Can Machine Learning Predict any Chaos in Tropical Cyclone Intensity?

Chanh Kieu[1]

[1]Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN 47405

Under Review: *27 January 2024*

* Correspondence to: Department of Earth and Atmospheric Sciences, GY4057A, Geological Building, Indiana University, Bloomington, Indiana 47405. Tel: 812-856-5704. Email: ckieu@indiana.edu.

Abstract

Whether tropical cyclones (TC) possess chaotic dynamics is an open question in current TC research. The existence of such chaotic dynamics is profound for TC model development and operational forecast, as it sets a limit on how much one can further improve intensity forecast skills or models in the future. Rapid advances of machine learning (ML) techniques and applications open up an opportunity to explore TC intensity chaos from a different angle. Building upon our recent results on the low-dimensional chaos of TC intensity, this study presents a novel use of ML models to quantify TC intensity chaos. By treating TC scales as input features for ML models, we show that TC intensity displays a limited predictability range of ~3 hours due to chaotic variability at the potential intensity (PI) equilibrium. This short predictability range for TC intensity is robust across ML architectures including deep neural networks (DNN), gated recurrent units (GRU), and long-short term memory (LSTM) examined in this study. Using the minimum central pressure as a metric for TC intensity could extend the predictability range to 5-6 hours, yet the limited predictability for TC intensity is still well captured in all ML models. As a result, the intrinsic variability of TC intensity related to low-dimensional chaos prevents intensity errors in any TC model from being arbitrarily reduced, regardless of how perfect a TC model or vortex initialization is. Our findings support the existence of chaotic dynamics at the PI limit and demonstrate an innovative way of applying ML to study atmospheric predictability.

Significance Statement

How much further one can improve tropical cyclone (TC) intensity in the future is an important yet challenging question in current TC research and prediction. Recent studies have suggested a potential existence of TC intensity chaos at the maximum intensity limit that puts a bound on the improvement of TC intensity forecast. In this study, a novel use of machine learning is presented to detect chaos in TC intensity under idealized conditions. The results reveal a very short predictability range of only ~3-6 hours after TCs reach their maximum intensity equilibrium. Due to this chaotic nature, TC intensity must possess an intrinsic variability that cannot be eliminated, which dictates a limit in TC intensity forecast accuracy even for perfect TC models and initial conditions.

Tropical cyclone intensity predictability

Searching for the limit in predicting tropical cyclone (TC) intensity has been a challenging problem in TC research and operation. One key difficulty in studying TC intensity predictability (TIP) is rooted in an open question of whether TC dynamics possesses chaos at any stage of TC development (Kieu and Rotunno 2022, Kieu et al. 2022, hereafter K22). For practical purposes, a TC intensity forecast must be issued from an early formation to the final dissipation stage, yet all predictability frameworks require a stationary attractor or fully-developed turbulent state such that statistical properties can be well-defined (e.g., Lorenz 1963, 1969, Kraichnan 1967, Leith 1971, Métais and Lesieur 1986, Vallis 2017). This fundamental requirement for chaotic dynamics explains confusingly different estimations for TIP, which varies from 3 hours to 7 days in previous studies (e.g., Hakim 2011, 2013, Emanuel and Zhang 2016, Kieu and Moon 2016 hereinafter KM16, Judt et al. 2016, Zhong et al. 2018).

Of all TC development stages, the only one that appears to meet the requirement for examining chaos is the maximum intensity state, known as the TC potential intensity (PI, Emanuel 1986, 2003). According to the PI theory, TCs will reach a steady state with a maximum intensity determined by environmental conditions. The existence of this PI state and its related stability have been extensively studied in numerous observational, theoretical, and modelling studies (e.g., Rotunno and Emanuel 1987, Holland 1997, Bryan and Rotunno 2009, Hakim 2011, 2013, Kieu 2015). However, whether a PI limit really exists is still inconclusive, as several modelling studies, e.g., by Smith et al. (2014, 2021) or Persing et al. (2019) showed that a TC vortex cannot maintain a steady state due to the transport of low angular momentum from upper levels to the surface. This process cuts off the supply of high angular momentum from the outer-core region and eventually weaken TC intensity, even under idealized environments.

Despite the controversial existence of the PI state, the fact that the maximum TC intensity can be captured and well maintained in very long integrations (e.g., Hakim 2011, Brown and Hakim 2013, KM16, K22) suggests that TC dynamics can settle down in a quasi-stationary equilibrium, if proper environments are designed. Such an equilibrium, hereinafter referred to as the PI equilibrium, offers a unique opportunity to quantify TIP in accordance with the current chaos theory. Specifically, the PI equilibrium helps define a reference climatology for TC intensity, on which one can measure an error growth over time. The range of predictability is then the maximum time interval at which a forecast distribution of TC intensity becomes indistinguishable from its PI climatology. Given a measure for such difference between intensity distributions, a predictability range can be then obtained by using, e.g., the decorrelation time, integrated time, or signal-noise ratios as studied in e.g., Lorenz (1969), Shukla (1981), Schneider and Griffies (1999), DelSole and Tipett (2007, 2009).

While real-time forecasts have strongly hinted at a possible limited predictability for TC intensity, quantifying the TIP range turns out to be difficult due to various ways that one can define a reference climatology for TC intensity. Note that predictability is not a universal measure, as it must be associated with one variable and a specific period during which a reference climatology is constructed. Thus, predictability can be different for different intensity metrics such as the maximum sustained wind (V_{max}), or the minimum central pressure (P_{min}). Because of this metric dependence, any estimation of TIP must be specific to an intensity metric and its climatology.

In the next, we will examine TIP within a framework of deterministic chaos, which is well suited for point-like intensity metrics such as V_{max} or P_{min}. Any extension to the multi-scale predictability framework would require a different approach and interpretation that are beyond the scope of this work, and so we will not present this extension hereinafter.

Intensity low-dimensional chaos

Given that the PI equilibrium is a possible state of TC intensity whose statistical properties are stationary for a meaningful intensity climatology, a natural question is how this equilibrium can help evaluate the predictability range for TC intensity. Along this line, KM16 presented an interesting approach based on a fidelity-reduced dynamics model proposed by Kieu (2015), Kieu and Wang (2017). Using TC scales obtained from a long integration of Rotunno and Emanuel (1987)'s axisymmetric model as dynamical variables, KM16 demonstrated that TC intensity always approaches a chaotic region in the phase space constructed from the scales of TC tangential wind, radial wind, and warm core anomaly. In this phase space, PI is no longer a single point but a bounded region with all properties of a typical chaotic attractor. A direct implication of this chaotic PI attractor is that TC intensity must possess some intrinsic variability, even for a perfect TC model under ideal conditions.

Of further importance about the existence of such a PI chaotic attractor is that TC intensity, once settling down in the PI equilibrium, should have limited predictability. KM16's direct estimation from a leading Lyapunov exponent in their axisymmetric simulations gives a predictability range of ~2-3 days. Any attempt to predict TC intensity beyond this limit would not be better than simply using an averaged PI value in a given environment. KM16's estimation of TIP differs from that in previous studies, which proposed a wide range from 3-9 hours in K22, 1-3 days in Hakim (2013) or Zhong et al. (2018) to even more than 7 days in Emanuel and Zhang (2016). Such a wide range of estimation for TIP is due not only to the dependence of PI on specific model dynamics, ocean basin, or environmental conditions, but also to how one defines the reference climatology for intensity.

Despite the inconclusive range for TIP, the possible existence of intensity chaos in a low-dimensional space is itself important from several perspectives. First, this low-dimensional attractor helps justify why forecasters can use only few bulk numbers such as V_{max}, P_{min}, storm size, or cloud top temperature to characterize a TC, instead of all possible details of TCs. This is also consistent with the fact that TC intensity models with only few degrees of freedom could capture well some broad properties of TC intensity as shown in previous studies (e.g., Emanuel 2012, DeMaria 2009, Wang et al. 2021, Schonemann and Frisius 2012, Kieu 2015, KM16).

Second, the existence of a low-dimensional attractor indicates that PI should not be represented by a single V_{max} value as in the current PI framework. Instead, the maximum intensity that a TC can get must vary within a range around the PI equilibrium as presented in Hakim (2011, 2013), regardless of how perfect an environmental condition or a TC model is. As a result, this intrinsic variability of TC intensity will act as a "noise" level in any TC intensity statistics that one has to take into account when detecting the change of PI under different climate conditions.

Third, the PI equilibrium is no longer just about V_{max}. Instead, PI has to be characterized by other features as well such as the warm core anomaly, the maximum radial wind (U_{max}), the radius of maximum wind (RMW), or the maximum eyewall vertical motion (W_{max}). This important property of the PI equilibrium was well demonstrated in Kieu (2015)'s TC-scale model, which showed strong fluctuation of V_{max} with time even when the initial condition for V_{max} is set to be exactly equal to the PI value, but with the warm core anomaly reduced by half. In this regard, any factor that can influence other dimensions of PI would cause strong fluctuation in V_{max}, regardless of whether V_{max} is equal to PI or not.

Given these implications of intensity chaos, a better understanding of its characteristics is needed so that a more accurate estimation of the TIP range can be obtained. Among several uncertainties in understanding the PI attractor, two important issues stand out. First, it is still not known if a PI attractor can actually be represented by few dimensions as examined in KM16. The phase-space reconstruction method in K22 is one way to directly search for the dimension of the PI attractor in terms of attractor invariants. This technique is powerful to examine low-dimensional chaos, but it contains significant subjectivity and is highly sensitive to data noise as discussed in, e.g., Kantz and Schreiber (2003). Second, the current estimation of the TIP range still varies widely among different models and methods. Thus, improving this TIP estimation is a warranted question that we want to address next, using machine learning.

Machine learning mapping

Broadly speaking, machine learning (ML) can be considered as a framework that can search for rules from data. Given an ML architecture, a measure of accuracy (i.e., a loss function), and input data, the rules can be obtained within a prescribed accuracy. The key advantage of ML in practical applications lies in its ability to learn rules from a vast amount of data without *a priori* knowledge, provided that the input data is sufficiently good[1]. With an inherently large volume of data, climate and weather prediction provide a great domain for ML applications, which justifies the surge of ML applications in atmospheric science recently.

Specifically for TC intensity, ML offers a unique way to study low-dimensional chaos. To set up a context for applying ML to our TC intensity chaos problem, we will focus on supervised ML, which requires a set of input data and corresponding targets (labels) for training an ML model. At a basic level, supervised ML models need a surjective mapping between an input training dataset ($\mathcal{T}$) and a target dataset ($\mathcal{L}$) (i.e., one $y \in \mathcal{L}$ will have at least one $x \in \mathcal{T}$) so that the training can be carried out. For a typical time-prediction problem (i.e., given a state of a system at one time $t = 0$, one needs to predict the state of the system at a later time $t = \tau$), this mapping can be considered as a propagator from a given initial condition to the later time τ. Mathematically, this propagator can be expressed as $x(\tau) = M(\tau)x(0)$, where $M(\tau)$ is the propagator from $t = 0$ to τ and $x(t)$ is the model state at time t.

For a full-physics model, $M(\tau)$ is nothing but a numerical model with governing equations integrated from $t = 0$ to $t = \tau$. For ML, $M(\tau)$ is a however nonlinear operator that is learned from a training dataset. In principle, the more data we have, the better an ML model can search for underlying rules and build $M(\tau)$ *without any physical equations*. Thus, we can feed an ML model

[1] A good set of training data should ensure several criteria including i) comprehensiveness, ii) relevancy, ii) consistency, and iv) uniformity.

with a large amount of data, and let it figure out the best possible relationships between $t = 0$ and $t = \tau$. For deep learning that is based on neutral networks, an ML model with sufficient layers and depth should capture a nonlinear mapping between two time slices, making it suitable for TC intensity prediction.

From this perspective, it is immediate that chaotic systems will pose a challenge to any ML model, because one input[2] may give totally different outcomes after reaching predictability limit T (i.e., one $x \in \mathcal{T}$ would give two different $y_1, y_2 \in \mathcal{L}$, $\forall \tau > T$). So, there exists no good mapping between the training and the label datasets, and ML models cannot learn any rule from data. This deterioration of ML models after entering the chaotic regime suggests, however, a very unique way to quantify predictability for chaotic systems. Specifically, we will search for a lead time T beyond which an ML model can no longer be trained from any input dataset, which gives us a direct estimation for a predictability range of the chaotic system. This approach is natural in the sense that an ML model should generally be able to predict the next state of a system from a given input, if the system remains predictable and sufficient training data is provided. Thus, ML models are a great tool for studying chaos.

Our use of ML to examine predictability as proposed above is in fact well suited for the TC intensity problem, as this approach serves two purposes: i) it verifies if a low-dimensional representation is sufficient for TC intensity, and ii) it helps estimate the TIP range in that low-dimensional phase space. Recall from the aforementioned discussions that the existence of low-dimensional chaos for TC intensity is currently questionable because we do not know if the few dimensions based on TC scales suffice to characterize TC intensity at the PI equilibrium. Also, whether TC intensity displays any chaos on these dimensions is still inconclusive. Our underlying hypothesis here is that a ML model can predict TC intensity in a low-dimensional phase space, whose dimensions correspond to the few TC scales, up to a certain lead time T. Beyond this lead time T, ML models can no longer be trained to predict TC intensity, thus revealing TC low-dimensional chaos and providing us a direct estimation for the TIP range.

The results in K22 provide a pathway to verify this hypothesis with ML. Specifically, we will assume that TC intensity can be described by four dimensions corresponding to four TC scales including $V_{max}, U_{max}, W_{max},$ and P_{min},. While it is not known in advance the exact dimension of

[2] Note here that two input x_1, x_2 are considered to be the same in a practical sense if their difference is within some measurement errors, $|x_1 - x_2| \leq \epsilon$..

the PI chaotic attractor, K22 suggested that a dimension of 4 should be sufficient to capture TC intensity chaos within the deterministic framework. As such, we will treat these four dimensions as input features for ML models and examine how these ML models can forecast intensity at different lead times, using the same output from a 100-day simulation with the CM1 model as in K22 (see Supplement Information for all details of our ML models and CM1 simulations). So long as ML training can still converge at a given τ, TC intensity will be considered to be predictable for that lead time as desired.

Detecting intensity chaos by machine learning

Given the current definition of TC intensity in terms of the maximum 10-m wind, we examine first the TIP range using V_{max} as a metric for TC intensity in three different ML models including deep neural networks (DNN), gated recurrent units (GRU), and long-short term memory (LSTM) model. Figure 1 shows the training absolute mean error as a function of iterations (epoch) for three forecast lead times including 3 minutes, 1 hour, and 3 hours. One notices first that the training errors rapidly decrease for τ =3 minutes in all three ML models, reaching a relative minimum error of ~ 1%, 5% and 7% for LSTM, GRU, and DNN models on a test data. Looking at the correlation between the ML-forecast and the true TC intensity for the test data in Fig. 2 (red points), it is apparent that all ML models could predict TC intensity variability for τ =3 minutes in the 4-dimensional phase space ($V_{max}, U_{max}, W_{max}, P_{min}$) as expected. This is interesting, as these ML models require a minimum number of input features (4 in our case here), yet they could produce good forecast of TC intensity based solely only a training data. In this regard, Figs. 1 and 2 confirm that a low-dimensional phase space could predict well TC intensity variability at least for a short lead time of 3 minutes, without any physical or governing equations.

At the 1-hour lead time, Fig. 1 shows however that all three ML models start losing their forecast accuracy quickly, and by 3 hours, all ML models cannot be trained any longer, with their errors roughly the same ~75-85% relative to the initial error value during the entire training period. Their predictions for the test set at the 3-hour lead time displays indeed almost completely no correlation to the true intensity (Fig. 2, blue dots). This result is consistent with the estimation from attractor invariants based on a leading Lyapunov exponent and the Sugihara-May correlation in K22, which also showed that TC intensity loses predictability in just ~3-6 hours at the PI equilibrium.

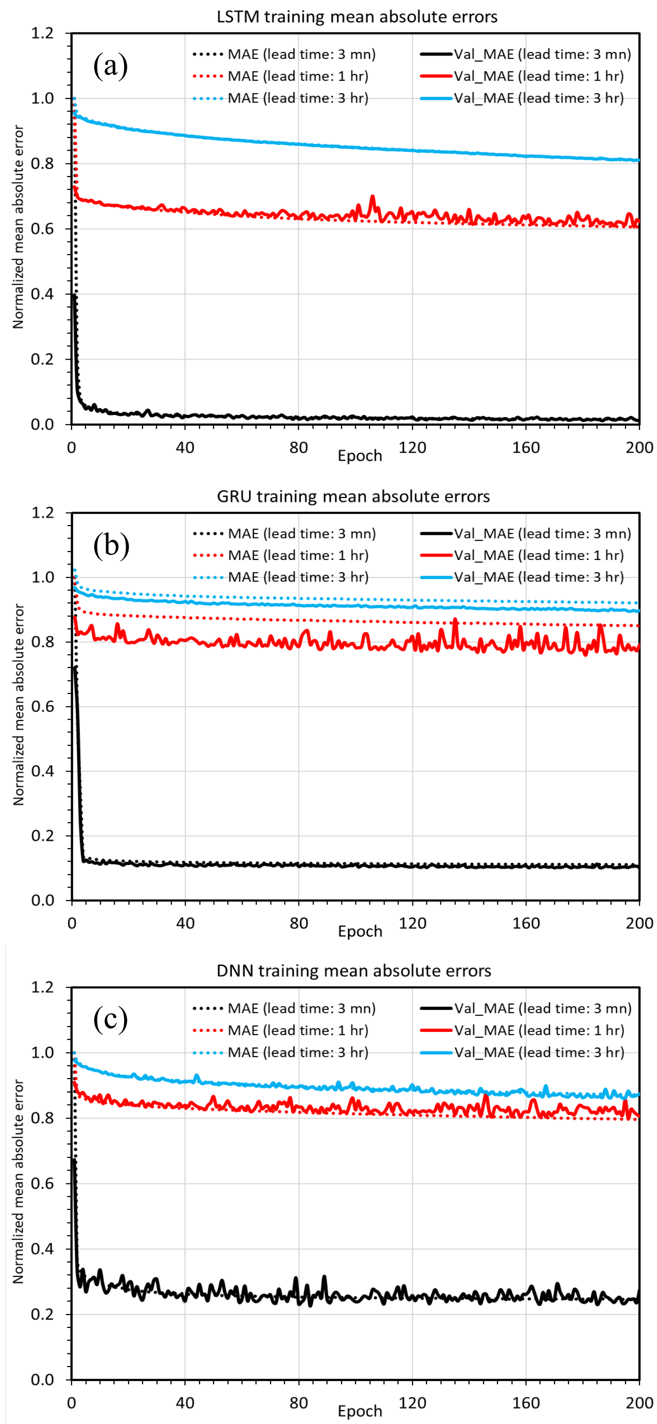

Figure 1. ML accuracy metric based on the mean absolute error (dotted lines) during the training process as a function of iterations (epochs) for three different ML models a) LSTM, b) GRU, and c) DNN at forecast lead times of $\tau = 3$ minutes (black), 1 hour (red), and 3 hours (blue). All absolute errors are normalized by the errors at the first iteration (epoch 1) for better comparison among different lead times. Solid lines denote the mean absolute errors for the corresponding validation dataset in each training process.

The dependence of these ML intensity predictions on forecast lead times is best seen when we compare their predictions to a reference (or climatology) forecast, which is taken to be a simple average of V_{max} at the PI equilibrium. Figure 3a shows the forecast skill of three ML models relative to this reference forecast as a function of lead times. It is apparent from Fig. 3a that ML models perform best for $\tau < 3$ hours. Beyond this, the ML-based prediction skill is no better (in terms of absolute errors) than a simple forecast using the average of V_{max} at the PI equilibrium. We emphasize here that such decaying of the ML forecast skill with τ does not hold true for any system. In fact, a simple experiment using purely random data for ML training would result in zero forecast skill at all lead times (not shown). On the other hand, for periodic systems, the forecast skill is a constant value of 1 for τ as discussed in, e.g., Sugihara and May (1990). In this regard, the decaying curve of the forecast skill shown in Fig. 3 is a manifestation of chaotic systems as captured by ML models.

Given the average $V_{max} = 84$ m s^{-1} with a standard derivation is ~ 7.5 m s^{-1} at the PI equilibrium, the TIP range obtained from Fig. 3a implies that TC intensity will vary indistinguishably within an internal of 84 ± 7.5 m s^{-1} in just 3 hours, even for a perfect TC model. This TIP range may be shortened further if stochastic forcings, asymmetric processes, or model internal errors are taken into account as discussed in Nguyen et al. (2020) or K22, which are however beyond the scope of our study here. Despite these issues, the results obtained from the ML models herein can at least support that TC intensity must have some intrinsic variability due to TC chaotic dynamics, which prevents TC intensity errors in any TC model or real-time forecast from being reduced indefinitely.

Among the three ML models, we note that LSTM appears to perform best in terms of the training and validation mean errors. Such a better performance of recurrent models over DNN models has more subtle implication in practical weather prediction. To see the significance of this recurrent networks, recall that an input at single time $t = 0$ is needed for DNN to make predictions at a lead time τ, very similar to a typical weather or climate model procedure. For LSTM/GRU models, the input data is however given over an interval of past data (i.e., $t \in [-N, 0]$) for a prediction at a lead time τ. This interval information allows these models to learn and memorize long-term dependencies in the past via multiple memory cells and gate controls as in LSTM or a single gate as in GRU models. The different performance between DNN and LSTM/GRU is seen more apparent at the longer lead times, as the past memory becomes increasingly more important when the data from a single time slice is no longer sufficient.

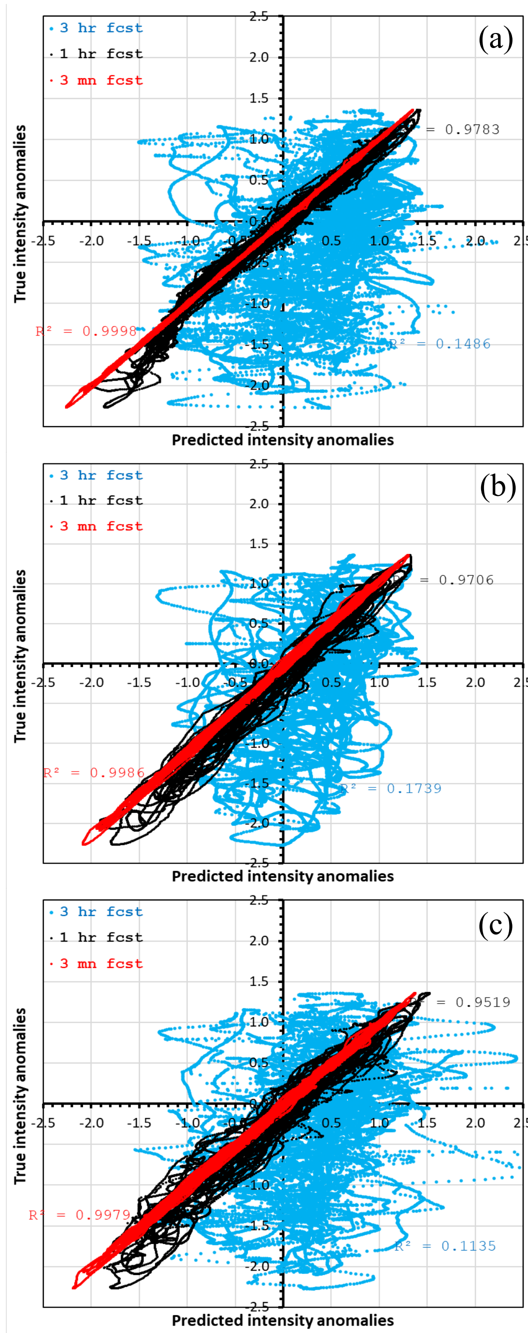

Figure 2. Scatter plots of the ML-predicted TC intensity anomaly (x-axis) and the CM1 true intensity anomaly (y-axis) for a test dataset taken between $t = 90 - 100$ days of the CM1 simulation at three lead times: τ =3 minutes (red), 1 hour (black), and 3 hours (blue) for a) LSTM, b) GRU, and c) DNN model. Note that TC intensity anomaly is relative to the average PI value of 84 ms^{-1} and normalized by its standard deviation $\sigma_V = 7.5$ m s^{-1}. The R values for each lead time best fit are also provided in each panel.

The use of extra information from a past interval to help improve future prediction as in LST/GRU presents a very different way of forecasting as compared to the traditional approach based on physical principles. To some extent, recurrent networks improve their prediction in the same way that four-dimensional data assimilation optimizes an initial condition over an interval instead of just one time slice. Despite this extra information from the past, the predictability of TC intensity could not be lengthened beyond 3 hours in both LSTM and GRU models as shown in Fig. 3. In fact, our sensitivity with deeper neural networks or a longer interval of past information for LSTM/GRU does not improve at all this TIP range, so long as TC intensity settles down in its PI equilibrium. Such a robust TIP estimation among ML models thus highly indicates the existence of TC intensity chaos at the PI equilibrium as previously speculated.

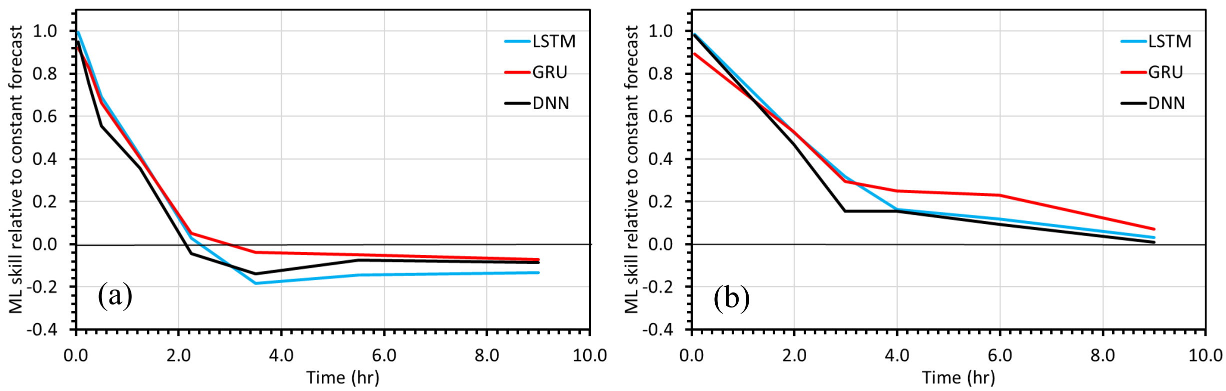

Figure 3. (a) Forecast skill of three ML models LSTM (blue), GRU (red) and DNN (black) as a function of lead time relative to the reference forecast that uses the average V_{max} value at the PI equilibrium, and (b) similar to (a) but using P_{min} for TC intensity. Here, the forecast skill is defined as $1-\frac{MAE_{ML}}{MAE_{ref}}$, where $MAE_{ML}$and MAE_{ref} are the mean absolute errors from the ML predictions and reference prediction of TC intensity over the test dataset, respectively.

Because predictability is metric-dependent, an apparent question is how the TIP range varies when using P_{min} for TC intensity instead of V_{max}. By applying the phase-space reconstruction for a P_{min} time series, K22 noticed that P_{min} possesses a smaller Lyapunov exponent and a longer Sugihara-May correlation. This suggests that P_{min} would take a longer time to approach its saturation limit, thus allowing for a longer range of predictability as noticed in previous studies (e.g., Magnusson et al. 2019, Klotzbach et al. 2020). Given the capability of our ML models, it is natural to extend the above analyses of ML forecast skill to P_{min}. In this regard, Fig. 3b shows the ML forecast skill of P_{min} as a function of lead time for three ML models, similar to Fig. 3a. It is of significance to observe that all ML models could capture similar decaying of the P_{min} forecast skill, but with a longer TIP range of ~5-6 hours as compared to 3 hours for V_{max}. The fact that these ML models could capture such a different predictability range between V_{max} and P_{min} is intriguing. Recall that V_{max} and P_{min} are highly correlated in terms of temporal variability due to their pressure-wind relationship. However, P_{min} represents the total mass at the storm center while V_{max} fluctuates more vigorously due to fine-scale processes at each model grid point. As such, P_{min} tends to better display a slow component of TC dynamics for which ML models could indeed detect, even when training data contains strong fluctuations from the wind field. From this perspective, using P_{min} for TC intensity could lengthen the range of intensity predictability for the operational forecast as previously noticed.

Regardless of intensity metrics, the above results demonstrate the existence of a TIP range consistent with the phase-space reconstruction reported in K22. While the practical application of this predictability range is somewhat restricted due to the requirement of a PI equilibrium, it indicates that intensity variability is an inherent part of TC dynamics in the world of TC numerical models. Thus, one needs to take into account intrinsic intensity variability when planning for the future improvement of any operational models.

Discussions

In this study, we presented a different use of ML to answer a question "*can machine learning detect any chaos in TC intensity*?". Our answer is "*Yes, it can*". This answer was obtained from a premise that TC intensity at the PI equilibrium can be described by a chaotic attractor in a low-dimensional phase space. By treating the dimensions of TC intensity attractor as input features for ML training, the skill of ML prediction can be estimated as a function of forecast lead times. Searching for a lead time that ML models can no longer provide skillful TC intensity prediction could establish the range of intensity predictability, which is ~2-3 hours based on V_{max}. The predictability range could be lengthened up to 5-6 hours if P_{min} is used for TC intensity instead of V_{max}, yet the limited predictability for TC intensity is still warranted in all ML models.

Our ML estimation of the predictability range is based on an assumption that TC intensity can be characterized by a four-dimensional phase space consisting of $(U_{max}, V_{max}, W_{max}, P_{min})$. How this TIP range changes in higher dimensions or with a different set of phase space variables remains elusive at present. There are several TC scales such as the radius of maximum wind, cloud top temperature, outflow temperature, or TC outer size that one could use to reconstruct TC intensity phase space. However, the insofar consistency among different ML models and estimation methods for TIP highly indicates that adding more dimensions or variables may not improve much the predictability range that is obtained in our ML models. From this perspective, our results present a unique use of ML for quantifying TIP. In fact, the approach of estimating a predictability range based on ML models as proposed herein is very generic and can be applied to any dynamical system. One can, for example, use the Lorenz 40-variable model to extract a time series of one variable in the chaotic regime and construct a ML model to search for the predictability range in a phase space of the delayed coordinates. Comparing this ML-based predictability range with that estimated from leading Lyapunov exponents or Sugihara-May correlation can help validate our approach, for which we will report in our upcoming study. So long as a dynamical system contains low-dimensional chaos, one can always use the data on those dimensions as input features for ML training to search for the range of predictability as designed.

Beyond the point-like intensity metrics such as V_{max} or P_{min}, one can also examine TIP from a multi-scale error growth framework as for turbulent systems. In this multi-scale framework, a prerequisite is the existence of a fully-developed homogeneous and isotropic state such that its energy spectrum and related error growth can be measured (Lorenz 1969, Leith and Kraichnan 1972, Métais and Lesieur 1986, Rotunno and Snyder 2007, Durran and Gingrich 2014). As discussed in Kieu and Rotunno (2022), TC dynamics is however generally nonhomogeneous, even at the quasi-stationary PI equilibrium. Unlike a homogenous turbulence for which all points are equally important, TCs possess an eye whose dynamics and thermodynamics are different from the rest. Using spectral analyses, Kieu and Rotunno (2022) showed in fact that the power spectrum of kinetic energy is different between these radial and azimuthal directions. In both directions, the error growth approaches a saturation limit between 9-18 hours, suggesting a limited predictability from the energy spectra perspective. Quantifying the TIP range in this multi-scale framework requires an error growth equation for each direction that is however beyond the scope of ML applications. Thus, we have not applied ML to studying TC intensity predictability within the multi-scale framework in this study.

Acknowledgments: This study is partially supported by the NSF (AGS # 2309929).

Data availability statement. The TC intensity time series used in this study are obtained from the same CM1 simulations as used in Kieu et al. (2022), which can also be directly accessed from a data repository: http://dx.doi.org/10.13140/RG.2.2.30264.01280.

References

Brown, B.R., and G. J. Hakim, 2013: Variability and predictability of a three-dimensional hurricane in statistical equilibrium. *J. Atmos. Sci.*, **70**, 1806–1820.

Bryan, G. H., R. Rotunno, 2009: The maximum intensity of tropical cyclones in axisymmetric numerical model simulations. *Mon. Wea. Rev*. **137**, 1770-1789.

DelSole, T., and M. K. Tippett, 2007: Predictability: Recent Insights from Information Theory. *Rev. Geophys.,* **45**, RG4002, doi:10.1029/2006RG000202.

DelSole, T., and M. K. Tippett, 2009: Average Predictability Time. Part II: Seamless Diagnosis of Predictability on Multiple Time Scales. *J. Atmos. Sci*., 66, 1188-1204.

DeMaria, M., 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev*., **137**, 68–82

Emanuel, K. A., 1986: An air–sea interaction theory for tropical cyclones. Part I: Steady-state maintenance. *J. Atmos. Sci*. **43**, 585–605.

Emanuel, K., and F. Zhang, 2016: On the predictability and error sources of tropical cyclone intensity forecasts. *J. Atmos. Sci*., **73**, 3739–3747.

_____, 2003. Tropical cyclones. *Ann. Rev. Earth Plan. Sci.* **31**, 75–104.

_____, 2012: Self-Stratification of tropical cyclone outflow. Part II: Implications for storm intensification. *J. Atmos. Sci*., **69**, 988–996

Durran, D.R. and M. Gingrich, 2014: Atmospheric predictability: why butterflies are not of practical importance. *J. Atmos. Sci*., **71**, 2476–2488.

Hakim, G. J., 2011: The mean state of axisymmetric hurricanes in statistical equilibrium. *J. Atmos. Sci*. **68**, 1364–1376.

____, 2013: The variability and predictability of axisymmetric hurricanes in statistical equilibrium. *J. Atmos. Sci*. **70**, 993–1005.

Holland, G. 1997: The Maximum Potential Intensity of Tropical Cyclones. *J. Atmos. Sci.* **54**, 2519–2541.

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. *Q.J.R. Meteorol. Soc*. **142**, 43–57. DOI: 10.1002/qj.2626.

Kantz, H., and T. Schreiber, 2003: Nonlinear Time Series Analysis. 2nd ed., Cambridge University Press, doi:10.1017/CBO9780511755798.

Keshavamurthy, K, and C. Kieu, 2021: Dependence of tropical cyclone intrinsic intensity variability on the large-scale environment. *Quarterly Journal of the Royal Meteorological Society*. **147**, 1606-1625.

Klotzbach, P. J., M. M. Bell, S. G. Bowen, E. J. Gibney, K. R. Knapp, and C. J. Schreck, 2020: Surface pressure a more skillful predictor of normalized hurricane damage than maximum sustained wind. *Bulletin of the American Meteorological Society*, **101**, 830 – 846.

Kraichnan, R.H., 1967: Inertial ranges in two-dimensional turbulence, *Physics of Fluids*, **10**, 1417–1423.

Kieu, C. Q., and Z. Moon, 2016: Hurricane intensity predictability. *Bulletin of the American Meteorological Society*, 97, 1847-1857

Kieu**,** C, Q., W. Cai, and L Fan. 2022: On the existence of low-dimensional chaos of the tropical cyclone intensity in an idealized axisymmetric simulation? *Journal of the Atmospheric Sciences,* **80**, 797-811.

Kieu, C. Q., and R. Rotunno, 2022: Characteristics of tropical-cyclone turbulence and intensity predictability. *Geophysical Research Letter.* **49***,* e2021GL096544.

Kieu, C. Q., 2015: Hurricane Maximum Potential Intensity Equilibrium. *Q. J. Royal Meteor. Soc*. **141**, 2471–2480. DOI:10.1002/qj.2556.

Kieu, C. Q. and Q. Wang, 2017: Stability of tropical cyclone equilibrium. *Journal of the Atmospheric Sciences,* **74**, 3591-3608.

Leith, C.E., 1971: Atmospheric predictability and two-dimensional turbulence, *J. Atmos. Sci*., **28**, 145–161.

Leith, C.E., and R.H. Kraichnan, 1972: Predictability of turbulent flows, *J. Atmos. Sci*, **29**, 1041–1058.

Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, **20**, 130-141.

____, 1969: The predictability of a flow which possesses many scales of motion. *Tellus.* **21**, 289-307.

____, 1990: Predictability–a problem partly solved. Predictability of Weather and Climate, ed. Tim Palmer and Renate Hagedorn. Published by Cambridge University Press.

Métais, O., and Lesieur M., 1986: Statistical Predictability of Decaying Turbulence. *J. Atmos. Sci.* **43**, 857–870.

Nguyen, P., C. Kieu, and W.T. Fan, 2020: Stochastic variability of tropical cyclone intensity at the asymptotic potential intensity equilibrium. *Journal of Atmospheric Sciences*, **77**, 3105-3118.

Schonemann, D., and Frisius, T. 2012. Dynamical system analysis of a low-order tropical cyclone model. *Tellus* A, **64**, 15817.

Sugihara, G., and R. May, 1990: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Journal of Economic Theory*, **344**, 734–741.

Persing, J., M. T. Montgomery, R. K. Smith and J. C. McWilliams, 2019: Quasi steady-state hurricanes revisited. *Tropical Cyclone Science Reviews*, **8**, 1-17.

Rotunno, R., and K. A. Emanuel, 1987: An air–sea interaction theory for tropical cyclones. Part II: Evolutionary study using a non-hydrostatic axisymmetric numerical model. *J. Atmos. Sci*. **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. *J. Atmos. Sci*., **65**, 1063–1076.

Schneider, T. and S. Griffies, 1999: A conceptual framework for predictability studies. *J. Climate*, **12**, 3133–3155.

Shukla, J., 1981: Dynamical predictability of monthly means. *J. Atmos. Sci*, **38**, 2547–2572.

Smith, R. K., G.Kilroy and M. T. Montgomery, 2021: Tropical cyclone life cycle in a three-dimensional numerical simulation. Quart. J. Royal Meteorol. Soc, 1 (739), 3373-3393

Smith, R. K., M. T. Montgomery and J. Persing, 2014: On steady-state tropical cyclones. Q. J. R. Meteorolo. Soc., 139, 1-15

Vallis, G. K., 2017: Atmospheric and Oceanic Fluid Dynamics, Cambridge University Press, 946p.

Wang, Y., Y. Li, and J. Xu, 2021: A new time-dependent theory of tropical cyclone intensification. *J. Atmos. Sci*., **78**, 3855–3865.

Zhong, Q., J. Li, L. Zhang, R. Ding, and B. Li, 2018: Predictability of Tropical Cyclone Intensity over the Western North Pacific Using the IBTrACS Dataset. *Mon. Wea. Rev*., **146**, 2741–2755.

Supplementary Information

Method and Data

a. Deep-learning models

Given the low dimension of feature vectors used for ML training, we present in this study several deep-learning models for TC intensity prediction. Specifically, three ML architectures including a deep neural network (DNN) model, a gated recurrent unit (GRU) model, and a long-short term memory (LSTM). The applications of these deep learning models have been rapidly grown due to the availability of more computational power, which helps accelerate their execution in practical problems. With four input features of TC scales and one real-value output corresponding to TC intensity, predicting TC intensity becomes a familiar supervised regression problem for which the above ML models are very well suitable.

Because the feature vectors are of dimension four, a design of 3 hidden layers with layer sizes of 32, 64, and 64 was used for DNN, followed by an output layer of size 1 that corresponds to TC intensity. Each neural layer was applied a standard ReLU activation, which helps ML models capture nonlinear effects as well as increase the interaction among layers. One could certainly design a deeper neural network for more complex relationship between input and output layers. However, our trial-and-error experiments with different neural designs showed very little improvement with more than 2 hidden layers for predicting TC intensity in such a low-dimensional input. As such, a fixed design of 32, 64, and 64 nodes is used, with further layer-sensitivity analyses provided.

For LSTM and GRU, these are recurrent neural models that require a data interval in the past to capture the memory in the training data. Our model architectures for these LSTM and GRU models thus need some additional setups. Specifically for these recurrent networks models, we used three layers of size 16, 32, and 64, with a dropout rate of 0.5. Technically, dropout is a type of regularizations that can help reduce overfitting. There is no particular formula to choose the value for this hyperparameter, other than empirical trials. In our intensity chaos problem, this dropout turns out to be an important for ensure the good model performance. For the past interval, we used 21 time slices, i.e., $t_i, i \in [-20,0]$, as input for LSTM/GRU models when predicting TC intensity at any given lead time.

All of these ML models employed the mean absolute error (MAE) metrics for the accuracy and the root mean squared errors for the loss function, with a fixed number of training epochs set to be 300. The standard optimizer for the gradient search based on the stochastic mini-batch learning method, the so-called Root Mean Squared Propagation (RMSprop), was applied in all training. Because of the different scales of the wind and pressure variables, all input data was scaled by the standard deviation around a mean value, which corresponds to the maximum intensity of the model vortex at the quasi-stationary equilibrium.

b. Data

In this study, the data from a CM1 model's 100-day simulations at 9-km resolution was used, which is model output from the CM1 simulations presented in Kieu et al. (2022). By applying a fixed Newtonian cooling relaxation of 2 K day^{-1}, K22 could obtain a long integration of 100 days that maintain well the quasi-stationary maximum intensity equilibrium, similar to Kieu and Moon (2016). Given this quasi-stationary simulation of TC intensity, the time series of several key TC scales including the maximum boundary-layer inflow U_{max}, the maximum wind speed V_{max} at the model lowest level, the maximum vertical motion in the eyewall region W_{max}, and the minimum central pressure P_{min} were output at a sampling frequency of 36 seconds for the purpose of training in ML models. With 100-day simulations and 36-s sampling frequency, a dataset of length 250,000 was therefore generated, which was split into training, validation, and test sets with a ratio of 90%, 5%, and 5%, respectively. To ensure that the CM1 model settled down in the PI equilibrium before training, the first 10 days of simulations were also discarded. All other details of this CM1 100-day simulation including model domain, physical parameterization, boundary and/or initial conditions can be found in K22.