Large-Scale Environmental Controls on the Seasonal Statistics of Rapidly Intensifying North Atlantic Tropical Cyclones

Ching Ho Justin Ng[1], Gabriel A. Vecchi[1,2,3]

[1]*Atmospheric and Oceanic Sciences (AOS), Princeton University*
Princeton, NJ 08540-6649, USA

[2]*Department of Geosciences, Princeton University*
Princeton, NJ 08540-6649, USA

[3]*Princeton Environmental Institute, Princeton University*
Princeton, NJ 08540-6649, USA

March 2020

Corresponding Author Address:

Ching Ho Justin Ng, Atmospheric and Oceanic Sciences (AOS), Princeton University, 300 Forrestal Road, Sayre Hall, Princeton, NJ 08540-6659, USA
Email: chjng@princeton.edu
ORCID: 0000-0003-2582-9459

Abstract

This study is concerned with the connections between the large-scale environment and the seasonal occurrence of rapid intensification (RI) of North Atlantic tropical cyclones. Physically-motivated statistical analysis using observations and reanalysis products suggests that for tropical cyclones over the open tropical North Atlantic, the interannual variability of the probability of storms undergoing RI is influenced by seasonal large-scale atmospheric and oceanic variables, but not so for storms over the Gulf of Mexico and western Caribbean Sea. We suggest that this differentiated response is due to the former region exhibiting a strong negative correlation between the seasonal anomalies of vertical wind shear and potential intensity. Differences in the mean climatology and subseasonal variations of the large-scale environment in these regions appear to play an insignificant role in the distinctive seasonal environmental controls on RI. We suggest that the interannual correlation of vertical wind shear and potential intensity is an indicator of seasonal predictability of tropical cyclone activity (including RI) across the tropics.

1. Introduction

Despite recent improvement in forecasting tropical cyclone (TC) genesis and tracks, prediction of their changes in intensity even on timescales of days has remained challenging (e.g. Elsberry et al. 2007, Rappaport et al. 2009, Elsberry 2014). Of particular scientific interest are those TCs that undergo rapid intensification (RI), owing to their widespread impacts on society and the relatively little lead time for preparation they may provide. Following existing literature (e.g. Kaplan and DeMaria 2003, Kaplan et al. 2010), an RI event is defined as a TC intensification of at least 30 knots (15.4 m/s) in 24 hours, separated by more than 24 hours from each other. Recently, Lee et al. (2016) suggested that the vast majority of global major storms rapidly intensified at least once during their lifetime, and highlighted their important role in the climatology of TCs. Moreover, the frequency of storms that intensify rapidly is projected to increase substantially in future climate (Emanuel 2017, Bhatia et al. 2018), and may have already increased in the North Atlantic (Bhatia et al. 2019), emphasizing the importance of improved predictions of rapidly intensifying TCs.

There has been continuing research on rapidly intensifying TCs at the storm scale and at weather forecast timescales (e.g. Bosart et al. 2000, Kaplan and DeMaria 2003, Lin et al. 2009, Kaplan et al. 2010, Shieh et al. 2013, Zhuge et al. 2015, Chen et al. 2017), while a few studies have attempted to study their predictability on seasonal timescales (e.g. Wang and Zhou 2008, Wang et al. 2017). Before the hurricane season starts, can the large-scale environment for the upcoming season be used to say something about the RI activity of TCs? The provision of seasonal predictions for RI events would serve to improve societal preparedness for rapidly intensifying tropical cyclones, complementing existing seasonal forecasts for TC frequency. Such examples include forecasts issued by the United States Climate Prediction Center (CPC), Geophysical Fluid Dynamics Laboratory (GFDL), Colorado State University (CSU), and the European Centre for Medium-Range Weather Forecasts (ECMWF), among others. To advance this goal, our approach here is to examine the relationship between seasonal statistics of the occurrence of RI in the North Atlantic and the large-scale environment on seasonal timescales. Predictions of the seasonal environment from numerical models could then be used for predictions of RI activity. Besides, knowledge about seasonal environmental controls on RI can connect with those about the environmental footprints of different modes of climate variability, such as the El Niño-Southern Oscillation (e.g. Gray 1984, Shapiro 1987), the Atlantic Meridional Mode (Chiang and Vimont 2004) and the North Atlantic Oscillation (Hurrell et al. 2003), which have been shown to exhibit seasonal predictability (Barnston et al. 2012, Vimont 2012).

A conceptual model for the large-scale environmental controls on intensity change of TCs, including RI, can be developed with the equation

$$dI(t)/dt = \tau^{-1}\left(I^* - I(t)\right) = f\left(\tau(t), I^*(t)\right) \qquad -(1)$$

(Lloyd et al. 2011), which considers how the time rate of change of TC intensity (dI/dt) is affected by the environmental potential intensity (I^*) and other environmental parameters known to influence TC intensity, such as vertical wind shear and tropospheric humidity, through the relaxation timescale τ. In this study, instead of considering the environment at each time instant t following the TC, the function f is now a function of the seasonal-mean large-scale environment, so that we consider the effects of the seasonal environment on the time tendency of intensity:

$$dI(t)/dt = \tilde{f}(seasonal\ environment) + \epsilon(t) \qquad -(2)$$

where $\epsilon(t)$ concerns with subseasonal fluctuations such as intraseasonal oscillations and weather-scale variations. For all the TCs in each season with its corresponding seasonal-mean environment, a distribution of dI/dt results from $\epsilon(t)$ at each time instant t following the TCs. Then, the equivalent question is whether the seasonal environment can change the statistics of the dI/dt distribution, in particular the proportion of the distribution that exceeds the threshold for RI.

To study the variability of the annual counts of RI instances (which we denote by $n(RI)$), we first note that part of its variance comes from the variance of the number of tropical cyclones (which we denote by N). Absent other information, one would expect more RI to occur simply when there are more tropical cyclones in a given year, all other aspects being equal. As is shown below, the variance of N only explains part of the variance of $n(RI)$, so that it is as important to understand the climate controls on the probability of each TC experiencing RI, *i.e.*, $p(RI) = n(RI)/N$. In other words, $n(RI)$ is dependent on the total number of storms, which may in turn be determined by factors outside of those influencing TC intensification, and $p(RI)$ can be considered as a 'normalized' measure of RI activity. For this reason, statistical relationships developed directly using the $n(RI)$ metric may not genuinely reflect large-scale environmental controls on RI. For example, in Wang et al. (2017) the seasonal large-scale atmospheric and oceanic conditions are linearly regressed to the RI number. A similar metric to $p(RI)$, the 'RI ratio', defined by the number of 24-hour intensity changes above 30 knots divided by total 24-hour intensity changes, was used in Bhatia et al. (2018) and Bhatia et al. (2019) for studying the response of RI to climate change and anthropogenic forcing. The month-to-month variations of RI ratio in the Northwest Pacific were also considered in Wang and Zhou (2008) and Ge et al. (2018), while Shu et al. (2012) described the variations of RI ratio with TC category in the Northwest Pacific. Since the statistical modeling for the number of tropical cyclones (N) over the North Atlantic has been performed in previous studies (Villarini et al. 2010, Vecchi et al. 2011, Murakami et al. 2016b), our approach in this study is to examine climate variability of $n(RI)$ through modeling the probability that a TC will experience RI. Any potential for prediction skill in RI probability can combine with existing seasonal outlooks of tropical cyclone frequency to provide additional information with regard to RI.

In this study, we will also investigate how RI activity varies within the North Atlantic basin. Previous studies on RI have focused on statistics for the entire basin or the Main Development Region (MDR).

For instance, linear regressions in Wang et al. (2017) were performed between the seasonal large-scale environment and RI number in the entire North Atlantic basin, and they suggested that certain variables in certain regions could be used to predict basinwide RI frequency. On the other hand, Klotzbach (2012) suggested that the number of RI occurrences is significantly higher during the La Niña phase and certain phases of the Madden-Julian Oscillation, for both the North Atlantic and MDR. However, few studies have considered sub-basin variability of RI activity, including those outside of the MDR. As the skill of seasonal TC forecasts has been shown to vary within the North Atlantic basin (Vecchi et al. 2014, Murakami et al. 2016a, Liu et al. 2018) and across the tropics (DeMaria et al. 2007, Vecchi et al. 2014, Zhang et al. 2016b), we will examine how the statistics of sub-basin RI probability and its variability are related to regional large-scale atmospheric and oceanic conditions on seasonal timescales.

This paper is structured as follows. Section 2 describes the data and methods employed in this study, while Section 3 analyzes the statistics of RI and its relationship with large-scale atmospheric and oceanic conditions. Sections 4 and 5 discuss sub-basin differences in this relationship, followed by some concluding remarks in Section 6.

2. Data & Methods

Sub-basin classification of RI activity in the North Atlantic is determined by a cluster analysis of tropical cyclone tracks, while the large-scale environmental controls on RI in each cluster are examined by statistical regression of large-scale climate variables on RI statistics. This section describes the datasets and climate reanalysis products studied, the cluster analysis and regression methods employed, and the methods to analyze subseasonal variations of certain variables.

2a. Tropical Cyclones and Large-Scale Variables

North Atlantic tropical cyclone data is obtained from the Atlantic Hurricane Database (HURDAT2, Landsea and Franklin 2013, version updated on April 11, 2017). HURDAT2 provides best-track positions, maximum sustained surface wind speed, minimum central pressure and wind radii of North Atlantic tropical cyclones at 6-hourly time intervals. Extratropical cyclones, tropical depressions and disturbances are removed prior to analysis. HURDAT2 data between 1979-2015 is studied (the satellite era, e.g. Rienecker et al. 2011, Truchelut et al. 2013).

Large-scale atmospheric and oceanic variables are obtained from various climate reanalysis products and observational datasets. Monthly-mean values of atmospheric variables including relative humidity, temperature and winds at various tropospheric levels are taken from MERRA (Rienecker et al. 2011) and the Japanese 55-year Reanalysis (JRA-55, Kobayashi et al. 2015), following, for example, Vimont and Kossin (2007), Hendricks et al. (2010), Vecchi et al. (2013), Wing et al. (2015), Zhang et al. (2016a), and Wang et al. (2017). Sea-surface temperature (SST) is taken from HadISST (Rayner et al. 2003) and the NOAA OISST Version 2 High Resolution Dataset (Reynolds et al. 2007), from which we calculate the relative sea-surface temperature (RELSST) defined as the difference between local SST and the tropical-mean SST (Vecchi and Soden, 2007). From MERRA and JRA-55, the vertical wind shear (VWS) is calculated as the magnitude of the vector difference between 200hPa and 850hPa winds. Potential intensity (PI) is calculated from the atmospheric profile in MERRA and JRA-55 following Bister and Emanuel (1998, 2002). PI is defined as $PI^2 = C_K/C_D \;\; T_S/T_0 \;\; (CAPE^* - CAPE)|_m$, where C_K is the exchange coefficient for enthalpy, C_D is the drag coefficient, T_S is the sea surface temperature, T_0 is the mean outflow temperature, $CAPE^*$ is the convective available potential energy of air lifted from saturation at sea level in reference to the environmental sounding, and $CAPE$ is that of boundary layer air. Both $CAPE$ terms are evaluated near the radius of maximum winds of the TC. Saturation deficit (SDEF), which provides a measure of the moist entropy deficit of the middle troposphere, is calculated as $SDEF = (s_b - s_m)/(s_0^* - s_b)$, where s_m, s_b, s_0^* are respectively the moist entropies of the middle troposphere, boundary layer (taken as 750hPa and 900hPa respectively, determined by a separate analysis of the tropical moist entropy profile), and the saturation moist entropy of the sea surface (Emanuel 2010). Tropical cyclone heat potential (TCHP) is calculated from the NCEP Global Ocean Data Assimilation System reanalysis (GODAS, Behringer et al. 1998), as the vertical integral of temperature from the sea surface to the depth of the 26ºC isotherm. We considered these variables as candidates for large-scale environmental controls on RI, since they have been suggested to influence TC activity and development in a number of studies, e.g. Ryan et al. (1992), Watterson et al. (1995), Bister and Emanuel (1998), Bister and Emanuel (2002), Wong and Chan (2004), Camargo et al. (2007), Vecchi and Soden (2007), and Tang and Emanuel (2012). Data for VWS, PI, SDEF, RELSST and TCHP in the period of 1980-2015 are studied.

2b. Cluster Analysis

To consider sub-basin variability of RI statistics within the North Atlantic, we performed a K-means cluster analysis on the HURDAT2 data, based on the location of genesis, lysis and maximum intensity of each TC, with the number of clusters (K) to be determined. A similar K-means clustering approach was used in Elsner (2003), Elsner and Liu (2003) and Ramsay et al. (2012), for TCs in the North Atlantic, Northwest Pacific and Southern Hemisphere basins respectively. In these studies, various TC statistics for each cluster were calculated, but not specifically related to RI events. TC genesis (lysis) is defined as the first (last) 6-hourly position with sustained wind speed above 34 knots (i.e. above tropical storm intensity), following, for example, Murakami et al. (2015) and Murakami et al. (2016b).

Based on the within-cluster spread (inertia) for each choice of K (not shown), we subjectively decided to choose $K \geq 4$, since the $K = 2, 3$ cases provide high marginal improvement in cluster spread. Meanwhile, there is an inherit upper bound to K since the number of data in each cluster decreases with K, which in turn increases noise when considering interannual variability and seasonal environmental controls on RI in the later sections. Thus, a value of K between 4 and 5 seems to be desirable. In deciding between these values, a cluster analysis is performed for both values, with the latter case further separating the cluster in the subtropics (not shown). Considering the relatively small number of RI instances in the subtropics, we decided to treat this as one cluster, hence performing the K-means cluster analysis with four clusters. The results of this cluster analysis are shown in Section 3a.

2c. Statistical Regressions

To explore statistical connections between the seasonal large-scale variables and regional RI probability for each TC cluster, we performed linear least-squares regressions between RI activity in each JJASON season between 1980-2015, and the 36-year time series of JJASON-mean atmospheric and oceanic variables, and calculated Spearman's correlation coefficients. In addition, to test the robustness of the linear least-squares regression, a linear median of pairwise slopes fit (Lanzante 1996) was also performed between the RI statistics and large-scale variables. The choice of the JJASON season follows from our analysis of HURDAT2 data that in the North Atlantic, all RI events in the period of 1980-2015 occurred between the months of June and November. This regression methodology is similar to that in Wang et al. (2017), in which they explored the relationship between the large-scale environment and RI number in the entire North Atlantic basin; here the correlations are also calculated for the $p(RI)$ metric and for the individual clusters. The significance of the Spearman correlations is determined by a bootstrap test with 1000 samples, using a two-sided 90% or 95% significance level.

Limitations are inherent in the assumptions underlying the above linear least-squares regression, which encourages the application of other generalized linear models in studying RI occurrence. In particular, the predicted value of $p(RI)$ from explanatory variables should inherently take values between 0 and 1. As this is not guaranteed by the linear regression model, and since RI is by its nature a binary process, we adopt the binary logistic regression, which assumes that $n(RI)$ follows a binomial distribution $B\left(N, p(RI)\right)$. The logistic regression uses the large-scale environmental variables as exogenous regressors (explanatory variables), with the endogenous (response) variable being n^* (number of successes) and $N - n^*$ (number of failures) for each year, where $n^*(RI)$ represents the number of RI instances which considers multiple RI instances of the same tropical cyclone as one. The metric $n^*(RI)$ is introduced because by virtue of the definition of RI, a TC can experience more than one RI instance if the RI occurrences are separated by more than 24 hours, so that $p(RI)$ can exceed unity (but rarely do so). The statistical package 'statsmodels' (Skipper and Perktold, 2010), with the binomial model family and the logit link function, performs the logistic regression using the iteratively reweighted least-squares (IRLS) method. The logit link function takes the form

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_i x_i \qquad -(3)$$

so that

$$p = \frac{1}{1 + \exp(-\beta_0 - \beta_i x_i)} \in [0, 1] \qquad -(4)$$

where positive values of β_i represent $p(RI)$ increases with the exogenous regressor x_i.

In the statistical modeling of $n(RI)$, given the nature of the data (counts), a Poisson regression model is adopted (Villarini et al. 2010), so that the expected value of $n(RI)$ is given by $\exp(\beta_0 + \beta_i x_i) \in [0, \infty)$. Similar to above, the statistical significance of the Spearman correlations between observed and predicted values is determined by a bootstrap test, while the skill of the regression models relative to climatology is calculated as

$$Skill = \left(1 - \frac{MSE_{prediction}}{MSE_{climatology}}\right) \times 100\% = \left(1 - \frac{\sum_{i=1980}^{2015}\left(n_{i,pred} - n_{i,obs}\right)^2}{\sum_{i=1980}^{2015}\left(\overline{n_{\iota,obs}} - n_{i,obs}\right)^2}\right) \times 100\% \qquad -(5)$$

where $\overline{n_{\iota,obs}} = \frac{1}{36}\sum_{i=1980}^{2015} n_{i,obs}$. Positive (negative) values represent predictions with skill above (below) climatology.

2d. Decomposition of Daily Variations

Daily data is used to examine the control of the seasonal environment on RI (Section 5). In particular, daily tropospheric profiles are obtained from MERRA reanalysis, for the period 1980-2015, from which VWS and PI (shown in Section 3b to significantly influence RI) are calculated. We decompose the daily values of VWS and PI as follows:

$$X(day) = \bar{X} + \langle X \rangle + X' = \bar{\bar{X}} + X_{SC} + \langle X \rangle + X', \qquad X = VWS, PI \qquad -(6)$$

where $\bar{X} = \bar{X}(month)$ represents the 1980-2015 JJASON climatological seasonal cycle, and can be further separated into a scalar climatological mean ($\bar{\bar{X}}$) and average seasonal cycle anomalies given by

$X_{SC} = X_{SC}(month)$. The interannual variability of the large-scale environment $\langle X \rangle = \langle X \rangle(year)$ is given by

$$\langle X \rangle(t_i) = \left(\frac{1}{N_{day}} \sum_{June\ 1, Year\ i}^{Nov\ 30, Year\ i} X \right) - \bar{\bar{X}} \quad \forall i \in [1980, 2015] \qquad - (7)$$

with $N_{day} = 183$ (number of days in each season), representing the JJASON-mean departure from climatology in each season. $X' = X'(day)$ represents subseasonal variations of the large-scale environment.

This decomposition enables us to distinguish between the large-scale environmental variations on seasonal and subseasonal (including weather) timescales. The difference between $X_{sub} = \bar{X} + X'$ and $X_{tot} = X_{sub} + \langle X \rangle = \bar{X} + \langle X \rangle + X'$ indicates the role of interannual variability $\langle X \rangle$. By construction, X_{sub} has the same JJASON seasonal-mean for each year, and has the same climatological mean as X_{tot}.

We then consider the VWS-PI phase space and compute the two-dimensional probability density function (PDF) of subseasonal (X_{sub}) and total (X_{tot}) VWS and PI values, denoted by PDF_{sub} and PDF_{tot} respectively. PDF_{sub} and PDF_{tot} are calculated using X_{sub} and X_{tot} values respectively, for each season and as area-averages for each TC cluster. These calculations involve Gaussian kernel density estimation, which is provided by the gaussian_kde function of the SciPy Python-based package (Jones at al. 2001), with the estimator bandwidth determined using Scott's Rule. In other words, we calculate

$$PDF_{sub} = PDF_{sub}(year) = p(VWS_{sub}, PI_{sub} | JJASON) \qquad - (8a)$$

$$PDF_{tot} = PDF_{tot}(year) = p(VWS_{tot}, PI_{tot} | JJASON) \qquad - (8b)$$

which represent the average large-scale environment in each cluster, for each season.

As will be discussed in Section 5, we consider the typical environment in which an RI event occurs. We calculate the PDF of VWS and PI values averaged over a 10ºx10º box, centered at the storm on the day an RI event occurs, for all RI occurrences in the period 1980-2015. The PDF is denoted as PDF_{RI}. When defining the day at which an RI event occurs, we consider the 6-hourly time evolution of TC intensity from HURDAT2, and compute the average date ($\bar{d}$, YY-MM-DD, where $\bar{\cdot}$ represents average over the variable $\cdot$), time ($\bar{t}$, UTC) and location ($\overline{lat}, \overline{lon}$) of all 6-hourly positions at which TC intensity increases by over 30 knots in its subsequent 24-hour period. Since the environmental parameters are obtained from daily-resolution MERRA time series, VWS and PI values at date $\bar{d}$, averaged over a 10ºx10º box centered at ($\overline{lat}, \overline{lon}$), is considered as a representation of the typical RI environment. The distribution of all such representations is used to construct PDF_{RI}.

Similarly, the typical environment in which a TC occurs (PDF_{TC}) is computed using VWS and PI values averaged over a 10ºx10º box, centered at every 6-hourly TC position, for all TC occurrences in the period 1980-2015. Since TC positions are obtained 6-hourly from HURDAT2, but VWS and PI values are at daily resolution, up to four TC positions may correspond to the same set of daily VWS and PI values from MERRA. In this case, however, this set of VWS and PI values is averaged over different 10ºx10º boxes following the 6-hourly TC best track, and will therefore still lead to different environments used to compute PDF_{TC}. In notation form, PDF_{RI} and PDF_{TC} could be written as:

$$PDF_{RI} = p(VWS, PI | 10^o \times 10^o centered\ at\ RI, 1980 - 2015) \qquad - (9a)$$

$$PDF_{TC} = p(VWS, PI | 10^o \times 10^o centered\ at\ TC, 1980 - 2015) \qquad - (9b)$$

Then, we calculate the convolution of PDF_{sub} and $PDF_{tot}$with PDF_{RI}, as a measure of how conducive the large-scale environment is to RI (how close the environment is to one in which RI occurs) in each season, i.e. we consider

$$C_{sub}(year) \equiv convolution(PDF_{sub}(year), PDF_{RI}) \qquad -(10a)$$

$$C_{tot}(year) \equiv convolution(PDF_{tot}(year), PDF_{RI}) \qquad -(10b)$$

The convolution framework is then extended from the North Atlantic to the Northeast Pacific and Northwest Pacific basins, to test the robustness of the statistical results. TC data for the Northeast Pacific, again for the period of 1980-2015, is obtained from the HURDAT2 dataset (Landsea and Franklin 2013), while that for the Northwest Pacific is obtained from best-track data compiled by the Joint Typhoon Warning Center, as part of the International Best Track Archive for Climate Stewardship dataset (IBTrACS, Knapp et al. 2010).

To further study the statistical sensitivity of rapid intensification to VWS and PI, PDF_{RI} is fitted to a two-dimensional hyperbolic tangent function of the form

$$PDF_{RI} = c_1 \cdot \tanh(c_2 VWS + c_3 PI + c_4 VWS \cdot PI + c_5) + c_6 \qquad -(11)$$

The motivations, assumptions and implications of this curve fitting are discussed in Section 5. The fitting is performed with the curve_fit function of SciPy, which adopts the Levenberg-Marquardt Algorithm for nonlinear least-squares curve fitting, optimizing the parameter values of the fitted function to minimize the sum of squared residuals between the fitted and actual data.

3. Seasonal Environmental Controls on RI

This section presents the statistical analysis of the seasonal large-scale environmental controls on RI for each North Atlantic TC cluster.

3a. Location and Statistics of Clusters

A K-means cluster analysis is first performed on HURDAT2 with four clusters, as described in Section 2b. These four clusters are plotted in Figure 1, indicating their regional distinction and hence potential insufficiency to categorize North Atlantic RI by entire-basin statistics. Cluster 0 is centered at the subtropics, while cluster 1 is centered at the Gulf of Mexico and western Caribbean Sea. Both clusters 2 and 3 contain tropical cyclones that form in the open tropical Atlantic; the former cluster tends to move along the subtropical ridge and dissipate (or undergo extratropical transition) at higher latitudes, while the latter tends to dissipate within the tropics. While these two clusters are centered in the open North Atlantic, a significant portion of these TCs that experience RI do make landfall along the East Coast of the United States (Figure 2). Table 1 displays the basic statistics for each TC cluster, including the number of TCs, the number of TCs that experience RI, and the percentage of TCs that experience RI. Although Kossin et al. (2010) used a different clustering method which involved a mixture of quadratic regression models, and performed their computation with TC data from different time periods, their four clusters also include a subtropical cluster, and tropical storms that are zonally separated into a subset of Gulf of Mexico storms and those that span the MDR, as is the case here.

3b. Sub-basin Environmental Controls

(i) Linear Regression

As mentioned in the introduction, the variance of $n(RI)$ is not entirely explained by that of N, so the potential predictability of $p(RI)$ is fundamental to the predictability of annual RI counts. Indeed, analysis of the basinwide annual RI counts gives a variance of 3.68, out of which 30% is from the variance of N, *i.e.*, $var(N(t) \cdot \bar{p}) = 1.11$, where $\bar{\cdot}$ represents the time-average over the annual counts. On the other hand, over 58% of the variance is associated with the probability of RI, *i.e.*, $var(p(t) \cdot \bar{N}) = 2.15$. Thus, we linearly regress the interannual variability of $n(RI)$ and $p(RI)$ in each cluster on the seasonal-mean large-scale atmospheric and oceanic variables, to study whether seasonal environmental anomalies have a significant role in controlling the statistics of seasonal RI occurrence.

The annual $n(RI)$ and $p(RI)$ in clusters 0 and 3 is frequently zero, giving a weak signal and large noise, and both the least-squares and median of pairwise slopes regressions give a constant response of $p(RI)$ to the variables (i.e. the regression line has zero slope, not shown). For these two clusters, it appears that there are too few RI occurrences to perform statistically sound analysis. A longer record with reliable RI data could help mitigate this issue. In addition, the behavior of subtropical TC rapid intensifications (in cluster 0, hereafter C0) may be influenced by tropical-extratropical interactions and transient baroclinicity; Murakami et al. (2016a) obtained a lower level of seasonal predictability for subtropical TCs than tropical ones in the North Atlantic for the GFDL CM2.5-FLOR model. On the other hand, since both clusters 2 and 3 tropical cyclones form in the open tropical North Atlantic, distinguishing between these TC clusters near their time of genesis might be difficult from a practical viewpoint, hence it might be worthwhile to combine the statistics for these two clusters (and this gives more robust statistical results). Hence, for the remainder of this study we will focus on the RI properties of cluster 1 (hereafter C1) and clusters 2 and 3 combined (hereafter C23).

Seasonal RI statistics of the open-ocean cluster (C23) are potentially predictable from some large-scale atmospheric and oceanic conditions. The median of pairwise slopes regression response of $p(RI)$ to the large-scale environmental variables for the entire North Atlantic (all clusters combined) is shown in Figure 3, while that for C23 is shown in Figure 4. Similar results hold for least-squares regressions (not shown). Figure 3 can be compared to Figure 9 of Wang et al. (2017), which show statistical regressions of annual RI number, for data between 1950-2014; Figure 4 suggests that high $p(RI)$ of C23 across seasons is associated with high seasonal PI, high RELSST and low VWS in the deep tropics, particularly below 20ºN. These correlations are significant at the 95% level in certain regions, for MERRA, JRA-55, HadISST and NOAA OISST. Similar results hold for linear regressions with the $n(RI)$ metric and with ASO-mean composites (not shown). MERRA SDEF, 850hPa relative humidity and TCHP do not show statistically significant correlations to RI probability for C23. On the other hand, the Gulf of Mexico cluster (C1) does not exhibit a statistically significant response to the regional environmental anomalies. Some correlation exists at regions away from the typical TC locations in C1 (Figure 1), and we interpret this as a coincidental correlation between variables as opposed to any physical reasons connecting the non-local environment to RI in this cluster. From a large-scale perspective, regional atmospheric and oceanic conditions less significantly impact RI for C1 than C23; RI statistics in C23 vary interannually via the influence of regional sea-surface temperature, potential intensity and vertical wind shear, and less so from atmospheric humidity.

(ii) Logistic Regression

Logistic regressions on the large-scale variables for C23 are shown in Figure 5: RELSST, VWS and PI are significantly correlated with RI successes/failures in some portions of the deep tropical Atlantic, though less ubiquitous than in the linear regression (particularly for VWS).

The interannual variability of RI counts and probability is modeled using Poisson regression and binary logistic regression, respectively, with PI and VWS as exogenous co-regressors, using the methodology outlined in Section 2c. The exogenous variables are averaged over 10ºN-30ºN, 100ºW-80ºW for C1 (see magenta box in Figure 4a), and over 10ºN-20ºN, 80ºW-20ºW for C23 (considering that this is the region which exhibits high $p(RI)$ correlation, see magenta box in Figure 4d). The results of the statistical modeling for C1 and C23 are shown in Figure 6. Logistic regression is unskillful relative to climatology in modeling RI probability of C1, while there is slight skill (14%) for C23. Poisson regression is able to provide some skill for annual RI counts in both clusters, with higher skill (27%) in C23. In addition, prediction of $n^{*}(RI)$ through separate predictions of N and $p(RI)$ shows promise in both C1 and C23, since prediction for RI counts calculated as the product of predicted RI probability and predicted TC counts shows comparable skill to the direct Poisson regression (Figure 6g-h).

In summary, statistical regressions suggest that for the tropical open-ocean cluster (C23), the probability of storms experiencing RI in each season is correlated with seasonal regional environmental anomalies, in particular those of vertical wind shear (VWS), relative sea-surface temperature (RELSST) and potential intensity (PI). Thus, this cluster exhibits potential for RI seasonal predictability through the seasonal large-scale environment. On the other hand, for the cluster of tropical cyclones in the Gulf of Mexico and western Caribbean Sea (C1), the seasonal large-scale environment does not exert significant controls on RI. Yet, on the seasonal timescale, why do these TCs at similar latitudes of the North Atlantic exhibit such distinctive relationships to the large-scale environmental anomalies? What is so 'special' about the Gulf of Mexico region that leads to the observed differences in RI environmental controls? We will explore into this question in the next section.

4. Hypotheses for Sub-basin Difference

In Section 3, we showed that on seasonal timescales, the statistics of RI activity in terms of probability and annual counts are significantly influenced by seasonal VWS, PI and RELSST, for the TC cluster in the Central/Eastern tropical North Atlantic (C23), but not for the Gulf of Mexico/Western Caribbean cluster (C1). Such a difference is also depicted in Figure 9c of Wang et al. (2017); the physical reasons for which, however, was not explored in their study. Why does there exist a difference in the seasonal large-scale environmental controls on RI between these TC clusters? Given the result of the statistical regressions above, and following some existing literature as discussed in Section 2a, we will focus on two environmental parameters shown to influence RI activity, namely VWS and PI, so that the physical model (Section 1) simplifies to the form $dI/dt = \tau^{-1}(PI - I)$, where $\tau = \tau(VWS)$.

We make the following hypotheses to explain the distinctive seasonal environmental controls on RI between C1 and C23:

1. The weaker seasonal environmental control on RI in C1 than C23 results from the seasonal anomalies of VWS and PI being less 'cooperative' (less negatively correlated) in C1 than C23.
2. We also hypothesize that there is a contribution from the climatology in C1 being overall more conducive to RI than C23, so that a larger magnitude of environmental changes is needed to influence seasonal variations of RI activity in C1. In other words, the seasonal RI statistics is less sensitive to seasonal environmental changes in C1 than C23, giving a weaker 'signal'. In C23, the more cooperative nature of the seasonal anomalies of VWS and PI, superimposed on its less RI-conducive climatology, is significant to influence RI activity on interannual timescales.
3. Our final hypothesis is that the reduced seasonal reproducibility of RI activity in C1 is driven by larger subseasonal (weather and intraseasonal timescale) fluctuations ('noise'), contributing to its weaker seasonal 'signal'. Subseasonal variations are smaller in C23, so that its seasonal large-scale environment can significantly impact seasonal RI statistics.

In Section 5, we will discuss the framework to examine these hypotheses, and establish the validity of some of these hypotheses in each subsection.

5. Discussion of Sub-basin Difference

This section examines each of the hypotheses presented in Section 4 by exploring the connections between the typical RI environment and variations of the large-scale environment in C1 and C23.

5a. Framework for Exploring Variations of VWS and PI

To examine the relationship between the seasonal large-scale environment and RI statistics, we study the variations of the large-scale environment in each cluster, again taking the average of 10ºN-30ºN, 100ºW-80ºW for C1, and that of 10ºN-20ºN, 80ºW-20ºW for C23. In particular, as described in Section 2d, we study the daily VWS and PI variations (denoted by X, where $X = VWS, PI$) decomposed into a climatology $\bar{X}$, an interannual variability $\langle X \rangle$, and a subseasonal X' component, for C1 (hereafter denoted by the subscript W for C1 in Western North Atlantic) and C23 (subscript E for C23 in Central/Eastern North Atlantic), from which the subseasonal (X_{sub}) and total (X_{tot}) variations are calculated. We build a framework to consider the VWS-PI phase space, and compute the subseasonal (PDF_{sub}) and total (PDF_{tot}) cluster-average VWS and PI values (the *average cluster environment* in each season), and the distribution of VWS and PI during RI events (PDF_{RI}, the *typical RI environment*).

We then hypothesize that RI activity in each cluster is connected to how close the *average cluster environment* is to the *typical RI environment*, and compute $C_{sub}(year)$ and $C_{tot}(year)$. As mentioned in Section 2d, these convolutions quantify the conduciveness of the seasonal large-scale environment in each cluster to RI. Indeed, comparing between PDF_{RI} and PDF_{TC} (Figure 7) suggests that during RI events, VWS (PI) is generally lower (higher) than the average TC environment. In other words, most RI occurs during the low-VWS and high-PI stages relative to the entire TC life cycle. Similar results are obtained using PDF_{RI} and PDF_{TC} trained with RI and TC cases from the North Atlantic, Northeast Pacific and Northwest Pacific combined.

5b. Role of Interannual Variability $\langle X \rangle$

We test our hypothesis by first comparing between the interannual variability terms $\langle X \rangle_W$ and $\langle X \rangle_E$. As can be seen in Figure 8, the seasonal anomalies of VWS and PI are more negatively correlated in C23 (correlation = -0.7586) than C1 (correlation = -0.2260). That is, these two variables tend to reinforce each other in C23, for example by becoming more RI-conducive by decreasing VWS and increasing PI. This convergence can potentially bring the climate towards the typical RI environment (high PDF_{RI} region, lower right in the VWS-PI phase space, see Figure 7c) in some years, and away from it in others.

Next, we test whether $\langle X \rangle$ plays a significant role in affecting the interannual variability of RI statistics. As illustrated in Section 2d, the difference between $C_{tot}(year)$ and $C_{sub}(year)$ indicates the role of $\langle X \rangle$. Figure 9 suggests that for C23, C_{tot} is significantly correlated with $n(RI)$ and $p(RI)$ over the period of 1980-2015 (Spearman correlation = 0.60, 0.44 respectively). This supports our hypothesis that the convolution, being a measure of how close the average cluster environment is to the typical RI environment, could be used as a proxy for RI activity. However, when the interannual variability of the environment is excluded, C_{sub} now has no significant relationship with both RI metrics, and RI predictability is lost. Also, the variance of the convolution decreases by about 9.8 times, as opposed to about 3.3 times for C1. The larger difference between C_{tot} and C_{sub} in C23 than C1 indicates a larger role of the environmental interannual variability in shifting VWS and PI in and out of the region favorable for RI (in the VWS-PI space). Therefore, in C23, seasonal timescale fluctuations of the environment $\langle X \rangle$ can explain the interannual variability of RI statistics: When the interannual variability of the environment pushes climate into (away from) RI-conducive region, leading to high (low) convolution C_{tot}, RI probability is high (low). Such results do not hold for C1; the correlation to RI metrics does not decrease from C_{tot} to C_{sub}.

PDF_{RI} is averaged over 10ºx10º boxes (see Section 2d), so the above calculation is repeated with PDF_{sub} and PDF_{tot} computed using values averaged at consecutive 10ºx10º boxes in each cluster (instead of averages over the entire cluster), to provide a more consistent comparison. In this case, X_{sub}, X_{tot} and $\langle X \rangle$ for C23 consist of VWS and PI values over six east-west oriented boxes, while those for C1 consist of four boxes. As in the previous calculation, VWS and PI are more negatively correlated in C23 than C1, and C_{tot} is significantly correlated with both RI counts and RI probability in C23 (not shown). In addition, similar results hold when PDF_{sub} and PDF_{tot} are calculated using VWS, PI values weighted by RI density in each cluster (instead of using box averages, Figure 10), and also hold for PDF_{RI} trained with RI from the North Atlantic, Northeast Pacific and Northwest Pacific combined (yielding 768 cases in total, Figure 11).

5c. Role of Mean Climatology $\bar{X}$

We now test the hypothesis that different states of climatology between C1 and C23 may lead to different responses of RI statistics to seasonal environmental changes. A cluster environment that is very close to high PDF_{RI} values will require larger changes in the environment to significantly affect the convolution, than one that lies away from the high PDF_{RI} values (e.g. lying at regimes where PDF_{RI} has a large gradient in the VWS-PI phase space). Figure 12 compares between the average cluster environments $\bar{X}_W$ and $\bar{X}_E$, and suggests that: (1) $\bar{\bar{X}}_E$ has higher VWS and lower PI than $\bar{\bar{X}}_W$; (2) $\bar{X}_E$ has higher VWS and lower PI at a larger portion of the season than $\bar{X}_W$. Thus, the climatology is indeed different between C1 and C23, with the latter being on average further away from RI-conducive values in the VWS-PI phase space, so that RI activity can potentially have a larger response to changes in the cluster environment. In other words, RI activity is potentially more sensitive to environmental changes in C23.

However, is the larger sensitivity of RI activity to seasonal environmental changes in C23 significant to provide environmental controls on RI seasonal statistics? If this were true, it would mean that if C23 had a lower sensitivity, just like in C1, then the interannual variability of RI statistics would be lost. In other words, if C23 had the less sensitive climatology of C1, does this different sensitivity change the significance of $\langle X \rangle$ in determining RI? For this purpose, we switch between the climatology of C1 and C23, and compare between the convolutions calculated with the C23 environment $\bar{X}_E + \langle X \rangle_E + X'_E$ and the hypothetical environment $\bar{X}_W + \langle X \rangle_E + X'_E$.

Calculations indicate that switching between C1 and C23 climatology leads to very different convolutions (correlation = -0.36), and the hypothetical convolution has no correlation with $n(RI)$ and $p(RI)$ (correlation = -0.15, -0.01 repsectively). However, considering PDF_{RI} and PDF_{tot} in the VWS-PI phase space suggests that in the hypothetical situation where C23 now has the C1 climatology, the new PDF_{tot} now lies in the phase space where PDF_{RI} drops to zero at very high PI and low VWS values. We interpret this as an artifact of the limited number of RI samples observed (134 samples in the period of 1980-2015): Such RI has not been observed since it is rare for TCs to be existent when the environment attains such extreme VWS and PI values. While an RI event at such extreme values has not been observed, we posit that the decrease of PDF_{RI} at these values is unphysical; PDF_{RI} should monotonically increase with PI and decrease with VWS. To correct this sampling issue, PDF_{RI} is fitted to a two-dimensional hyperbolic tangent function of VWS and PI (Figure 13), using the methodology described in Section 2d. The curve fitting as shown is performed using $VWS \in [0,40]$, $PI \in [30, 81]$

as the support; these bounds are chosen to be close to the observed environmental VWS and PI limits, and to minimize the variance of the parameter estimates. The choice of these bounds only affects the curve fitting parameters but not the results of the convolutions presented above. Other bounds give a similar shape to the fitted hyperbolic tangent curve as shown, though the values of the fitted PDF carry no physical meaning since the choice of the bounds is artificial. This curve fitting to the hyperbolic tangent function also assumes the following: (1) Favorability to RI monotonically increases with PI and decreases with VWS, less so at extreme values than others and remains bounded; (2) The simultaneous VWS decrease and PI increase is more conducive to RI than to changes in only one variable; (3) The c_4 term is added to include any possible nonlinearities in the variation of RI statistics with VWS and PI. The convolutions of the fitted PDF_{RI} with PDF_{sub} and PDF_{tot} behave similarly to those shown in Figure 9 (not shown), and the standard deviation of the parameter estimates amounts to 0.8-4% of the parameter values.

Having attempted to correct the sampling issue, the hypothetical convolution (with East-West climatology switched) now looks like the before-switch convolution (correlation = 0.95), and is significantly correlated with $n(RI)$ and $p(RI)$ (correlation = 0.52, 0.40 respectively). This implies that even with C1 climatology, the negatively correlated $\langle X \rangle_E$ can explain interannual variations in RI statistics in C23, indicating the importance of the interannual variability of VWS and PI over the climatology. On the other hand, while we previously showed that $\langle X \rangle_W$ is not significant to explain RI interannual variability in C1, this also holds if C1 had a less conducive C23 climatology (comparing between $\bar{X}_W + \langle X \rangle_W + X'_W$ and $\bar{X}_E + \langle X \rangle_W + X'_W$), again indicating that the state of climatology plays a secondary role compared to the interannual variability of VWS and PI. The importance of $\langle X \rangle$ only breaks down at very extreme climatological values, for example when $\overline{PI}_E$ is increased by at least 30 m/s (not shown), which is much larger than the projected PI increase over the 21st century (e.g. Vecchi and Soden 2007, Yu et al. 2010, Camargo 2013, Knutson et al. 2013, Sobel et al. 2016, Vecchi et al. 2019).

5d. Role of Subseasonal Variations $\boldsymbol{X'}$

Lastly, we study the impact of subseasonal environmental variability on the interannual variability of RI statistics, and test whether the significance of $\langle X \rangle_E$ depends on X'_E. For this purpose, we consider the hypothetical situation in which C23 had the subseasonal variation of C1, and compare between the convolutions calculated with $\bar{X}_E + \langle X \rangle_E + X'_E$ and $\bar{X}_E + \langle X \rangle_E + X'_W$ (i.e. switching between C1 and C23 subseasonal variations). These convolutions are very similar to each other (correlation = 0.92, Figure 14a), and the exclusion of $\langle X \rangle$ from $\bar{X} + X'$ decreases variance of the convolution by 10.48 times, similar to results shown previously. In addition, similar results hold when comparing the original convolution with that calculated with $\bar{X}_E + \langle X \rangle_E + X'_{2010,E}$ (using subseasonal variations in say year 2010 for all years, Figure 14b). This highlights the importance of the negatively correlated $\langle X \rangle$ in C23 over X' (and the interannual variability of X') in explaining interannual variations of RI statistics.

In summary, the argument presented in this section suggests that the negatively correlated seasonal anomalies of VWS and PI are significant controls of seasonal RI statistics in C23, while the mean climatology and subseasonal variations of the large-scale environment are not important, at least for 'realistic' magnitudes. Meanwhile, VWS and PI vary less consistently interannually in C1, so that $\langle X \rangle_W$ is not a significant control on the interannual variability of seasonal RI statistics. This results in a relatively larger importance of subseasonal variations of the environment or storm-scale dynamics in controlling RI in C1 than C23 – and, accordingly, lower seasonal predictability. This finding is consistent with that of Kossin et al. (2010), in which they suggested that TCs in the Gulf of Mexico are strongly modulated by the Madden-Julian Oscillation, which is known to modulate TCs on subseasonal timescales (e.g. Maloney and Hartmann 2000).

6. Concluding Remarks

This study is concerned with the large-scale environmental controls on the seasonal statistics of RI of North Atlantic TCs. Statistical analysis indicates that on interannual timescales, increased probability of TCs undergoing RI in the open tropical North Atlantic (C23) is significantly related to increased seasonal relative sea-surface temperature (RELSST), increased potential intensity (PI), and decreased vertical wind shear (VWS). However, tropical cyclones in the Gulf of Mexico and western Caribbean Sea (C1) and the Sargasso Sea (C0) do not exhibit such behavior.

We demonstrated that the interannual variability of RI statistics in C23 is significantly controlled by the *negatively correlated* ('cooperative') interannual variability of VWS and PI, due to which seasonal anomalies of VWS and PI act constructively to change the probability that the environment will move in and out of the 'RI-favorable' space (region of high PDF_{RI} in the VWS-PI space). At least for 'realistic' magnitudes, changes in the mean climatology, which we hypothesized might have lead to different sensitivities of RI activity to environmental changes, and changes in subseasonal variations of the large-scale environment are not significant to alter the seasonal environmental controls on RI. Making assumptions on the likelihood of RI at very extreme VWS and PI values beyond the currently observed range, the importance of the interannual variability of VWS and PI breaks down at extreme climatological values beyond future climate projections, which suggest that projected changes in climate may not affect RI seasonal predictability assuming small changes in the interannual variability of the large-scale environment.

On the other hand, for the Gulf of Mexico cluster (C1), the seasonal anomalies of VWS and PI are weakly correlated and do not significantly control RI statistics on seasonal timescales, suggesting the potential importance of subseasonal environmental variability and storm-scale dynamics in controlling

RI. In the former sense, RI occurrences over this region could be mostly considered as 'weather-related' events.

This study provides a step towards RI seasonal predictability by exploring the statistical sensitivity of RI to large-scale environmental anomalies. Our statistical framework is developed using two climate predictors, while future study could include more RI predictors such as sea-surface height, high-altitude divergence, and others used by Kaplan et al. (2010). In addition, the analysis presented in this study could be repeated with inclusion of the most recent hurricane seasons, as new data becomes available. The finding here that a significant seasonal environmental control on RI is determined by the negative correlation of VWS and PI seasonal anomalies, invites applications to other basins in which the distribution of RI activity could be compared to that of the correlation between seasonal VWS and PI in these basins. In addition, VWS and PI are among the factors discussed in Camargo et al. (2007) to influence genesis of tropical cyclones, while Tang and Emanuel (2012) modeled the probability of TC genesis as a function of the ventilation index (which depends on VWS, PI and entropy deficit). These suggest that the interannual correlation between VWS and PI may be extended to one involving other environmental anomalies to understand the seasonal predictability of TC activity across the tropics. Noting the significance of VWS and PI interannual variability in controlling RI in the Central/Eastern tropical North Atlantic, another application is to investigate whether coupled climate models could be used to predict VWS and PI seasonal anomalies in this region as a means of RI seasonal predictability, in both present and future climates. In particular, our findings highlight the importance of assessing future changes in the correlation between seasonal anomalies of VWS and PI, as opposed to changes in their absolute values, in speculating the potential of future RI seasonal predictability through climate change simulations. Lastly, work is underway to understand the physical mechanisms behind the observed strong (weak) negative correlation between VWS and PI in C23

(C1), and more generally to understand physical controls of the spatial distribution of such a correlation across global ocean basins.

In light of the societal impact of rapidly intensifying TCs, this research is aimed at improving our understanding of their predictability on seasonal timescales through exploring their statistical connections with large-scale atmospheric and oceanic conditions. It is apparent that improved forecasts for the seasonal statistics of rapidly intensifying TCs that occur closer to land would be more beneficial to society than those further away. The results from this study show that on seasonal timescales, RI experienced by TCs furthest away from land (C23) are the most predictable, some of which indeed making landfall on the eastern sector of the Caribbean and North American region (Figure 2). The seasonal statistics of RI events that occur closer to the East Coast of the United States (C0), and those that affect population in Central America, Mexico and southern United States (C1), seem to be less predictable than their counterparts in the open tropical North Atlantic.

Acknowledgements

This work is supported in part by National Oceanic and Atmospheric Administration (NOAA) Grants NA14OAR4320106 and NA18OAR4320123. Gabriel A. Vecchi is supported by the Carbon Mitigation Initiative (CMI) at Princeton University.

References

1. Barnston, A.G., Tippett, M.K., L'Heureux, M.L., Li, S. and DeWitt, D.G., 2012: Skill of real-time seasonal ENSO model predictions during 2002-11: Is our capability increasing?. Bulletin of the American Meteorological Society, 93, 631-651
2. Behringer, D.W., Ji, M. and Leetmaa, A., 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. Monthly Weather Review, 126, 1013-1021
3. Bhatia, K., Vecchi, G., Murakami, H., Underwood, S. and Kossin, J., 2018: Projected response of tropical cyclone intensity and intensification in a global climate model. Journal of Climate, 31, 8281-8303
4. Bhatia, K.T., Vecchi, G.A., Knutson, T.R., Murakami, H., Kossin, J., Dixon, K.W. and Whitlock, C.E., 2019: Recent increases in tropical cyclone intensification rates. Nature communications, 10, 635
5. Bister, M. and Emanuel, K.A., 1998: Dissipative heating and hurricane intensity. Meteorology and Atmospheric Physics, 65, 233-240
6. Bister, M. and Emanuel, K.A., 2002: Low frequency variability of tropical cyclone potential intensity 1. Interannual to interdecadal variability. Journal of Geophysical Research: Atmospheres, 107, ACL-26
7. Bosart, L.F., Bracken, W.E., Molinari, J., Velden, C.S. and Black, P.G., 2000: Environmental influences on the rapid intensification of Hurricane Opal (1995) over the Gulf of Mexico. Monthly Weather Review, 128, 322-352
8. Camargo, S.J., Emanuel, K.A. and Sobel, A.H., 2007: Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. Journal of Climate, 20, 4819-4834
9. Camargo, S.J., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. Journal of Climate, 26, 9880-9902

10. Chen, X., Wang, Y., Zhao, K. and Wu, D., 2017: A numerical study on rapid intensification of Typhoon Vicente (2012) in the South China Sea. Part I: Verification of simulation, storm-scale evolution, and environmental contribution. Monthly Weather Review, 145, 877-898
11. Chiang, J.C. and Vimont, D.J., 2004: Analogous Pacific and Atlantic meridional modes of tropical atmosphere-ocean variability. Journal of Climate, 17, 4143-4158
12. DeMaria, M., Knaff, J.A. and Sampson, C., 2007: Evaluation of long-term trends in tropical cyclone intensity forecasts. Meteorology and Atmospheric Physics, 97, 19-28
13. Elsberry, R.L., Lambert, T.D. and Boothe, M.A., 2007: Accuracy of Atlantic and eastern North Pacific tropical cyclone intensity forecast guidance. Weather and Forecasting, 22, 747-762
14. Elsberry, R.L., 2014: Advances in research and forecasting of tropical cyclones from 1963-2013. Asia-Pacific Journal of Atmospheric Sciences, 50, 3-16
15. Elsner, J.B., 2003: Tracking hurricanes. Bulletin of the American Meteorological Society, 84, 353-356
16. Elsner, J.B. and Liu, K.B., 2003: Examining the ENSO-typhoon hypothesis. Climate Research, 25, 43-54
17. Emanuel, K., 2010: Tropical cyclone activity downscaled from NOAA-CIRES reanalysis, 1908-1958. Journal of Advances in Modeling Earth Systems, 2
18. Emanuel, K., 2017: Will global warming make hurricane forecasting more difficult?. Bulletin of the American Meteorological Society, 98, 495-501
19. Enfield, D.B. and Mayer, D.A., 1997: Tropical Atlantic sea surface temperature variability and its relation to El Niño-Southern Oscillation. Journal of Geophysical Research: Oceans, 102, 929-945
20. Ge, X., Shi, D. and Guan, L., 2018: Monthly variations of tropical cyclone rapid intensification ratio in the western North Pacific. Atmospheric Science Letters, 19, e814
21. Gray, W.M., 1984: Atlantic seasonal hurricane frequency. Part I: El Niño and 30mb quasi-biennial oscillation influences. Monthly Weather Review, 112, 1649-1668

22. Hendricks, E.A., Peng, M.S., Fu, B. and Li, T., 2010: Quantifying environmental control on tropical cyclone intensity change. Monthly Weather Review, 138, 3243-3271

23. Hurrell, J.W., Kushnir, Y., Ottersen, G. and Visbeck, M., 2003: An overview of the North Atlantic Oscillation. The North Atlantic Oscillation: Climatic significance and environmental impact, 134, 1-35

24. Jones E, Oliphant E, Peterson P, et al. SciPy: Open Source Scientific Tools for Python, 2001

25. Kaplan, J. and DeMaria, M., 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. Weather and Forecasting, 18, 1093-1108

26. Kaplan, J., DeMaria, M. and Knaff, J.A., 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. Weather and Forecasting, 25, 220-241, doi:10.1175/2009WAF2222280.1

27. Klotzbach, P.J., 2012: El Niño-Southern Oscillation, the Madden-Julian Oscillation and Atlantic basin tropical cyclone rapid intensification. Journal of Geophysical Research: Atmospheres, 117

28. Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J. and Neumann, C.J., 2010: The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data. Bulletin of the American Meteorological Society, 91, 363-376

29. Knutson, T.R., Sirutis, J.J., Vecchi, G.A., Garner, S., Zhao, M., Kim, H.S., Bender, M., Tuleya, R.E., Held, I.M. and Villarini, G., 2013: Dynamical downscaling projections of twenty-first-century Atlantic hurricane activity: CMIP3 and CMIP5 model-based scenarios. Journal of Climate, 26, 6591-6617

30. Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H. and Miyaoka, K., 2015: The JRA-55 reanalysis: General specifications and basic characteristics. Journal of the Meteorological Society of Japan. Ser. II, 93, 5-48

31. Kossin, J.P., Camargo, S.J. and Sitkowski, M., 2010: Climate modulation of North Atlantic hurricane tracks. Journal of Climate, 23, 3057-3076

32. 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

33. Lanzante, J.R., 1996: Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. International Journal of Climatology: A Journal of the Royal Meteorological Society, 16, 1197-1226

34. Lee, C.Y., Tippett, M.K., Sobel, A.H. and Camargo, S.J., 2016: Rapid intensification and the bimodal distribution of tropical cyclone intensity. Nature communications, 7, 10625

35. Lin, I.I., Chen, C.H., Pun, I.F., Liu, W.T. and Wu, C.C., 2009: Warm ocean anomaly, air sea fluxes, and the rapid intensification of tropical cyclone Nargis (2008). Geophysical Research Letters, 36

36. Liu, M., Vecchi, G.A., Smith, J.A., Murakami, H., Gudgel, R. and Yang, X., 2018: Towards Dynamical Seasonal Forecast of Extratropical Transition in the North Atlantic. Geophysical Research Letters, 45, 12-602

37. Lloyd, I.D., Marchok, T. and Vecchi, G.A., 2011: Diagnostics comparing sea surface temperature feedbacks from operational hurricane forecasts to observations. Journal of Advances in Modeling Earth Systems, 3

38. Maloney, E.D. and Hartmann, D.L., 2000: Modulation of hurricane activity in the Gulf of Mexico by the Madden-Julian oscillation. Science, 287, 2002-2004

39. Murakami, H., Vecchi, G.A., Underwood, S., Delworth, T.L., Wittenberg, A.T., Anderson, W.G., Chen, J.H., Gudgel, R.G., Harris, L.M., Lin, S.J. and Zeng, F., 2015: Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. Journal of Climate, 28, 9058-9079

40. Murakami, H., Vecchi, G.A., Villarini, G., Delworth, T.L., Gudgel, R., Underwood, S., Yang, X., Zhang, W. and Lin, S.J., 2016a: Seasonal forecasts of major hurricanes and landfalling tropical cyclones using a high-resolution GFDL coupled climate model. Journal of Climate, 29, 7977-7989
41. Murakami, H., Villarini, G., Vecchi, G.A., Zhang, W. and Gudgel, R., 2016b: Statistical-dynamical seasonal forecast of North Atlantic and US landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model. Monthly Weather Review, 144, 2101-2123
42. Rappaport, E.N., Franklin, J.L., Avila, L.A., Baig, S.R., Beven, J.L., Blake, E.S., Burr, C.A., Jiing, J.G., Juckins, C.A., Knabb, R.D. and Landsea, C.W., 2009: Advances and challenges at the National Hurricane Center. Weather and Forecasting, 24, 395-419
43. Ramsay, H.A., Camargo, S.J. and Kim, D., 2012: Cluster analysis of tropical cyclone tracks in the Southern Hemisphere. Climate Dynamics, 39, 897-917
44. Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C. and Kaplan, A., 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. Journal of Geophysical Research: Atmospheres, 108
45. Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S. and Schlax, M.G., 2007: Daily high-resolution-blended analyses for sea surface temperature. Journal of Climate, 20, 5473-5496
46. Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.K. and Bloom, S., 2011: MERRA: NASA's modern-era retrospective analysis for research and applications. Journal of climate, 24, 3624-3648
47. Ryan, B.F., Watterson, I.G. and Evans, J.L., 1992: Tropical cyclone frequencies inferred from Gray's yearly genesis parameter: Validation of GCM tropical climates. Geophysical Research Letters, 19, 1831-1834
48. Seabold, S. and Perktold, J., 2010: Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference, 57, 61

49. Shapiro, L.J., 1987: Month-to-month variability of the Atlantic tropical circulation and its relationship to tropical storm formation. Monthly Weather Review, 115, 2598-2614

50. Shieh, O.H., Fiorino, M., Kucas, M.E. and Wang, B., 2013: Extreme rapid intensification of Typhoon Vicente (2012) in the South China Sea. Weather and Forecasting, 28, 1578-1587

51. Shu, S., Ming, J. and Chi, P., 2012: Large-scale characteristics and probability of rapidly intensifying tropical cyclones in the western North Pacific basin. Weather and Forecasting, 27, 411-423

52. Smirnov, D. and Vimont, D.J., 2011: Variability of the Atlantic Meridional Mode during the Atlantic hurricane season. Journal of Climate, 24, 1409-1424

53. Sobel, A.H., Camargo, S.J., Hall, T.M., Lee, C.Y., Tippett, M.K. and Wing, A.A., 2016: Human influence on tropical cyclone intensity. Science, 353, 242-246

54. Tang, B. and Emanuel, K., 2012: A ventilation index for tropical cyclones. Bulletin of the American Meteorological Society, 93, 1901-1912

55. Truchelut, R.E., Hart, R.E. and Luthman, B., 2013: Global identification of previously undetected pre-satellite-era tropical cyclone candidates in NOAA/CIRES Twentieth-Century Reanalysis data. Journal of Applied Meteorology and Climatology, 52, 2243-2259

56. Vecchi, G.A. and Soden, B.J., 2007: Effect of remote sea surface temperature change on tropical cyclone potential intensity. Nature, 450, 1066

57. Vecchi, G.A., Zhao, M., Wang, H., Villarini, G., Rosati, A., Kumar, A., Held, I.M. and Gudgel, R., 2011: Statistical-dynamical predictions of seasonal North Atlantic hurricane activity. Monthly Weather Review, 139, 1070-1082

58. Vecchi, G.A., Fueglistaler, S., Held, I.M., Knutson, T.R. and Zhao, M., 2013: Impacts of atmospheric temperature trends on tropical cyclone activity. Journal of Climate, 26, 3877-3891

59. Vecchi, G.A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A.T., Zeng, F., Anderson, W., Balaji, V., Dixon, K. and Jia, L., 2014: On the seasonal forecasting of regional tropical cyclone activity. Journal of Climate, 27, 7994-8016
60. Vecchi, G.A., Delworth, T.L., Murakami, H., Underwood, S.D., Wittenberg, A.T., Zeng, F., Zhang, W., Baldwin, J.W., Bhatia, K.T., Cooke, W. and He, J., 2019: Tropical cyclone sensitivities to CO2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. Climate Dynamics, 53, 5999-6033
61. Villarini, G., Vecchi, G.A. and Smith, J.A., 2010: Modeling the dependence of tropical storm counts in the North Atlantic basin on climate indices. Monthly Weather Review, 138, 2681-2705
62. Vimont, D.J. and Kossin, J.P., 2007: The Atlantic Meridional Mode and hurricane activity. Geophysical Research Letters, 34
63. Vimont, D.J., 2012: Analysis of the Atlantic Meridional Mode using linear inverse modeling: Seasonality and regional influences. Journal of Climate, 25, 1194-1212
64. Wang, B. and Zhou, X., 2008: Climate variation and prediction of rapid intensification in tropical cyclones in the western North Pacific. Meteorology and Atmospheric Physics, 99, 1-16
65. Wang, C., Wang, X., Weisberg, R.H. and Black, M.L., 2017: Variability of tropical cyclone rapid intensification in the North Atlantic and its relationship with climate variations. Climate Dynamics, 49, 3627-3645
66. Watterson, I.G., Evans, J.L. and Ryan, B.F., 1995: Seasonal and interannual variability of tropical cyclogenesis: Diagnostics from large-scale fields. Journal of climate, 8, 3052-3066
67. Wing, A.A., Emanuel, K. and Solomon, S., 2015: On the factors affecting trends and variability in tropical cyclone potential intensity. Geophysical Research Letters, 42, 8669-8677
68. Wong, M.L. and Chan, J.C., 2004: Tropical cyclone intensity in vertical wind shear. Journal of the atmospheric sciences, 61, 1859-1876

69. Yu, J., Wang, Y. and Hamilton, K., 2010: Response of tropical cyclone potential intensity to a global warming scenario in the IPCC AR4 CGCMs. Journal of Climate, 23, 1354-1373
70. Zhang, W., Vecchi, G.A., Murakami, H., Villarini, G. and Jia, L., 2016a: The Pacific meridional mode and the occurrence of tropical cyclones in the western North Pacific. Journal of Climate, 29, 381-398
71. Zhang, W., Vecchi, G.A., Murakami, H., Delworth, T., Wittenberg, A.T., Rosati, A., Underwood, S., Anderson, W., Harris, L., Gudgel, R. and Lin, S.J., 2016b: Improved simulation of tropical cyclone responses to ENSO in the western North Pacific in the high-resolution GFDL HiFLOR coupled climate model. Journal of Climate, 29, 1391-1415
72. Zhuge, X.Y., Ming, J. and Wang, Y., 2015: Reassessing the use of inner-core hot towers to predict tropical cyclone rapid intensification. Weather and Forecasting, 30, 1265-1279

Tables

Table 1: The number of TCs, the number of RI events, and the ratio between these two variables, in each TC cluster. Cluster analysis is performed with the K-means method, for HURDAT2 data between 1979-2015

	Number of TCs, N	Number of RI events, $n(RI)$	$p(RI) = n(RI)/N$
Cluster 0	134	23	17%
Cluster 1	158	59	37%
Cluster 2	82	58	71%
Cluster 3	79	21	27%

Figure Captions

Fig. 1 TC density for each cluster, defined by the number of TC 6-hourly positions per year over a 10°x10° box centered at each location. Also shown are the average locations of genesis, maximum intensity and lysis for each cluster

Fig. 2 Tracks of TCs that experience RI, in each cluster in the North Atlantic. Crosses indicate landfall positions of these TCs, as indicated in the HURDAT2 dataset. The title in each subplot indicates the number of such landfalling RI TCs, and the percentage of landfall TCs that experience RI

Fig. 3 Slope of the linear (median of pairwise slopes) regression of JJASON-mean MERRA PI, HadISST RELSST and MERRA VWS with RI probability, for all TCs in the North Atlantic. Regressions at 95% statistical significance are shaded in crosses, while those at 90% are shaded in dots

Fig. 4 Slope of the linear (median of pairwise slopes) regression of JJASON-mean MERRA PI, HadISST RELSST and MERRA VWS with RI probability, for (a-c) C1 and (d-f) C23. The magenta boxes indicate areas over which averages for C1 and C23 are computed. (g-i) As in (d-f), but for PI, RELSST and VWS from JRA-55 and NOAA OISST. Regressions at 95% statistical significance are shaded in crosses, while those at 90% are shaded in dots

Fig. 5 As in Figure 4d-f, but for logistic regression with the slope parameter β_1 plotted

Fig. 6 Observed and predicted (a-b) RI probability ($p^*(RI)$, predicted with the binary logistic regression), (c-d) RI counts ($n^*(RI)$, predicted with the Poisson regression) and (e-f) TC counts (N, predicted with the Poisson regression), for C1 and C23. (g-h) Prediction for RI counts in C1 and C23 calculated as the product of predicted RI probability and predicted TC counts

Fig. 7 The typical large-scale environment in which RI and TCs exist. The green shading and contours show the two-dimensional probability density function (a) PDF_{RI}, for all 134 RI occurrences in C1 and C23 during the period of 1980-2015, and (b) PDF_{TC}, for all TCs in C1 and C23 during the same period. (c) The difference between PDF_{RI} and PDF_{TC}. (d-f) As in (a-c), but for all 768 RI occurrences and all TCs in the North Atlantic, Northeast Pacific and Northwest Pacific combined

Fig. 8 Interannual variability $\langle X \rangle$ of VWS and PI values in C1 (left) and C23 (right), where each dot represents one JJASON season. Seasonal anomalies of VWS and PI are more negatively correlated in C23 than C1

Fig. 9 Reconstruction of RI seasonal statistics without and with interannual variability of the large-scale environment, as given by C_{sub} (blue) and C_{tot} (red) respectively, for C1 (left) and C23 (right), over the 36-year period of 1980-2015. See text for a description of these convolutions. The tables below show the lag-zero Spearman correlation of C_{sub} and C_{tot} with seasonal $p(RI)$ and $n(RI)$ statistics in each cluster. Statistical significance is computed with the Student's t-test using 36-2=34 degrees of freedom, as provided by the SciPy statistical package

Fig. 10 As in Figures 8 and 9, but with PDF_{sub} and PDF_{tot} calculated using VWS and PI values weighted by RI density for each cluster, instead of box averages over the cluster region

Fig. 11 As in Figure 9, but with PDF_{RI} calculated using all RI events in the North Atlantic, Northwest Pacific and Northeast Pacific combined (shown in Figure 7d)

Fig. 12 Two-dimensional probability density function PDF_{RI} (green contours, as in Figure 7a) in the VWS-PI space, superimposed by that of the 1980-2015 MERRA climatological JJASON seasonal cycle $\bar{X}$ (blue) and scalar-mean JJASON climatology $\bar{\bar{X}}$ (red), where $X = VWS, PI$, averaged for C1 (left) and C23 (right)

Fig. 13 Best fit of the North Atlantic PDF_{RI} (shown in Figure 7a) to the hyberbolic tangent function in the VWS-PI space, as described in Section 2d

Fig. 14 Impact of subseasonal environmental variability (X') on seasonal RI statistics in C23. (a) Convolution between PDF_{RI} and the two-dimensional PDF computed from $\bar{X}_E + \langle X \rangle_E + X'_E$ (blue) or $\bar{X}_E + \langle X \rangle_E + X'_W$ (red), where X'_E and X'_W represents subseasonal variations in C23 and C1 respectively. See text for a full explanation of these variables. (b) As in (a), but the red plot indicates the convolution computed using subseasonal variations in C23 in 2010 for all years ($\bar{X}_E + \langle X \rangle_E + X'_{2010,E}$)