 4410–4419, doi:10.1175/2009MWR2863.1.

Zhang, F., and D. Tao, 2013: Effects of vertical wind shear on the predictability of tropical cyclones. *J. Atmos. Sci.*, **70**, 975–983, doi:10.1175/JAS-D-12-0133.1.

——, and Y. Weng, 2015: Predicting hurricane intensity and associated hazards. *Bull. Amer. Meteor. Soc.*, **96**, 25–33, doi:10.1175/BAMS-D-13-00231.1.

——, N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594, doi:10.1175/JAS4028.1.

——, M. Minimide, and E. E. Clothiaux, 2016: Potential impacts of assimilating all-sky satellite radiances from GOES-R on convection-permitting analysis and prediction of tropical cyclones through EnKF assimilation of reconnaissance aircraft observations. *Geophys. Res. Lett.*, **43**, 2954–2963, doi:10.1002/2016GL068468.

Zhang, Y., Z. Meng, F. Zhang, and Y. Weng, 2014: Predictability of tropical cyclone intensity evaluated through 5-yr forecasts with a convection-permitting regional-scale model in the Atlantic basin. *Wea. Forecasting*, **29**, 1003–1022, doi:10.1175/WAF-D-13-00085.1.