



# On the growth of intensity forecast errors in the operational hurricane weather research and forecasting (HWRF) model

Chanh Kieu<sup>1</sup>  | Kushal Keshavamurthy<sup>1</sup>  | Vijay Tallapragada<sup>2</sup> | Sundararaman Gopalakrishnan<sup>3</sup> | Samuel Trahan<sup>2</sup>

<sup>1</sup>Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, Indiana, USA

<sup>2</sup>Environmental Modeling Center, NWS/NOAA/NCEP, College Park, Maryland, USA

<sup>3</sup>Atlantic Oceanographic and Meteorological Laboratory, NOAA Hurricane Research Division, Miami, Florida, USA

## Correspondence

Chanh Kieu, Department of Earth and Atmospheric Sciences, GY428A Geological Building, Indiana University, Bloomington, Indiana 47405, USA.  
Email: ckieu@indiana.edu

This study examines the growth of tropical cyclone (TC) intensity forecast errors and related intensity predictability for the NOAA operational Hurricane Weather Research and Forecasting (HWRF) model. Using operational intensity forecasts during the 2012 to 2016 seasons, two conditions for a limited range of TC intensity predictability are demonstrated, which include (a) the existence of an intensity error saturation limit, and (b) the dependence of the intensity error growth rate on storm intensity during TC development. By stratifying intensity errors based on different initial intensity bins, it is shown that TC intensity error growth rate is relatively small ( $\sim 0.3 \text{ kt h}^{-1}$ ) at the early stage of TC development, but it quickly increases to  $\sim 1 \text{ kt h}^{-1}$  during TC intensification. Of further importance is that the intensity error saturation varies in the range of 14–18 kt in different ocean basins, thus suggesting the potential dependence of the intensity predictability on large-scale environment. Additional idealized experiments with the HWRF model confirm the saturation of intensity errors, even under a perfect model scenario. The existence of the intensity error saturation together with the finding of a faster error growth rate for higher intensity suggests that the TC dynamics possesses an inherent limited predictability, which prevents us from reducing the intensity errors in TC dynamical models below a certain threshold.

## KEYWORDS

hurricane intensity, HWRF model, intensity error growth, intensity forecasting, practical predictability, tropical cyclones

## 1 | INTRODUCTION

The predictability of tropical cyclone (TC) intensity is an open problem. Similar to many atmospheric systems, TC dynamics may possess some inherent chaotic properties that are not fully understood at present due to various nonlinear dynamic and thermodynamic feedbacks. Among several difficulties related to the definition of the TC intensity predictability, the main obstacle to our current understanding of intensity predictability lies perhaps in the fact that this problem is practically time-dependent. That is, one has to examine the growth of intensity forecast errors at all stages of TC development, starting from the early tropical depression stage to the final dissipation stage, for which the traditional predictability

formalism based on the stationary statistics cannot be applied. Thus, examining the characteristics of TC intensity errors at different stages of TC development is necessary to quantify the TC intensity predictability for operational applications.

Except for stochastic forcings, TC intensity forecast errors in dynamical models are generally attributed to two major factors, which include initial condition errors and model errors (Davis *et al.*, 2010; Gopalakrishnan *et al.*, 2012; Tallapragada *et al.*, 2014; Judt *et al.*, 2016). For regional dynamical models, lateral boundary conditions provided by a global forecasting system introduce another source of uncertainty, which could lead to large intensity differences if a model storm is guided into a wrong environment, even with a perfect regional model

and initial conditions (e.g. DeMaria, 2010; Tien *et al.*, 2013). These sources of error strongly evolve in time due to the non-linear nature of the atmosphere, and it is difficult to separate their relative roles in real-time intensity forecasts.

The problem of intensity errors is further complicated by the fact that TCs are multi-scale systems in which errors at different scales grow differently. In Lorenz's predictability framework, errors at the convective scale quickly grow, saturate, and then propagate to larger scales until the entire spectrum of errors is saturated, which underline the limited predictability of atmospheric multi-scale flows (e.g. Lorenz, 1969; 1990; Leith, 1971; Rotunno and Snyder, 2007; Palmer *et al.*, 2014; Durran and Gingrich, 2014). This predictability paradigm, the so-called statistical predictability in Métais and Lesieur (1986), is based on a fundamental property that the background flow is statistically homogeneous and stationary so that errors can saturate after a certain time. However, for TCs, such a stationary background spectrum is not applicable, because TCs always evolve with time in favourable environmental conditions until they reach the maximum intensity limit. (Kieu, 2015; Kieu and Wang, 2017). A study of a two-dimensional turbulence model by Métais and Lesieur (1986) showed indeed that the predictability range in their turbulent model could change by as much as 50% by simply relaxing the stationary assumption for the bounding error curve. This sensitivity of predictability on the background dynamics thus highlights the key difficulty in defining the TC intensity predictability for practical purposes.

Because of this non-stationary nature of TC background dynamics, our approach to understand the TC intensity error growth during TC development must be different from that during the quasi-stationary stage as examined in Hakim (2013). Specifically, one needs to rely on the so-called transient<sup>1</sup> (or non-central) property of a deterministic dynamical system to examine the growth of intensity errors during the TC intensifying period (Lorenz, 1963). In this deterministic framework, the predictability of a dynamical system can be studied using the familiar properties of chaotic dynamics such as the existence of positive leading exponents, or the boundedness and denseness of chaotic attractors.

The current practice of using point-like metrics such as the maximum sustained 10 m wind ( $V_{MAX}$ ) or the minimum central pressure ( $P_{MIN}$ ) to measure TC intensity offers a unique context to examine the characteristics of intensity errors from the perspective of deterministic dynamics. Indeed, the verification of TC intensity forecast errors shows intriguingly that TC intensity errors do appear to approach a saturation

limit regardless of the ocean basins or modelling systems (Franklin and Brown, 2008; DeMaria, 2010; Tallapragada *et al.*, 2014; 2015; Bhatia and Nolan, 2015; 2016; Kieu and Moon, 2016, hereinafter KM16). For example, the 4–5 day intensity error saturation in operational numerical models has stayed around 15–17 kt in the Eastern Pacific (EPAC), ~14–16 kt in the North Atlantic (NATL), or ~16–18 kt in the Northwestern Pacific (WPAC) basin (KM16). Although the intensity forecast skill has been gradually improved over last several decades, the reduction in intensity errors in operational dynamical models turns out to be much smaller than the track error reduction. The persistence of this small reduction in the intensity error saturation among different modelling systems and ocean basins therefore raises a natural question of whether and why TC intensity errors approach a saturation limit rather than linearly increasing with forecast lead times.

In the context of idealized simulations, KM16 addressed this question by examining the property of the TC intensity at the maximum potential intensity ( $MPI^2$ ) limit from the perspective of deterministic dynamics. Using Rotunno and Emanuel (1987)'s axisymmetric model, KM16 demonstrated the existence of a chaotic attractor at the MPI equilibrium that possesses an intrinsic variation of ~16 kt in a long integration under the perfect model assumption. Similar to the Lorenz three-variable model, the sensitive dependence of the TC intensity error growth on initial conditions inside this bounded MPI attractor imposes a limit on our ability to reduce the  $V_{MAX}$  errors at 4–5 day lead times. KM16's further estimation of the predictability range for TC intensity forecasts based on the  $V_{MAX}$  metric suggests that the range of the TC intensity predictability is around 3 days, and it may be shorter for initially stronger TCs. This range of predictability is consistent with the estimation in Hakim (2013) and Brown and Hakim (2013), which is based on the analyses of both the inverse linear and nonlinear models at the quasi-stationary equilibrium.

If the above chaotic nature of TC dynamics proposed in KM16 is correct, an immediate consequence is that the intensity error growth should be different between the mature stage and the transient stage (i.e. the intensification period) for which analyses of the growth rate inside the MPI attractor in KM16 could not fully capture. As a step to further examine the TC intensity predictability and the TC intensity error growth during the intensification period, analyses of real-time intensity error growth rate for two operational models currently maintained by the NOAA National Centers for Environmental Prediction (NCEP) – the Hurricane Weather Research and Forecasting (HWRF) and the Geophysical Fluid Dynamics Laboratory (GFDL) models – will be presented in this study. Our main objectives of this study are (a) to quantify the dependence of the intensity error growth rate on the storm intensity

<sup>1</sup>We follow the Lorenz (1963) definition and use the phrase “transient stage” to refer to the non-central trajectories of a dynamical system that are approaching but are not contained in a limiting trajectory. By this definition, the transient stage of TC development would correspond to the intensification stage of TCs. Note that this transient property, which is related to the non-central trajectories, differs from the transient growth of errors along an orbit inside an attractor, which is represented by local Lyapunov vectors (Lorenz, 1996; Trevisan and Pancotti, 1998).

<sup>2</sup>In this study, the MPI limit is understood as the maximum tangential wind in a statistical quasi-stationary sense rather than an instantaneous maximum value (Hakim, 2011)

during TC development, and (b) to examine how the intensity error saturation varies in different basins and the possible implications of this error saturation for practical intensity predictability (PIP). Addressing these issues will help confirm the nature of TC chaotic dynamics that is not currently well understood, thus revealing the potential limit of operational models in reducing intensity errors in future.

The rest of this work will be presented as follows: section 2 presents necessary conditions for the TC intensity to have a limited predictability. Section 3 discusses the methodology to obtain intensity error growth rates from real-time forecasts at different stages of TC development. Analyses of intensity errors and the related growth rate are then presented in section 4. Section 5 discusses results from idealized experiments with the HWRf model under a perfect model scenario, and some concluding remarks are given in the final section.

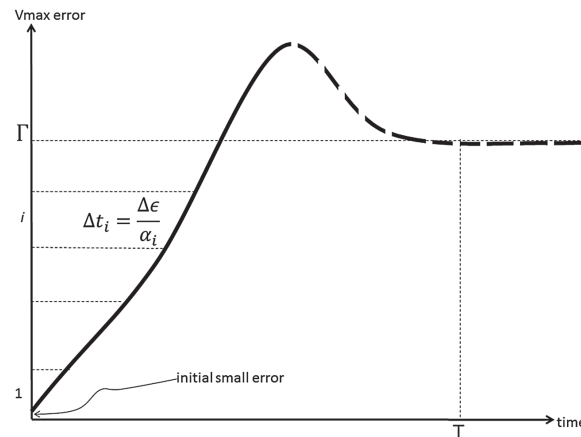
## 2 | CONDITIONS FOR TC INTENSITY LIMITED PREDICTABILITY

Because our examination of the TC intensity predictability in this study is from the perspective of a deterministic framework, an important requirement to be considered for any chaotic dynamical system is the existence of a bounded attractor whose size determines the averaged differences between two randomly initial states. In Lorenz's three-variable model, the existence of this saturation limit is guaranteed by the Lorenz attractor that any initial condition will be quickly pulled in and subsequently trapped inside this attractor. Together with the sensitive dependence on the initial condition (i.e. a positive leading Lyapunov exponent), this bounded property of the Lorenz attractor leads to a strong consequence that the time interval  $T$  required for an initial error  $\epsilon_0$  to grow and approach the saturation limit  $\Gamma$  will be given by

$$T \sim \ln \frac{\Gamma}{\epsilon_0} \quad (1)$$

As seen from Equation 1, the saturation limit  $\Gamma$  severely constrains our ability to lengthen the range of predictability  $T$ ; a 10-time reduction in the initial error  $\epsilon_0$  will only help lengthen the range of predictability by just a factor of  $\ln(10) \sim 2.3$ . As such, the existence of a stationary error saturation limit, which represents the size of the attractor, is the first condition that one needs to quantify when studying the predictability limit for TC intensity. Whether an error saturation limit exists for the TC intensity problem and, if so, how this intensity error saturation varies with space and time, is the first condition that we have to determine.

Assuming the existence of an error saturation  $\Gamma$  for TC intensity, a second important issue is how fast an initial error will approach this saturation limit. In principle, one could have a system in which an initial error could grow extremely slowly such that it would take forever to reach saturation. In this case, the range of predictability is so long that the



**FIGURE 1** Illustration of the intensity error evolution during the developing stage (black solid curve) for an initial small error at  $t = 0$ . In the interval  $i$ , the growth rate is assumed to be  $\alpha_i$ , and  $\Delta t_i$  denotes the time required for the error in the  $i$ th interval to increase by an increment  $\Delta \epsilon$ . Here,  $T$  denotes the total time that the initial error would take to asymptotically approach the saturation limit  $\Gamma$ . The thick dashed line represents the duration when the growth rate slows down as the error approaches the saturation limit  $\Gamma$  inside the MPI attractor. Note that the overshooting of the  $V_{\text{MAX}}$  error is to highlight the error growth during the transient stage, which differs from the error growth inside the attractor

system can be technically considered predictable. From the dynamical standpoint, the error growth required for a system to possess limited predictability therefore needs to satisfy certain properties, which are ultimately related to the existence of a positive leading Lyapunov exponent. For the TC intensity forecast problem, such a requirement of a positive Lyapunov exponent is, nonetheless, not applicable during the intensification stage, because the Lyapunov exponents are not defined outside the attractor (Lorenz, 1996; Legras and Vautard, 1996; Trevisan and Pancotti, 1998). Thus, another condition for the intensity error growth rate during the TC intensification stage needs to be defined differently.

To facilitate our subsequent analyses, we establish in this section a requirement for the TC intensity error growth rate during the transient stage that could ensure the limited predictability. Specifically, we assert that a necessary condition for an initial intensity error to approach an error saturation  $\Gamma$  in a finite time is that the error growth rate has to be larger for a higher intensity during TC development. Indeed, let us divide the range of the error saturation limit  $[0, \Gamma]$  into  $N$  intervals with an increment of error  $\Delta \epsilon = \Gamma/N$  as illustrated in Figure 1. Assume that the error growth rate within each interval is  $\alpha_i$ , the time interval  $\Delta t_i$  that the error will grow in an interval  $i$  is then given by  $\Delta t_i = \Delta \epsilon / \alpha_i$ . Thus, the total time for the initial error  $\epsilon_0$  to reach  $\Gamma$  is

$$T = \frac{\Delta \epsilon}{\alpha_1} + \frac{\Delta \epsilon}{\alpha_2} + \dots + \frac{\Delta \epsilon}{\alpha_N} = \frac{\Delta \epsilon}{\alpha_0} \sum_{i=1}^{i=N} \frac{1}{\lambda^i}, \quad (2)$$

where we have assumed that the successive growth rate  $\alpha_i = \lambda \alpha_{i-1} = \dots = \lambda^i \alpha_0$ . Apparently, the geometrical series (2) will converge to a finite value for  $N \rightarrow \infty$  if and only if  $\lambda^{-1} < 1$  or equivalently  $\lambda > 1$ . This implies that the error

growth rate at the later time  $\alpha_i$  has to be faster than the error growth rate at the previous time  $\alpha_{i-1}$  so that  $T$  is finite. Otherwise, the series (2) will not converge; i.e. there would exist an orbit that is stable at some point along the orbit such that any small initial error starting at this point would subside over time, thus precluding the existence of a chaotic attractor (see Lorenz (1963) for a more precise definition of stable orbits and related properties). Because TCs strengthen with time during their intensifying period, the condition  $\lambda > 1$  justifies our above statement about the faster intensity error growth rate for higher intensity.

We stress that our aforementioned statement will not hold for all stages of TC development. This is because the growth rate has to eventually slow down as the error approaches a saturation limit. As a result, our statement about a faster intensity error growth rate for higher intensity should be applied only to the intensifying period. After a storm reaches its mature stage, our statement should be no longer valid (dashed line in Figure 1).

In summary, we propose two conditions that one needs to verify if the TC intensity predictability is limited:

- (a) the existence of a stationary intensity error saturation, and
- (b) a faster intensity error growth rate for higher intensity during TC intensification.

Our hypothesis is that these conditions are intrinsic properties of TC dynamics, which underline the characteristic of the intensity error curve in current operational TC models, regardless of how good the models are. This is an important hypothesis that needs to be carefully verified, because it would imply that a part of the current intensity error saturation is not rooted in model deficiencies but actually in the TC dynamics. Thus, the ability to improve TC intensity forecast accuracy will have some barrier that we may not overcome in future, even if all favourable environmental conditions are maintained and the TC models are perfect.

### 3 | METHODOLOGY

#### 3.1 | Real-time intensity analysis

With the two conditions required for examining the TC predictability as established in section 2, we present in this section our methodology to quantify these conditions from real-time forecasts. Unlike idealized simulations for which one can readily control environmental factors governing TC development, real-time TC intensity forecasts suffer from several issues that make them difficult to directly verify the above conditions and so some specific assumptions must be made.

First, real-time forecasts contain all cycles with different initial intensity at all stages of TC development. Therefore, general intensity error statistics will include both the intensifying and the weakening phases, which do not characterize the intrinsic TC intensity error growth during the intensification

stage as needed. This is because any change in TC intensity due to the movement of TCs into a different environment or making a landfall would not represent the intrinsic growth of TC intensity, but a manifestation of changes in the large-scale environment. For example, a weakening phase of a storm can be well predicted if its landfalling track is correctly forecast. Such a track-dependent intensity change represents the dependence of TC intensity on large-scale conditions rather than the TC intrinsic intensity predictability (IIP) that we wish to examine in this study.

Second, real-time forecasts contain influences from many external factors such as dry air intrusion, vertical wind shear, or upper-level trough interactions which may produce intensity fluctuations sufficiently larger than those associated with the IIP. There is no effective way to isolate these external factors in real-time intensity forecasts, except in the idealized framework. Because of these uncontrollable factors, any predictability information extracted from real-time forecasts should represent the practical predictability of TC intensity rather than the IIP. While our ultimate goal is to determine the IIP, the modest objective of this study is to maximally extract the information of the PIP from the real-time data.

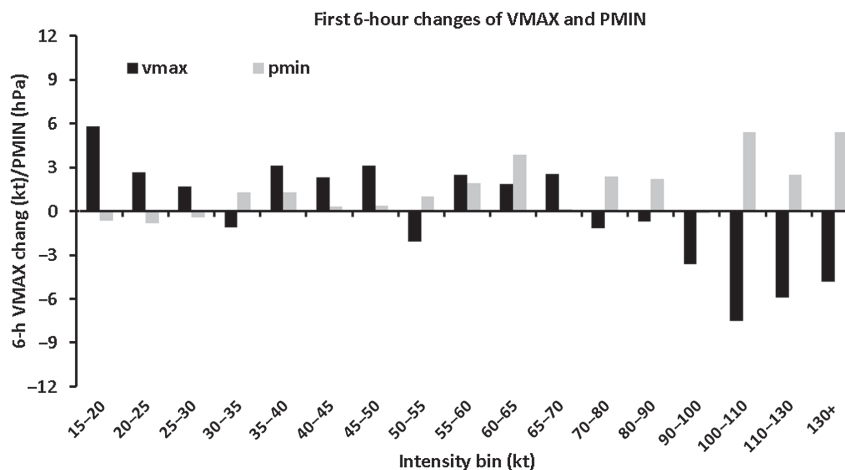
Given the above issues with the real-time forecasts, it is thus critical to minimize the impacts of the weakening cycles related to TC movement or inimical environment before one can obtain useful PIP information. To this end, we first divide the entire life cycle of a TC into intensifying periods, quasi-stationary periods, and weakening periods, using the initial 12 h change of the maximum 10 m sustained wind ( $V_{MAX}$ ) for each cycle provided in the post-analysis best track dataset (bdecks). For each forecasting cycle, the first 12 h change of the observed  $V_{MAX}$  in the bdecks is calculated for this initial time, i.e.  $\Delta V = V_{MAX}(12\text{ h}) - V_{MAX}(0\text{ h})$ . All cycles with  $\Delta V > 0$  are identified as intensifying cycles and then stratified into six different initial intensity bins, based on the initial value of  $V_{MAX}$  in the ranges 25–45, 45–65, 65–85, 85–105, 105–125, and 125–180 kt as reported in the bdecks.

With these intensity bins, an intensity error growth rate, denoted by  $\epsilon$ , for each bin is calculated as a change of the model forecast  $V_{MAX}$  error over a time interval of 18 h, i.e.

$$\epsilon = \frac{Error_{V_{MAX}}(18\text{ h}) - Error_{V_{MAX}}(0\text{ h})}{18\text{ h}}, \quad (3)$$

where  $Error_{V_{MAX}}(18\text{ h})$  and  $Error_{V_{MAX}}(0\text{ h})$  are the  $V_{MAX}$  errors between the model forecast intensity and the observed intensity at  $t = 18\text{ h}$  and at the initial time, respectively. The choice of a fixed time interval of 18 h in the above error growth calculation is to minimize the impacts of model spin-down or spin-up, which often takes place during the first 6–12 h of model integration (Tallapragada *et al.*, 2014; Pu *et al.*, 2016). On the other hand, it is short enough to ensure that the intensity error saturation is not reached. The latter requirement is necessary because of our focus on the error growth during the transient period before TC intensity errors





**FIGURE 2** Stratification of the first 6 h change of the maximum 10 m wind ( $V_{MAX}$ , kt, black) and the minimum surface pressure ( $P_{MIN}$ , hPa, grey) for the operational HWRf model during 2012–2016 seasons in the NATL basin

experience saturation, which can result in an underestimation of  $\epsilon$ , as will be shown later.

Along with analyses of the error growth rate  $\epsilon$ , the time interval required for TC intensity errors to approach saturation is also estimated for each intensity bin. To be specific, an error saturation time will be defined to be the first forecast lead time, denoted by  $\tau$ , at which  $\epsilon < 0.1 \text{ kt h}^{-1}$ . This definition of  $\tau$  is more flexible than the traditional signal-to-noise ratio, as it allows for computing  $\tau$  in any basin and any intensity bin regardless of the exact value of the intensity error saturation  $\Gamma$ . Similar to the analyses of  $\epsilon$ , all calculations of  $\tau$  are carried out only for intensifying cycles to ensure that favourable conditions for the TC development are maintained.

Forecasts used in this study were all taken from the HWRf and GFDL models for the WPAC, the NATL, and the EPAC basins during 2012–2016, which are archived in real-time forecast files (adeck) in the Automated Tropical Cyclone Forecasting System (ATCF) format (Sampson and Schrader, 2000). Along with these real-time forecasts, the two most recent upgrade versions of the HWRf model in 2015 (H215; Tallapragada *et al.*, 2015) and in 2016 (H216; Biswas *et al.*, 2016) were also employed to further isolate the inhomogeneity in the annual upgrades of the operational models. For verification, the bdecks provided by the National Hurricane Center (NHC) and Joint Typhoon Warning Center (JTWC) were utilized. Except for a slight initial difference between the real-time TC records (known as TCvitals) and the post-analysis best track, the overall initial TC intensity is well matched between the model and the observation, thus allowing us to effectively stratify the statistics of the model intensity errors according to the observed intensity.

Among several caveats to our approach for real-time intensity analyses, we should highlight at this point an important difference between structure errors and strength errors that our methodology could not fully capture. Technically, a structure error refers to an initial error in the three-dimensional (3D) vortex structure, whereas a strength error specifically refers to the error of TC initial strength represented by

point-like metrics such as the  $V_{MAX}$  or the minimum central pressure ( $P_{MIN}$ ). Although structure errors are essentially part of an initial condition problem, errors related to the vortex initial structure are however not well reflected in the initial  $V_{MAX}$  errors. One could have in principle a zero initial strength error but an incorrect vortex initial structure such as a too weak warm core or a too high outflow level. Such an initial vortex with a perfect match between the model  $V_{MAX}$  and the observed  $V_{MAX}$  but a wrong vertical structure can cause any model to experience rapid initial adjustment and produce large intensity errors (Gopalakrishnan *et al.*, 2012; Tallapragada *et al.*, 2014; Kieu, 2015), even if the model is perfect. Therefore, characterizing the TC intensity by the point-like metrics such as  $V_{MAX}$  or  $P_{MIN}$  alone is generally not sufficient, because these point-like metrics could not fully represent the TC 3D structure as discussed in Vukicevic *et al.* (2013).

To see the impacts of structure errors on the intensity forecasts, Figure 2 shows an analysis of initial intensity adjustment during the first 6 h obtained from the HWRf model during the 2012–2016 seasons in the NATL basin. One notices in Figure 2 the persistent spin-down of the model vortex for the cycles with an initial intensity of 100 kt and above, despite the good match between model  $V_{rmMAX}$  and observed  $V_{rmMAX}$  at initial time. As discussed in Tallapragada *et al.* (2014), this initial spin-down of strong intensity cycles is systematically related to improper vortex structures that the current HWRf initialization scheme is not able to handle, no matter how closely the model initial  $V_{MAX}$  can match the observed  $V_{MAX}$ .

In practice, the problem of TC structure initialization is very hard to address and is mostly represented by a discrete vortex depth indicator in case the observation of the TC 3D structure is not available (Trahan and Sparling, 2012). Hence, these strength and structure errors are not separable and become increasingly important at the strong intensity stage. In the current HWRf model, it is ultimately the structure errors that are responsible for the model initial spin-down for an initial intensity  $> 100 \text{ kt}$ , even when the model initial

$V_{MAX}$  could match well the observed  $V_{MAX}$  (Figure 2). In this case, the 18 h intensity error growth for strong storms could be just a manifestation of both the initial strength errors and the structure errors and there is no effective way to separate them. Because of this, one should be cautioned when analyzing the intensity error growth for intensity bins  $> 100$  kt.

### 3.2 | Idealized experiments

While real-time analyses could capture actual intensity error saturation for different basins, an apparent issue with the real-time verification is the different environmental condition in different basins, which makes it hard to isolate the nature of the intensity error growth. To this end, the idealized configuration of the HWRF model (version 3.7) was used to further examine the intrinsic growth of TC intensity errors. The model was configured with identical settings as in Kieu *et al.* (2016), but with double-nested domains at a horizontal resolution of 9 and 3 km instead of triple-nested domains due to the requirement of a large number of integrations in this study.

All idealized experiments were designed on an  $f$ -plane centred at  $12.5^{\circ}\text{N}$  in a quiescent environment with sea surface temperature fixed at 300K. The model was initialized with the Jordan (1958) mean tropical sounding and a weak vortex that has the maximum surface wind of  $20\text{ m s}^{-1}$  and the radius of maximum wind (RMW) of 90 km with 61 vertical levels. Under a perfect model scenario, a control (CTL) experiment was integrated for 5 days to serve as a reference trajectory. The CTL experiment captures a rapid intensification of  $\sim 50$  kt per 24 h during the 12 to 48 h period, followed by a quasi-stationary stage after 3 days into integration (Figure 9).

A set of sensitivity experiments was then conducted with different realizations of random wind perturbations added to the CTL trajectory starting from  $t = 12$  h and ended at  $t = 72$  h at an interval of 6 h to mimic the impacts of random errors at different stages of TC development. Unlike the idealized experiments in Nguyen *et al.* (2008) in which random perturbations were added only at the initial time, our design introduces random perturbations at different stages of the vortex intensification. By quantifying the subsequent growth of these errors at different stages, it is possible to assess how the intensity error growth depends on the model vortex strength without the severe issue of the model spin-up or spin-down as in the operational mode. For the CTL trajectory in this study, these instants of time correspond well to the intensity bins analyzed in the real-time analyses.

To increase the representativeness of the sampling, a range of random wind perturbation amplitudes between  $0.5$  and  $5\text{ m s}^{-1}$  at intervals of  $0.5\text{ m s}^{-1}$  were added to the CTL trajectory at each stage of the model vortex development. A total of 110 experiments were therefore conducted. (The random perturbations were added at 11 times during the vortex development between 12 and 72 h.) In this perfect model scenario, the spread of the intensity deviations relative to the CTL

trajectory can represent the growth of intensity errors due to random perturbations in the HWRF model as expected.

## 4 | TC INTENSITY ERROR CHARACTERISTICS

### 4.1 | Intensity error saturation

With the first requirement related to the existence of an error saturation, Figure 3 shows the verification of real-time intensity forecasts during the intensifying period for both the HWRF and GFDL models during 2012–2016 in the three major basins. Except for the NATL basin, one notices in Figure 3 that TC intensity errors generally grow from 0 to 72 h, and quickly slow down and approach a saturation limit at 4–5 day lead time. A closer examination of Figure 3 shows that both the HWRF and GFDL models appear to capture a similar behaviour of the error saturation with a lower saturation value in the EPAC basin ( $\sim 14$ – $17$  kt), and the largest value in the WPAC basin ( $\sim 18$ – $20$  kt). Assuming that the observational error of  $V_{MAX}$  is  $\sim 7.5$  kt (Torn and Snyder, 2012; Landsea and Franklin, 2013) and it is independent from model forecast errors, it can be then obtained that the intensity error saturation  $\Gamma$  is  $\sim 12$ – $14$  kt in the EPAC basin, and  $\Gamma \sim 14$ – $18$  kt in the WPAC basin. That both modelling systems with different treatments of dynamical cores, nonlinear coupling, physical parametrizations, and domain configurations display the same characteristics of error growth among the three basins indicates that the TC intensity error saturation is not the same, but depends upon the ambient environment that TCs are embedded in.

While the error saturations are apparent in the WPAC and EPAC basins, it is intriguing to see that the intensity errors in the NATL basin appears to exhibit somewhat different characteristics with a linear growth up to 4–5 day lead time, especially in the GFDL model. This linear growth is consistent with the reports in DeMaria *et al.* (2014), Emanuel and Zhang (2016), Bhatia and Nolan (2016), and Bhatia *et al.* (2017) that are based on different verification periods for the NATL basin during the 2009–2015 and 2011–2015 seasons. At first sight, the linear curve seems to suggest no error saturation limit in the NATL basin as in other two basins. However, more detailed examination of the TC climatology in the NATL basin reveals that the linear growth of intensity errors turns out to be specifically pertinent to the NATL basin due to the dominance of weak storms with prolonged lifecycles extending beyond  $35^{\circ}\text{N}$  in this basin during the 2012–2016 seasons.

To confirm such impacts of weak storms on the intensity error saturation, Figure 4 shows the intensity statistics for cycles with an initial intensity of Category 1 and above<sup>3</sup> as

<sup>3</sup>The choice of Category 1 and above for the strong storm statistics is solely because of the lack of hurricane cases during 2012–2016 in the NATL basins.

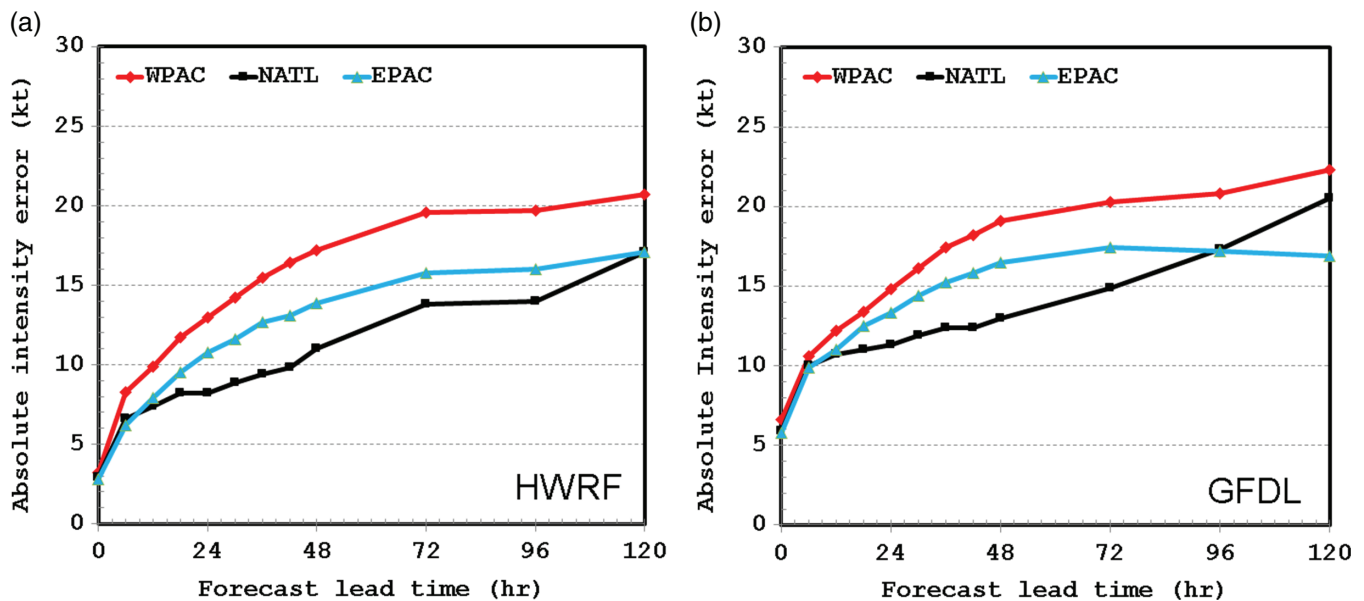


FIGURE 3 (a) Verification of the real-time intensity forecasts for the HWR model in the NATL (red), WPAC (black), and EPAC (blue) basins during 2012–2016. (b) is as (a), but for the GFDL model [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

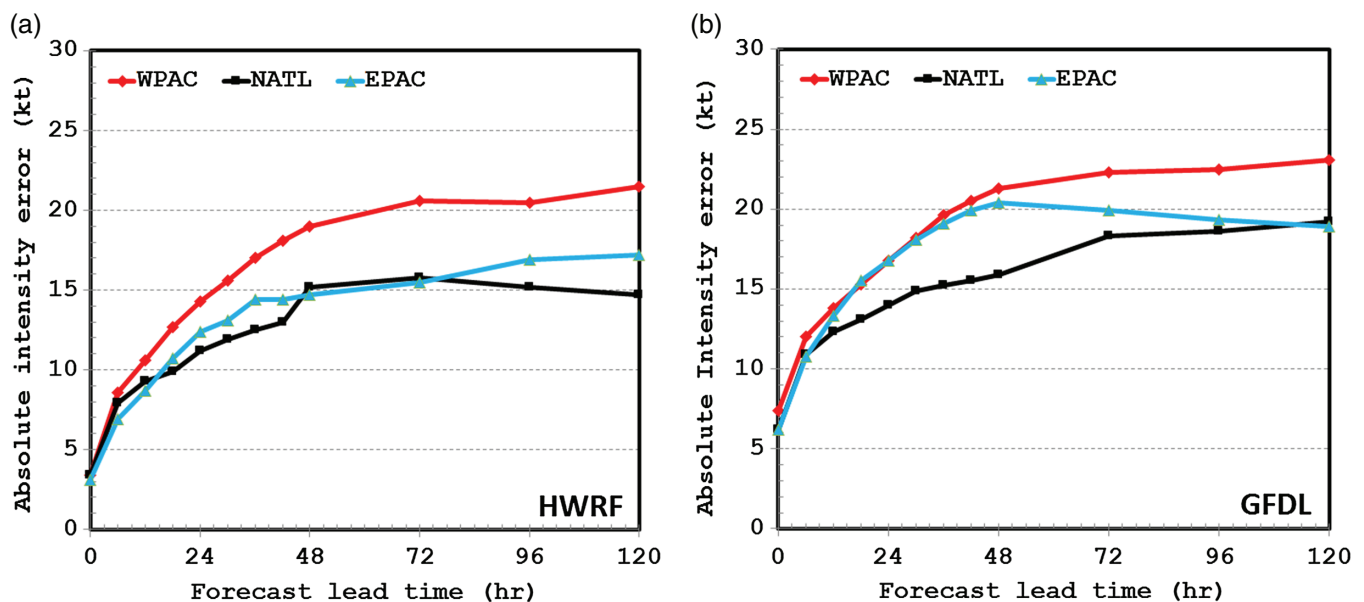


FIGURE 4 As Figure 3, but for TCs that are of Category 1 and above [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

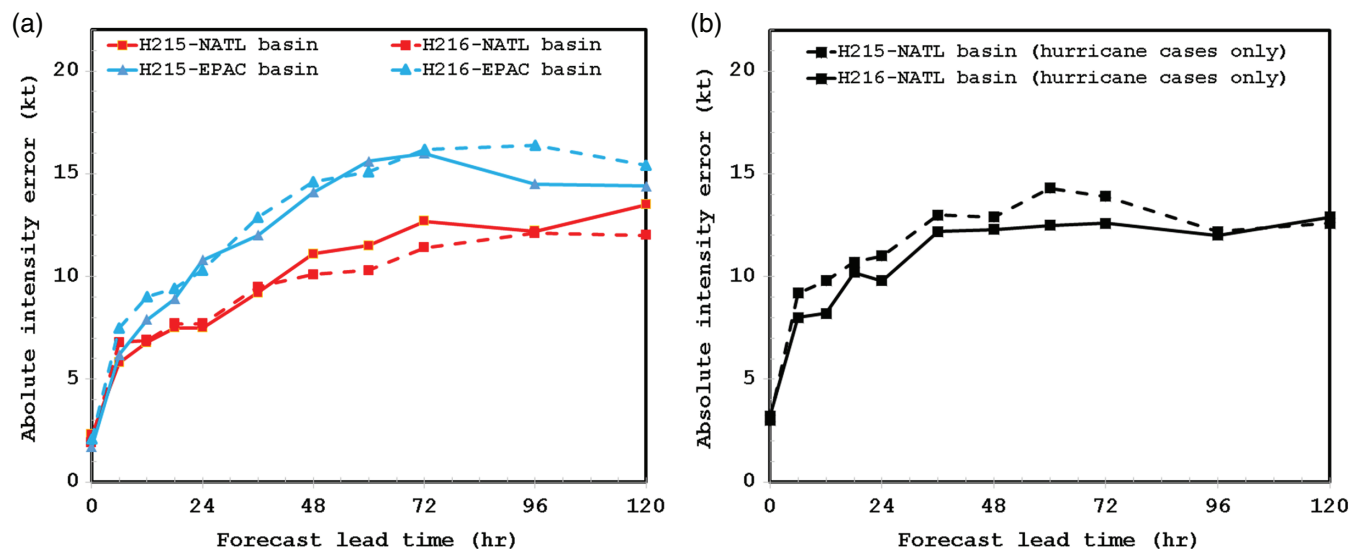
indicated in the best track data for both the HWR and GFDL models. Similar to the verification for intensifying cycles shown in Figure 3, all cycles with landfalling points or extratropical cyclone transition classification reported in the best track data are also excluded in this analysis of strong storms to ensure that the development of these strong intensity cycles would be least influenced by the storm track or unfavourable conditions.

It is of interest to see in Figure 4 a large shift in the intensity characteristics for the NATL basin, with a more apparent

error saturation around 14–15 kt for the HWR model and  $\sim$  17–18 kt for the GFDL model. In contrast, the same verification of Category 1 and above storms in the EPAC or WPAC does not show much change in the intensity error saturation curve. This insensitivity of the statistics in the EPAC and WPAC basins indicates the minimal effect of weak storms on the error saturation in these two basins where there are many more strong storms than in the NATL basin. Because the majority of TCs in the NATL basin during 2012–2016 period were weak storms, it is therefore clear that the full statistics for the NATL basin could not reveal the error saturation as shown in Figure 3.

Of further significance is that the error saturation for strong storms in the NATL basin shown in Figure 4 is slightly

Our try with Category 4 and above shows even more evidently the saturation of intensity errors for stronger TCs, but the statistics are not significant. (Of 70 TCs during 2012–2016 NATL seasons, there are only 27 major hurricanes with four reaching Category 4 and above).



**FIGURE 5** (a) As Figure 3, but for the retrospective experiments during the 2012–2015 seasons using the 2015 version (solid) and the 2016 version (dashed) of the HWRf model in the NATL (red) and the EPAC (blue) basins; (b) is as (a), but for the TC cases in the NATL basin that are of Category 1 and above [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

smaller than that in the EPAC basin ( $\sim 14$ – $17$  kt for the HWRf model), but is considerably smaller than that in the WPAC basin (similar results appear in e.g. Bhatia and Nolan, 2013; Bhatia *et al.*, 2017). Although the observed  $V_{\text{MAX}}$  in the WPAC basin has more uncertainty due to lack of the reconnaissance observation that could affect the  $V_{\text{MAX}}$  error saturation in this basin, it should be noted that the different error saturation in different basins indicates a profound role of large-scale conditions on the MPI attractor. How the MPI attractor depends on the environmental parameters such as sea surface temperature, tropical troposphere temperature, or stratification is still elusive at present. However answering this question would require sensitivity experiments that are outside the scope of the operational error statistics in this study.

While the real-time verification of the HWRf and GFDL models could consistently capture the intensity error growth and subsequent saturation, an issue with real-time forecasts is that operational models are annually upgraded. Therefore, the verification shown in Figures 3 and 4 contains five different model versions in five different years during 2012–2016. This inhomogeneity in the HWRf model configuration makes it hard to fully justify the saturation of intensity errors. To remove the impacts of different model versions on the characteristics of the intensity error growth, Figure 5 shows the intensity verification for the HWRf model obtained from the H215 and H216 retrospective experiments. These two retrospective experiments share the same period from 2012 to 2015 for both the EPAC and the NATL basins<sup>4</sup>, and the homogeneous verifications can therefore be compared directly.

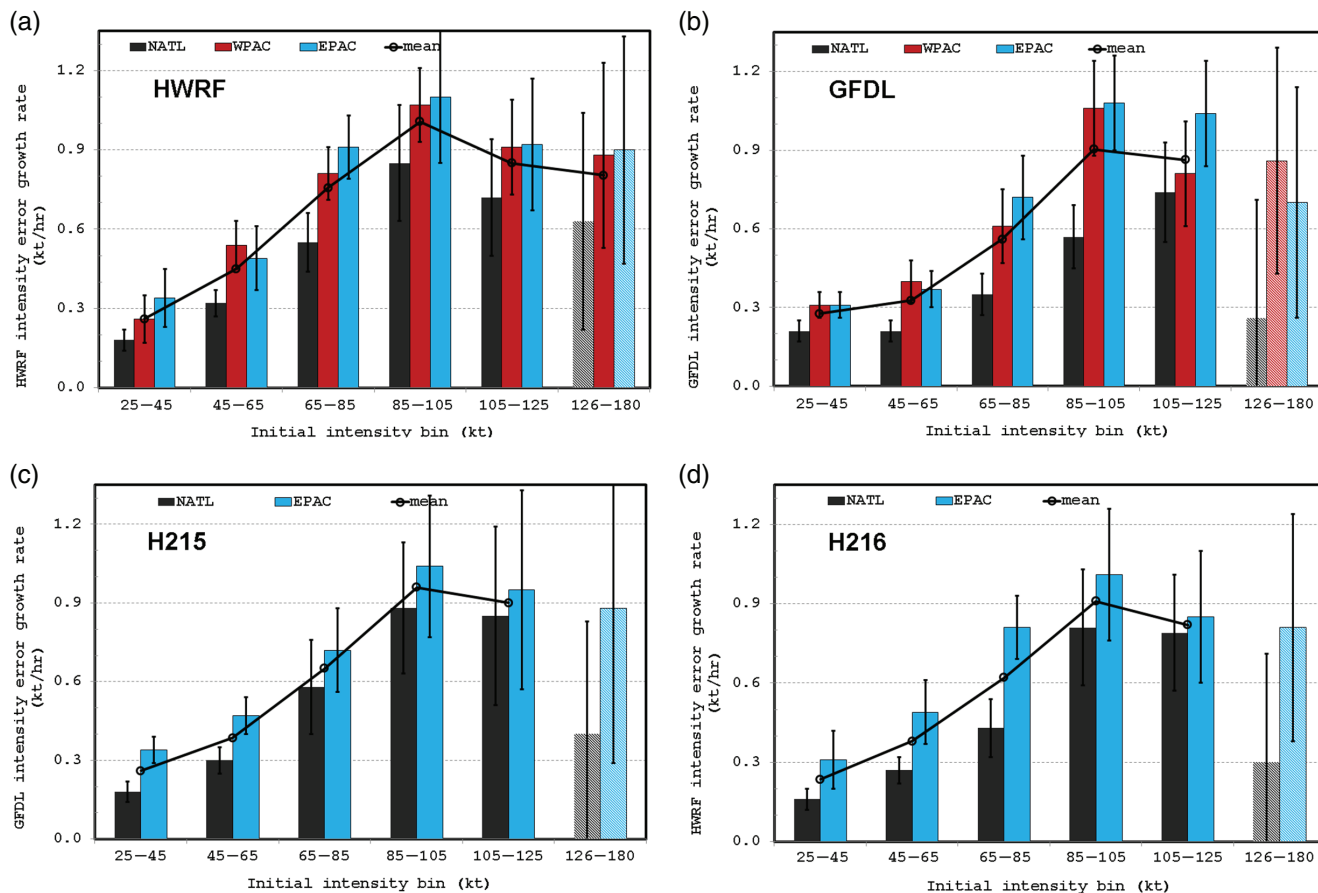
Similar to the real-time verification, one notices in Figure 5a again the rapid intensity error growth during the first 2 days, followed by the error saturation  $\sim 12$ – $13$  kt in the NATL basin and  $\sim 14$ – $15$  kt in the EPAC basin for both H215 and H216. Although the H216 version has a slight improvement in the NATL basin, the similar intensity error curves among all retrospective experiments and the real-time forecasts reiterate the unique properties of the intensity error growth curve as obtained from real-time statistics. Indeed, our analyses of intensity errors for several other operational models all show analogous error growth with the smallest saturation limit in the NATL basins and the largest in the WPAC basin, although the exact values of the error saturation slightly vary from one model to the other (not shown).

Because of the contribution from weak storms in the NATL basin during the 2012–2015 seasons, Figure 5b shows additional intensity verification for TCs of Category 1 and above, similar to the stratification shown in Figure 4 for real-time forecasts. It is again evident from this strong storm verification that the intensity errors exhibit a saturation of 13 kt after just 2 days into integration in the NATL basin. This strong storm verification highlights the different error growth between weak and strong storms as captured in real-time forecasts, and supports the existence of the intensity error saturation.

As discussed in section 3, the existence of the intensity error saturation demonstrated in Figures 3–5 is by no means sufficient to conclude that TC intensity has limited predictability. However, the fact that the error saturation exists is itself important, because it demonstrates that a necessary condition for such a limited intensity predictability is warranted, irrespective of its exact value. We now turn to the second condition which implies that limited predictability must be related to how fast an intensity error would evolve during the TC intensification period.

<sup>4</sup>Due to operational constraint and priorities, both the H215 and H216 upgrades of the HWRf model did not have full retrospective testing for the WPAC basin during 2012–2015. So, only experiments in the EPAC and the NATL basins are presented in this retrospective analysis.





**FIGURE 6** Intensity error growth rate (columns, knot h) as a function of intensity bins obtained from the real-time TC intensity forecasts during 2012–2016 in the NATL basin (black), the EPAC basin (blue), and the WPAC basin (red) for (a) the HWRf model, (b) the GFDL model, (c) the H215 retrospective experiment, and (d) H216 retrospective experiments during 2012–2014. Error bars denote 95% confidence intervals [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

## 4.2 | Intensity error growth rate

Given the condition related to the error growth rate  $\epsilon$  as discussed in section 2, Figure 6a shows  $\epsilon$  as a function of the initial intensity bins, which is obtained from the HWRf forecasts for the 2012–2016 seasons. It is apparent in Figure 6a that  $\epsilon$  indeed exhibits a strong dependence on the storm strength, with a faster growth rate for higher initial intensity. On average,  $\epsilon$  is  $\sim 0.3 \text{ kt h}^{-1}$  for the 25–45 kt bin, and increases to  $\sim 1.1 \text{ kt h}^{-1}$  for intensity bins larger than 85 kt during TC development (Figure 6a). Although the value of  $\epsilon$  is slightly different among the three basins, such an increase of  $\epsilon$  with storm strength is very persistent in all basins and for both operational models. Despite some differences in defining the intensity error growth rate and intensity bins, it can be seen in Figure 6 that the faster error growth rate for higher storm initial strength is consistent with a similar analysis for GFDL model in the NATL basin as shown in Bhatia and Nolan (2013). A direct implication of this property is that the HWRf model will more quickly undergo a large intensity error as TCs intensify, making it harder for the model to predict TC intensity with time, at least from the perspective of the  $V_{\text{MAX}}$  metric. Note that this behaviour of the error growth rate is not only seen in TCs, but also applies to other

nonlinear systems with limited predictability (e.g. the growth of different eigenmode modes in the Eady problem in Farrell (1982)).

Another important point in Figure 6 is that  $\epsilon$  appears to reach a limit after the  $V_{\text{MAX}} > 85 \text{ kt}$ . Specifically,  $\epsilon$  increases quickly at first but then stays nearly constant around  $1.1 \text{ kt h}^{-1}$  for the 85–105 kt and 102–125 kt bins, and is somewhat smaller for the 125–180 kt bin. Although the growth rate for the most intense bin 125–180 kt has a large uncertainty due to a small sample size, the upper bound on  $\epsilon$  at the mature stage accords with the existence of a chaotic attractor at the MPI limit recently proposed in KM16. As discussed in KM16, the existence of a chaotic MPI attractor implies that any error will quickly occupy the entire attractor due to (a) the sensitive dependence on initial conditions inside the attractor, and (b) the boundedness of the attractor<sup>5</sup>. From this perspective, the saturation of  $\epsilon$  seen in Figure 6a is an indication that TCs enter the chaotic attractor at their mature stage. As a result, the same error growth inside the attractor reflects the consistent

<sup>5</sup>Similar to study by KM16, an MPI attractor is defined as a bounded region in the phase space surrounding the MPI equilibrium that a TC will eventually settle at its quasi-stationary stage.

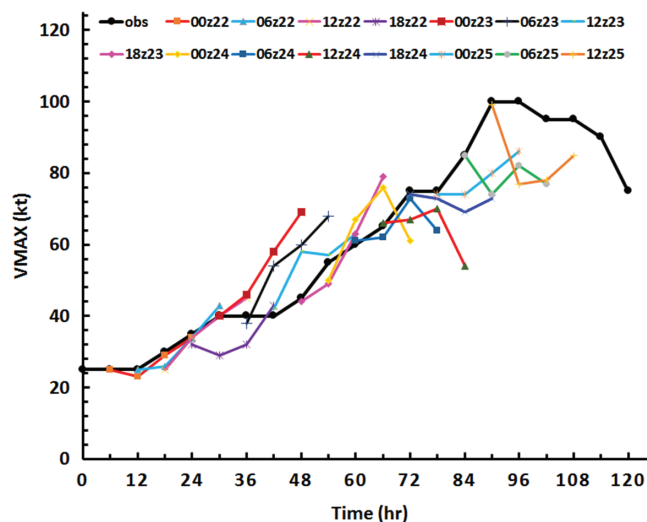
leading Lyapunov exponent of the attractor, regardless of TC initial conditions.

Although the upper bound of  $\epsilon$  is in accordance with the existence of a chaotic attractor, we should recall that an actual intensity error in a regional model is not a pure result of initial condition errors, but depends on several other factors including (a) unbalance adjustment due to structure errors, (b) model internal errors, and (c) lateral boundaries from global models. Of the three, the initial unbalance is most significant for intensity bins  $> 105$  kt, which accounts for a large  $\epsilon$  due to the model spin-down during the first 6–12 h (cf. Figure 2). Since our definition of  $\epsilon$  is based on a fixed interval 0–18 h, the initial adjustment issue is not too severe, at least for initially weak storms. It is thus reasonable to expect that the increase of  $\epsilon$  with the TC strength seen in Figure 6 is a general property of TC dynamics for initial intensity bin  $< 85$  kt, for which the HWRF real-time statistics are sufficiently robust.

Regarding model errors and lateral boundary condition errors, there is no simple way to quantify these errors as a model is integrated in time, especially for the lead times longer than 3 days (Buizza, 2006; Tien *et al.*, 2013). Nevertheless, the fact that there are several different operational TC models in different ocean basins together with various retrospective experiments suggests a pathway to examine the relative roles of the initial condition errors and the model errors. Specifically, different models have different treatment of dynamical cores, physical parametrizations, numerical approximations, boundary conditions, or resolutions, which may produce different feedbacks and error growth behaviour with time. Assuming the same initial intensity errors, the variation of  $\epsilon$  among these models may provide some insight into the nature of the intensity error growth related to initial condition errors and model errors.

In this regard, Figure 6b–d show the error growth analysis for the GFDL model similar to that for the HWRF model during the same 2012–2016 period, along with the other two retrospective experiments H215 and H216. Despite different model configurations and nonlinear interaction among error sources, the main characteristic of  $\epsilon$  is very consistent between the GFDL and different versions of the HWRF model, with the same faster error growth rate for higher intensity. Similarly, the different error growth rates in all three basins also accord well among the HWRF model upgrades, with the lowest growth rate in the NATL basin and the fastest growth rate in the WPAC basin. Except for the issue with a limited sample size for the 125–180 kt bin, the overall consistency of  $\epsilon$  among the HWRF, the GFDL, H215, H216, and the axisymmetric model presented in KM16 indicates that the increase of  $\epsilon$  with the storm intensity during the transient period is an inherent property of TC dynamics regardless of modelling systems or ocean basins.

To demonstrate the implication of the above error growth characteristic, Figure 7 shows an example of real-time intensity forecasts for 14 continuous cycles of hurricane *Sandy*

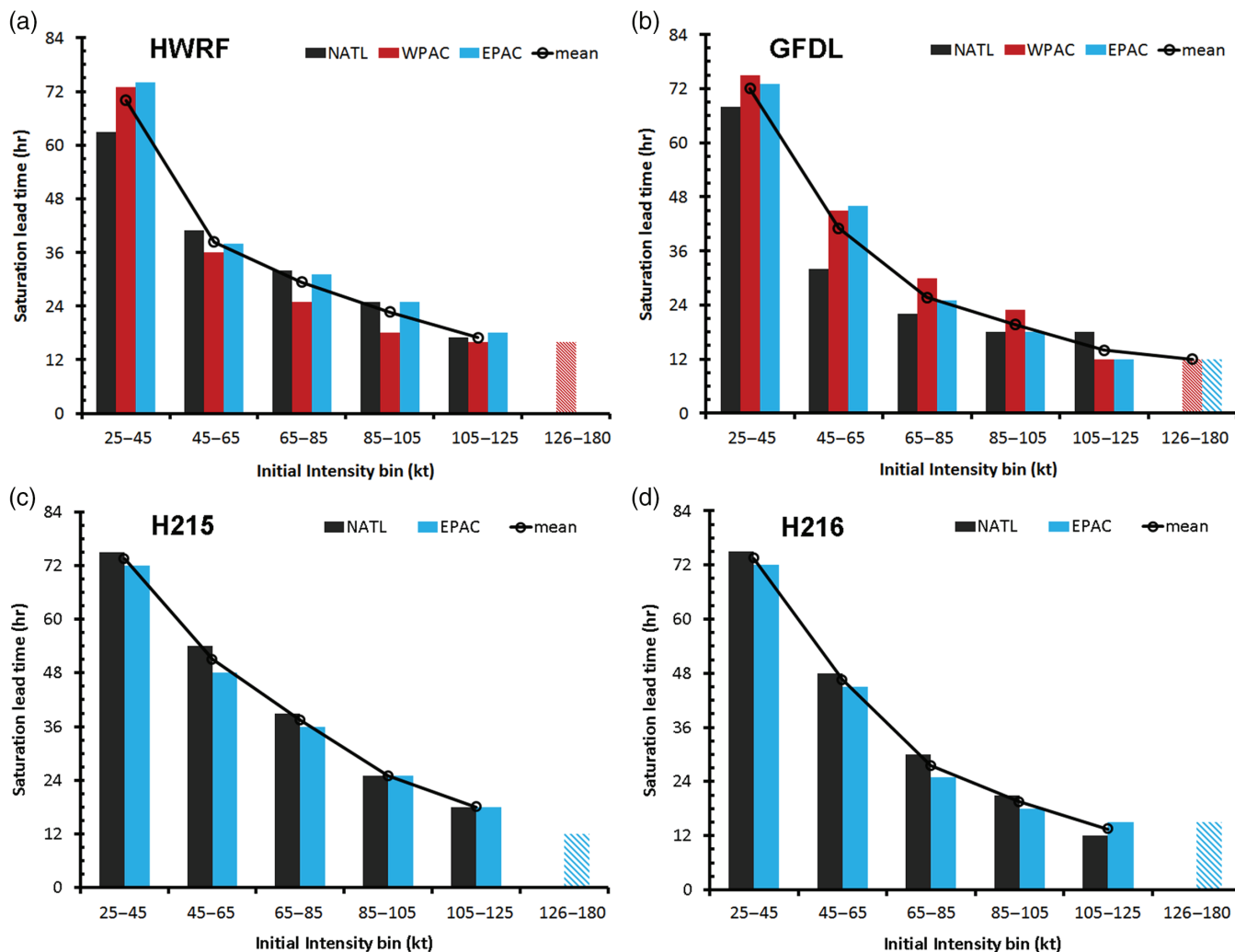


**FIGURE 7** Real-time intensity forecasts of hurricane *Sandy* (2012) during its developing stage from 0000 UTC on 22 to 1200 UTC on 25 October 2012 obtained from the operational HWRF model. Note that only the first 18 h of the HWRF intensity forecasts are plotted for each cycle to facilitate comparison with the error growth rate in Figure 6 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

(2012) by the HWRF model, starting from 0000 UTC on 22 October to 1200 UTC on 25 October 2012. For the sake of error growth visualization, only the first 18 h of each cycle are plotted to facilitate the comparison with the error growth rate in Figure 6. As seen from this example of *Sandy* forecasts, the  $V_{MAX}$  errors indeed evolve relatively slowly for the early cycles from 0000 to 1200 UTC on 22 October, with an averaged 18 h error growth of  $\sim 10$  kt. As *Sandy* reached a strength of 40 kt between 1800 UTC 22 and 1200 UTC 23, the  $V_{MAX}$  error growth becomes faster with an averaged 18 h growth rate of  $\sim 16$  kt. From 1800 UTC 23 to 0600 UTC 24 during which *Sandy* became stronger, HWRF exhibits mixed behaviour with one cycle overestimating the intensity whereas other cycles intensify but quickly decay. As soon as *Sandy* reached  $> 85$  kt after 1200 UTC 24, all cycles show a strongly negative bias with a systematical spin-down during first 18 h. From these forecasts of hurricane *Sandy*, one can see that the error growth rate clearly depends on the storm initial intensity with larger error growth for higher intensity. This dependence of the error growth rate on the storm initial intensity explains the progressively larger spread of the composite intensity time series seen in Figure 7. Apparently, the  $V_{MAX}$  error spread for this case of *Sandy* is consistent with the overall statistics shown in Figure 6, which reveals the intricate nature of the intensity error growth during TC development. From an operational forecasting perspective, one thus expects to see a larger spread of an intensity ensemble with time, which render intensity forecasts more challenging as TCs intensify.

#### 4.3 | Intensity error saturation time

With both conditions related to the existence of the error saturation and the dependence of the error growth rate on storm



**FIGURE 8** As Figure 6, but for the intensity error saturation time  $\tau$  (h) as a function of intensity bins [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

intensity obtained from the real-time statistics, an immediate consequence is that the time interval during which an initial intensity error will approach the saturation limit  $\Gamma$  must become shorter as TCs evolve. This is because a larger growth rate would allow an error to approach  $\Gamma$  faster, assuming the same value for  $\Gamma$ .

Figure 8 shows  $\tau$  as a function of TC intensity bins for the 2012–2016 real-time verification. As anticipated,  $\tau$  is markedly shortened for higher intensity in both the HWRF and the GFDL models; it is about 72 h for the intensity  $<65$  kt, and quickly reduces to  $\sim 18$ – $20$  h for the intensity  $>85$  kt in all three ocean basins. This reduced saturation time implies that a dynamical model would have a shorter time ( $\sim 18$ – $20$  h) to predict intensity for initially stronger storms than for initially weaker storms ( $\sim 72$  h). In this regard, intensity forecasts based on dynamical models would become indistinguishable from a climatological forecast after reaching the error saturation if the right environment for intensification is ensured.

The above interpretation of a shorter forecast range for higher intensity is consistent with the faster error growth for higher intensity seen in Figure 6, but caution is advised. Note again that the intensity of an actual storm is sensitive

to landfall, vertical wind shear environment, or SST conditions. Therefore, the PIP of a real-time intensity forecast may depend more crucially on TC movement than on the inherent chaotic properties in practice. Any intensity change due to a predictable track forecast must be distinguished from the IIP associated with the saturation time  $\tau$  shown in Figure 8, which is applied only to the TC internal dynamics under a fixed environmental condition.

Similar to analysis of the error growth rate in Figure 6, the estimation of  $\tau$  presented in Figure 8 is admittedly crude for the strongest intensity bin 120–180 kt due to the small sample size. In operational practice, a TC may have no time to reach its saturation, let alone stay at the mature stage, for more than 18 h so that  $\tau$  can be evaluated. Therefore, the intensity errors obtained from real-time verification are not just due to the TC internal dynamics, but also due to model errors and incorrect track forecasts that the stratification method presented in section 3 could not completely filter.

In addition to the aforementioned track-dependence issue, it is generally hard to attribute the error growth either to the initial conditions inside the chaotic attractor or to model initial balance adjustment for TCs of Category 4 and above. Thus,

the interpretation of the TC intensity error saturation for the intensity bins  $>120$  kt is subject to large uncertainties, which can be seen by the large error bars in Figures 6 and 8. The best approach to examine the intensity error saturation time at the high-intensity limit is to select only TC cases that develop and maintain their mature stage for several days, because this long quasi-stationary stage would allow us to better estimate the error growth rate as well as  $\Gamma$  with minimal inadvertent impacts of the environment. However, this condition results in a sample size which is too small to achieve significant statistics, and so is not presented here.

## 5 | HWRF IDEALIZED EXPERIMENTS

In the most optimistic scenario of TC model development, one could argue that future model advances could lead to a perfect TC model with a perfect vortex initialization scheme. However, random fluctuations at all model grid points are unavoidable regardless of how perfect a modelling system or vortex initialization is. This type of random error underlines the intrinsic predictability limit proposed by Lorenz (1963; 1969). Despite various modelling and theoretical work on atmospheric predictability to determine the existence of such an intrinsic limit for TC intensity, the degree to which random errors can affect the intensity predictability has not been adequately addressed due to the subtle distinction between the transient stage and the MPI stage. As always, this type of the random error growth depends not only on the TC intrinsic dynamics but also on specific TC models and assumptions.

To help quantify the intrinsic limit of TC intensity errors for the HWRF model from this perspective, we analyze in this section a set of idealized experiments with the HWRF model in which random wind errors are introduced at different stages of TC development to see how these random errors will evolve and become saturated with time under a perfect model scenario. This is an important issue to address, because it sets a threshold on the lowest absolute VMAX error that the HWRF model can be improved. As emphasized in section 3, our design of adding random perturbations to the model reference state in this study is not applied at the initial time as in Nguyen *et al.* (2008). Instead, the random errors are introduced along the trajectory of TC development during the vortex intensification. This design allows us to examine how the intensity error growth rate varies with storm intensity similar to real-time forecasts, which could not be addressed in the study by Nguyen *et al.* (2008).

Figure 9a shows the growth of the  $V_{MAX}$  errors relative to the CTL trajectory during the entire development of the model vortex. It is of interest to notice from these idealized experiments that the  $V_{MAX}$  errors increase rapidly at first but then approach a saturation limit after the model vortex reaches its peak intensity very similar to real-time verification, irrespective of when and how the wind perturbations are added along the CTL trajectory. Although the  $V_{MAX}$  variation could

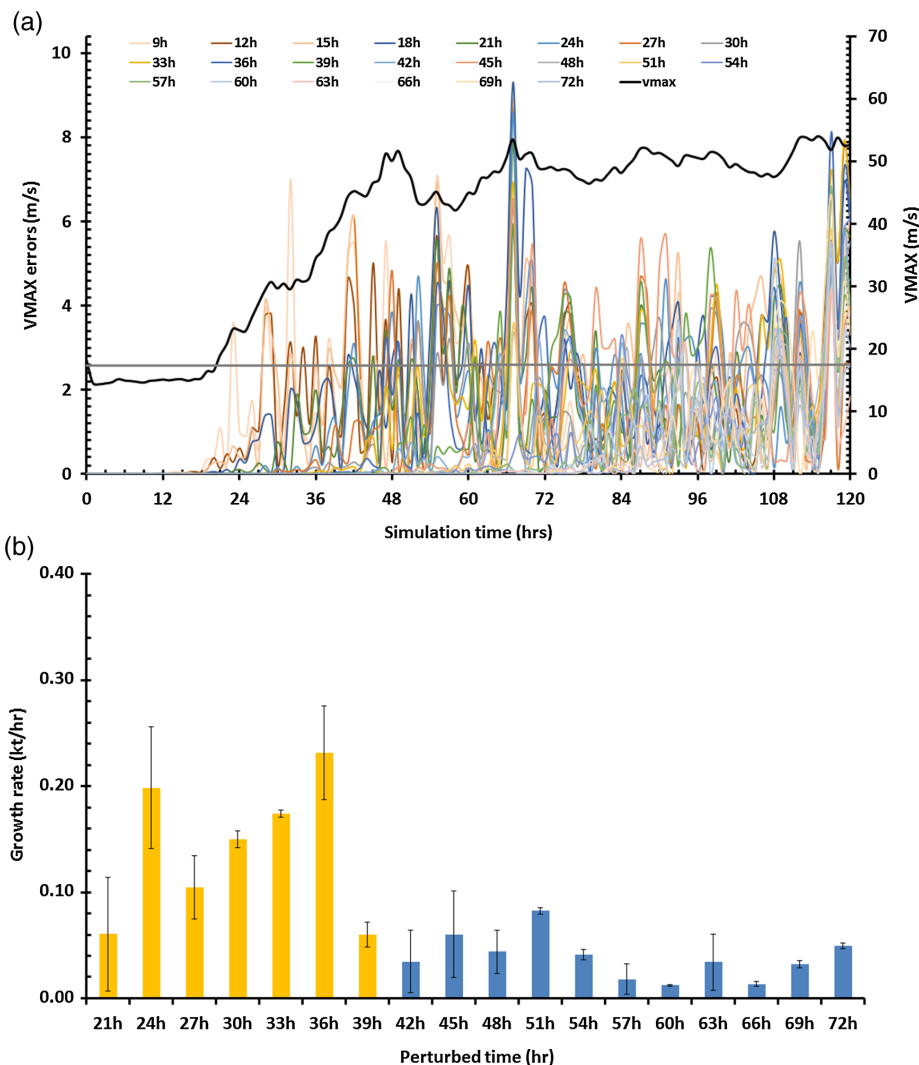
sporadically reach  $7\text{--}9\text{ m s}^{-1}$  ( $\sim 15\text{--}16$  kt) as seen in Figure 9a, the average  $V_{MAX}$  error does not grow indefinitely but is confined within a range of  $2\text{--}5\text{ m s}^{-1}$  (a standard deviation is  $3\text{ m s}^{-1}$  in our current experiment set-up).

Similar to the faster error growth for higher intensity as in real-time forecasts (cf. Figure 6), we also notice a faster growth rate of random errors for higher intensities during the idealized vortex intensification from 12 to 48 h (Figure 9b). This consistency between real-time verification and idealized experiments strongly suggests that the dependence of the error growth on the storm intensity is a general property of the TC dynamics.

Although the exact value of the error saturation may vary in different large-scale environments, as seen in real-time statistics for different basins, it should be noted that the existence of the error saturation associated with the random perturbations in these idealized experiments is the most significant result here. This is because the idealized error saturation reveals the lowest absolute  $V_{MAX}$  error that one could obtain with the current HWRF model in a given environment, assuming a perfect model and vortex initialization. In reality, the saturation  $V_{MAX}$  error is significantly larger (cf. Figures 3 and 5) due to various issues such as model deficiencies, boundary conditions, the model spin-up/spin-down issues as well as observational errors (Torn and Snyder, 2012; Landsea and Franklin, 2013). Nonetheless, the lowest bound of the intensity variation as seen in Figure 9 could at least indicate the limit in the intensity forecast accuracy that one could expect to achieve for the HWRF model in future, based on the  $V_{MAX}$  metric.

To look further into the distribution of the wind variation beyond the point-like  $V_{MAX}$  metric, Figures 10 and 11 show the horizontal and vertical distributions of the wind speed standard deviations for the transient and the quasi-stationary stages as obtained from an ensemble experiment in which random wind perturbations are specifically added to the CTL trajectory at  $t = 18$  h into integration. As anticipated, the variation of the wind speed quickly grows in magnitude during the transient stage (Figure 10a,b), and it is mostly confined within the storm central region with no further change in magnitude during the quasi-stationary stage. However, it is of more significance to notice that the largest wind speed variation is not located at the radius of maximum wind (RMW), but instead inside the eye region (Figure 10b–d). Specifically, a wavenumber-1 structure of the wind speed variation in the eye region with the peak amplitude more than  $8\text{ m s}^{-1}$  is coherently captured during the quasi-stationary period. Unlike the transient stage during which the intensity variation mostly amplifies, the intensity variation is nearly unchanged during the quasi-stationary stage, as seen in both vertical and horizontal cross-sections (cf. Figures 10c,d and 11c,d). Although there are some fluctuations in the distribution of these wind speed variations, the depth of the largest area of wind variation appears to extend up to  $\sim 500$  hPa (Figure 11),





**FIGURE 9** (a) Time series of the  $V_{MAX}$  obtained from the control experiment (black solid curve,  $m s^{-1}$ ) using the HWRF idealized configuration, and the ensemble of the  $V_{MAX}$  deviations (thin coloured curves,  $m s^{-1}$ ) relative to the control experiment using different realizations of random wind perturbations at different stages of the model vortex development. The horizontal grey line denotes the standard deviation of the error variations during the quasi-stationary stage. (b) The averaged 18 h growth rate of the  $V_{MAX}$  errors ( $kt hr^{-1}$ ) obtained from the idealized experiments with random wind perturbation added at different stages of the model vortex development as indicated on the x-axis. Yellow columns denote the intensifying stage, and blue columns denote the growth rate at the mature stage. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

revealing the high sensitivity of the TC inner core to random perturbations.

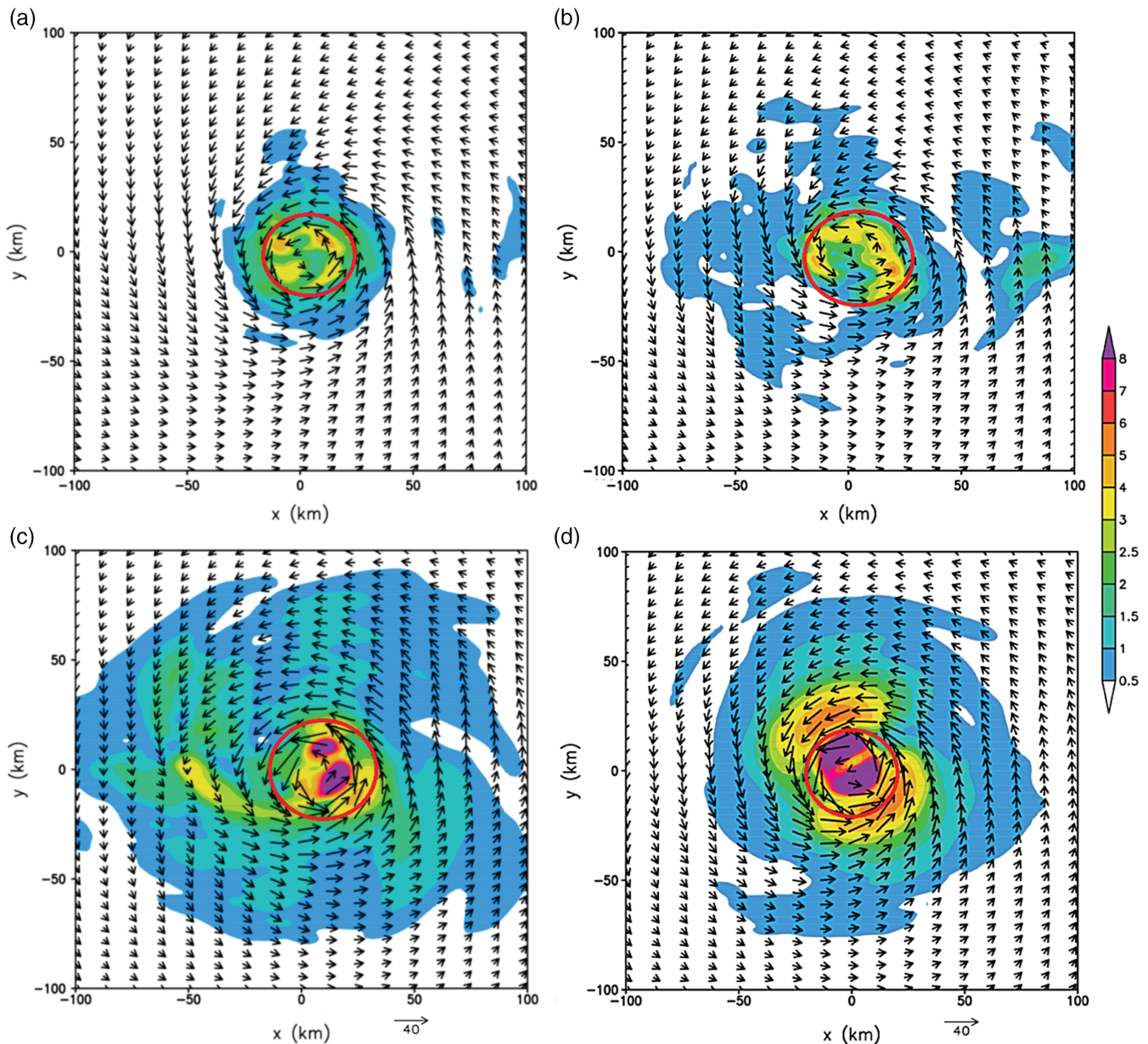
At the RMW location, we notice that the wind speed variation is less than half the magnitude of the wind variation in the eye region, which is consistent with the average of the  $V_{MAX}$  variation in the range  $3\text{--}4 m s^{-1}$  seen in Figure 9a. In this regard, the use of a single point-like  $V_{MAX}$  metric at the RMW to describe the variability of the storm intensity is insufficient to capture the variability of the entire TC structure, as discussed in Vukicevic *et al.* (2013).

## 6 | DISCUSSION AND CONCLUSION

In this study, the characteristics of TC intensity error growth in operational models and their implications for practical intensity predictability were presented, using the operational

intensity forecasts by the HWRF model and the HWRF FY2015 and FY2016 retrospective experiments. Given the current practice of TC intensity forecasting whereby intensity predictions need to be issued at all stages of TC development, two conditions on TC intensity errors that are required to quantify the intensity predictability were examined: (a) the existence of a stationary intensity error saturation, and (b) the dependence of the intensity error growth rate on TC initial intensity during the intensification stage. The first condition ensures that an initial error will grow and eventually approach a bounded limit, while the second condition ensures that the time required for the initial error to approach the saturation limit is finite. These two conditions allow us to effectively examine the range of TC intensity predictability for any numerical model.

Analyses of real-time intensity forecasts for the three major ocean basins using different versions of the HWRF and the

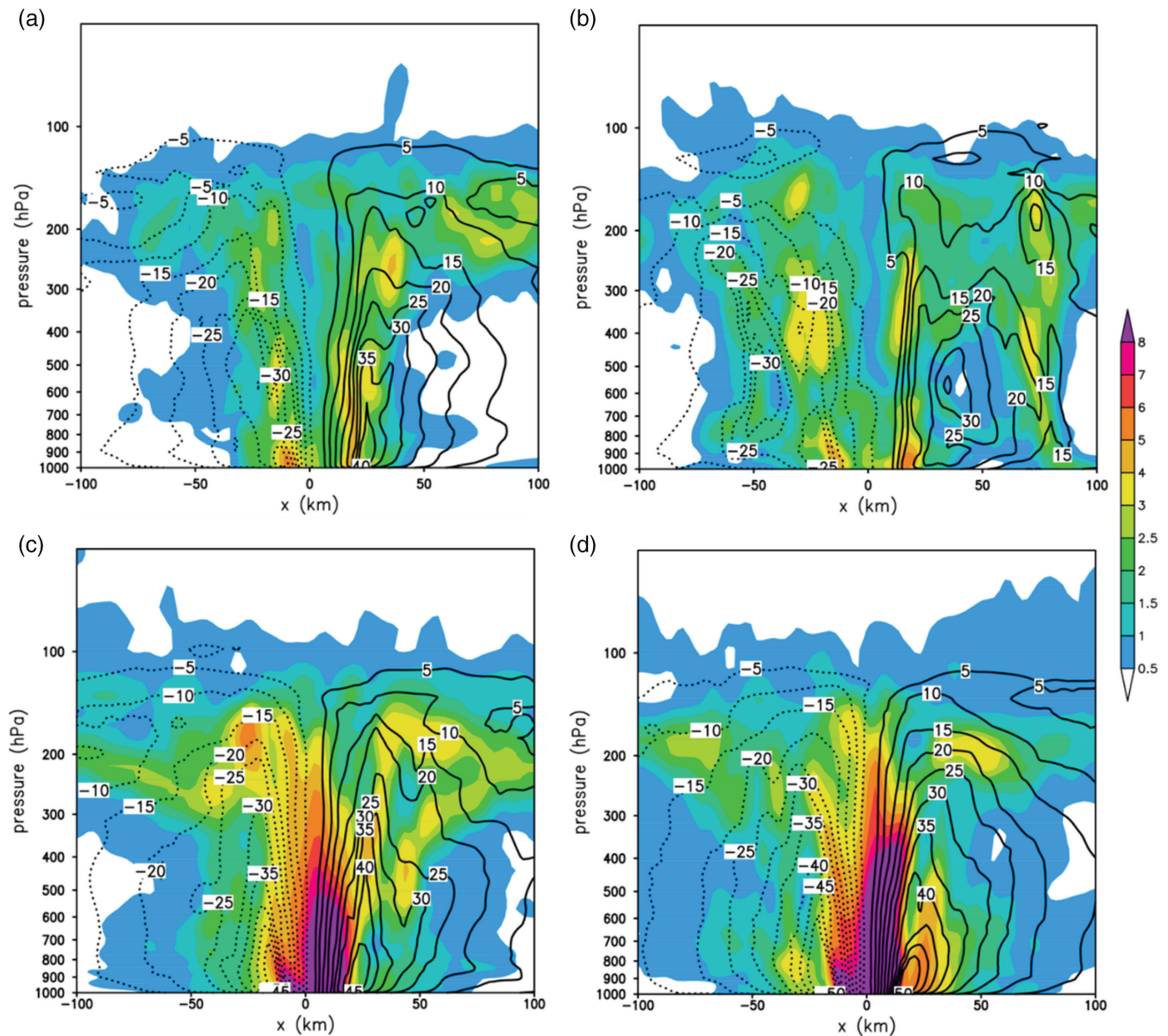


**FIGURE 10** Horizontal cross-sections of the 10m wind (vectors) during the transient stage valid at (a)  $t = 36$  h, (b)  $t = 48$  h, and during the quasi-stationary stage valid at (c)  $t = 72$  h, and (d)  $t = 84$  h into integration, obtained from the control experiment. The ensemble-averaged standard deviations of the wind speed (shading,  $\text{m s}^{-1}$ ) are obtained from experiments with random wind perturbations added to the control trajectory at 12 h into integration. The solid red circles denote the radius of maximum surface wind at each time [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

GFDL models demonstrated the existence of an intensity error saturation at 4–5 day lead times, which varies in the range 12–18 kt. While this saturation of the intensity errors appears to be trivial from a practical perspective given the bounded maximum intensity that a TC can attain in a prescribed environment, the existence of such an error saturation implies a much more profound property of the TC chaotic dynamics that is currently not understood. Specifically, it is not known if a chaotic MPI attractor exist, and if so, what is the size of this attractor and how fast an initial error inside this attractor would evolve. These questions are open at present due to the interference of observational errors and various external factors. In this regard, our results about the error saturation in real-time analyses not only helps confirm the

existence of the chaotic MPI attractor as speculated in KM17, but also indicates the size of this attractor. In particular, we found that the error saturation is basin-dependent, with the lowest value in the North Atlantic basin ( $\sim 12$ – $14$  kt) and the highest value in the northwestern Pacific basin ( $\sim 16$ – $18$  kt). Such different intensity error saturation in different basins as obtained from the HWRF model strongly suggests the HWRF's intensity predictability limit is not universal, but it changes from one basin to another depending on the large-scale conditions.

While the exact value of the error saturation is still inconclusive due to the limited sample size, the observational errors, the model specifics as well as basin characteristics, the existence of such an error saturation as inferred from



**FIGURE 11** As Figure 10, but for the west–east vertical cross-section through the storm centre of the  $V_{MAX}$  deviations (shading,  $m s^{-1}$ ) with respect to the control run, and the tangential wind from the control run (contours,  $m s^{-1}$ ) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

the HWRf real-time analyses is most apparent for strong TCs as they approach their mature stage. From a dynamical system perspective, the existence of such an error saturation suggests that the TC dynamics have an internal property that may prevent us from reducing the absolute intensity errors below a certain threshold, as speculated in KM16. This is a significant finding, because it indicates that a large part of the current intensity errors in operational TC models should be attributed to the intrinsic property of TC dynamics, rather than to model deficiencies. In fact, the error saturation presented in this study is not limited to dynamical models such as the HWRf or GFDL models. Our examination of other statistical-dynamical models used for operational intensity forecasts captures similar error behaviours as in the HWRf model at 4–5 day lead time. However, the error growth in these statistical-dynamical models during the transient

period is somewhat different due to underlying statistical assumptions in these models that are beyond the TC dynamics examined in this study.

By additionally stratifying the real-time intensity forecast errors based on different intensity bins, it was found that the HWRf intensity errors grow consistently faster for higher intensities during the intensification period. Using  $V_{MAX}$  as a metric to measure the growth of intensity errors, we found that the HWRf intensity error growth rate is about  $0.3 \text{ kt h}^{-1}$  for an intensity bin of 25–45 kt, and approaches  $\sim 1.1 \text{ kt h}^{-1}$  for intensity bins  $>85\text{--}100 \text{ kt}$ . For larger intensity bins, the  $V_{MAX}$  error growth rate tends to be smaller but the exact value of this saturation growth rate is very uncertain due to the limited sample size as well as the prevailing bias of the model spin-down associated with unbalanced adjustment at the high-intensity limit.



Given the existence of an intensity error saturation and a faster error growth rate for higher intensities during the intensifying period, we propose that the range of the TC intensity predictability is limited. Our examination of the time interval in which an initial error approaches the saturation limit indeed showed a shorter time to reach the saturation for higher intensities. Assuming that (a) a favourable environment is ensured such that a TC can reach and maintain its MPI limit for the error to grow and become saturated, and (b) the MPI attractor is chaotic such that any error will quickly grow and occupy the entire attractor, this result would imply that the range of intensity predictability for stronger storms should be reduced as they intensify. For Category 1 storms and above, intensity errors approach the saturation limit in <18 h, thus restraining our ability to forecast TC intensity at a longer range. This result is consistent with the estimation based on a linear inverse model presented in figure 11 of Hakim (2013), and the doubling time of 3 h within the axisymmetric framework recently reported in Kieu and Moon (2016). Of course, our conclusion of a limited predictability for TC intensity in this study is applied only to the  $V_{MAX}$  metric and the HWRF model under an assumption that favourable environments are well maintained. It is entirely possible that predicting the phase of the TC development such as rapid intensification or rapid weakening, or use of different intensity metrics, could give a different range of predictability, which this work could not address.

In a further attempt to minimize the impacts of environmental factors and observational errors on the estimation of error saturation in real-time forecasts, idealized experiments with the HWRF model under a perfect model scenario were also performed. By implementing a scheme that allows for adding random perturbations at different stages of the model intensification, it was again demonstrated that the HWRF model possesses an inherent error saturation very similar to that obtained from real-time forecasts, even under the perfect model scenario regardless of when and how the random perturbations are added. However, the saturation limit obtained from the HWRF idealized experiments is  $\sim 3 \text{ m s}^{-1}$ , which is substantially smaller than that obtained from real-time statistics that we don't fully understand.

On the one hand, the smaller error saturation as obtained from the HWRF idealized experiments implies that a large fraction of intensity errors must be associated with other factors such as model errors, vortex initialization, or model physics that we can hope to further reduce in future. On the other hand, the idealized error saturation reveals the minimum absolute intensity threshold that one can reduce in the HWRF model with the current configuration. The exact value of this lowest intensity error threshold as obtained from the HWRF idealized experiments is by no means conclusive, because this threshold depends on the large-scale environment and specific model settings. In addition, our assumptions of stationary intensity error statistics as derived from the 5-day simulations, together with the idealized environment needed to ensure the

statistics of intensifying cycles, are not always applicable in practice. As such the saturation limit of  $\sim 3 \text{ m s}^{-1}$  is specific to the current idealized HWRF experiments, and it may change in future upgrades of the HWRF model. More details of the dependence of the intensity error saturation on the ambient environment will be presented in future studies.

## ACKNOWLEDGEMENTS

This work was supported by the NOAA's Hurricane Forecast Improvement Project (HFIP) (award NA16NWS4680026), the ONR Young Investigator Program (N000141812588), and the Indiana University Faculty Research Fund. Thanks are also extended to Zhan Zhang and Lin Zhu at NCEP/EMC for their help with the data from the H215/H216 retrospective experiments. We wish to also thank three reviewers for their constructive comments and suggestions, which have helped improve this work substantially.

## ORCID

Chanh Kieu  <http://orcid.org/0000-0001-8947-8534>

Kushal Keshavamurthy  <http://orcid.org/0000-0003-1804-4178>

## REFERENCES

- Bhatia, K.T. and Nolan, D.S. (2013) Relating the skill of tropical cyclone intensity forecasts to the synoptic environment. *Weather and Forecasting*, 28, 961–980.
- Bhatia, K.T. and Nolan, D.S. (2015) Prediction of Intensity Model Error (PRIME) for Atlantic basin tropical cyclones. *Weather and Forecasting*, 30(6), 1845–1865. <https://doi.org/10.1175/WAF-D-15-0064.1>.
- Bhatia, K.T. and Nolan, D.S. (2016) Tropical cyclone intensity model improvement: better models or easier forecasts? In: *32nd Conference on Hurricanes and Tropical Meteorology*, San Juan, Puerto Rico
- Bhatia, K.T., Nolan, D.S., Schumacher, A.B. and DeMaria, M. (2017) Improving tropical cyclone intensity forecasts with PRIME. *Weather and Forecasting*, 32, 1353–1377.
- Biswas, M.K., Carson, L., Newman, K., Bernardet, L. and Holt, C. (2016) *Hurricane Weather Research and Forecasting (HWRF) model: community HWRF users' guide v3.8a*. Available at: [http://www.dtcenter.org/HurrWRF/users/docs/users\\_guide/HWRF\\_v3.8a\\_UG.pdf](http://www.dtcenter.org/HurrWRF/users/docs/users_guide/HWRF_v3.8a_UG.pdf) [Accessed 10th September 2018]
- Brown, B.R. and Hakim, G.J. (2013) Variability and predictability of a three-dimensional hurricane in statistical equilibrium. *Journal of the Atmospheric Sciences*, 70, 1806–1820.
- Buizza, R. (2006) The ECMWF ensemble prediction system. In: Palmer, T.N. and Hagedorn, R. (Eds.) *Predictability of Weather and Climate*. Cambridge University Press, pp. 459–488.
- Davis, C., Wang, W., Dudhia, J. and Torn, R. (2010) Does increased horizontal resolution improve hurricane wind forecasts? *Weather and Forecasting*, 25(6), 1826–1841. <https://doi.org/10.1175/2010WAF2222423.1>.
- DeMaria, M. (2010) Tropical cyclone intensity change predictability estimates using a statistical-dynamical model. In: *29th Conference on Hurricanes and Tropical Meteorology, Tucson, AZ*. Available at [https://ams.confex.com/ams/29Hurricanes/techprogram/paper\\_167916.htm](https://ams.confex.com/ams/29Hurricanes/techprogram/paper_167916.htm) [Accessed 10th September 2018]
- DeMaria, M., Sampson, C.R., Knaff, J.A. and Musgrave, K.D. (2014) Is tropical cyclone intensity guidance improving? *Bulletin of the American Meteorological Society*, 95, 387–398.



- Durran, D.R. and Gingrich, M. (2014) Atmospheric predictability: why butterflies are not of practical importance. *Journal of the Atmospheric Sciences*, 71(7), 2476–2488. <https://doi.org/10.1175/JAS-D-14-0007.1>.
- Emanuel, K. and Zhang, F. (2016) On the predictability and error sources of tropical cyclone intensity forecasts. *Journal of the Atmospheric Sciences*, 73(9), 3739–3747. <https://doi.org/10.1175/JAS-D-16-0100.1>.
- Farrell, B.F. (1982) The initial growth of disturbances in a baroclinic flow. *Journal of the Atmospheric Sciences*, 39(8), 1663–1686. [https://doi.org/10.1175/1520-0469\(1982\)039<1663:TIGODI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1982)039<1663:TIGODI>2.0.CO;2).
- Franklin, J.L. and Brown, D.P. (2008) Atlantic hurricane season of 2006. *Monthly Weather Review*, 136(3), 1174–1200. <https://doi.org/10.1175/2007MWR2377.1>.
- Gopalakrishnan, S.G., Goldenberg, S., Quirino, T., Zhang, X., Marks, F., Yeh, K.S., Atlas, R. and Tallapragada, V. (2012) Toward improving high-resolution numerical hurricane forecasting: influence of model horizontal grid resolution, initialization, and physics. *Weather and Forecasting*, 27(3), 647–666. <https://doi.org/10.1175/WAF-D-11-00055.1>.
- Hakim, G.J. (2011) The mean state of axisymmetric hurricanes in statistical equilibrium. *Journal of the Atmospheric Sciences*, 68, 1364–1376.
- Hakim, G.J. (2013) The variability and predictability of axisymmetric hurricanes in statistical equilibrium. *Journal of the Atmospheric Sciences*, 70, 993–1005.
- Jordan, C.L. (1958) Mean soundings for the West Indies area. *Journal of Meteorology*, 15, 91–97.
- Judt, F., Chen, S.S. and Berner, J. (2016) Predictability of tropical cyclone intensity: scale-dependent forecast error growth in high-resolution stochastic kinetic-energy backscatter ensembles. *Quarterly Journal of the Royal Meteorological Society*, 142, 43–57. <https://doi.org/10.1002/qj.2626>.
- Kieu, C. (2015) Hurricane maximum potential intensity equilibrium. *Quarterly Journal of the Royal Meteorological Society*, 141, 2471–2480. <https://doi.org/10.1002/qj.2556>.
- Kieu, C., Tallapragada, V., Zhang, D.L. and Moon, Z. (2016) On the development of double warm-core structures in intense tropical cyclones. *Journal of the Atmospheric Sciences*, 73, 4487–4506.
- Kieu, C.Q. and Moon, Z. (2016) Hurricane intensity predictability. *Bulletin of the American Meteorological Society*, 97, 1847–1857.
- Kieu, C.Q. and Wang, Q. (2017) Stability of tropical cyclone intensity equilibrium. *Journal of the Atmospheric Sciences*, 74, 3591–3608.
- Landsea, C.W. and Franklin, J.L. (2013) Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, 141, 3576–3592.
- Legras, B. and Vautard, R. (1996) A guide to Liapunov vectors. In: *Seminar on Predictability*, 4–8 September 1995. ECMWF, Reading, UK, pp. 143–156. <https://www.ecmwf.int/sites/default/files/elibrary/1995/10702-guide-liapunov-vectors.pdf> [Accessed 12th September 2018]
- Leith, C.E. (1971) Atmospheric predictability and two-dimensional turbulence. *Journal of the Atmospheric Sciences*, 28, 145–161.
- Lorenz, E.N. (1963) Deterministic non-periodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141.
- Lorenz, E.N. (1969) The predictability of a flow which possesses many scales of motion. *Tellus A*, 21A, 289–307.
- Lorenz, E.N. (1990) Predictability – A problem partly solved. In: Palmer, T.N. and Hagedorn, R. (Eds.) *Predictability of Weather and Climate*. Cambridge, UK: Cambridge University Press, pp. 40–58.
- Lorenz, E.N. (1996) Predictability – A problem partly solved. In: *Seminar on Predictability*, 4–8 September 1995. ECMWF, Reading, UK, pp. 1–18. <https://www.ecmwf.int/sites/default/files/elibrary/1995/10829-predictability-problem-partly-solved.pdf> [Accessed 12th September 2018]
- Métais, O. and Lesieur, M. (1986) Statistical predictability of decaying turbulence. *Journal of the Atmospheric Sciences*, 43, 857–870.
- Nguyen, V.S., Smith, R.K. and Montgomery, M.T. (2008) Tropical-cyclone intensification and predictability in three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 134, 563–582.
- Palmer, T.N., Doring, A. and Seregin, G. (2014) The real butterfly effect. *Nonlinearity*, 27, 123–141.
- Pu, Z., Zhang, S., Tong, M. and Tallapragada, V. (2016) Influence of the self-consistent regional ensemble background error covariance on hurricane inner-core data assimilation with the GSI-based hybrid system for HWRF. *Journal of the Atmospheric Sciences*, 73, 4911–4925.
- Rotunno, R. and Emanuel, K.A. (1987) An air–sea interaction theory for tropical cyclones. Part II: evolutionary study using a nonhydrostatic axisymmetric numerical model. *Journal of the Atmospheric Sciences*, 44, 542–561.
- Rotunno, R. and Snyder, C. (2007) A generalization of Lorenz’s model for the predictability of flows with many scales of motion. *Journal of the Atmospheric Sciences*, 65, 1063–1076.
- Sampson, C.R. and Schrader, A.J. (2000) The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bulletin of the American Meteorological Society*, 81, 1231–1240.
- Tallapragada, V., Kieu, C., Kwon, Y., Trahan, S., Liu, Q., Zhang, Z. and Kwon, I.H. (2014) Evaluation of storm structure from the operational HWRF during 2012 implementation. *Monthly Weather Review*, 142, 4308–4325. <https://doi.org/10.1175/MWR-D-13-00010.1>.
- Tallapragada, V., Kieu, C., Trahan, S., Liu, Q., Wang, W., Zhang, Z., Tong, M., Zhang, B., Zhu, L. and Strahl, B. (2015) Forecasting tropical cyclones in the western North Pacific basin using the NCEP operational HWRF model. Model upgrades and evaluation of real-time performance in 2013. *Weather and Forecasting*, 31, 877–894. <https://doi.org/10.1175/WAF-D-14-00139.1>.
- Tien, D.D., Ngo-Duc, T., Mai, H.T. and Kieu, C. (2013) A study of the connection between tropical cyclone track and intensity errors in the WRF model. *Meteorology and Atmospheric Physics*, 122(1–2), 55–64. <https://doi.org/10.1007/s00703-013-0278-0>.
- Torn, R.D. and Snyder, C. (2012) Uncertainty of tropical cyclone best-track information. *Weather and Forecasting*, 27, 715–729.
- Trahan, S. and Sparling, L. (2012) An analysis of NCEP tropical cyclone vitals and potential effects on forecasting models. *Weather and Forecasting*, 27(3), 744–756. <https://doi.org/10.1175/WAF-D-11-00063.1>.
- Trevisan, A. and Pancotti, F. (1998) Periodic orbits, Lyapunov vectors, and singular vectors in the Lorenz system. *Journal of the Atmospheric Sciences*, 55(3), 390–398. [https://doi.org/10.1175/1520-0469\(1998\)055<0390:POLVAS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<0390:POLVAS>2.0.CO;2).
- Vukicevic, T., Uhlhorn, E., Reasor, P. and Klotz, B. (2013) A novel multiscale intensity metric for evaluation of tropical cyclone intensity forecasts. *Journal of the Atmospheric Sciences*, 71(4), 1292–1304. <https://doi.org/10.1175/JAS-D-13-0153.1>.

**How to cite this article:** Kieu C, Keshavamurthy K, Tallapragada V, Gopalakrishnan S, Trahan S. On the growth of intensity forecast errors in the operational hurricane weather research and forecasting (HWRF) model. *Q J R Meteorol Soc.* 2018;144:1803–1819. <https://doi.org/10.1002/qj.3344>