Here we explore the relationship between the global climatological characteristics of tropical cyclones (TCs) in climate models and the modeled large-scale environment across a large number of models. We consider the climatology of TCs in 30 climate models with a wide range of horizontal resolutions. We examine if there is a systematic relationship between the climatological diagnostics for the TC activity [number of tropical cyclones (NTC) and accumulated cyclone energy (ACE)] by hemisphere in the models and the environmental fields usually associated with TC activity, when examined across a large number of models. For low-resolution models, there is no association between a conducive environment and TC activity, when integrated over space (tropical hemisphere) and time (all years of the simulation). As the model resolution increases, for a couple of variables, in particular vertical wind shear, there is a statistically significant relationship in between the models’ TC characteristics and the environmental characteristics, but in most cases the relationship is either nonexistent or the opposite of what is expected based on observations. It is important to stress that these results do not imply that there is no relationship between individual models’ environmental fields and their TC activity by basin with respect to intraseasonal or interannual variability or due to climate change. However, it is clear that when examined across many models, the models’ mean state does not have a consistent relationship with the models’ mean TC activity. Therefore, other processes associated with the model physics, dynamical core, and resolution determine the climatological TC activity in climate models.

1. Introduction

Since the 1970s climate models have been known to simulate tropical cyclone–like structures (Manabe et al. 1970; Bengtsson et al. 1982; Haarsma et al. 1993). These models have been used for projections of tropical cyclone (TC) activity under anthropogenic climate change (Broccoli and Manabe 1990; Bengtsson et al. 1996) and their use for such projections continues to this day, using low- (Camargo 2013; Tory et al. 2013; Chand et al. 2017) and high-horizontal-resolution models (Murakami et al. 2012b; Knutson et al. 2013; Manganello et al. 2014; Bhatia et al. 2018; Bacmeister et al. 2018). Another common use of climate models is for TC dynamical forecasts on seasonal (Vitart et al. 2001; Camargo and Barnston 2009; Manganello et al. 2016; Camp et al. 2019;
G. Zhang et al. 2019; W. Zhang et al. 2019) and sub seasonal time scales (Lee et al. 2018; Camp et al. 2018; Gregory et al. 2019; Zhao et al. 2019). A recent review on this topic is provided by Camargo and Wing (2016).

Both TC future projections, as well as the subseasonal and seasonal forecasts, are dependent on the ability of the models to simulate TC climatological characteristics. Various types of bias correction procedures can be applied to the model output (Camargo and Zebiak 2002; Camargo and Barnston 2009; Camp et al. 2018). However, these regional bias corrections cannot be used to obtain an unbiased global TC climatology and could lead to errors in TC projections and forecasts (Murakami et al. 2014; Lee et al. 2018; W. Zhang et al. 2019).

The relationship of TC genesis with large-scale environmental conditions has been studied since the late 1940s, which was summarized recently in Emanuel (2018). First, Palmén (1948) showed using surface data and soundings that Atlantic hurricanes typically form over ocean with temperatures above 27°C and within 5° latitude of the equator, in regions that are conditionally unstable, while in other regions and/or seasons the tropics are typically stable. Then, Gray (1979) developed an empirical relationship between genesis and climatological conditions of the environment, identifying six environmental conditions necessary for genesis: ocean thermal energy, low-level relative vorticity, vertical wind shear, Coriolis parameter, relative humidity of the troposphere, and a measure of instability of the atmosphere. Since then, many other empirical genesis indices have been developed (DeMaria et al. 2001; Emanuel and Nolan 2004; Emanuel 2010; Tippett et al. 2011; Bruyère et al. 2012; Camargo et al. 2014), making improvements and modifications on the original predictors by Gray. These modifications include using potential intensity instead of sea surface temperature (SST) (Emanuel and Nolan 2004; Emanuel 2010), determining a threshold effect for vorticity (Tippett et al. 2011), considering the saturation deficit of the midtroposphere instead of relative humidity (Emanuel 2010; Camargo et al. 2014), and adding the vertical velocity as an additional predictor (Murakami and Wang 2010). More recently, Tang and Emanuel (2012a,b) proposed a ventilation index combining humidity, shear, and potential intensity and have shown that it influences both genesis and intensification of TCs. Furthermore, Emanuel (2000), Wing et al. (2007), and Kossin and Camargo (2009) showed a relationship between observed TC intensity and potential intensity, supporting the inclusion of potential intensity in TC indices.

Given this rich history of relating environmental conditions to TC genesis and intensification, it is not surprising that when analyzing the climatology of TCs in climate models, the scientific community would make the assumption that there is a relationship between the model environmental fields climatology and the TC climatology in the model. We might expect these relationships to be valid in models, because they are valid in observations. Furthermore, if we look at the geographic and temporal variability within a single dataset (either model or observations/reanalysis), there is indeed a clear relationship between the model environment and TC activity. For example, genesis indices are typically able to reproduce the global climatological TC pattern, as well as the seasonal and interannual variability in individual TC basins in observations (Camargo et al. 2007a; Tippett et al. 2011) and models (Camargo et al. 2007b; Camargo et al. 2013, 2014, 2016). Therefore, typically biases in modeled TC climatology are explained in the literature through the bias in the climatology of the large-scale environmental conditions in these models (Manganello et al. 2012; Camargo et al. 2016).

However, we are not aware of a study that shows that this hypothesis is actually valid. In fact, Reed et al. (2015) and Vecchi et al. (2019) showed that the TC activity in two versions of the same model could not be explained by the differences in their large-scale environment. In Reed et al. (2015) the model differences were due to different dynamical cores, whereas in Vecchi et al. (2019) they are due to different horizontal resolutions. Nevertheless, Vecchi et al. (2019) found that the responses in the models of different resolution could be reconciled after accounting for the sensitivity of pre-TC synoptic disturbances to changing climate, in addition to changes in large-scale environmental factors. We do know that if there are changes in the environment within a given model, due to climate variability or climate change, that the TC activity will change accordingly; this is why there is skill in dynamical seasonal forecasts (Camargo and Barnston 2009; Vitart 2009; Vecchi et al. 2014), for instance. Therefore, if we examine one specific model, generally there is a geographical and/or temporal relationship of the large-scale environment and TC activity. It is not obvious that this relationship actually leads to a coherent relationship across many models between the models’ mean climatological conditions and the models’ TC climatology. If such a relationship existed, it would help explain the large differences in TC climatology characteristics among models.

While increasing model horizontal resolution is known to improve the ability of climate models to simulate TCs (Murakami and Sugi 2010; Manganello et al. 2012, 2014; Walsh et al. 2013; Strachan et al. 2013; Wehner et al. 2014; Roberts et al. 2015, 2018), resolution alone is not a solution for model biases. Models with the same or very similar horizontal resolution can have very different TC climatology characteristics (Camargo et al. 2013; Shaevitz et al. 2014).
It has been shown that besides resolution, the TC climatology in climate models is sensitive to model physics (Vitart et al. 2001; Reed and Jablonowski 2011; Kim et al. 2012; Murakami et al. 2012a; Zhao et al. 2012; Duvel et al. 2017), dynamical core (Reed et al. 2015), and coupling to the ocean (Zarzycki 2016; Scoccimarro et al. 2017; Li and Sriver 2018), as well as the tracking algorithm used to identify TCs in the model output (Horn et al. 2014; Zarzycki and Ullrich 2017). It is clear that complex processes in the model determine the formation and intensification of TCs.

Therefore, it is important to analyze the role of the climatological large-scale environment in determining the model TC climatology. The question we explore here is: do models with a climatological large-scale environment that is more conducive to TC genesis and intensification (e.g., higher values of potential intensity and lower values of vertical wind shear) have a TC climatology with more frequent and intense TCs? Similarly, if a model climatological environment is drier than other climate models, is this model’s TC climatology less active than other models? These are simple questions, but they have not been explored systematically across multiple models.

We will consider 30 climate model simulations of TCs, from three multimodel ensembles, spanning a variety of models’ horizontal resolution, physics, and dynamical core, as well as TC tracking algorithms. There are 14 model simulations from phase 5 of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2012), 6 from the U.S. CLIVAR Hurricane Working (HWG) (Walsh et al. 2015) dataset, and 10 model simulations from a collaborative effort for a NOAA funded project that is part of the NOAA Model Diagnostics Task Force (MDTF) (Maloney et al. 2019). Similarly to what was done in the analysis of the TCs in the HWG project (Shaevitz et al. 2014; Daloz et al. 2015; Nakamura et al. 2017; Ramsay et al. 2018), we are considering the tracking provided by each modeling group as part of the model package.

This is an ensemble of opportunity; that is, we use the model simulations and TC tracks that are available to us, as they are. These model simulations were not produced for this purpose. Therefore, there are caveats in our analysis that we need to be aware of, such as the dependence on the models’ TC tracking schemes and thresholds. Our assumption is that if there are significant differences among the models’ climatologies, these would be larger than those due to tracking scheme. As discussed in Horn et al. (2014), the sensitivity to differences in TC tracking schemes is more important for low-resolution models and weak storms; as the model resolution increases, the sensitivity to tracking routine decreases. We note that in most of the 30 models, with the exception of three cases, one of two tracking algorithms was used to track TCs. The TCs in all low-resolution models were tracked with the same tracking routine, so differences among them arise from other sources. This should help mitigate some of the tracking sensitivity in our analysis. Another issue that we should be aware of is that the model simulations in these multimodel ensembles do not consider the same periods and have different lengths. Furthermore, the CMIP5 multimodel ensemble consists of coupled ocean–atmospheric simulations (CMIP5) and in the other two ensembles, the simulations are forced with fixed SST. Finally, the HWG simulations are forced with climatological SST (i.e., the same SST for every year of the simulation varying only monthly), while the NOAA-MDTF simulations are forced with yearly varying monthly SST.

We will examine the climatological environmental fields that are typically associated with TC genesis and intensification among these models and determine if there is a robust relationship across models between the TC climatology and these environmental fields.

In section 2, we list the model simulations and data used in our analysis, as well as the large-scale environmental fields considered. Our results are given in section 3, and the conclusions in section 4.

2. Models, data, and diagnostics

a. Models

The models used in this analysis are from three different multimodel ensembles. The first set of models is from CMIP5 multimodel ensemble. The TCs in the historical coupled simulations (1850–2005) of 14 low-resolution models have been tracked using the Camargo–Zebiak algorithm (Camargo and Zebiak 2002), with the same thresholds globally. One ensemble member was used for each model. We only considered the period 1971–2000 so that we have a similar period and number of years as the observations and other multimodel ensembles. Various aspects of the TC activity in those simulations have been discussed in detail in Camargo (2013), as well as in Tang and Camargo (2014), Kossin et al. (2016), and Nakamura et al. (2017). Table 1 has a list of the CMIP5 models analyzed, including their references, resolution, and here are referred to as C1 to C14.

The second set of models is from the U.S. CLIVAR Hurricane Working Group (HWG) multimodel ensemble simulations for the current climate. Details about the HWG models and simulations are described in Walsh et al. (2015). The models in this ensemble have higher horizontal resolution (0.25° to 1.25°) and were all forced with the same fixed climatological SST for the present
climate for the period (1985–2001). Various aspects of these simulations have been discussed in the literature (e.g., Shaevitz et al. 2014; Horn et al. 2014; Wang et al. 2014; Scoccimarro et al. 2014; Villarini et al. 2014; Daloz et al. 2015; Camargo et al. 2016; Han et al. 2016; Nakamura et al. 2017; Ramsay et al. 2018). Table 3 has a list of the HWG models, their resolution, and references, and they are called W1 to W6 in this manuscript. Only the models that had all the output necessary for our analysis were included here. Each modeling group contributed a different number of years for this project, varying from 10 to 20 years. All available years were considered in our analysis.

The third set of models is a contribution to our project as part of the NOAA Model Diagnostics Task Force (MDTF; Maloney et al. 2019). Various modeling groups agreed to contribute their existing simulations to this effort. Subsets of these simulations have been used for developing and testing process-oriented diagnostics for tropical cyclones in climate models, as described in Kim et al. (2018), Wing et al. (2019), and Moon et al. (2020). These models typically have resolutions of 0.5° and 0.25°, with exception of the model simulations that were performed for the MDTF (1°). Note that the HiRAM simulations that are part of this group were originally a contribution for the HWG, but with observed monthly varying SST. The list of the models in this group is given in Table 3 and they are named P1–P10. The CAM5-SE simulations used a variable-resolution grid, with resolution of 0.25° in the North Atlantic and 1° in the rest of the globe.

Note that the GISS-C180 employs a development version of GISS ModelE3 from early 2018 with a resolution of 0.5°. The dynamical core used in this study (Putman and Lin 2007) is the same as that used in GISS ModelE2. The parameterizations in this version of the model were also used by Cesana et al. (2019), who outline some updates to the physics distinguishing E3 from E2. Stratiform hydrometeors in E3 evolve within the two-moment microphysics framework of Gettelman and Morrison (2015), and water cloud fraction and cloud water mixing ratio are both diagnosed from a triangular probability density function. The turbulence model is based on Bretherton and Park (2009). The moist convection scheme retains the overall E2 structure but incorporates numerous updates to downdrafts, entrainment, and microphysics, and now features cold pools (Del Genio et al. 2015).

As mentioned above, the model simulations were not designed for this analysis, but rather we are using as many models as possible in it. For each case, we used the model output of as many years of the simulation as they were available to us, and the TC tracking routine that was used by each modeling group. This way we are considering the model and tracking routine as a package. It should be emphasized though that all low-resolution and two high-resolution models (C1–C14, W4, P7, P8, and P9) have been tracked with the Camargo–Zebiak tracking algorithm. Furthermore, most of the high-resolution models are tracked with the Vitart/Zhao algorithm,
b. Data

We compare the models’ environmental fields with those produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim) dataset (Dee et al. 2011), which is available from 1979 to the present. Here we consider the period 1981–2010 for the climatology, as it has the largest overlap with all the models’ climatology.

The TC observations are based on the best track datasets of the National Hurricane Center (NHC) for the Atlantic and eastern North Pacific (Landsea and Franklin 2013) and the Joint Typhoon Warning Center (JTWC) (Chu et al. 2002) for the other basins. The best track datasets from these two agencies were chosen, as they have consistent time averaging of 1 min for the maximum wind speed. We consider the observations in the period 1981–2010. We also consider TCs tracked in the ERA-Interim reanalysis using the Murakami and Sugi (2010) tracking method, as described in Murakami et al. (2014) for the period 1981–2010. Similarly to reanalysis, this period was chosen due to the largest overlap across all models.

c. Diagnostics

We use two diagnostics to represent the TC climatology: the number of TCs (NTC) and the accumulated cyclone energy (ACE). We only consider TCs that form in the tropics (30°S–30°N). For observations, we consider only TCs that reach at least tropical storm intensity (i.e., surface wind speeds of at least 17 m s⁻¹). In contrast, for the models we did not use an additional threshold in the storm’s intensity, as it is standard to have thresholds in the models’ tracking schemes, which are typically dependent on the models’ horizontal resolution (Walsh et al. 2007). In the models we excluded storms that form in the South Atlantic and southeast Pacific (east of 250°), as they are not present in either the NHC or the JTWC observed datasets in that period. Furthermore, we want to match the environmental fields to TC formation areas and the environmental conditions in those regions are not conducive to TC activity and would bias our results.

ACE is defined as $\Sigma u^2$ for all 6-hourly time steps, where $u$ is the maximum sustained surface wind speed. In observations, only time steps for which the surface winds reach at least 35 kt are included in the ACE calculation, following the definition of Bell et al. (2000). For the models, we used a modified version, including the wind speed at all time steps, as in Camargo (2013). This is particularly important in the case of low-resolution models, which generate very weak storms. Furthermore, Davis (2018) showed that from a dynamical perspective, 0.25° (the finest resolution considered here) should not produce a realistic number of category 4 and 5 storms in the absence of larger wind radii or supplementary parameterization. Zarzycki et al. (2017) showed that integrated quantities as ACE are less sensitive to differences in tracking algorithms than TC counts. The reason for that is that all trackers are typically able to track the most intense long-lived storms, which contribute most significantly to ACE. Therefore, our analysis of ACE has smaller uncertainty due to track sensitivity than NTC.

Track density is calculated by counting the number of TCs passing in each grid point in the 6-hourly tracks. For each model, the track density is normalized by the

<table>
<thead>
<tr>
<th>Model</th>
<th>Name</th>
<th>Resolution</th>
<th>Reference</th>
<th>Tracking</th>
<th>Type</th>
<th>No. of years</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFDL AM2.5-AMP</td>
<td>P1</td>
<td>0.5°</td>
<td>Delworth</td>
<td>Murakami</td>
<td>SST</td>
<td>20</td>
</tr>
<tr>
<td>GFDL CM2.5-FLOR</td>
<td>P2</td>
<td>0.5°</td>
<td>Vecchi</td>
<td>Murakami</td>
<td>Coupled</td>
<td>20</td>
</tr>
<tr>
<td>GFDL HiRAM</td>
<td>P3</td>
<td>0.5°</td>
<td>Zhao</td>
<td>Zhao/Vitart</td>
<td>SST</td>
<td>20</td>
</tr>
<tr>
<td>CMCC-CAM5</td>
<td>P4</td>
<td>0.25°</td>
<td>Scoccimarro/Cherchi</td>
<td>Zhao/Vitart</td>
<td>SST</td>
<td>5</td>
</tr>
<tr>
<td>CAM5-FV</td>
<td>P5</td>
<td>0.25°</td>
<td>Neale/Wehner</td>
<td>Zhao/Vitart</td>
<td>SST</td>
<td>20</td>
</tr>
<tr>
<td>CAM5-SE</td>
<td>P6</td>
<td>0.25°/1°</td>
<td>Neale/Zarzycki</td>
<td>TempestExtremes</td>
<td>SST</td>
<td>20</td>
</tr>
<tr>
<td>GFDL AM4-MDTF</td>
<td>P7</td>
<td>1°</td>
<td>Zhao18</td>
<td>Camargo and Zebiak</td>
<td>SST</td>
<td>5</td>
</tr>
<tr>
<td>CCSM4-MDTF</td>
<td>P8</td>
<td>1°</td>
<td>Gent</td>
<td>Camargo and Zebiak</td>
<td>SST</td>
<td>5</td>
</tr>
<tr>
<td>NASA GEOS5</td>
<td>P9</td>
<td>0.5°</td>
<td>Rienecker/Molod</td>
<td>Camargo and Zebiak</td>
<td>SST</td>
<td>20</td>
</tr>
<tr>
<td>GISS-C180</td>
<td>P10</td>
<td>0.5°</td>
<td>Schmidt</td>
<td>Zhao/Vitart</td>
<td>SST</td>
<td>20</td>
</tr>
</tbody>
</table>
number of simulation years. The track density was calculated using a common grid for models and observations, namely a uniform 4° grid.

We examined various standard environmental fields that are typically used to determine if the environment is conducive for TC formation and intensification; in particular, the components of genesis indices (Gray 1979; Emanuel and Nolan 2004; Emanuel 2010; Tippett et al. 2011; Camargo et al. 2014) and the ventilation index (Tang and Emanuel 2012b,a), as well one genesis index combining several fields together. All these variables were computed from monthly-mean fields in models and observations. We only show a subset of the analyzed fields here. They are as follows:

- Vertical shear: magnitude of the vertical wind shear between 200 and 850 hPa
- Relative humidity at 600 hPa
- Column relative humidity: ratio of the column integrated water vapor path and the column saturated water vapor path (Bretherton et al. 2004; Camargo et al. 2014)
- Omega at 500 hPa
- Relative vorticity at 850 hPa
- Potential intensity (PI): theoretical maximum intensity that a TC can reach based on the local thermodynamics conditions, as defined in (Emanuel 1988), following the calculations of Bister and Emanuel (2002).
- Tropical cyclone genesis index (TCGI): empirical genesis index following the formulation of Tippett et al. (2011) and Camargo et al. (2014).

This formulation of TCGI was chosen due to facilitate a comparison with NTC, as the integrated value of TCGI gives the predicted NTC by the index. This is not the case for most genesis indices. Furthermore, Menkes et al. (2012) have shown that TCGI has a performance similar to or even superior to other genesis indices. The climatology of all environmental variables is calculated using either 30 years or all years available if the number of years is smaller than 30. When we integrate the environmental field climatologies, the years were chosen as to have the maximum overlap among the models. Figure 3 shows the track density, using all years available in each case (varying from 5 to 30 years) using a grid box of size of 4° for all models. It is clear from Figs. 1–3 that many low-resolution models (C1–C14) have a very unrealistic climatology of TC-like storms, with very few storms and in many cases no TC-like storms in some basins, especially in the Atlantic. This is not restricted to the CMIP5-type models; these strong biases are still present in some of the HWG (e.g., W2, W6) and MDTF models (e.g., P8), all of which have 1° resolution. Furthermore, in observations the TC activity in the Southern Hemisphere is about half of the Northern Hemisphere, and many models do not reproduce this difference.

These model biases can be seen in more detail in the distributions of NTC and ACE per year for all models, reanalysis, and observations shown in Fig. 4. It is clear that the models in the HWG (W1–W6) and and MDTF (P1–P10) ensembles simulate a number of TCs much closer to observations than the CMIP5 models. The exception of the HWG models is W2, which used the observed threshold for defining TCs in that model (Wehner et al. 2015), while other models of the same resolution use a resolution-dependent threshold (Walsh et al. 2007). Some of the high-resolution models actually produce too many storms compared with observations, in particular P2, P3, and W5. In contrast, most models and the reanalysis have ACE values that are too low compared with observations, indicating that their TCs are too weak, which could be expected based on their horizontal resolution (Davis 2018). The only exceptions are P4, P5, and W1, which have a bias toward high values of ACE. These three models are different versions of the CAM5 model at 0.25°, which indicates some specific characteristic of this particular model that leads to strong storms. Various studies (Zarzycki 2016; Scoccimarro et al. 2017; Li and Sriver 2018) showed that this bias can be
improved by coupling the atmospheric model to an ocean model, instead of using fixed SSTs.

b. Dependence on horizontal resolution

The first point we examine is how the models’ TC climatology is dependent on model horizontal resolution. Figure 5 shows scatterplots between NTC and ACE with model horizontal resolution for the tropics and by hemisphere. There is some dependence of NTC on model resolution, with higher values of NTC and ACE as the model horizontal resolution increases. This relationship is stronger for models with resolution finer than 1°. In contrast, for low-resolution models there is a much weaker relationship between NTC and ACE as the model horizontal resolution increases. While most high-resolution models are able to replicate the observed behavior of a higher level of TC activity in the Northern Hemisphere than in the Southern Hemisphere, this is not the case for low-resolution models, which have similar levels of activity in both hemispheres. (Note that Fig. S4 in the online supplemental material is similar to Figs. 5c–f, but instead of separating the models by resolution, the models are separated by tracking routine. Both figures are very similar, showing that this analysis is not sensitive to the tracking algorithm used.)

Figure 6 shows scatterplots of high percentiles of TC maximum surface wind speed with resolution. These are computed for all model TCs in the Northern Hemisphere for either low-resolution (e.g., models with 2°) or high-resolution (e.g., models with 0.5°) models. The spread can be large for ACE as well, especially across high-resolution models. If we separate NTC and ACE by hemisphere (Figs. 5c–f), there is a similar behavior in both hemispheres.
tropics in ASO. The values of the 99th, 95th, and 90th percentiles of the distribution for each model are shown. The 99th percentile regression line is the steepest one in the top panel, showing a stronger dependence on resolution for the most intense storms in low-resolution models. Similarly to what was already noted for NTC and ACE, models with similar resolution can have very different values of maximum wind speed, making clear that resolution is not the only factor that determines how intense the models’ TCs can be.

c. Environmental fields

We now examine the biases in the climatology of the models’ environmental fields associated with TC activity. Figures 7–9 show the anomalies of the vertical shear, relative humidity at 600 hPa and PI compared with ERA-Interim climatology, which is also shown in all figures. The anomalies in the Northern Hemisphere are calculated in ASO and in the Southern Hemisphere in JFM.

There is large range in the anomalies of the models for the vertical shear (Fig. 7). While some models have very small biases across the globe, such as P2, P3, and P6, others have large anomalies. P7 has large positive anomalies in the tropics, in particular near the date line. Furthermore, in many models the tropical Atlantic vertical shear is too strong (P4, P5, P7, W1, C7, C8, C9, C10, C12, and C13). In contrast, the vertical shear is too weak in the north Indian Ocean in a few models, in particular P4, W1, W2, W6, C10, and C12. It should be noted though that the north Indian Ocean has a minimum of TC activity in ASO due to the high wind shear associated with the Indian monsoon.

In the case of relative humidity (Fig. 8), the values in various models are too high across both hemispheres, such as P4, P7, P9, P10, W3, W5, C4, and C12. In contrast, other models tend to have negative biases in some regions and
positive in others (P3, W1, W6, C2, C5, C10). Many models have their largest biases in relative humidity in the equatorial region (P3, W1, W2, C2), in particular in the central Pacific.

Potential intensity (PI) anomalies are shown in Fig. 9. A few models’ biases stand out in this case; while P4 PI is too high in both hemispheres, C12 PI is too low. In contrast, P5 and P6 have strong negative anomalies in the Northern Hemisphere only. Many CMIP5 models (C) show too strong values of PI in the eastern Pacific in both hemispheres (C4, C5, C6, C8, C9, C10, C11, C12, C13, and C14). As this type of bias is not present in any of the models forced with fixed SST (most P models and W models), this bias is probably related to coupling.

For completeness we show similar plots for omega at 500 hPa, relative vorticity at 850 hPa, and TCGI in Figs S1, S2, and S3 in the online supplemental material. In the case of Omega (Fig. S1), many P and W models have positive biases in the Indo-Pacific equatorial region, with exception of P7 and P10, which show negative biases in the same region. In contrast, the relative vorticity biases (Fig. S2) have dipole anomaly patterns in both hemispheres, indicative of a shift in location of the vorticity in the models. Models typically have positive biases in TCGI (Fig. S3) in the regions of maximum TC activity, and negative biases outside of that region.

In an attempt to quantify these results, Fig. 10 shows scatterplots of the spatial correlations and root-mean-square error (RMSE) of these environmental fields in both hemispheres (in the tropics and over the ocean), relative to the ERA-Interim reanalysis. It is clear across the panels that the CMIP5 (C) models typically have lower correlations and higher RMSE than the P and W models. This is not surprising, given that the C models have lower resolution and are coupled, which tend to lead to large biases. This is particularly true for the spatial correlations of relative humidity, potential intensity, and TCGI.
Another interesting result is that for both omega Fig. 10d and relative vorticity Fig. 10e, there is a clear separation for all model types by hemisphere, with lower RMSE in the Northern Hemisphere and higher in the Southern Hemisphere. Interestingly, the spatial correlation in the Northern Hemisphere reaches lower values than in the Southern Hemisphere.

While in the case of vertical shear Fig. 10a there is an almost linear relationship between RMSE and correlations, with low RSME values associated with high correlations, and the opposite for high RSME. However, this is not the case for other variables. In particular, for the relative humidity (Fig. 10b), many models have high spatial correlations, but a large range of RSME values, indicating that the models can replicate the reanalysis pattern well, but not its magnitude. This is also typically the case for TCGI (Fig. 10f) for P and W models, but not for all C models. While the P and W models have high correlations and low RMSE for PI (Fig. 10c), with exception of one
outlier, C models have much lower spatial correlations, probably related to the biases in the eastern Pacific noted above.

d. Relationship of environmental fields and TC climatology

We next examine whether there is a relationship between climatological environmental fields and climatological TC activity in the models—that is, if a model has a more conducive environment for TC formation and/or intensification, does it have more TCs or are there more TCs that reach higher intensity values?

To examine this question we integrated the climatological environmental fields in the tropics in the season of interest (ASO in the Northern Hemisphere, JFM in the Southern Hemisphere) for each model and related
this to the corresponding NTC and ACE, as described in section 2. Given the very different range of values in NTC and ACE for low-resolution and high-resolution models, we split each scatterplot in two, one for low-resolution models (C models) and another for high-resolution models (W and P models). The resulting scatterplots for NTC are given in Figs. 11 and 12, for the Northern and Southern Hemisphere, respectively. Similar figures for ACE in each hemisphere are shown in Figs. 13 and 14. In each panel the linear fit and corresponding correlation coefficient are also shown.

Figures 11–14 make clear that there is no coherent relationship between the mean environmental conditions across the models and the mean TC climatology. For instance, large values of midlevel relative humidity are important for tropical cyclogenesis (Gray 1979; Emanuel and Nolan 2004). Nolan et al. (2007) and Rappin et al. (2010) found that reducing the free troposphere saturation deficit is critical for intensification. While low-resolution models with higher climatological relative humidity do generate more TCs in the Northern Hemisphere (Fig. 11a), that is not the case in the Southern Hemisphere (Fig. 12a), or for high-resolution models (Figs. 11b, and 12b) (actually, the opposite relationship is observed). Similarly, while there is a positive relationship between relative humidity and ACE in...
both hemispheres for low-resolution models (Figs. 13a and 14a), this is not the case for high-resolution models (Figs. 13b and 14b). In Figs. 11–14 we only show our results using the midlevel relative humidity, but similar plots were obtained using saturation deficit and column relative humidity.

In the case of relative vorticity, we would expect a higher number of TCs for models with higher mean climatological relative vorticity values (Gray 1979; Emanuel and Nolan 2004), as these models potentially could have more disturbances that lead to more TCs. Tippett et al. (2011) showed that the relationship of relative vorticity to genesis has a threshold beyond which higher values of vorticity are not related to more frequent cyclogenesis. While there is a positive relationship with NTC and ACE in the Northern Hemisphere for low-resolution models (Figs. 11c and 13c) and for ACE for high-resolution models in the Northern Hemisphere (Fig. 13d), the opposite occurs for low-resolution models in the Southern Hemisphere for NTC and ACE (Figs. 12c and 14c) and
for NTC in high-resolution models in the Southern Hemisphere (Fig. 12d), and there is no relationship at all in the other cases.

Vertical wind shear has a strong control on the climatology of TCs (Gray 1968), with developing storms tending to form under low values of vertical wind shear (McBride and Zehr 1981; Tang and Emanuel 2010). Large-scale vertical wind shear also tends to weaken tropical cyclones (DeMaria and Kaplan 1994; Tang and Emanuel 2010). For the vertical wind shear [Figs. 11–14, panels (e) and (f)], there is a decrease in NTC and/or ACE with the magnitude of the climatological vertical wind shear, as expected. But the relationship is weak and in one case (Fig. 12e) the relationship is the opposite.

There is a strong relationship between observed TC intensity and PI (Emanuel 2000; Wing et al. 2007; Kossin and Camargo 2009), with higher values of PI corresponding...
to stronger TCs in a large range of time scales and spatial scales. Empirically, genesis is rarely observed for PI values below 40 m s\(^{-1}\) (Emanuel 2018) and PI is used in various genesis indices (Emanuel and Nolan 2004; Emanuel 2010; Camargo et al. 2014), as well as being one of the components of the ventilation index (Tang and Emanuel 2012a,b). While for low-resolution models there is a positive relationship between PI and NTC, and PI and ACE [Figs. 11–14, panels (g)], the same is not true for high-resolution models [panels (h)], with a decrease of NTC in the Northern Hemisphere and of ACE in both hemispheres, in contrast to observations.

Murakami and Wang (2010) added vertical velocity as an additional predictor to the Emanuel and Nolan (2004) genesis index, arguing that a high frequency of TC genesis correspond to areas with large upward motion and that this vertical motion was not fully taken into account in the original genesis index. Zhao and Held (2012) explored the relationship of TC activity with various environmental variables using one of the climate models from our study (P3) and found that the strongest relationship was with vertical velocity at 500 hPa. Furthermore, the same authors argued in Held and Zhao (2011) that the atmospheric vertical mass flux can be useful in understanding the reduction of TC hurricane activity in their idealized climate change experiments. However, Camargo et al. (2014) did not find a coherent response of vertical velocity with this reduction in genesis in a perfect model experiment.
Here we find a positive relationship between vertical velocity and TC activity diagnostics only for low-resolution models [Figs. 11–14, panels (i)], not for high-resolution models [Figs. 11–14, panels (j)].

We also show scatterplots of NTC and ACE with one genesis index, namely TCGI developed by Tippett et al. (2011). We performed the same analysis with other versions of this index, using other predictors, as discussed in Camargo et al. (2014) (not shown). Similar to other genesis indices (e.g., Camargo et al. 2005; Camargo 2013; Wehner et al. 2015), there is not a strong relationship between the model climatological
TC activity and the climatological values of these indices in the same models, although the relationship between NTC and genesis indices seemed to improve with horizontal resolution for a few models (Camargo et al. 2005). However, changes in TC activity due to climate variability (e.g., El Niño–Southern Oscillation or volcano activity) are indeed reflected in changes in these indices (Camargo et al. 2005; Pausata and Camargo 2019; Camargo and Polvani 2019). Overall, similar to other variables, we do not obtain a coherent response of the mean climatological TCGI across models in our analysis [Figs. 11–14, panels (k) and (l)].

To try to examine if the lack of a robust relationships between NTC and ACE is influenced by the models’ resolution, we repeated this analysis for two different groups of models: in the first one, only models with resolution between 1.4° and 0.75° are considered, and in the second only models with resolution of 0.5° or higher. The results (shown in Figs. S2–S9) are very consistent with the ones shown above: there is no robust relationship between the environmental variables and NTC and ACE, even for models with similar resolutions and excluding models with unrealistic TC climatology. The only exception was found for models with resolution of 0.5° and higher, in

Fig. 11. Scatterplots of mean climatological NTC and environmental fields for the tropical Northern Hemisphere in ASO; C (CMIP5) models are shown in blue, W (HWG) models in red, P (process-oriented) models in green, reanalysis (R) in black. Statistically significance correlations at the 95th significance levels are shown in bold.
which NTC and ACE decrease with increasing vertical shear and the relationship is significant in 3 of the 4 cases examined. We repeated this analysis grouping the analysis by tracking routine for the models tracked with either the Camargo–Zebiak or Zhao/Vitart tracking algorithms. The results are similar (Figs. S13–S16), with no robust relationship between the climatological environmental variables and NTC and ACE for these two tracking algorithms across resolutions. This is a good indication that our results are not sensitive to tracking routine used, especially in the case of ACE (Zarzycki and Ullrich 2017).

4. Conclusions

It is common in the literature to try to explain biases in climate models’ TC climatology using biases in these models’ environmental variables climatology. However, as far as we are aware there is no study that shows that such relationship is actually valid. Here we explore this relationship using 30 climate models from three different multimodel ensembles at various resolutions. We show that there is no coherent relationship between the mean state of these models, represented here by a large number of environmental variables usually associated...
with TC activity, and the mean TC model climatology. In particular, there is no universal relationship between the simulated large-scale environment and TC activity, as while there are some relationships between environment and TC activity in certain classes of models, these relationships are not consistent across all models. This lack of coherent relationship between the environment and TC activity occurs even if only models with similar resolution are considered, and by excluding models with low-resolution and unrealistic TC climatology.

Our results are not surprising, given the large number of studies that explored the sensitivity of model TC climatology to various model characteristics (e.g., model physics, dynamical core, tracking methodology, etc.). However, given the widespread use of the large-scale environmental fields as an explanation to models’ TC biases, it is important to show that this standard practice is actually not valid. To understand model TC climatological biases more in-depth diagnostics are necessary, such as the process-based diagnostics developed by Kim et al. (2018), Wing et al. (2019), and Moon et al. (2020). Additional information may also be gained by working to understand the response of pre-TC synoptic disturbances, or “TC seeds,” in addition to the large-scale environmental impact on TC genesis (Vecchi et al. 2019).

FIG. 13. Scatterplots of mean climatological ACE and environmental fields for the tropical Northern Hemisphere in ASO; C (CMIP5) models are shown in blue, W (HWG) models in red, P (process-oriented) models in green, reanalysis (R) in black. None of the correlations are statistically significant at the 95th significance level.
Furthermore, TCs do not respond passively to the large-scale environment; they can influence it (e.g., Sobel and Camargo 2005), although the exact magnitude and nature of this influence is not completely understood. This could be another reason why there is a lack of relationship between the simulated environment and TCs—the modeled environment might be partially a consequence of the modeled TC activity and this interaction might be model dependent.

One of the caveats of our analysis is that it can be sensitive to the tracking algorithm used by each modeling group. In the case of low-resolution models, this is not an issue, as the same tracking algorithm was used across models. The differences in tracking algorithm could potentially influence our results for high-resolution models, but as the sensitivity to tracking algorithm is not as critical for strong TCs and high-resolution models, we expect our results to be robust.

It is important to stress that our analysis was restricted to the relationship between the models’ climatological environmental conditions and models’ TC climatology. This does not preclude the existence of such relationship in nature. Furthermore, our results do not have implications for the ability of climate models to simulate TC variability, in particular the modulation of TC activity.
by modes of climate variability (e.g., El Niño–Southern Oscillation or the Madden–Julian oscillation). It is well established that the variability of TC activity in models has the correct association with these climate modes, even if the model TC climatology is incorrect (Shaevitz et al. 2014; Wang et al. 2014; Han et al. 2016; Lee et al. 2018). Similarly, the response of the TC activity in models to climate change is not affected by our conclusions.

Acknowledgments. This work is a contribution to the process-oriented diagnostic effort of the NOAA (National Oceanic and Atmospheric Administration), MAPP (Modeling, Analysis, Predictions and Projections) Model Diagnostics Task Force, who also contributed with the simulations of two models in this study. This work was supported by NOAA’s Climate Program Office’s Modeling, Analysis, Predictions, and Projections program through Grant NA15OAR4310087. The authors thank all the members of U.S. CLIVAR Hurricane Working Group (HWG) for their contribution to this significant effort, in particular those who produced the model simulations used in this study. We would also like to thank Naomi Henderson for managing the HWG dataset. SJC, AHS, JDOS, and MK acknowledge support of NASA Grant 80NSSC17K0196. ES acknowledges support from the project PRIMAVERA, Grant Agreement 641727 of the Horizon 2020 research program. GAV is supported in part under NOAA Award NAOAR4320123. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, or the U.S. Department of Commerce.


REFERENCES


——, and L. M. Polvani, 2019: Little evidence of reduced global tropical cyclone activity following recent volcanic eruptions. npj Climate Atmos. Sci., 2, 14, https://doi.org/10.1038/S41612-019-0070-Z.


