Earth and Space Science

RESEARCH ARTICLE
10.1029/2020EA001415

Tropical Cyclone Characteristics in the MERRA-2 Reanalysis and AMIP Simulations

Zoe S. Aarons[1,2], Suzana J. Camargo[3], Jeffrey D. O. Strong[3,4], and Hiroyuki Murakami[5,6]

[1]Bowdoin College, Brunswick, ME, USA, [2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA, [3]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA, [4]AIR Worldwide, Boston, MA, USA, [5]National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [6]University Corporation for Atmospheric Research, Boulder, CO, USA

Key Points:

- Comparison of tropical cyclone activity in one reanalysis and its corresponding climate model simulation
- No significant improvement in the tropical cyclone activity climatological characteristics in the reanalysis compared to the climate model
- Tropical cyclones in this model are weaker than in models of similar horizontal resolution

Supporting Information:

- Supporting Information S1

Correspondence to:

S. J. Camargo,
suzana@ldeo.columbia.edu

Citation:

Aarons, Z. S., Camargo, S. J., Strong, J. D. O., & Murakami, H. (2021). Tropical cyclone characteristics in the MERRA-2 Reanalysis and AMIP simulations. *Earth and Space Science*, *8*, e2020EA001415. https://doi.org/10.1029/2020EA001415

Received 19 AUG 2020
Accepted 16 JAN 2021

Abstract This study evaluates the tropical cyclone (TC) activity in two high-resolution data sets—MERRA-2 Reanalysis (Modern-Era Retrospective Analysis for Research and Applications, Version 2) and MERRA-2 AMIP (Atmospheric Model Intercomparison Project). These data sets use the same atmospheric model, the Goddard Earth Observing System Model, Version 5 (GEOS-5) during the same period. However, while MERRA-2 AMIP is a free-running atmospheric simulation forced only with sea surface temperature (SST), MERRA-2 Reanalysis uses an advanced data assimilation system to include a large variety of data sets. Thus, we analyze (1) the sensitivity of TC activity to the model forcing, (2) how well the TCs in both data sets replicate observed TC characteristics, (3) the sensitivity of these results to tracking schemes and thresholds. Standard diagnostics such as the number of tropical cyclones and their intensity distribution are very similar in the AMIP model and the reanalysis. TCs in both data sets are weaker than observed, as is typical for the spatial resolution of these global models. Overall, the use of data assimilation in the MERRA-2 Reanalysis does not lead to a significantly better TC climatology than in AMIP. Furthermore, comparison of the MERRA-2 Reanalysis to two other reanalysis data sets shows that MERRA-2 generates fewer, but more intense TCs, than those reanalysis products.

1. Introduction

Past work has shown that General Circulation Models (GCMs) have the ability to generate Tropical Cyclones (TCs) with similar structures and statistics to observed storms (e.g., Camargo et al., 2005; Camargo & Wing, 2016; Shaevitz et al., 2014). Prior research in this field has focused on comparing the TC characteristics in individual models, multimodel ensembles, or reanalysis data sets to each other as well as the historical record (e.g., Camargo, 2013; Hodges et al., 2017; Murakami, 2014; Schenkel & Hart, 2012; Shaevitz et al., 2014; Wehner et al., 2015). However, to date, there has been no study that analyzed the differences between a reanalysis and a GCM when both have exactly the same setup.

Because reanalysis data sets are built using a similar underlying GCM, reanalysis data sets can be used as a historical baseline for GCMs (Murakami, 2014). But reanalysis data sets do not perfectly mirror observations. Most notably, reanalysis data sets underestimate the intensity of TCs (Hodges et al., 2017; Murakami, 2014). As the intensity of model TCs is closely linked to the horizontal model resolution, high-resolution models are needed to capture the fine scale processes involved in cyclogenesis and intensification of TCs (Walsh et al., 2007). For instance, in the case of the North Atlantic, Reale et al. (2009) estimated that at least horizontal resolutions of 20–30 km are necessary to accurately simulate TCs due to presence of the Saharan air layer and steep moisture gradients. Because biases are model dependent, a pair of reanalyses and GCM data sets developed using exactly the same formulation may have the potential to offer unique insight into the factors that can affect TC characteristics in models. Here we analyze the impacts of the data assimilation on the TC activity in a reanalysis data set by comparing with a companion GCM simulation forced only with SST prescribed to observed values, i.e., in our comparison there is no impact of the model formulation in simulating TCs, only how the model was forced. Because the reanalysis assimilates historical data, we hypothesize that the reanalysis should more accurately replicate observed storm characteristics than the model.

To examine this relationship between reanalysis and GCM, we analyze the MERRA-2 Reanalysis (Modern-Era Retrospective Analysis for Research and Applications, Version 2) and a MERRA-2 AMIP

(Atmospheric Model Intercomparison Project) simulation—referred to as MERRA-2 and AMIP, respectively (Gelaro et al., 2017; Molod et al, 2015). AMIP is a free-running GCM forced with only SST and sea-ice concentration. In contrast, the MERRA-2 reanalysis assimilates observed data into the GCM (Gelaro et al., 2017; Rienecker et al., 2008, 2011). By comparing TCs tracked in these data sets to each other and the historical record, we evaluate the sensitivity of the model TCs to the use of data assimilation, as well as how well both AMIP and MERRA-2 TCs replicate observed storms characteristics. The use of multiple tracking schemes throughout the analysis provides insight as to what differences are consistent across tracking schemes, and which vary as a result of tracking scheme rather than differences between the model and reanalysis.

Section 2 describes the methods and data used. The results are given in section 3, including a comparison of the MERRA-2 reanalysis to other reanalysis data sets, as well as an analysis of their environmental fields. Discussion and conclusions are presented in section 4.

2. Methods and Data

2.1. Data Sets

The observed historical data are from the National Hurricane Center (Atlantic and Eastern North Pacific) and Joint Typhoon Warning Center (Western North Pacific, North Indian Ocean and southern hemisphere) best-track data sets (Chu et al., 2002; Landsea & Franklin, 2013), in the period 1980–1999, which matches the model simulations.

AMIP consists of free-running simulations performed using the NASA Goddard Earth Observing System Model Version 5 (GEOS-5). Although there are 10 ensemble members for AMIP, we were only able to use one member in our analysis, as the high-frequency data (6-hourly) necessary to track the TCs was only available for one ensemble member (Collow et al., 2017). The model is forced with historical SST and the sea-ice boundary conditions (Bosilovich et al., 2015). The GEOS model has a c180 cubed-sphere grid which equates to a $0.5° \times 0.625°$ latitude-longitude grid and 72 vertical levels (Gelaro et al., 2017), and uses a Relaxed Arakawa-Schubert convection scheme (Moorthi & Suarez, 1992). These specifications are exactly the same as the GEOS model used for the MERRA-2 reanalysis. Therefore, as mentioned above, the only difference between the AMIP simulation and the MERRA-2 reanalysis is the addition of observations into MERRA-2 through data assimilation (Gelaro et al., 2017), which also gives an overview of the data assimilation system and its performance. The MERRA-2 data assimilation scheme is very similar to the original MERRA system (Gelaro et al., 2007; Rienecker et al., 2008, 2011), using a three-dimensional variational data algorithm with a number of important updates, including, e.g., the analysis algorithm and the observing system.

Analysis of the TC activity in the European Centre for Medium-Range Weather Forecasts Interim Reanalysis (ERA-Interim) and Japanese 55-Year Reanalysis (JRA-55) (Dee et al., 2011; Ebita et al., 2011; Kobayashi et al., 2015) will also be carried out. The resolutions of ERA-Interim and JRA-55 6-hourly data set are $1.5° \times 1.5°$ and $1.25° \times 1.25°$ latitude and longitude respectively (Dee et al., 2011; Kobayashi et al., 2015). Both utilize a four-dimensional variational data assimilation. JRA-55 also generates and uses artificial wind profiles at the TC observed locations (Ebita et al., 2011; Hatsushika et al., 2006), which are not included in either MERRA-2 or ERA-Interim.

2.2. Tracking Schemes

TC tracks are obtained by applying two different tracking schemes, the Camargo-Zebiak and the Geophysical Fluid Dynamics Laboratory (GFDL) TSTORMS (Camargo & Zebiak, 2002; Zhao et al., 2009) to the model and reanalysis 6-hourly data output. Although the details are slightly different between the two trackers, they have similar ingredients. In both the Camargo-Zebiak and TSTORMS tracking schemes, TCs are initially identified by the collocation of a local minimum in sea level pressure, a local maximum in column air temperature, and a local maximum in relative vorticity. However, the two tracking schemes employ different algorithms to detect these characteristics as described in more detail in Appendix A. The thresholds for these features are model-dependent and derived from probability distribution functions in the Camargo-Zebiak scheme. These snapshots are stitched together by the tracking routine and the TC

Table 1
Wind Speed Thresholds (m/s) Used in this Study for MERRA-2, JRA-55 (Murakami, 2014), and ERA-Interim (Murakami, 2014) Data Sets

MERRA-2	JRA-55	ERA-Interim
15.2	14	13.5

wind speed is recorded along the track (Camargo & Zebiak, 2002). A storm is defined by the presence of an integrated warm core anomaly using four pressure levels (850, 700, 500, and 250 hPa) and a threshold of 1.4 °C, a local minimum in pressure, and local maximum in both surface wind speed and 850 hPa vorticity, with thresholds of 9.9 and 15.2 m/s for the wind speed and 6.7×10^{-5} and 5.0×10^{-5} s^{-1} for vorticity—the first for the detection portion of the algorithm, the second for tracking (Camargo & Zebiak, 2002). The detection part of the algorithm also requires that the storm satisfies these conditions for 2 days. The tracking portion is applied in two ways—following the standard Camargo-Zebiak scheme (referred to as CZ) and with an additional wind speed requirement (referred to as CZ+). For CZ, the 15.2 m/s wind speed threshold must be met at least once over the course of the storm. For CZ+, the storm must meet this threshold for 3 days (not required to be consecutive). As noted in Wing et al. (2019), this modification removes the large number of weak storms and reduces the overall number of TCs. The same 15.2 m/s wind speed threshold was applied to TSTORMS (referred to as TS) and all tracking schemes required that TCs form in the tropics (30°S to 30°N). The wind speed threshold of 15.2 m/s is in the range recommended for a 50-km grid model by Walsh et al. (2007).

The TC tracks for ERA-Interim and JRA-55 are from Murakami (2014) and utilize the tracking scheme by Murakami and Sugi (2010). As the tracks in these reanalyses data sets have been obtained by a different tracking algorithm, we expect there to be some sensitivity to the tracking algorithm used, especially for the weaker storms (Horn et al., 2014). However, these tracks would give at least an overview of how the MERRA-2 TC climatology compares with TCs in other reanalyses data sets. And similar to the Camargo-Zebiak and TSTORMS tracking schemes, the Murakami and Sugi (2010) scheme identifies a warm core and uses wind speed, vorticity, and duration thresholds. The wind speed thresholds for each reanalysis meet the criteria presented by Walsh et al. (2007), where the threshold is determined by the horizontal resolution of the reanalysis data set (Table 1).

The tracking schemes used in this study are dependent on wind thresholds, which is a standard approach for tracking TCs in models (e.g., Vitart et al., 1997), following the tracking algorithm originally developed by Bengtsson et al. (1995). Here, the wind speed thresholds were obtained using the nominal, rather than the effective, resolution despite the fact that grid spacing is only a rough approximation of the effective resolution of a model (Klaver et al., 2020). Recently, some caveats of the dependence of wind speed thresholds on model resolution have been discussed in Davis (2018) and Moon et al. (2020a). Furthermore, Roberts et al. (2020) discussed discrepancies between TC intensity based on the lowest model level and 10 m winds for some models, suggesting sensitivity to the boundary layer scheme used. We note that there are various tracking algorithms that are not dependent on wind speed thresholds, such as TRACK (Hodges, 1994; Hodges et al., 2017), TempestExtremes (Ullrich & Zarzycki, 2017), and Tory et al. (2013), which identify storms using vorticity (on a common spectral grid), sea level pressure, and regions of enhanced vorticity deformation, respectively.

2.3. Diagnostics

Typical diagnostics will be used to examine number, distribution, and intensity of model and reanalysis TCs. The number of tropical cyclones (NTC) will be used to assess the frequency per year. Accumulated cyclone energy (ACE), which is calculated by summing the square of the 6-hourly wind speeds of TCs, is a standard diagnostic of TC activity, integrating TC intensity, frequency, and duration (Camargo & Sobel, 2005). These diagnostics will be applied globally and in individual TCs basins: South Indian (SI), Australian (AU), South Pacific (SP), North Indian (NI), Western North Pacific (WNP), Central North Pacific (CNP), Eastern North Pacific (ENP), and Atlantic (ATL) (Figure 1). A two-sample Kolmogorov-Smirnov test with a 95% confidence interval is applied to assess the significance of these metrics. Besides these TC metrics, we analyzed environmental variables typically associated with TCs, such as vertical shear, relative humidity, vorticity. Furthermore, we calculated the Emanuel's potential intensity (Emanuel, 1988), which is the theoretical maximum intensity that a TC can reach based on the local environmental conditions. We used the standard formulation of PI given in Bister and Emanuel (2002).

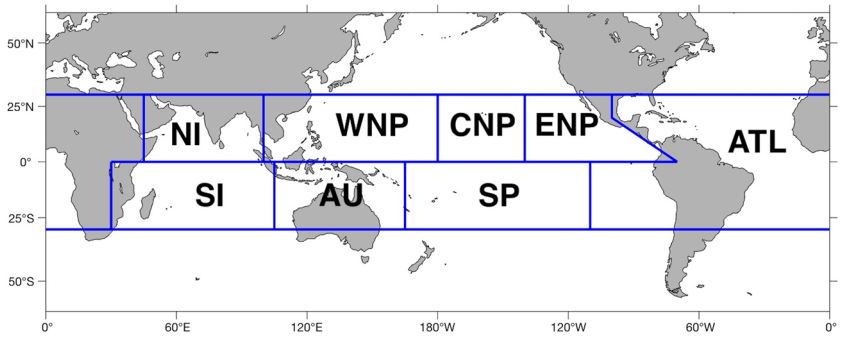

Figure 1. Ocean basins used to categorize location of tropical cyclones.

El Niño Southern Oscillation (ENSO) diagnostics are calculated using the Oceanic Niño Index from the National Weather Service Climate Prediction Center (Kousky & Higgins, 2007). El Niño and La Niño seasons are determined by hemisphere, where the ENSO events for ASO (August to October) are used for the Northern Hemisphere and the JFM (January to March) season is considered for the Southern Hemisphere. Mean values of NTC and ACE are calculated for neutral, El Niño, and La Niño seasons. Anomalies are then calculated by subtracting the neutral conditions from the El Niño and La Niño data sets. ACE and TC density can then be calculated by summing the ACE and NTC values per 5° × 5° box and then dividing by the number of El Niño or La Niño seasons in each hemisphere. By taking the difference between the anomaly during El Niño and La Niña seasons, positive values indicate an increase in the NTC or ACE during El Niño and negative values represent an increase during La Niña.

3. Results

3.1. Sensitivity of Tracking Schemes

The choice of threshold values has a large impact on the number of cyclones detected by tracking algorithms, including the Camargo-Zebiak tracking scheme (Figure 2). Among the thresholds considered, altering the temperature threshold (warm core) has the least sensitivity in the resulting NTC, while wind speed has the greatest impact. The number of TCs detected by CZ+ decreases by more than 100 storms per year as a result of raising the wind speed threshold. Past studies have recognized the need for resolution-dependent criteria for wind speed thresholds (Davis et al., 2018; Moon et al., 2020a; Walsh et al., 2007). As noted in the section 2;, this study uses a wind speed threshold of 15.2 m/s in agreement with Walsh et al. (2007). While this choice significantly decreases the NTC, it will also reduce the number of weak storms. When vorticity, temperature, and duration are also considered, the combined impact of threshold values becomes even greater. Kim et al. (2020) applied the Camargo-Zebiak tracking scheme but considered lower thresholds than this study. Their threshold values are also lower than those used in models with much lower horizontal resolution (e.g., Camargo, 2013). When applying similar thresholds globally as those used in Kim et al. (2020) for the western North Pacific leads to the detection of 2,000 more storms than CZ+ over the period 1980–1999. Additionally, the significance of the duration threshold is clear in the comparison of CZ+ and CZ (Figure 3). Raising the duration threshold results in significantly fewer storms, of which a greater percentage are more intense. And, as expected, the duration threshold increases not only the intensity of storms but the average length of storm tracks (Figure 4). It is clear that the selection of thresholds plays a significant role in the resulting storm tracks. This study chooses to select high threshold values in order to eliminate the detection of numerous weak and short-lived TCs and therefore produce storms with characteristics that are, on average, more similar to observed TCs and in line with the thresholds used in other studies in the literature.

3.2. Climatology: AMIP Versus MERRA-2

Using the Camargo-Zebiak tracking scheme, AMIP generates a higher global NTC than MERRA-2 for the period 1980–1999 for all tracking schemes (Figure 3). All tracking schemes for MERRA-2 as well as CZ+ and TS for AMIP track significantly fewer storms than observed. In contrast, AMIP CZ tracks a similar

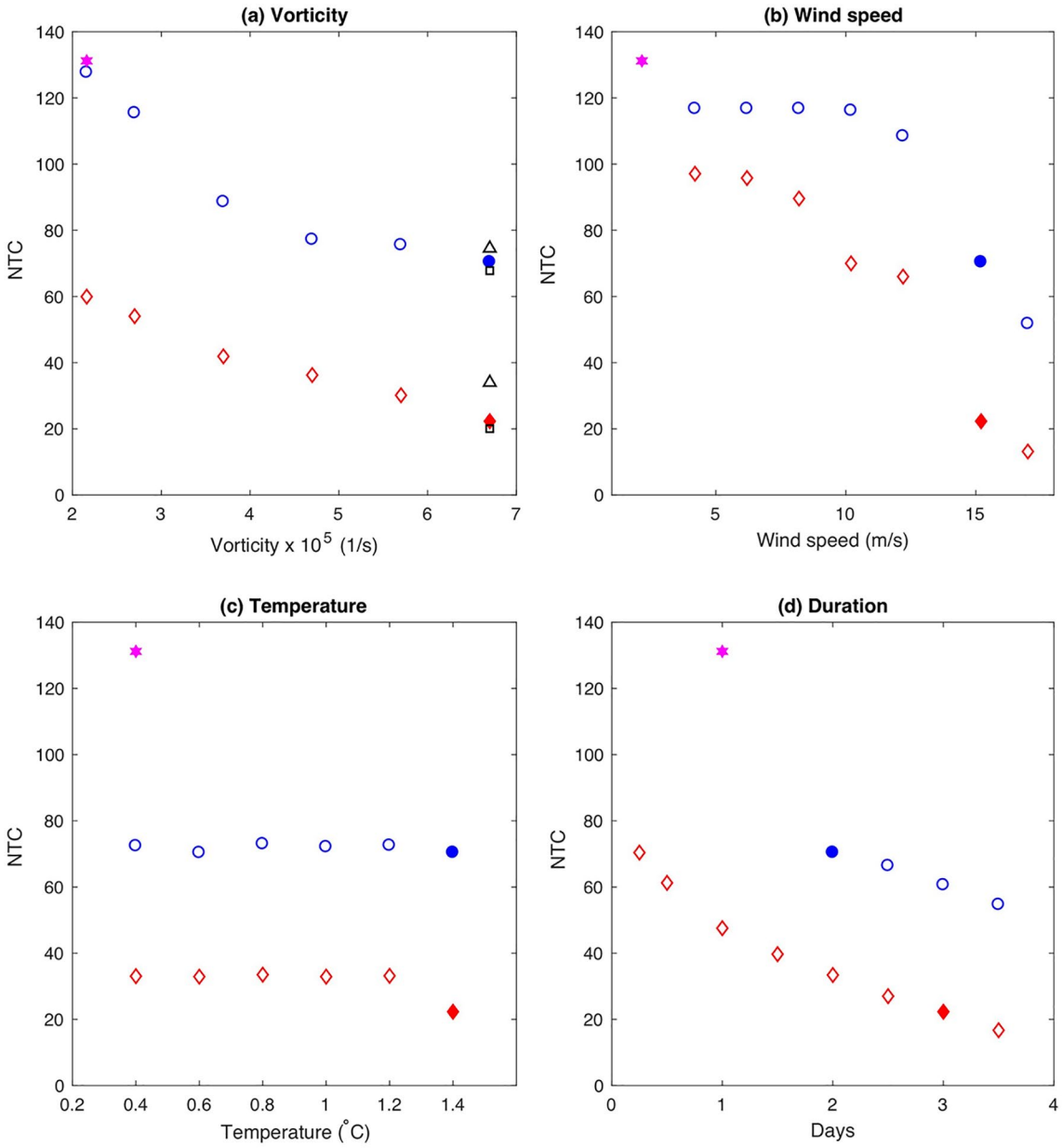

Figure 2. Impact of varying tracking scheme threshold values on annual mean NTC for MERRA-2 for 1980–1999. Pink stars represent the threshold values used in Kim et al. (2020) for WNP. Filled blue circles and red diamonds depict the thresholds for CZ and CZ+, respectively. Unfilled blue circles and red diamonds represent thresholds for CZ and CZ+ where a single threshold is altered at a time in the detection and tracking parts of the algorithm: (a) vorticity (detection), (b) wind speed (tracking), (c) temperature (detection), and (d) duration (tracking). Black squares and triangles show the impact of lowering (4.5×10^{-5} s^{-1}) or raising (5.5×10^{-5} s^{-1}) the vorticity threshold in only the tracking portion of the algorithm for CZ and CZ+. NTC, number of tropical cyclones; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; WNP, Western North Pacific. CZ, Camargo-Zebiak.

number of storms as in observations. This is a result of the lower wind speed duration thresholds applied in the CZ tracking scheme compared to CZ+, which allows the detection of higher number of weak storms, increasing the total NTC, but shifting the intensity distribution toward lower wind speeds (Wing et al., 2019). In contrast, while MERRA-2 CZ mean NTC is greater than CZ+ and TS due to the lower thresholds, it is lower than observations and AMIP CZ.

At the basin scale, AMIP produces a greater NTC than MERRA-2, while MERRA-2 never outperforms AMIP, and both MERRA-2 and AMIP detect fewer storms than observed (Figure 5 and Table S1). AMIP has

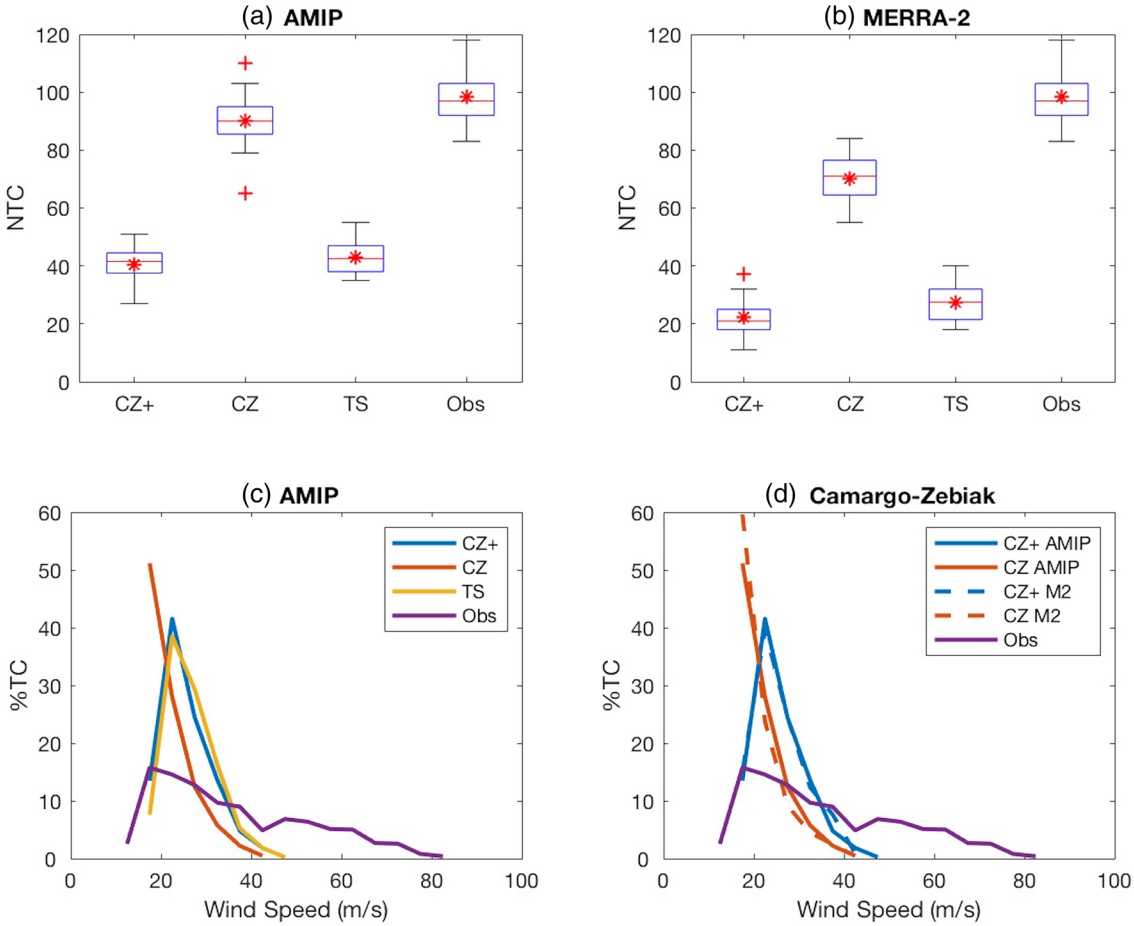

Figure 3. Global number of TCs per year in observations and tracked by the CZ+, CZ, and TS tracking schemes for (a) AMIP and (b) MERRA-2. Maximum wind speed (m/s) of TCs (%) in observations and model products using (c) all schemes in AMIP and using (d) both Camargo-Zebiak tracking schemes for AMIP and MERRA-2 (M2). For each boxplot (a,b), the red line indicates the median and the top and bottom lines of the box represent the 75th and 25th percentiles respectively. The whiskers represent the maximum and minimum ($w \approx +/-2.7\sigma$) while the red crosses denote extremes (defined as greater than $q_3 + w \times (q_3 - q_1)$ where q_1 and q_3 are the 25th and75th percentiles of the data, respectively). The red stars denote the distribution mean. TC, tropical cyclone; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; AMIP, Atmospheric Model Intercomparison Project; CZ, Camargo-Zebiak.

a significantly higher average NTC per year than MERRA-2 for all tracking schemes in the Australian basin. In four other basins, two tracking schemes detect a higher NTC for AMIP than MERRA-2—CZ+ and TS in the South Indian and North Indian as well as CZ+ and CZ in the Western North Pacific and Atlantic. There is no significant difference among the tracking schemes in the average NTC for the South Pacific, Central North Pacific, and Eastern North Pacific basins. Additionally, MERRA-2 has significantly fewer TCs than observed in all basins for CZ+ and TS. CZ detects significantly more storms than observed in the North Indian, while it finds no significant difference in the South Indian, South Pacific, and Central North Pacific and significantly fewer storms in the remaining four basins. AMIP has significantly fewer storms than observed for all tracking schemes in the Western North Pacific, Eastern North Pacific, and Atlantic. While AMIP outperforms MERRA-2 in terms of the NTC, both are unable to accurately represent the observed global NTC across tracking schemes and basins. This number of TCs in AMIP is lower than that of models with similar resolution (Camargo et al., 2020; Wing et al., 2019). However, despite the significant differences between the NTC detected in AMIP, MERRA-2, and observations, the relationship between the percent distribution of storms across basins is less clear (Figure 6). There is no consistent pattern evident in the model and reanalysis or across tracking schemes and, with the exception of TS for MERRA-2 in the Western North Pacific, AMIP, and MERRA-2 do not deviate from the observations by more than around 10%. And the increased sensitivity to tracking scheme in the WNP is attributed to the fact that the greatest number of storms occur in this region. Furthermore, the relationships are not coherent for both NTC and distribution across

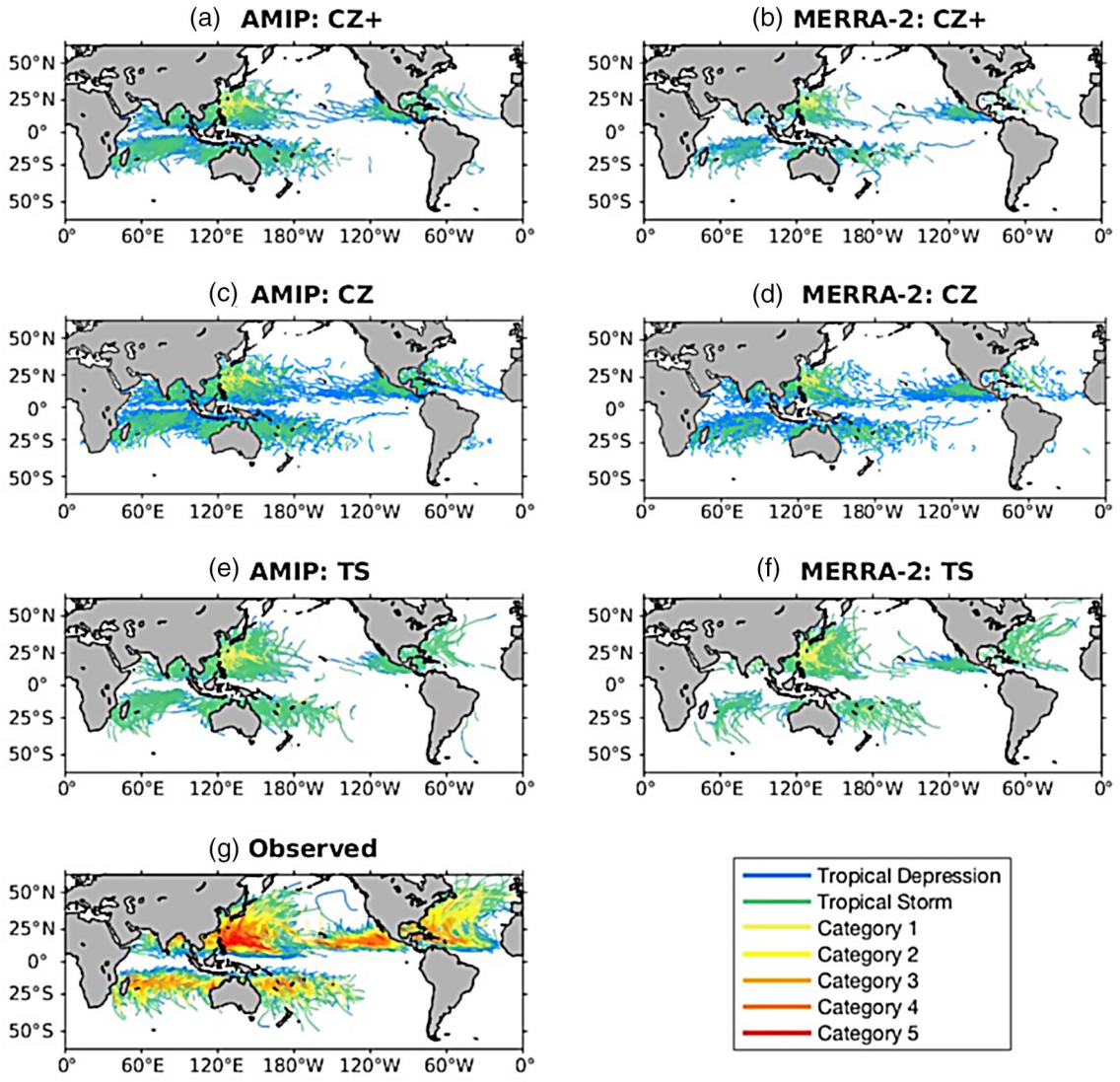

Figure 4. Tropical cyclone tracks for 1980–1999 in observations (g) and identified using the CZ+ (a,b), CZ (c,d), and TS (e,f) tracking schemes in the AMIP (left column) and MERRA-2 (right column) data sets. Colors represent the intensity of storms as classified under the Saffir-Simpson hurricane wind scale. AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; CZ, Camargo-Zebiak.

basins. For instance, while AMIP produces more storms than MERRA-2 in the Australian basin, it does not have a higher percentage of storms in this basin. The distribution of storms indicates that while AMIP and MERRA-2 underperform in generating TC frequency, they may still be capable of more accurately assessing other elements of the observed storm distribution and traits.

When considering the latitude of TC genesis rather than the number of storms per basin, both AMIP and MERRA-2 predict, on average, a greater percentage of storms than observed in the Southern Hemisphere (Figure 7). The high percentage of TCs detected by the CZ+ and CZ tracking schemes from 0-5°S may be a result of the model bias in which TCs to form too close to the equator or a bias of the tracking algorithm in falsely detecting non-TCs, e.g., tropical disturbances. In the North Hemisphere, MERRA-2 and AMIP maxima are close to the observed maximum around 15°N. Yet both the model and reanalysis exhibit fewer TCs than observed in the peak range of TC genesis—10–15°N. The only exception is TS for MERRA-2 which produces a similar number of TCs to observed in this latitude range and also is the only tracking scheme with fewer storms than observed in the latitude range 0–10°S. From the equator to 10°N, AMIP relatively accurately mirrors the observations while the percentage of TC genesis in MERRA-2 is too low. Therefore,

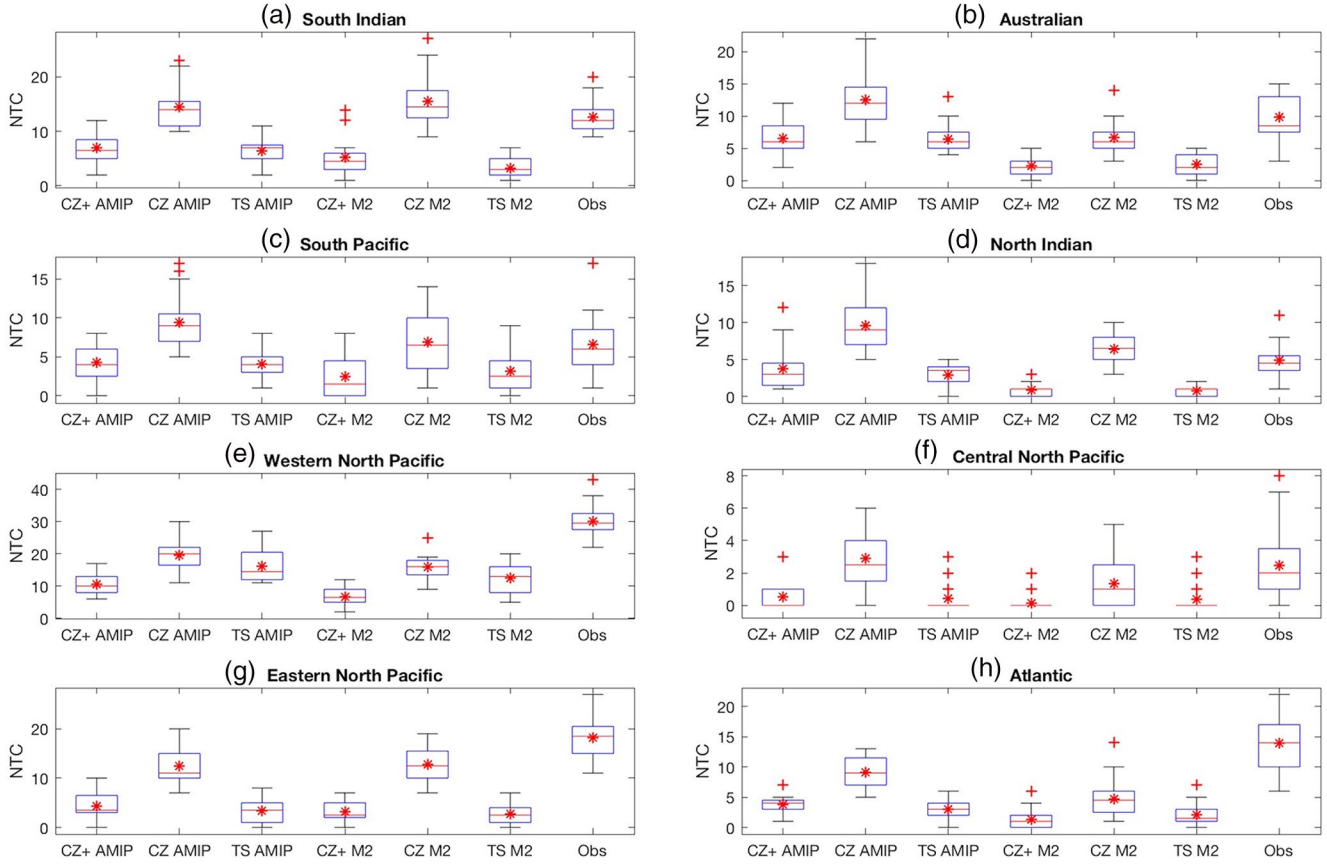

Figure 5. Boxplot of mean number of TCs per year in eight basins from 1980 to 1999 in AMIP, MERRA-2 (M2), and observations. All three tracking schemes (CZ+, CZ, and TS) are considered for both AMIP and M2. Storms are assigned an ocean basin by the latitude and longitude of the TC's genesis location. Boxplot design is the same as Figure 3. CZ, Camargo-Zebiak; AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; TC, tropical cyclone.

while close to the observed distribution, both the AMIP and MERRA-2 percentage of TCs is too high in the Southern Hemisphere and too low in the Northern Hemisphere.

AMIP and MERRA-2 are both capable of reproducing the observed NTC seasonal cycle (Figure 8). The peaks in seasonal cycle found in the model and reanalysis are flatter compared to observations, particularly in the North Pacific and Atlantic basins, due to a the lower global NTC. When the percentage of storms, rather than total number, is considered, the agreement with the observations is evident across most basins (Figure S2). Despite differences in the number of storms, the majority of the tracking schemes demonstrate statistically significant positive correlations with the observed seasonal variations in all basins except the Central North Pacific, where only CZ and TS for MERRA-2 are significantly correlated with observations. In the Australian, South Pacific, Eastern North Pacific, and Atlantic, all tracking schemes are significantly correlated with observations and, for each of these basins, at least three of the tracking schemes have correlation coefficients above 0.80. The South Indian and Western North Pacific each have significant correlations for five tracking schemes and also have high correlation coefficients. The North Indian, which contains two seasonal peaks, has the lowest correlation coefficient values. This is a common issue in many climate models (e.g., Shaevitz et al., 2014), one of the possible reasons is that the models produce similar structures for TCs and monsoon depressions in the region, which cannot be distinguished by the tracking schemes. With the exception of the North Indian, tracking schemes applied to MERRA-2 outperform or perform equally to their AMIP counterparts. And overall, more significant correlation coefficients are reported for MERRA-2 than AMIP. Despite the inability of AMIP and MERRA-2 to fully reproduce seasonal TC patterns in the North Indian, both the model and reanalysis perform well across basins, with MERRA-2 outperforming AMIP.

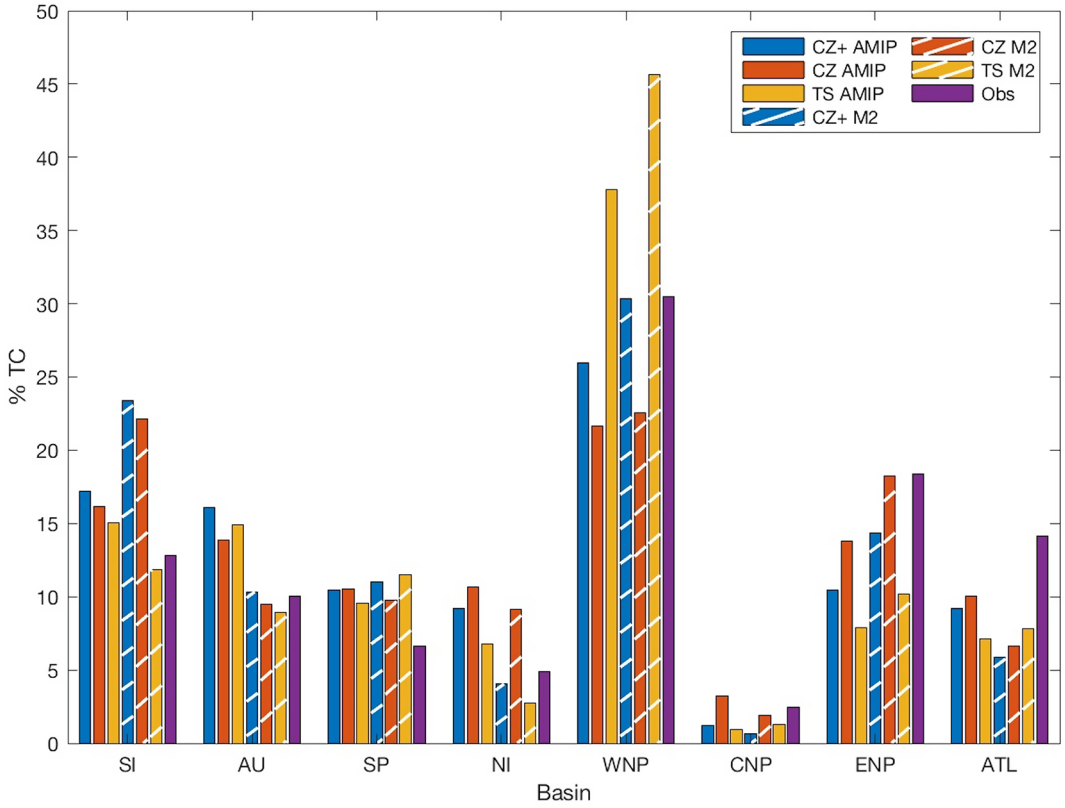

Figure 6. Percentage of TCs in each basin during 1980–1999 for the AMIP, MERRA-2 (M2), and observed data sets for the CZ+, CZ, and TS tracking schemes. Percentage is calculated using the mean number of TCs per year in each basin: (NTC_{basin}/NTC_{globe}) * 100. CZ, Camargo-Zebiak; AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; TC, tropical cyclone; NTC, number of tropical cyclones.

On interannual timescales, AMIP and MERRA-2 largely fail to reproduce the observed NTC annual variability (Figure 9). In the majority of basins, no tracking schemes for AMIP or MERRA-2 has statistically significant correlations with the observed timeseries. In the Western North Pacific, only TS AMIP has a significant correlation coefficient (0.51), while in the Atlantic the significant correlations are in the range 0.46–0.56. In the Atlantic for MERRA-2 the correlation values are 0.48 (CZ+) and 0.56 (TS), which are both much lower than the value of 0.84 reported by Kim et al. (2020). Besides lower thresholds for the Camargo-Zebiak tracking scheme Kim et al. (2020) considers a longer time period (1980–2016) than this study. In the South Pacific, where the mean NTC is closer to observed, the interannual variability is significantly correlated with the observations using all tracking schemes, with correlation coefficients in the range from 0.69 to 0.75, which are also higher than in all other basins. Therefore, MERRA-2's performance in the South Pacific basin shows the greatest ability to replicate the interannual variability of the observations.

Although AMIP and MERRA-2 fail to fully reproduce the observed storm count annual variability, they are capable of replicating the modulation of ENSO on ACE density (Figure 10). All tracking schemes show a positive ACE anomaly, indicative of an increase in ACE during El Niño events, in the Australian and South Pacific basins as observed (e.g., Camargo et al., 2010). And CZ+ for MERRA-2, TS for AMIP, and CZ for both MERRA-2 and AMIP reproduce the negative anomaly in the Atlantic ACE, corresponding to active La Niña seasons. CZ+ MERRA-2, TS AMIP, CZ AMIP, and CZ MERRA-2 also reproduce the observed southwestward/northeastward shift of ACE in the Western North Pacific (Camargo & Sobel, 2005). The increase in TC activity in El Niño years in the eastern North Pacific (Camargo et al., 2008) does not appear clearly here in observations, probably due to the small number of ENSO events in the period, but it is very clear using the TS and CZ AMIP tracking. Though neither AMIP nor MERRA-2 reproduce TC

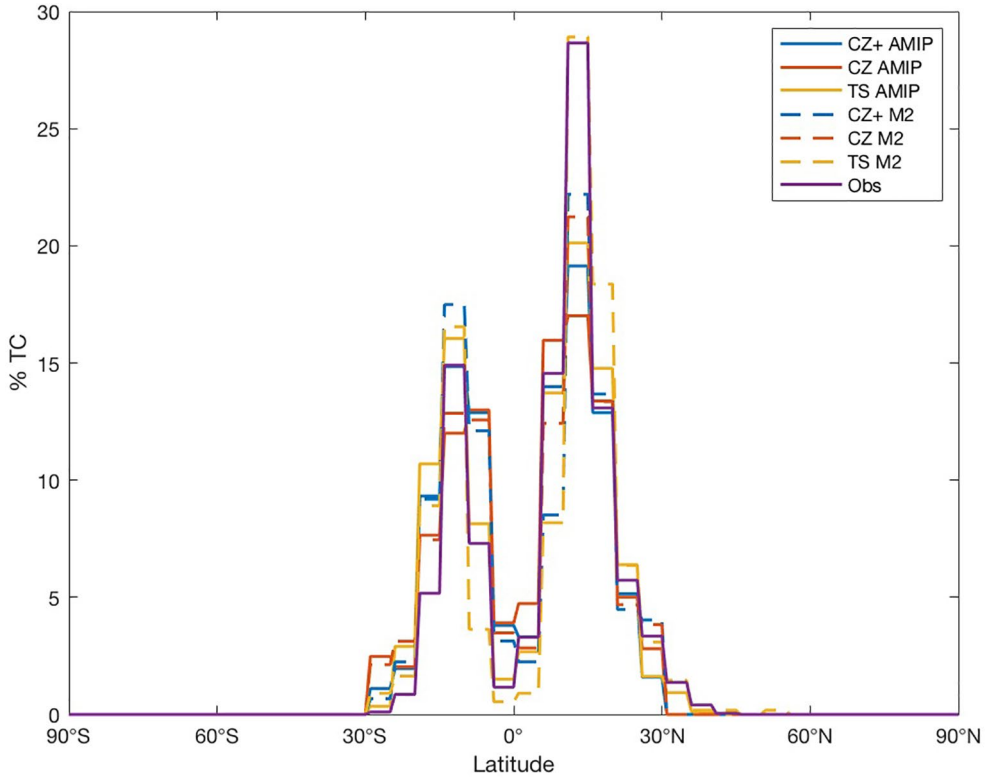

Figure 7. Latitude of TC genesis in AMIP, MERRA-2 (M2), and observations for the CZ+, CZ, and TS tracking schemes from 1980 to 1999. Percent of TCs found at each latitude is calculated for 5° bins. AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; CZ, Camargo-Zebiak; TC, tropical cyclone; AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications.

track density anomalies that reflect the impact of ENSO in TC occurrence very well (Figure S1), likely a result of the low number of storms detected overall, there is some similarity with observations. However, as there were only nine ENSO events during the study period (6 El Niño and 5 La Niña events in the Northern Hemisphere and 7 El Niño and 4 La Niña events in the Southern Hemisphere), this could hinder somewhat the ability to properly assess the ENSO modulation. So, while MERRA-2 and AMIP show some ability to reproduce the modulation of ENSO on ACE, this does not occur homogenously across basins and tracking schemes.

As in the case of the number of TCs, AMIP and MERRA-2 have lower ACE values than observed for all basins, with the exception of CZ+ AMIP in the North Indian Ocean and CZ AMIP in the South Pacific and North Indian Oceans (Figure 11 and Table S2). As ACE is an integrated measure of all storms wind speeds throughout their lifetime, these results are in agreement data set with the distribution of winds shown in Figure 3. Neither MERRA-2 nor AMIP has any storms with maximum wind speeds of >50 m/s while the observed contains TCs that reach >80 m/s. This difference in intensity is to be expected because to simulate storms with such wind speeds, a finer horizontal resolution is necessary than found in the model or reanalysis (Davis, 2018). Yet the intensity of the AMIP storms is still much weaker than other models with similar resolution (Camargo et al., 2020; Moon et al., 2020b; Shaevitz et al., 2014; Wing et al., 2019). For example, AMIP (using the CZ+ tracking scheme) exhibits a median TC maximum intensity that is about 5 m/s less than HiRAM (Geophysical Fluid Dynamics Laboratory High-Resolution Atmosphere Model) (Shaevitz et al., 2014). While AMIP has higher ACE values than MERRA-2 for CZ+ in all basins except the Central North Pacific and Eastern North Pacific, the wind speed distributions for AMIP and MERRA-2 do not differ significantly (Figure 3d). Therefore, the higher ACE value of AMIP compared to MERRA-2 for the CZ+ tracking scheme is a result of a higher number of TCs rather than the intensity of individual storms.

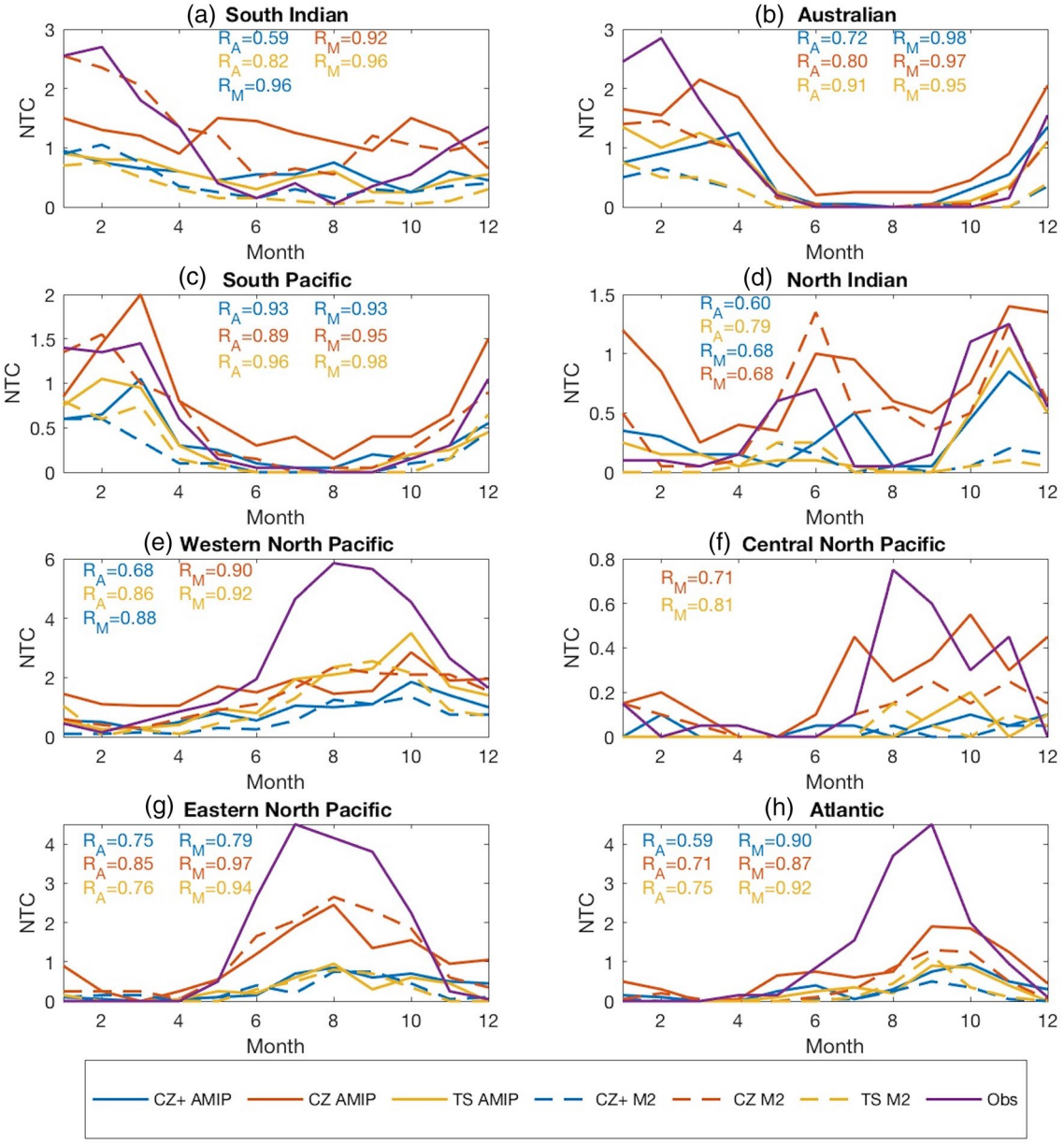

Figure 8. Average NTC per month in AMIP, MERRA-2 (M2), and observations for the CZ+, CZ, and TS tracking schemes from 1980 to 1999. Correlation coefficients between tracked TCs and observations are shown in same color as monthly data (with subscript denoting AMIP versus MERRA-2) when significant using a 95% confidence interval. AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; CZ, Camargo-Zebiak; TC, tropical cyclone; NTC, number of tropical cyclones.

While CZ+ identified higher NTC values for AMIP than MERRA-2 in more basins than CZ or TS, there are more basins with significant differences in ACE values using CZ, namely higher ACE values for AMIP than MERRA-2 in seven basins. While using CZ+ and TSTORMS, this occurs in 6 basins and four basins, respectively. Because CZ only detected a higher NTC for AMIP than MERRA-2 in three basins, the significant differences in ACE values are likely a result of differences in the strength of TCs across basins, rather than just the number of storms. The numerous low intensity storms detected by CZ lead to overall greater ACE values than the other tracking schemes.

MERRA-2 storms have, on average, a shorter duration than AMIP (Figure 12). For each tracking scheme, MERRA-2 contains a greater percentage of short-lived storms. There are also differences across tracking

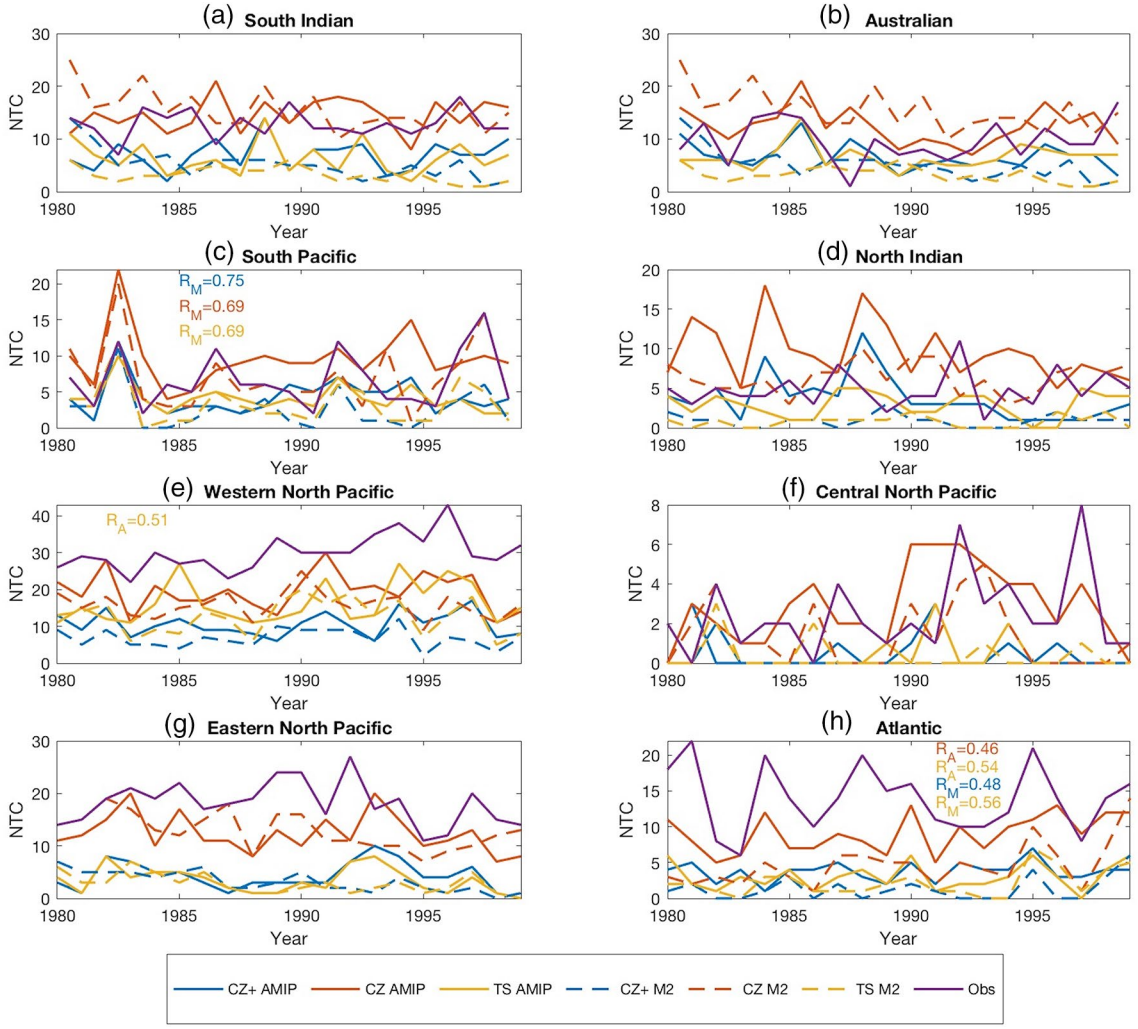

Figure 9. NTC per year in AMIP, MERRA-2 (M2), and observations for the CZ+, CZ, and TS tracking schemes from 1980 to 1999. For basins in the Northern Hemisphere, TCs are summed from January to December. For the Southern Hemisphere, years are calculated as July-June and are plotted from 1980.5 to 1998.5. Correlation coefficients between tracked TCs and observations are shown in same color as monthly data (with subscript denoting AMIP versus MERRA-2) when significant using a 95% confidence interval. TC, tropical cyclone; CZ, Camargo-Zebiak; AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; TC, tropical cyclone; NTC, number of tropical cyclones.

schemes, with TS producing the briefest storms and CZ+ detecting those with longer average durations. CZ AMIP and CZ+ MERRA-2 have distributions that best mirror the observed data. Although MERRA-2 storms have shorter duration than AMIP, neither the model nor the reanalysis reproduces well the observed storm duration distribution, in particular the location of the peak and the tail of long-lived storms.

3.3. Climatology: Reanalysis Data Sets

Let us compare now the TC activity in MERRA-2 with other reanalysis data sets. MERRA-2 detects, on average, fewer TCs per year per basin compared to JRA-55, ERA-Interim, and the observations (Figure 13). While all MERRA-2 tracking schemes do not have a statistically significant difference in NTC compared to ERA-Interim in the Central North Pacific and TS does not have a statistically significant difference to JRA in the South Pacific, CZ+ and TS otherwise track fewer NTCs than JRA-55 and ERA-Interim across the rest of the basins (Table S3). CZ only detects significantly fewer storms than both JRA-55 and ERA-Interim in the Australian and Western North Pacific basins. JRA-55 and ERA-Interim also have NTC values that are closer to the obser-

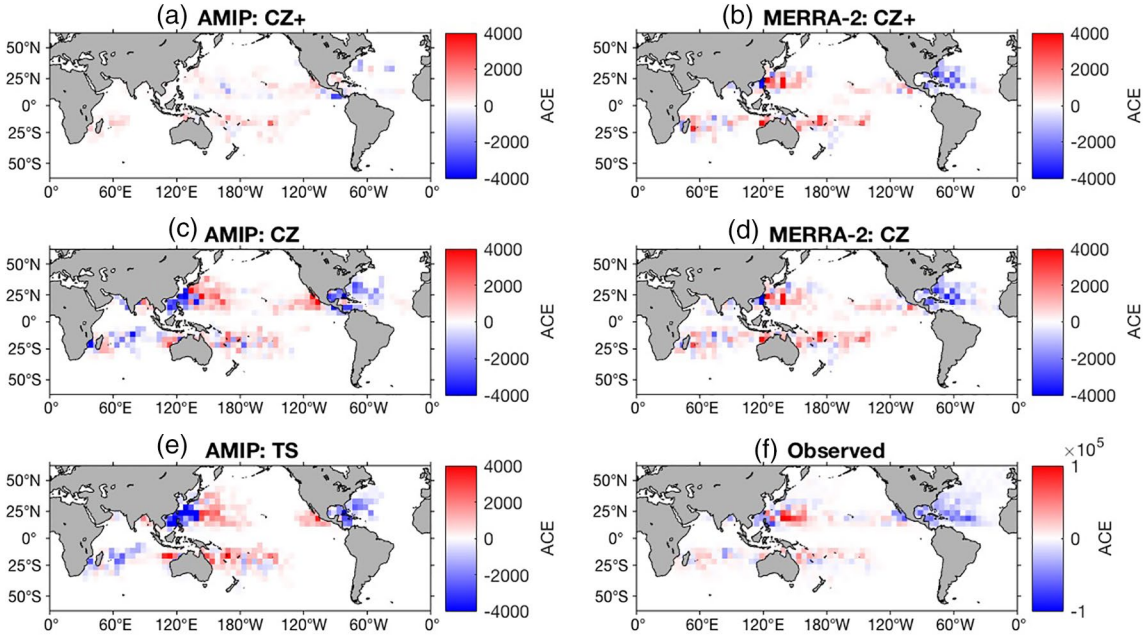

Figure 10. Anomalies in ACE density (m^2/s^2) during ENSO in AMIP, MERRA-2, and observations for the CZ+, CZ, and TS tracking schemes from 1980 to 1999. ACE density is the summed ACE values per 5° × 5° box. The anomaly values are then calculated by taking the difference between the anomaly during El Niño years and during La Niña years so that positive is an increase in ACE during El Niño and negative is an increase during La Niña. CZ, Camargo-Zebiak; ACE, accumulated cyclone energy; AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; ENSO, El Niño Southern Oscillation.

vations than those of MERRA-2. Specifically, ERA-Interim NTC is similar to the observations in the South Indian, Australian, South Pacific, and Western North Pacific basins, while JRA-55 only differs significantly from the observed NTC in the Atlantic. As the previous MERRA reanalysis, JRA-55 outperforms both MERRA and MERRA-2 in detecting TCs (Murakami, 2014). The improved NTC found by JRA-55 is likely due to the use of artificial wind profiles, derived from observed storm data, in the assimilation (Ebita et al., 2011; Hatsushika et al., 2006). When the percentage of storms per basin is considered, MERRA-2 looks much more similar to the other reanalysis data sets and to the observed data (Figure S3). Thus, MERRA-2 has a reasonable distribution of storms per basins but too few TCs when compared to JRA-55, ERA-Interim, and observations (cf., Figures 6 and 10). These results contrast those from past studies that found that MERRA-2 produced a comparable if not greater number of storms than JRA-55 and ERA-Interim across individual basins and a larger total global NTC (Hodges et al., 2017; Roberts et al., 2020). However, it is important to note that in Roberts et al. (2020) there are large differences in the mean number of TCs identified in MERRA-2 using TRACKS and TempestExtremes, 64.8 (53.9) and 41.2 (25.5) in the northern (southern) hemisphere, respectively (see Figures 1 and 2 of Roberts et al., 2020). These large differences among tracking algorithms are very suggestive of the occurrence of a high number of short-lived and weak storms, which are extremely sensitive to the tracking algorithm used, as discussed in Horn et al. (2014). Table 1c in Murakami (2014) also shows a large range in the mean number of TCs for the same reanalyses data sets, when tracked by different tracking algorithms. In the most extreme case for the MERRA reanalysis, one study obtained a global mean of 9.9 TCs (Walsh et al., 2007), while the another obtained 83.6 (Murakami, 2014), with a third one obtaining a value in between (39.3 in Strachan et al. (2013)). Therefore, a large discrepancy in the number of TCs due to use of different tracking algorithm in the same data set has been clearly well documented in the literature before.

Like for the NTC, MERRA-2 never exhibits significantly larger ACE values than JRA-55 and ERA-Interim (Figures 13i and 13j; Table S4). And MERRA-2, JRA-55, and ERA-Interim all have significantly smaller ACE values than observed (Tables S2 and S4). The CZ+ and TS tracking schemes for MERRA-2 result in ACE values that are significantly lower than those of JRA-55 in the majority of basins, and TS MERRA-2 results in ACE values that are significantly less than ERA-Interim in the majority of basins (Table S4). But the TS

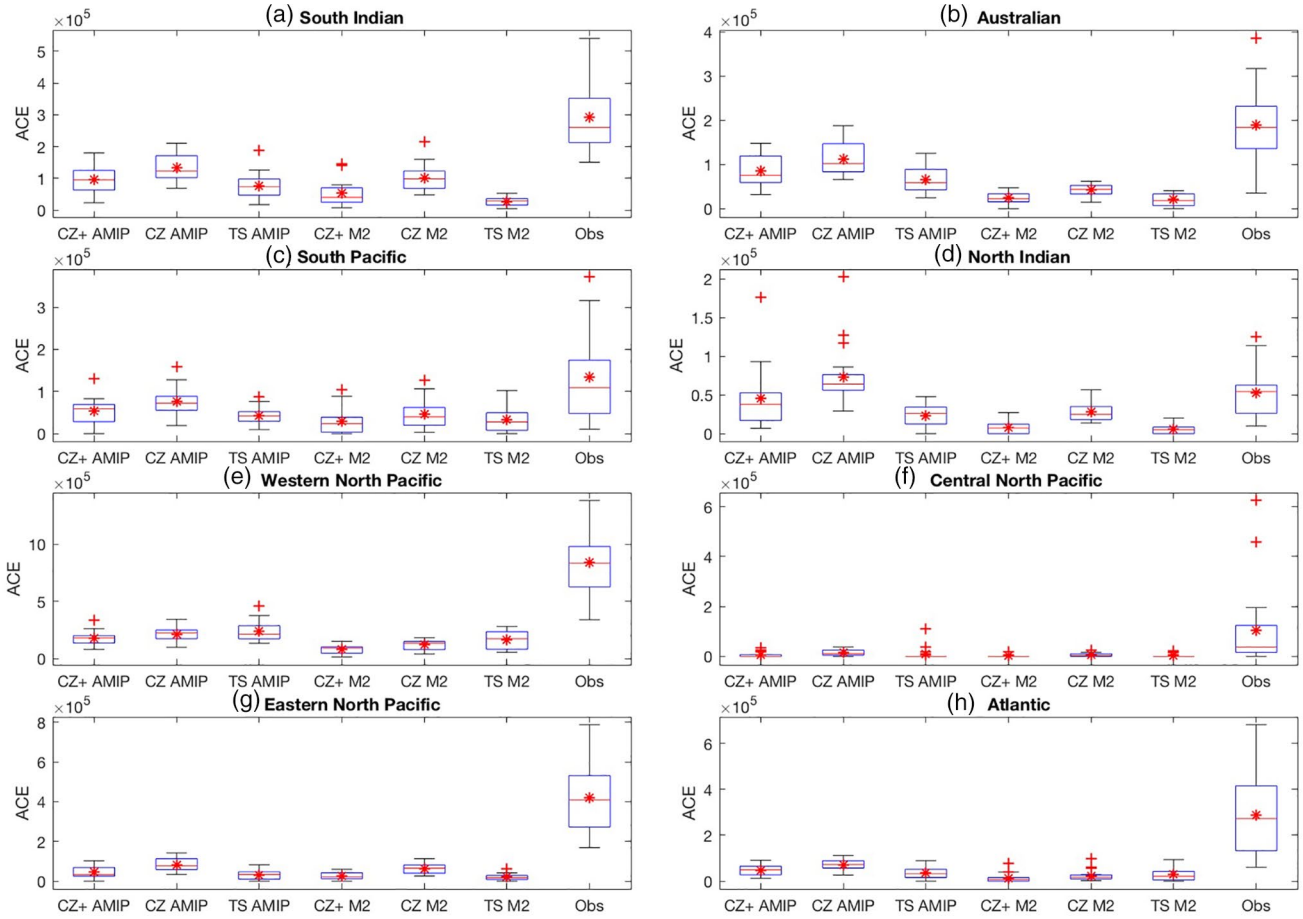

Figure 11. Boxplot of mean ACE (m^2/s^2) per year in eight basins from 1980 to 1999 in the AMIP, MERRA-2 (M2), and observed data sets for the CZ+, CZ, and TS tracking schemes. Boxplot definitions as in Figure 3. ACE, accumulated cyclone energy; CZ, Camargo-Zebiak; AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; ACE, accumulated cyclone energy; CZ, Camargo-Zebiak.

tracking scheme results in no significant difference with JRA-55 in the Western North Pacific basin. This is particularly notable given that all tracking schemes for MERRA-2 detected fewer storms than JRA-55 and ERA-Interim in the Western North Pacific (Table S2). The lack of significant difference in ACE values between TS MERRA-2 and the other reanalysis data sets can be explained by the intensity of wind speeds found in these reanalyses (Figure 14). TS MERRA-2 outperforms JRA-55 and ERA-Interim by generating, on average, more intense storms than the other two reanalyses. This is likely a result of the finer spatial resolution of the MERRA-2 reanalysis compared to JRA-55 and ERA-Interim (Dee et al., 2011; Ebita et al., 2011; Gelaro et al., 2017; Kobayashi et al., 2015). TS MERRA-2 has a peak in the 20–30 m/s range while JRA-55 and ERA-Interim peak at around 10–20 m/s. MERRA-2 also detects TCs with higher maximum windspeeds than ERA-Interim, which is in agreement with the results of Roberts et al. (2020). However, Roberts et al. (2020) find that, using the same tracking scheme for MERRA-2 and ERA-Interim, the two data sets both have the greatest number of storms in the 15–20 m/s range. The wind speed thresholds applied to MERRA-2 in this study reveal that, despite JRA-55 and ERA-Interim having more storms than TS MERRA-2, the TS MERRA-2 TCs are more intense, which leads the reanalysis to have similar ACE values despite differing in the average number and intensity of TCs. CZ+ and CZ MERRA-2 also produce more intense storms than JRA-55 and ERA-Interim. For the MERRA-2 CZ+ tracking scheme, where the NTC was significantly less than in the other reanalyses, the differences in TC intensity result in fewer cases where ACE is significantly less than JRA-55 and ERA-Interim. Because CZ is characterized by numerous weak storms, it does not follow this same pattern.

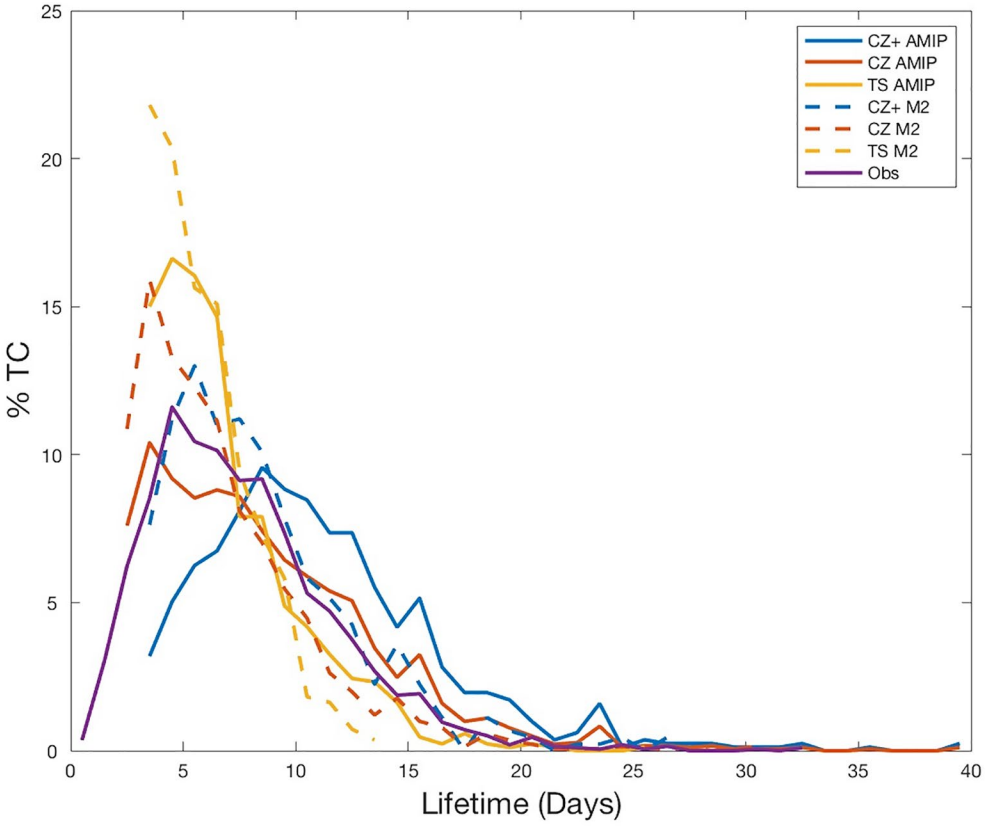

Figure 12. Lifetime (days) of TCs (%) in the AMIP, MERRA-2 (M2), and observed data sets for the CZ+, CZ, and TS tracking schemes from 1980 to 1999. TC, tropical cyclone; AMIP, Atmospheric Model Intercomparison Project; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; CZ, Camargo-Zebiak.

3.4. Environmental Fields

While the same SST values are used to force both AMIP and MERRA-2, the simulated environmental fields differ between the model and reanalysis (Figure 15). There is a strong latitudinal gradient in vorticity between AMIP and MERRA-2 in the Eastern North Pacific where MERRA-2 has stronger vorticity at lower latitudes than AMIP. A similar gradient, although not as intense, is found in the Western Pacific. AMIP also exhibits a higher potential intensity in the tropics while MERRA-2 finds a higher potential intensity in the midlatitudes. However, despite the higher values of potential intensity in the AMIP data set, AMIP does not simulate stronger TCs than MERRA-2 and the distribution of maximum wind speeds in the two data sets is not statistically different. AMIP also has larger values of vertical wind shear in the key regions of the Western North Pacific and the Eastern North Pacific and greater relative humidity in the Western North Pacific as well as in the Atlantic, particularly the Caribbean. High values of vertical wind shear in general inhibit the genesis and growth of TCs—particularly weaker storms (DeMaria, 1996). Yet AMIP generates more storms than MERRA-2 with comparable wind speeds. In contrast, the higher relative humidity could be potentially related to the greater number of storms found in these regions by AMIP compared to MERRA-2. Given the conflicting results, these climatological large-scale environmental variables do not help explain the differences in TC climatology for AMIP and MERRA-2. This is agreement with the results of Camargo et al. (2020), who found no relationship between the mean climatological environmental variables and the models mean TC climatology (mean NTC and ACE values globally and per hemisphere).

4. Conclusions

This study shows that there is no significant advantage in the performance of the MERRA-2 reanalysis compared to the free-running AMIP model in the representation of tropical cyclone climatology. Despite

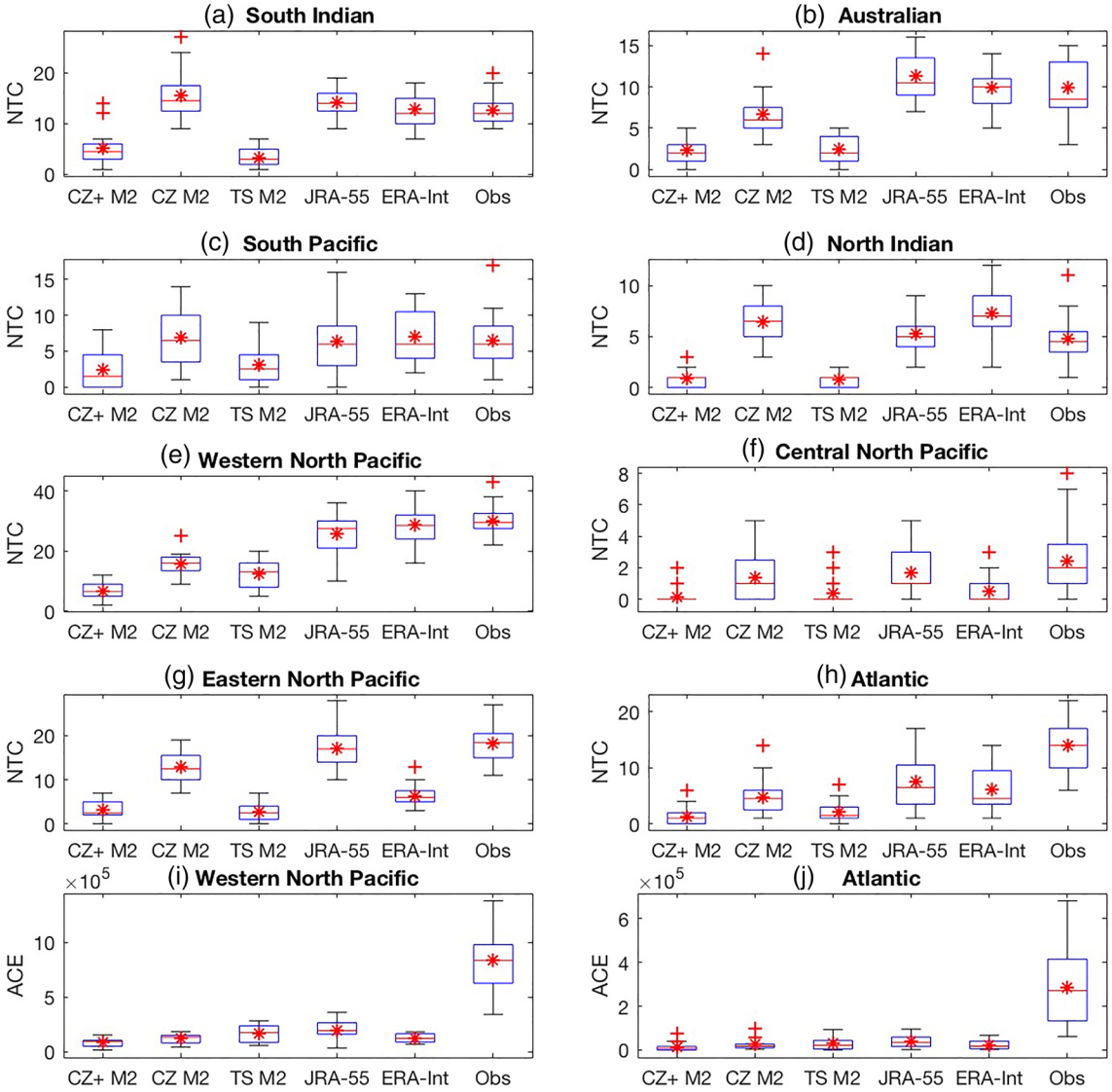

Figure 13. Boxplot of number of TCs in (a) South Indian, (b) Australian, (c) South Pacific, (d) North Indian, (e) Western North Pacific, (f) Central North Pacific, (g) Eastern North Pacific, and (h) Atlantic as well as ACE (m^2/s^2) in (i) Western North Pacific and (j) Atlantic for MERRA-2 (M2), JRA-55, ERA-Interim and observations for 1980–1999. The MERRA-2 data is shown for three tracking schemes—CZ+, CZ, and TS. Boxplot definitions as in Figure 3. TC, tropical cyclone; ACE, accumulated cyclone energy; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; CZ, Camargo-Zebiak; JRA, Japanese 55-Year Reanalysis.

the historical data assimilated into the reanalysis, MERRA-2 generates fewer TCs than AMIP and the model and reanalysis simulate TCs with similar intensity. MERRA-2 also offers no clear advantage over AMIP in reproducing the NTC monthly and interannual variability. Because AMIP and MERRA-2 use the exact same atmospheric model with a greater amount of data included in the data assimilation for MERRA-2, we expected a TC record that more closely mirrors the observations in MERRA-2 than AMIP, but that is not the case. AMIP and MERRA-2 have similar biases, e.g. in both cases the TCs are weaker than in observations and in models of similar resolution to AMIP (Moon et al., 2020b; Shaevitz et al., 2014; Wing et al., 2019). This is further highlighted by the comparison of MERRA-2 to other reanalysis data sets as, while MERRA-2 simulates storms with higher intensities than the other reanalyses, it struggles to generate enough storms over the 1980–1999 period compared to ERA-Interim and JRA-55, unless low thresholds are considered. This indicates there may be either an issue in the data assimilation scheme or that there is a fundamental issue with the model TC climatology characteristics associated with the model physics, with the caveat that our results are sensitive to the tracking algo-

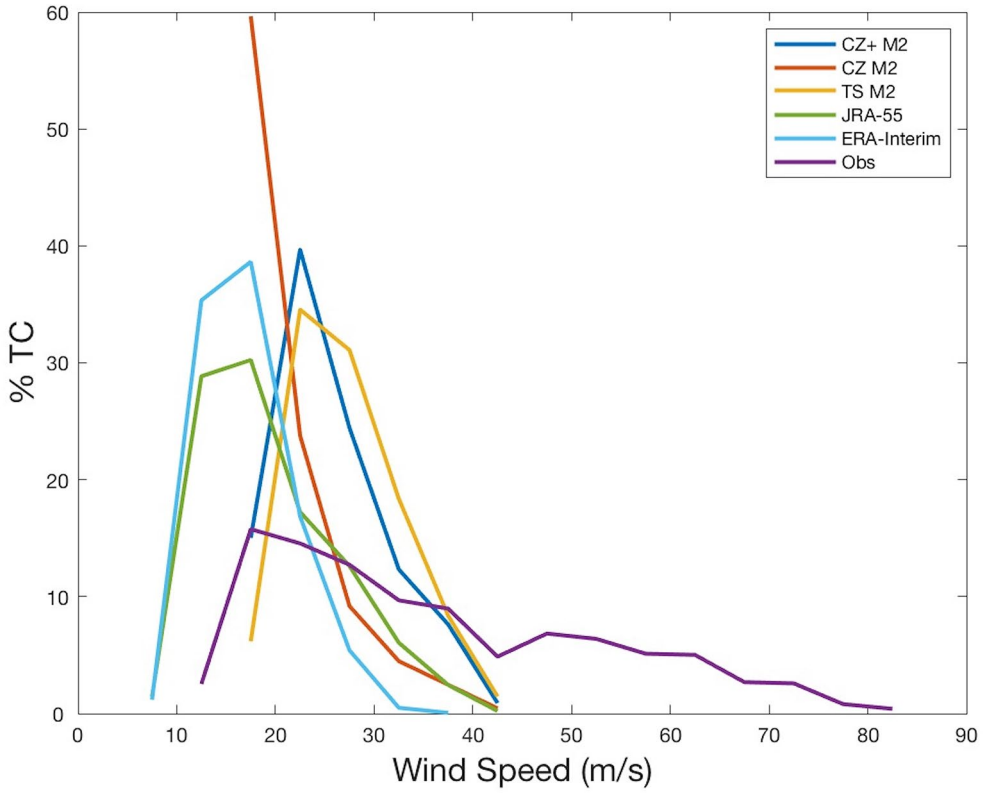

Figure 14. Maximum wind speed (m/s) of TCs (%) in the MERRA-2 (M2), JRA-55, ERA-Interim, and observed data sets from 1980 to 1999. The MERRA-2 data is shown for three tracking schemes – CZ+, CZ, and TS. CZ, Camargo-Zebiak; TC, tropical cyclone; MERRA2, Modern-Era Retrospective Analysis for Research and Applications; JRA, Japanese 55-Year Reanalysis.

rithm used in each case. More sophisticated diagnosis (e.g., Kim et al., 2018; Moon et al., 2020b; Wing et al., 2019) and/or specially designed simulations would be necessary to gain a better understanding of what is behind these results.

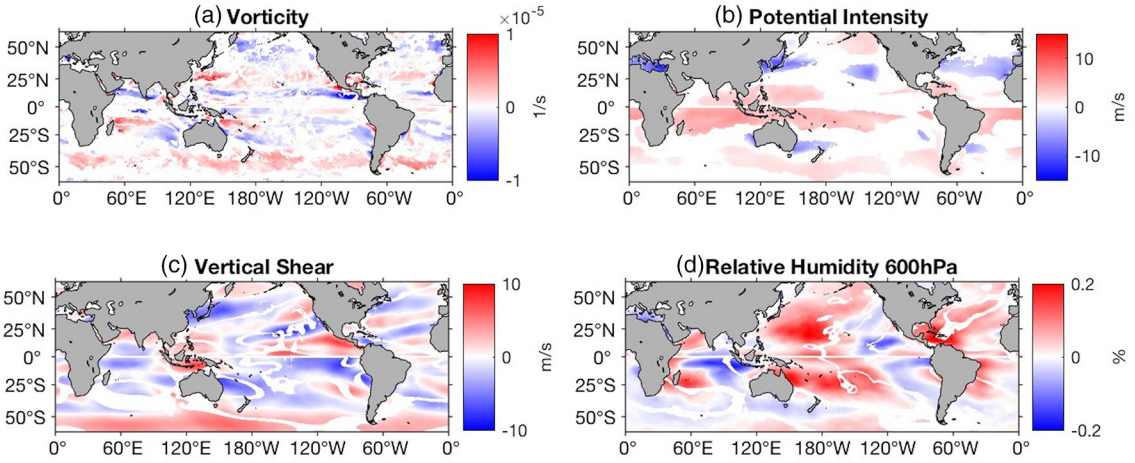

Figure 15. Difference in the environmental fields for (a) vorticity, (b) potential intensity, (c) vertical shear, and (d) relative humidity at 600 hPa between the NASA AMIP Model and MERRA-2 Reanalysis averaged over 1980–1999. Differences calculated by subtracting MERRA-2 values from AMIP environmental field. Values shown only where difference between AMIP and MERRA-2 is significant as calculated by a two-samples *t* test with a 95% confidence interval. MERRA2, Modern-Era Retrospective Analysis for Research and Applications; AMIP, Atmospheric Model Intercomparison Project.

Appendix A: CZ and TSTORMS Tracking Algorithms

While both CZ and TSTORMS were designed to track features with TC characteristics models and have many similarities, they have been developed by different groups, using different languages and the algorithms are clearly distinct.

As described in detail in Camargo and Zebiak (2002), there are various steps required in tracking TCs. First, in the detection part, the CZ algorithm (i) identifies all grid points with vorticity values above model-dependent vorticity threshold; (ii) requires that the maximum surface (or 1,000 hPa) winds in a region (size is resolution dependent) centered around the vorticity maximum exceed a model-dependent wind speed threshold; (iii) finds the SLP minimum on the same region; (iv) requires that the temperature anomaly averaged over the region and three pressure levels (300, 500, 700h Pa) exceeds a model-dependent temperature anomaly threshold; (v) requires that the local temperature anomaly in the region is greater at 300 hPa than at 850 hPa; (vi) requires that the mean wind speed averaged over the region at 850 hPa is larger than at 300 hPa. In the second step of the CZ algorithm, the identified grid points are then connected if they are less than a certain distance (~6°) from the center of the previous time step analyzed (for 6-hourly data). The center of the storm is defined as the minimum SLP in the region of interest. If the storm last at least 1.5 days (not necessarily consecutive), then the storm is considered a model TC. The third part of the algorithm is storm tracking, the first position and time in which the storm meets the detection criteria is taken and then using 850 hPa vorticity, the storm is tracked backward and forward in time, using a new center defined by the vorticity centroid and using only a relaxed vorticity threshold. The tracks are then compared to keep only one track belonging to each storm. The final TC tracks are obtained by smoothing the tracks that connected the vorticity centroid maxima.

What is described above is the standard Camargo and Zebiak algorithm. As mentioned above and in Wing et al. (2020), for these MERRA-2 and MERRA-AMIP simulations we imposed an additional requirement in order to exclude a high number of very weak short-living storms present, by considering only the TCs with surface winds that reached at least 15.2 m/s during their lifetime (CZ). In the case of CZ+ we additionally required that this wind threshold should be maintained for at least 3 days.

The TSTORMS algorithm has many similarities with the CZ algorithm, but there are also significant differences. We use here a summary of the description given in Zhao et al. (2009). The first step of the TSTORMS is the identification of storms: (i) identify relative vorticity maxima exceeding 1.6×10^{-4} s^{-1} within areas of 6° × 6° latitude and longitude; (ii) the local minimum of sea level pressure within a distance of 2° latitude or longitude is defined as the center of the storm; (iii) the local maximum surface winds is recorded; (iv) the local maximum temperature averaged between 300 and 500 hPa is defined as the center of the warm-core, which must be at least 1 °C warmer than the surrounding local mean and cannot be farther than 2° from the storm center. The second step is the storm tracking, which is done by connecting storm centers located in the following 6-h time step within a distance of 400 km. If there is more than one possibility, the one closest to the storm is chosen. And if there is still more than one possibility, preference is given to the storms that are to the west and poleward of the current location. To qualify as a storm, the track has to last at least 3 days and the lifetime maximum wind speed has to be greater than 15.2 m/s during at least 3 days (not necessarily consecutive).

Data Availability Statement

Monthly mean data from 10 ensemble members of the AMIP simulation and daily data from one ensemble member in which the TCs were tracked are available at https://portal.nccs.nasa.gov/datashare/gmao_m2amip/. Observed best-track data sets are available from the National Hurricane Center (https://www.nhc.noaa.gov/data/) and the Joint Typhoon Warning Center (https://www.metoc.navy.mil/jtwc/jtwc.html?best-tracks). The JRA-55 data are available at https://jra.kishou.go.jp/JRA-55/index_en.html#jra-55; and the ERA-Interim at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim. The ONI values are available at: https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php.

Acknowledgments
We thank NSF Grant OCE17-57602 for support of Zoe Aarons as a summer intern at the Lamont Summer Intern Program for Undergraduates at the Lamont-Doherty Earth Observatory of Columbia University during the summer of 2018. Suzana J. Camargo acknowledges partial funding from NOAA MAPP Grants NA15OAR4310095, NA16OAR4310079, and NA18OAR4310277. Suzana J. Camargo and Jeffrey D. O. Strong acknowledge support from NASA Grant 80NSSC17K0196. MERRA-2 data are available at http://disc.gsfc.nasa.gov/. The authors thank Tim Marchok and Yitian Qian for comments on an earlier version of this manuscript.

References

Bengtsson, L., Botzet, M., & Esch, M. (1995). Hurricane-type vortices in a general circulation model. *Tellus, Series A*, *47*, 175–196.

Bister, M., & Emanuel, K. A. (2002). Low frequency variability of tropical cyclone potential intensity: 1. Interannual to interdecadal variability. *Journal of Geophysical Research*, *107*(D24), 4801. https://doi.org/10.1029/2001JD000776

Bosilovich, M. G., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., et al. (2015). MERRA-2: Initial evaluation of the climate(NASA/TM-2015-14606, Vol. *43*). Greenbelt, MD:National Aeronautics and Space Administration, Goddard Space Flight Center. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160005045.pdf

Camargo, S. J (2013). Global and regional aspects of tropical cyclone activity in the CMIP5 models. *Journal of Climate*, *26*(24), 9880–9902. https://doi.org/10.1175/JCLI-D-12-00549.1

Camargo, S. J., Barnston, A. G., & Zebiak, S.E. (2005). A statistical assessment of tropical cyclone activity in atmospheric general circulation models. *Tellus, Series A*, *57*, 589–604. https://doi.org/10.3402/tellusa.v57i4.14705

Camargo, S. J., Giulivi, C. F., Sobel, A. H, Wing, A. A., Kim, D., Moon, Y., et al. (2020). Characteristics of model tropical cyclone climatology and the large-scale environment. *Journal of Climate*, *33*, 4463–4487. https://doi.org/10.1175/JCLI-D-19-0500.1

Camargo, S. J., Robertson, A. W., Barnston, A. G., & Ghil, M. (2008). Clustering of eastern North Pacific tropical cyclone tracks: ENSO and MJO effects. *Geochemistry, Geophysics, Geosystems*, *9*, Q06V05. https://doi.org/10.1029/2007GC001861

Camargo, S. J., & Sobel, A. H. (2005). Western North Pacific tropical cyclone intensity and ENSO. *Journal of Climate*, *18*(15), 2996–3006. https://doi.org/10.1175/JCLI3457.1

Camargo, S. J., Sobel, A. H., Barnston, A. G., & Klotzbach, P. J. (2010). The influence of natural climate variability, and seasonal forecasts of tropical cyclone activity. In J. C. L. Chan, & J. D. Kepert (Eds.), Global perspectives on tropical cyclones: From science to mitigation (2nd ed.). (Vol. 4, pp. 325–360). Singapore: World Scientific.

Camargo, S. J., & Wing, A. A. (2016). Tropical cyclones in climate models. *WIREs Climate Change*, *7*(2), 211–237. https://doi.org/10.1002/wcc373

Camargo, S. J., & Zebiak, S. E. (2002). Improving the detection and tracking of tropical cyclones in atmospheric general circulation models. *Weather and Forecasting*, *17*(6), 1152–1162. https://doi.org/10.1175/1520-0434(2002)017<1152:ITDATO>2.0.CO;2

Chu, J. H., Sampson, C. R., Levine, A. S., & Fukada, E. (2002). *The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945–2000* (Rep. NRL/MR/7540-02-16). Honolulu, HI: Joint Typhoon Warning Center.

Collow, A. B. M., Mahanama, S. P., Bosilovich, M. G., Koster, R. D., & Schubert, S. D. (2017). An evaluation of teleconnections of the United States in an ensemble of AMIP simulations with the MERRA-2 configuration of the GEOS atmospheric model (NASA/TM-2017-104606, Vol. 47). Greenbelt, MD: National Aeronautics and Space Administration, Goddard Space Flight Center.

Davis, C. A. (2018). Resolving tropical cyclone intensity in models. *Geophysical Research Letters*, *45*, 2082–2087. https://doi.org/10.1002/2017GL076966

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

DeMaria, M. (1996). The effect of vertical shear on tropical cyclone intensity change. *Journal of the Atmospheric Sciences*, *53*(14), 2076–2088. https://doi.org/10.1175/15200469(1996)053<2076:TEOVSO>2.0.CO;2

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., et al. (2011). The Japanese 55-year reanalysis "JRA-55": An interim report. *Scientific Online Letters on the Atmosphere*, *7*, 149–152. https://doi.org/10.2151/sola.2011-038

Emanuel, K. A. (1988). The maximum intensity of hurricanes. *Journal of Atmospheric Sciences*, *45*(7), 1143–1155. https://doi.org/10.1175/1520-0469(1988)045<1143:TMIOH>2.0.CO;2

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Hatsushika, H., Tsutsui, J., Fiorino, M., & Onogi, K. (2006). Impact of wind profile retrievals on the analysis of tropical cyclones in the JRA-25 reanalysis. *Journal of the Meteorological Society of Japan*, *84*, 891–905. https://doi.org/10.2151/jmsj.84.891

Hodges, K. I. (1994). A general method for tracking analysis and its application to meteorological data. *Monthly Weather Review*, *122*, 2573–2586.

Hodges, K., Cobb, A., & Vidale, P. L. (2017). How well are tropical cyclones represented in reanalysis datasets?. *Journal of Climate*, *30*(14), 5243–5264. https://doi.org/10.1175/JCLI-D-16-0557.1

Horn, M., Walsh, K., Zhao, M., Camargo, S. J., Scoccimarro, E., Murakami, H., et al. (2014). Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. *Journal of Climate*, *27*, 9197–9213. https://doi.org/10.1175/JCLI-D-14-00200.1

Kim, D., Moon, Y., Camargo, S. J., Wing, A. A., Sobel, A. H., Murakami, H., et al. (2018). Process-oriented diagnosis of tropical cyclones in high-resolution GCMs. *Journal of Climate*, *31*(5), 1685–1702. https://doi.org/10.1175/JCLI-D-17-0269.1

Kim, H., Lee, M.-I., Kim, S., Lim, Y.-K., Schubert, S. D., & Molod, A. (2020). Representation of tropical cyclones by the Modern-Era retrospective analysis for research and applications version 2. *Asia-Pacific Journal of Atmospheric Sciences*, *57*, 35–49. https://doi.org/10.1007/s13143-019-00169-y

Klaver, R., Haarsma, R., Vidale, P. L., & Hazeleger, W. (2020). Effective resolution in high resolution global atmospheric models for climate studies. *Atmospheric Science Letters*, *21*(4), e952. https://doi.org/10.1002/asl.952

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kousky, V. E., & Higgins, R. W. (2007). An alert classification system for monitoring and assessing the ENSO cycle. *Weather and Forecasting*, *22*(2), 353–371. https://doi.org/10.1175/WAF987.1

Landsea, C. W., & Franklin, J. L. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, *141*(10), 3576–3592. https://doi.org/10.1175/MWR-D-12-00254.1

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA toMERRA2. *Geoscientific Model Development*, *8*(5), 1339–1356. http://doi.org/10.5194/gmd-8-1339-2015

Moon, Y., Kim, D., Camargo, S. J., Wing, A. A., Reed, K. A., Wehner, M. F., & Zhao, M. (2020a). A horizontal resolution-dependent wind speed adjustment factor for tropical cyclones in climate model resolutions. *Geophysical Research Letters*, *46*, e2020GL087528. https://doi.org/10.1029/2020GL087528

Moon, Y., Kim, D., Camargo, S. J., Wing, A. A., Sobel, A. H., Murakami, H., et al. (2020b). Wind and thermodynamic structures of tropical cyclones in global climate models and their sensitivity to horizontal resolution. *Journal of Climate*, *33*(4), 1575–1595. https://doi.org/10.1175/JCLI-D-19-0172.1

Moorthi, S., & Suarez, M. J. (1992). Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models. *Monthly Weather Review*, *120*(6), 978–1002. https://doi.org/10.1175/1520-0493(1992)120<0978:RASAPO>2.0.CO;2

Murakami, H. (2014). Tropical cyclones in reanalysis data sets. *Geophysical Research Letters*, *41*, 2133–2141. https://doi.org/10.1002/2014GL059519

Murakami, H., & Sugi, M. (2010). Effect of model resolution on tropical cyclone climate projections. *Scientific Online Letters on the Atmosphere (SOLA)*, *6*, 73–76. https://doi.org/10.2151/sola.2010-019

Reale, O., Lau, W. K., Kim, K. M., & Brin, E. (2009). Atlantic tropical cyclogenetic processes during SOP-3 NAMMA in the GEOS-5 global data assimilation and forecast system. *Journal of the Atmospheric Sciences*, *66*(12), 3563–3578. https://doi.org/10.1175/2009JAS3123.1

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, *24*(14), 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1

Rienecker, M. M., Suarez, M. J., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., et al. (2008). The GEOS-5 Data Assimilation System - Documentation of Versions 5.0.1, 5.1.0, and 5.2.0. (NASA/TM-2008-104606, Vol. *27*). Greenbelt, MD:National Aeronautics and Space Administration Goddard Space Flight Center. Retrieved from https://gmao.gsfc.nasa.gov/pubs/docs/GEOS5_104606-Vol27.pdf

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B, Caron, L. P., et al. (2020). Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multimodel ensemble. *Journal of Climate*, *33*(7), 2557–2583. https://doi.org/10.1175/JCLI-D-19-0639.1

Schenkel, B. A., & Hart, R. E. (2012). An examination of tropical cyclone position, intensity and intensity life cycle within atmospheric reanalysis datasets. *Journal of Climate*, *25*(10), 3453–3475. https://doi.org/10.1175/2011JCLI4208.1

Shaevitz, D. A., Camargo, S. J., Sobel, A. H., Jonas, J. A., Kim, D., Kumar, A., et al. (2014). Characteristics of tropical cyclones in high-resolution models in the present climate. *Journal of Advances in Modeling Earth Systems*, *6*(4), 1154–1172. https://doi.org/10.1002/2014MS000372

Strachan, J., Vidale, P. L., Hodges, K., Roberts, M., & Demory, M.-E. (2013). Investigating global tropical cyclone activity with a hierarchy of AGCMs: the role of model resolution. *Journal of Climate*, *26*, 133–152. https://doi.org/10.1175/JCLI-D-12-00012.1

Tory, K. J., Chand, S. S., Dare, R. A., & McBride, J. L.. (2013). The Development and Assessment of a Model-, Grid-, and Basin-Independent Tropical Cyclone Detection Scheme. *Journal of Climate*, *26*, (15), 5493–5507. http://dx.doi.org/10.1175/jcli-d-12-00510.1.

Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, *10*(3), 1069. https://doi.org/10.5194/gmd-10-1069-2017

Vitart, F., Anderson, J. L., & Stern, W. F. (1997). Simulation of Interannual Variability of Tropical Storm Frequency in an Ensemble of GCM Integrations. *Journal of Climate*, *10*, (4), 745–760. https://doi.org/10.1175/1520-0442(1997)010<0745:soivot>2.0.co;2

Walsh, K. J. E., Fiorino, M., Landsea, C. W., & McInnes, K. L. (2007). Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *Journal of Climate*, *20*(10), 2307–2314. https://doi.org/10.1175/JCLI4074.1

Wehner, M., Prabhat, K. A., Stone, D., Collins, W. D., & Bacmeister, J. (2015). Resolution dependence on future tropical cyclone projections of CAM5.1 in the U.S. CLIVAR Hurricane Working Group idealized configurations. *Journal of Climate 28*(10), 3905–3925. https://doi.org/10.1175/JCLI-D-14-00311.1

Wing, A. A, Camargo, S. J., Sobel, A. H., Kim, D., Moon, Y., Murakami, H., et al. (2019). Moist static energy budget analysis of tropical cyclone formation and intensification in high-resolution climate models. *Journal of Climate*, *32*(18), 6071–6095. https://doi.org/10.1175/JCLI-D-18-0599.1

Zhao, M., Held, I. M., Lin, S. J., & Vecchi, G. A. (2009). Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *Journal of Climate*, *22*(24), 6653–6678. https://doi.org/10.1175/2009JCLI3049.1