# An Assessment of a Model-, Grid-, and Basin-Independent Tropical Cyclone Detection Scheme in Selected CMIP3 Global Climate Models

K. J. TORY, S. S. CHAND, R. A. DARE, AND J. L. MCBRIDE

*Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Victoria, Australia*

(Manuscript received 30 July 2012, in final form 10 December 2012)

ABSTRACT

A novel tropical cyclone (TC) detection technique designed for coarse-resolution models is tested and evaluated. The detector, based on the Okubo–Weiss–Zeta parameter (OWZP), is applied to a selection of Coupled Model Intercomparison Project, phase 3 (CMIP3), models [Commonwealth Scientific and Industrial Research Organisation Mark, version 3.5 (CSIRO-Mk3.5); Max Planck Institute ECHAM5 (MPI-ECHAM5); and Geophysical Fluid Dynamics Laboratory Climate Model, versions 2.0 (GFDL CM2.0) and 2.1 (GFDL CM2.1)], and the combined performance of the model and detector is assessed by comparison with observed TC climatology for the period 1970–2000. Preliminary TC frequency projections are made using the three better-performing models by comparing the detected TC climatologies between the late twentieth and late twenty-first centuries. Very reasonable TC formation climatologies were detected in CSIRO-Mk3.5, MPI-ECHAM5, and GFDL CM2.1 for most basins, with the exception of the North Atlantic, where a large underdetection was present in all models. The GFDL CM2.0 model was excluded from the projection study because of a systematic underdetection in all basins. The above detection problems have been reported in other published studies, which suggests model rather than detector limitations are mostly responsible. This study demonstrates that coarse-resolution climate models do in general produce TC-like circulations with realistic geographical and seasonal distributions detectable by the OWZP TC detector. The preliminary projection results are consistent with the published literature, based on higher-resolution studies, of a global reduction of TCs between about 6% and 20%, with a much larger spread of results (about +20% to −50%) in individual basins.

## 1. Introduction

The enormous social and economic impact of tropical cyclones (TCs) around the world has led to strong interest in potential changes in TC behavior in a warming climate. The potentially serious consequences of more frequent or more intense TCs have created high demand for regionally specific projections of future TC behavior. Given the possibility of significant preemptive investment in adaptation programs, it is especially important that the uncertainty in the projections is well understood and well communicated. It is difficult to quantify the uncertainty regarding the ability of models to reproduce accurate climatologies of TC-like circulations and the ability of TC detectors to identify these circulations accurately (e.g., Tory et al. 2013a). While this uncertainty is largely unquantifiable at present, confidence in TC projections has been improved by considering the similarity of results from multiple models (e.g., Camargo et al. 2005; Emanuel et al. 2008; Knutson et al. 2010; Walsh et al. 2013). It follows that similar improved confidence could be gained from the application of different TC detectors to these models. In this paper, we document, in unprecedented detail, the results of a new and alternative TC detection technique (Tory et al. 2013a,b).

Global climate models (GCMs) have been generating TC-like circulations for decades (e.g., Manabe et al. 1970). Subsequent studies have demonstrated that some GCMs are capable of reproducing reasonably realistic TC structure, climatology, and interannual variability (e.g., Bengtsson et al. 1982, 1995; Wu and Lau 1992; Vitart et al. 1997; Camargo et al. 2005; Walsh et al. 2013; Strachan et al. 2013), despite even the highest-resolution GCMs not being able to resolve TC structures at

*Corresponding author address:* Kevin Tory, Centre for Australian Weather and Climate Research, GPO Box 1289, Melbourne, VIC 3001, Australia.
E-mail: k.tory@bom.gov.au

DOI: 10.1175/JCLI-D-12-00511.1

observed scales. While this resolution issue remains there will always be some subjectivity in determining which TC-like circulation should be classified as a TC and which should not. As discussed in Tory et al. (2013a), traditional TC detection techniques require threshold values of specific TC characteristics (e.g., lower troposphere relative vorticity, warm core, wind speed) to be met for a TC to be declared. In such techniques the thresholds have been adjusted to give the best match between the GCM and observed TC climatologies. These adjusted thresholds form the definition of the model TC, that is, they define the GCM TC. Both the GCM TC definition and the climatology of GCM TC-like circulations contain some error. Tuning to best reproduce observed TC climatology helps minimize the combined error, but in doing so, the GCM TC definition is adjusted to compensate for any model- or basin-specific biases or idiosyncrasies, which imposes a model- or basin-specific bias to the GCM TC definitions.

Arguably, the greatest model bias is associated with the GCM grid resolution, which effectively imposes a limit on the horizontal gradients of wind speed. As a result, GCM TC-like circulations tend to be much larger than observed TCs and have weaker maximum wind speeds. Given that observed TCs are largely defined by their maximum wind speed, the lower wind speeds of GCM TC-like circulations require a revised wind speed threshold to define the GCM TC in order to account for grid-resolution dependency. Walsh et al. (2007) proposed a semiobjective[1] technique for determining the TC wind speed threshold as a function of grid resolution, which, when applied across model basins and models, helps minimize the above-mentioned error compensation. However, many models still show varying skill in reproducing TC climatologies from basin to basin. Additionally, because some thresholds were adjusted to best reproduce TC climatology, it is not clear whether the GCM definition is inappropriate in the underperforming basin or whether the GCM over- or underproduces TC-like circulations.

Bengtsson et al. (2007) avoided grid-dependent thresholds by tracking warm-cored circulations that exceed an 850-hPa relative vorticity threshold applied to data degraded to a common coarse resolution. However, this detection method did not distinguish between tropical depressions and TCs (Bengtsson et al. 2007, p. 402). Strachan et al. (2013) refined the warm-core criteria to reduce the number of tropical depressions included in the detections.

In their detection scheme, Tory et al. (2013a,b) attempted to avoid the model resolution bias by using only large-scale parameters, resolvable in contemporary GCMs. They chose also to avoid the above-mentioned error compensation by developing and tuning their detection technique in reanalysis data, which is applied directly to a selection of Coupled Model Intercomparison Project, phase 3 (CMIP3), GCMs in this paper. While the Okubo–Weiss–Zeta parameter (OWZP) detection technique showed good performance (Tory et al. 2013a), it was not possible (and it was not their intention) to judge which detection scheme performed the best. Instead, the authors suggested that all methods could be used in parallel. The very different philosophies behind each method suggest their respective strengths and weaknesses should differ, and together they should help reduce the uncertainty of TC projections. In this paper we document the performance of the OWZP TC projection technique in four CMIP3 GCMs and consider the projected changes in TC frequency in each model.

The remainder of the paper is structured as follows. Section 2 outlines various datasets used in this study. Section 3 summarizes the OWZP TC detection and tracking methodology. In section 4 the OWZP TC detection results are presented for the current climate from a selection of CMIP3 models, and in section 5 these results are compared with equivalent future climate results to provide projections into possible changes in TC frequency in a warming climate. The results are summarized in section 6.

## 2. Models and observational data

The GCMs assessed in this paper were obtained from the World Climate Research Programme (WCRP) CMIP3 multimodel dataset described by Meehl et al. (2007). The WCRP CMIP3 dataset comprises 24 ocean–atmosphere coupled models that were used for the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC). Since the main emphasis of this study is to demonstrate the performance of the OWZP TC detection scheme in climate models rather than to make future projections of TC activity, we applied the detection scheme to the four models with daily data[2] readily available to us at the time (see Table 1 for the names and host institutes of these models; model descriptions and associated references

[1] While the technique can be applied objectively, subjectivity is unavoidable in the design, and subjective adjustments have been made in its application [as discussed in the introduction of Tory et al. (2013)].

[2] The data are daily averaged, which results in smeared circulations elongated in the direction of storm motion.

TABLE 1. List of CMIP3 models investigated in this study.

| Model | Model institution (country) | Atmosphere resolution | Ocean resolution |
|---|---|---|---|
| Commonwealth Scientific and Industrial Research Organisation Mark, version 3.5 (CSIRO-Mk3.5) | CSIRO Marine and Atmospheric Research (Australia) | 1.9° × 1.9° | 0.8° × 1.9° |
| Max Planck Institute ECHAM5 (MPI-ECHAM5) | Max Plank Institute (Germany) | 1.9° × 1.9° | 1.5° × 1.5° |
| Geophysical Fluid Dynamics Laboratory Climate Model, version 2.0 (GFDL CM2.0) | Geophysical Fluid Dynamics Laboratory (United States) | 2.0° × 2.5° | 0.3°–1° × 1° |
| Geophysical Fluid Dynamics Laboratory Climate Model, version 2.1 (GFDL CM2.1) | Geophysical Fluid Dynamics Laboratory (United States) | 2.0° × 2.5° | 0.3°–1° × 1° |

can be found at http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php). The late-twentieth-century simulations (between 1970 and 2000) are used in this study to assess the TC climatology as determined by the OWZP TC detector. While not the main purpose of this paper, we comment on projected changes in TC frequency using the OWZP TC detector. Because of relatively small frequency changes, only the most extreme greenhouse gas scenario (A2) and the longest projection time period are considered (i.e., the late twenty-first century, 2081–2100). A more detailed TC frequency projection study is currently underway using all available CMIP5 data.

For current climate TC verification, the International Best Track Archive for Climate Stewardship (IBTrACS) database is used (Knapp et al. 2010; http://www.ncdc.noaa.gov/oa/ibtracs). The IBTrACS data are a comprehensive compilation of quality-controlled global TC best-track data sourced from various meteorological organizations and agencies around the world. The IBTrACS database contains the storm location, maximum sustained wind, and central pressure, available at 6-hourly intervals (i.e., 0000, 0600, 1200, and 1800 UTC), during the lifetime of a TC. In this study, only data at 24-h intervals (0000 UTC) are considered to be consistent with the temporal resolution of the climate model data. We choose to define a TC as any system in the IBTrACS database that reached the 10-min maximum sustained wind speed of at least 17 m $s^{-1}$ at any 0000 UTC time during its lifetime. The genesis location is assumed to be the position where the wind speed first reaches 17 m $s^{-1}$. To compare the GCM TC detections with the most reliable observational TC data, only the twentieth-century satellite era (i.e., 1970–2000) is considered in this paper.

## 3. Detection and tracking methodology

The OWZ variable that underpins the OWZP TC genesis parameter was introduced in Tory et al. (2013b), and the detection and tracking algorithms were described in detail in Tory et al. (2013a,b). A brief overview of the methodology is given here. As mentioned in the introduction, the philosophy behind the OWZP TC detection method differs from conventional TC detectors that attempt to directly identify TC circulations. The OWZP method instead attempts to identify the large-scale environment in which a TC will form, and a TC is declared if the environment remains favorable for a sufficient period of time. OWZ is used to identify regions of near solid-body rotation in the low to mid-troposphere [which Tory et al. (2013b) argue is necessary for TC formation]. It takes the place of relative vorticity used in other detection schemes and genesis parameters.

As a TC declaration is dependent on favorable formation conditions persisting for a number of model time periods, tracking of precursor disturbances is required just to identify TCs. The OWZP TC detection is a two-step process, in which 1) circulations with the dynamic potential to support TC formation are identified and tracked before 2) TCs are determined by a more rigorous dynamic and thermodynamic assessment of the storm tracks over a set period of time. For a circulation to be included in a storm track, a set of ''initial'' thresholds must be satisfied at two or more neighboring grid points (step 1). Then a set of ''core'' thresholds must be met at three consecutive time periods somewhere in the storm track before a TC is declared (step 2). The two sets of thresholds used in this paper are provided in Table 1 of Tory et al. (2013a), labeled ''Initial'' and ''Criterion 1.''

## 4. Evaluation of model TC climatology

The thresholds and other conditions used in the OWZP TC detection method were developed and tuned in European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) reanalyses (Tory et al. 2013a). They are applied directly to the CMIP3 models in this paper without any further adjustment. This ensures that the detection method is

independent of the GCMs.[3] While this approach minimizes error compensation between detection and model errors, which occurs when the detection method is tuned to best reproduce TC climatology, it cannot separate the detection error from the model error. Thus, our analysis of the CMIP3 TC climatologies using the OWZP TC detection method reflects the combined performance of the models and detection method. TC formation locations, TC frequency, and seasonal and interannual variability are considered in the performance assessment of the following CMIP3 models: CSIRO-Mk3.5, MPI-ECHAM5, GFDL CM2.0, and GFDL CM2.1 (see Table 1 for model details).

### *a. Geographical distribution of TC genesis positions and frequency*

The geographical distribution of observed and modeled TC formation locations are shown in Figs. 1 and 2 for the period 1970–2000. Each formation position is denoted by a black dot in Fig. 1, and for ease of comparison in the higher-density formation regions, smoothed spatial density maps are presented in Fig. 2. The spatial density maps are obtained by first binning TCs into a regular 2.5° × 2.5° grid and then smoothing the gridded data by generating probability density functions (PDFs) computed from anisotropic Gaussian functions (e.g., Ramsay et al. 2008; Chand and Walsh 2009).

As can be seen in Figs. 1a and 2a, the most concentrated area of TC genesis globally is in the eastern North Pacific (ENP) basin around 15°N, 110°W. In the western North Pacific (WNP), genesis is spread over a large area, with maxima occurring around the Philippines and extending into the central Pacific. A secondary maximum is also observed in the South China Sea. TCs in the north Indian (NI) basin mainly form in the Bay of Bengal region. In the North Atlantic (NA) basin, TC genesis locations are widely distributed over a large area with predominant formations in the Gulf of Mexico and the region spanning part of the Caribbean Sea and the tropical NA basin (the so-called main development region: 6°–18°N, 20°–60°W). In the South Pacific (SP) basin, TCs usually form in the Coral Sea region (between 150°E and 180°) with a rapid decrease eastward. In the Australian region, maximum genesis density occurs in the Gulf of Carpentaria and off the northwest coast in the Timor Sea. TCs in the south Indian (SI) basin are spread across the region with maximum formation around about 70°E longitude.

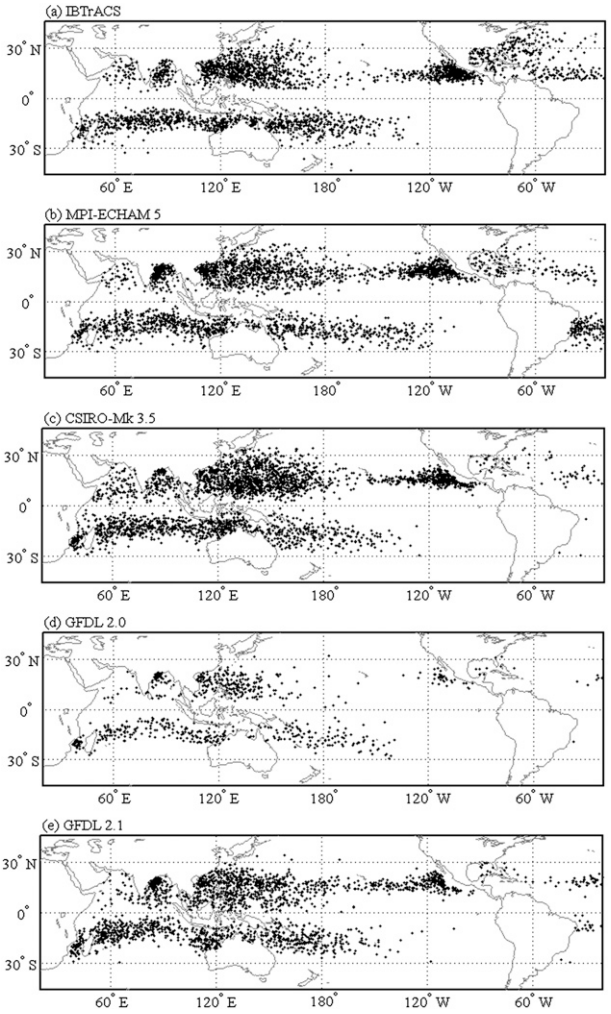


FIG. 1. Genesis positions of TCs (a) observed in IBTrACS and those detected using the OWZP technique in (b) MPI-ECHAM5, (c) CSIRO-Mk3.5, (d) GFDL CM2.0, and (e) GFDL CM2.1 over the period 1970–2000.

The observed spatial patterns are generally well replicated in the four CMIP3 models (Figs. 1b–e, 2b–e), with OWZP detections in the two somewhat higher-resolution models (1.9° × 1.9°, MPI-ECHAM5 and CSIRO-Mk3.5; Figs. 1b,c and 2b,c) reproducing the formation density reasonably well. However, a closer inspection on regional scales highlights biases and deficiencies. The mean annual TC formation frequency for each 2.5° latitude and longitude band is given in Fig. 3. In MPI-ECHAM5 there is a poleward bias in the detected formation maxima (Fig. 3a) with an overdetection of formation particularly in the Bay of Bengal (80°–90°E) in the NI basin (Figs. 1b, 2b, and 3c) and east of about 130°W in the SP basin (Figs. 1b, 2b, and 3b). Also evident is a marked deficiency of TC detections in the NA basin, particularly poleward of about 25°N (Figs. 1b, 2b).

[3] It was necessary to use relative humidity and specific humidity on the 925-hPa level, instead of 950 hPa because that level was not available in the daily CMIP3 pressure-level data.

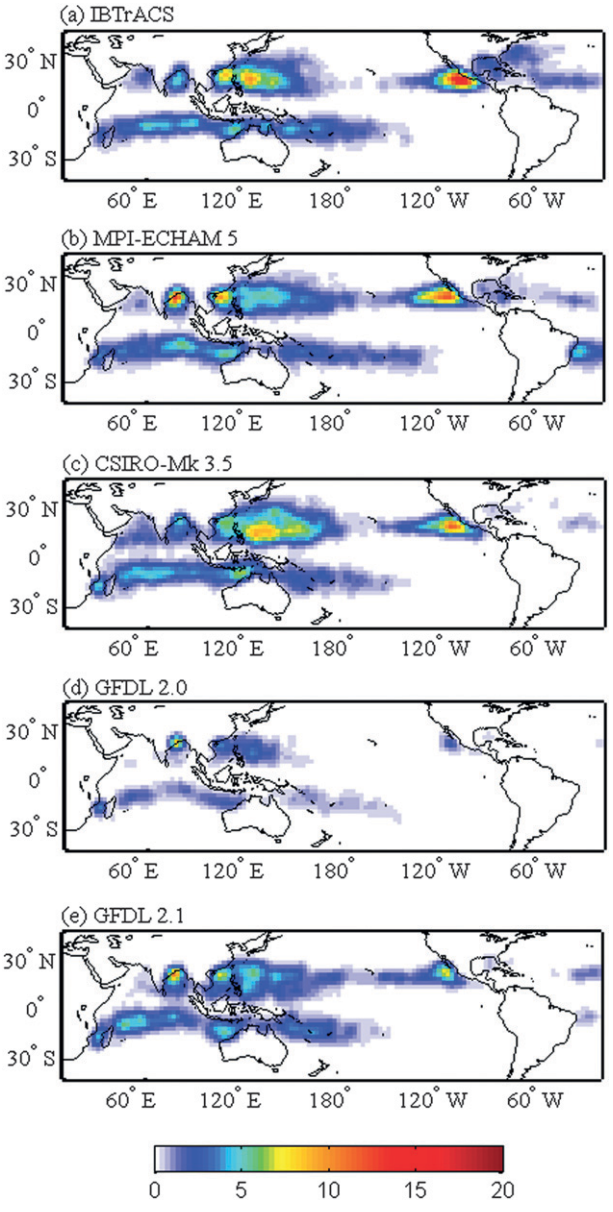


FIG. 2. Anisotropic Gaussian density distribution of TC genesis positions in a 2.5° × 2.5° grid box for TCs (a) observed in IBTrACS and OWZP detected in (b) MPI-ECHAM5, (c) CSIRO-Mk3.5, (d) GFDL CM2.0, and (e) GFDL CM2.1 over the period 1970–2000.

South Atlantic TC formation is also detected in MPI-ECHAM5, where TCs rarely form. Elsewhere, the TC formation distribution compares well with the observations.

The TC formation detections in CSIRO-Mk3.5 are similar to MPI-ECHAM5, with the main difference being an even greater deficit of NA TCs (Figs. 1c, 2c). The WNP basin is slightly overdetected in CSIRO-Mk3.5 in both formation density and areal spread (Fig. 2c), and a slight underdetection is evident for the ENP basin (Fig. 2c, 3c; 120°–100°W). While MPI-ECHAM5 appears to have a slightly better representation of TCs in the ENP and NA basins compared to CSIRO-Mk3.5, the latter model is free from the latitudinal bias evident in MPI-ECHAM5 (Fig. 3a).

Fewer TCs were detected in the two slightly coarser models (2.0° × 2.5°, GFDL CM2.0 and GFDL CM2.1), with a significant underdetection in the GFDL CM2.0 model (Figs. 1d, 2d). This result is consistent with Walsh et al. (2013), who noted a difference in detection rate of more than a factor of 2, despite the models having the same resolution and convection scheme. While the GFDL CM2.1 model (Figs. 1e, 2e) does not suffer from the same underdetection issues as GFDL CM2.0, a notable deficiency in the formation density is evident in the ENP and NA basins. The formation densities in the Southern Hemisphere basins generally compare well with observations, although a handful of South Atlantic TCs have been detected.

With the exception of GFDL CM2.0, TC formation locations are in general realistically distributed in all basins except for the NA. Other studies that used lower-resolution models (e.g., Camargo et al. 2005) have also noted similar deficiencies in the NA basin. This could be because of an inability to adequately represent African easterly waves that provide the majority of TC precursors in the NA basin (e.g., Thorncroft and Hodges 2001). More recent investigations using very fine resolution models (e.g., Chauvin et al. 2006; Zhao et al. 2009) have noted a considerable improvement in model TCs over the NA basin. On the other hand, our preliminary CMIP5 studies have shown that the CSIRO-Mk3.6 model, with the same grid spacing as CSIRO-Mk3.5, produces more realistic numbers of TCs in the NA basin.

Excluding the GFDL CM2.0 model, these results demonstrate that the GCMs generally reproduce observed TC climatology reasonably well as detected by the OWZP detection technique. As the OWZP was developed independently of the GCMs, we assume any model and detector error cancellation is small, and the good replication of the observed TC climatology indicates good performance of both the detector and model. Next, we further evaluate the GCM and OWZP TC detector performance with more quantitative comparisons between modeled and observed TCs.

### b. *Climatological means*

The mean number of model-detected and observed TCs per year (NTCs) for the globe, the two hemispheres, and each ocean basin over the period 1970–2000 is shown in Fig. 4. The boundaries for the six TC basins used in this study are defined in Fig. 4a. Statistical significance tests for evaluating the difference between the

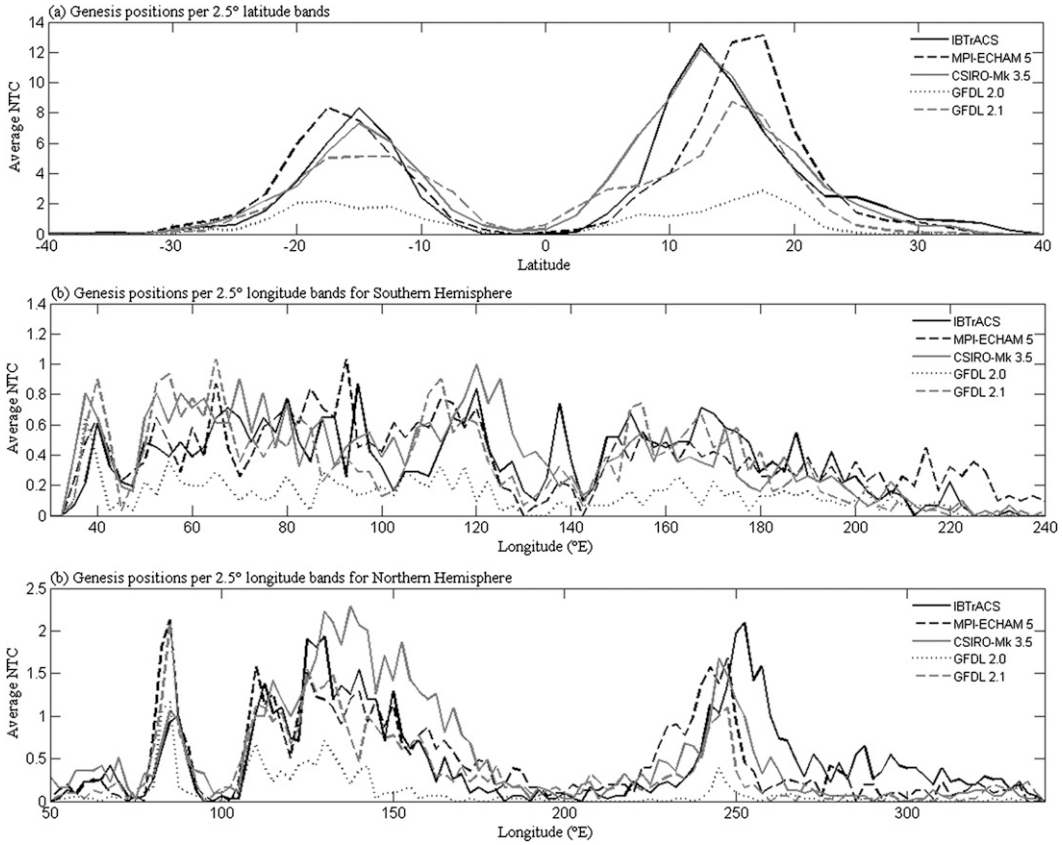


FIG. 3. Distribution of TC genesis positions expressed in terms of a mean annual frequency for each 2.5° (a) latitude band, (b) Southern Hemisphere longitude band, and (c) Northern Hemisphere longitude band. Solid black line represents observed TCs.

mean values of TCs in observations and model data are conducted using the bootstrap resampling method (Efron and Tibshirani 1991). This method is based on the assumption that the empirical density function associated with the 30 years of data is a reasonable estimate of the unknown population density function. In this study, the two sets of data are resampled separately 1000 times and the associated mean statistics for each bootstrap sample are calculated, giving a total of 1000 bootstrap distribution sample means for each set of data. The two distributions are then compared using their respective 95% confidence intervals. If bootstrap confidence intervals for the two distributions overlap, then the means of the two sets of data are considered statistically similar.

On average, about 88 TCs formed per year globally from 1970 to 2000 (Fig. 4b). Of these, about 34% formed in the Southern Hemisphere (Fig. 4c) and 66% in the Northern Hemisphere (Fig. 4d). The observed climatological mean NTCs are well reproduced by the OWZP TC detections in the CSIRO-Mk3.5 and MPI-ECHAM5 models, both globally and in the two hemispheres (Figs. 4b–d). The GFDL CM2.1 model detections better represent the observed climatology in the Southern Hemisphere than in the Northern Hemisphere. In the Northern Hemisphere TCs are underdetected in the WNP, ENP, and NA basins. Because of the substantial underdetection of TCs in the GFDL CM2.0 model in all TC basins, this model will be mostly excluded from the following discussion.

At a basin-wide scale, the model-detected TC numbers are statistically similar to the observed TC numbers (at the 95% significance levels) in the SP basin (Fig. 4e) and in the SI basin (Fig. 4f). In the NI basin (Fig. 4g), only the number of GFDL CM2.1 model-detected TCs is statistically similar to that observed, while NI TCs are overdetected in CSIRO-Mk3.5 and MPI-ECHAM5. An overdetection might be expected here given a similar overdetection by the OWZP detector when applied to the ERA-Interim reanalyses (Tory et al. 2013a). In the WNP basin (Fig. 4h), the MPI-ECHAM5 and GFDL CM2.1 detection distributions appear to reproduce well the observed climatology, while there is an overdetection in CSIRO-Mk3.5. The numbers of CSIRO-Mk3.5 and

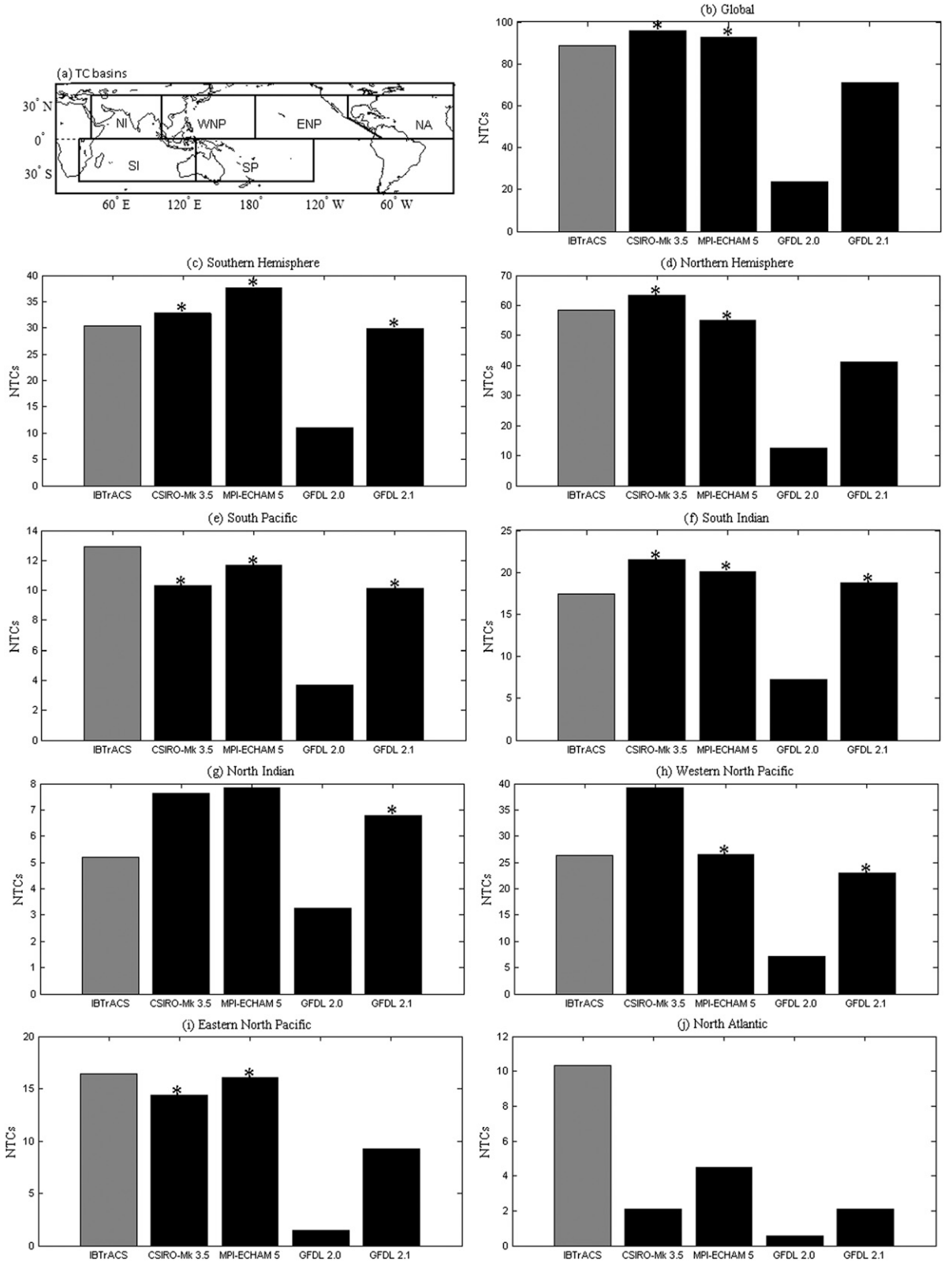


FIG. 4. (a) Boundaries for the six TC basins. (b)–(j) Comparison of observed (gray) and model mean TC frequency globally, hemispherically, and in the individual basins. Asterisk (plus) indicates 95% (90%) statistical significance level.

TABLE 2. Percentage of total global TCs present in each basin and hemisphere. The global ratio of observed to model-detected TCs is given in the last column. Basin acronyms are as described in text; SH is Southern Hemisphere and NH is Northern Hemisphere.

| | SP | SI | NI | WNP | ENP | NA | SH | NH | Ratio |
|---|---|---|---|---|---|---|---|---|---|
| OBS | 14.6 | 19.6 | 5.9 | 29.7 | 18.6 | 11.6 | 34.2 | 65.8 | 1.00 |
| CSIRO-Mk3.5 | 10.4 | 23.5 | 8.0 | 40.8 | 15.0 | 2.2 | 34.0 | 66.0 | 1.08 |
| MPI-ECHAM5 | 12.7 | 21.7 | 8.5 | 28.7 | 17.4 | 4.9 | 40.5 | 59.5 | 1.04 |
| GFDL CM2.0 | 15.8 | 30.8 | 13.9 | 30.4 | 6.2 | 2.5 | 47.1 | 52.9 | 0.26 |
| GFDL CM2.1 | 14.2 | 26.4 | 9.5 | 32.4 | 13.1 | 2.9 | 42.0 | 58.0 | 0.80 |

MPI-ECHAM5 model-detected TCs are statistically similar to those observed in the ENP basin (Fig. 4i). The systematic TC underdetection in the NA basin noted above is clearly evident in Fig. 4j.

It is likely that some of the difference between the observed and detected TC climatology is due to a systematic bias in each combination of model and detection scheme. If the differences can be largely explained by such a bias, then it would be expected that the percentage of basinwide TC numbers to global TC numbers should match between the models and observations. These percentages are listed in Table 2. The relative hemispheric-detected TC distribution in the CSIRO-Mk3.5 model differs from the observed distributions by only 0.2%. No such close agreement exists for the other models, which suggests detected and observed climatology differences are not dominated by systematic biases. A comparison of the observed and modeled TC percentages within individual basins leads to the same conclusion for all models. The highest percentage of TCs are observed in the WNP basin at about 30%, which is similar to that found in all models except CSIRO-Mk3.5, which has nearly 40% detected in that basin. The lowest observed TC percentage is in the NI basin with about 6%. All models demonstrate a higher NI basin percentage than this amount, which perhaps should be expected given the overprediction of the detector in that basin (mentioned above). However, the smallest percentage in all models is the NA basin rather than the NI. The 11.6% of global TCs observed in the NA basin is not even closely matched by the models, which range from about 2% to 5%. This further highlights the difficulty of these models in matching the NA basin TC climatology. The relatively small observed fraction of NA TCs means this significant NA underdetection does not have a severe impact on the global or Northern Hemisphere climatology.

### c. Seasonal variability

In this section we compare the model-detected to observed TC seasonal variability to test how well climate models capture TC formation frequency from month to month throughout the year. Figure 5 shows the average number of model-detected and observed TCs in the six TC basins around the globe. On the whole, the model-detected TCs capture the phase of seasonal variability quite well. The main difference in these plots is the amplitude associated with biases mentioned in the previous subsection. The TC season in the SP (Fig. 5a) and SI (Fig. 5b) basins, for example, extends from November to April, with peak activity between January and March. However, on rare occasions, and possibly tied to the El Niño–Southern Oscillation (ENSO), TC formations in these basins can be observed as early as October and as late as June (e.g., Ramage and Hori 1981). These features of seasonal variability are well captured in all models examined in this study.

Over the WNP basin, TCs are observed throughout the year, with high activity in boreal summer months peaking around the month of August (e.g., Chan 2005). This seasonal feature is very well reproduced in the OWZP detections from the CSIRO-Mk3.5 and MPI-ECHAM5 models (Fig. 5d). However, the two GFDL detections show a slight bias in peak activity toward the months of September and October.

The observed peak TC activity in the ENP basin occurs from July to September (Fig. 5e), with few TCs occurring before June and after October. This pattern is very realistically replicated in the MPI-ECHAM5–detected TCs. The detected TCs in the CSIRO-Mk3.5 and GFDL CM2.1 models slightly underestimate the peak activity, with a 1-month late bias in the latter.

In the NA basin, the observed TC activity peaks around September, with a few formations likely as early as May and as late as December (Fig. 5f). All models (except GFDL CM2.0) are able to realistically replicate this seasonal variability, but with substantial underestimation of the peak activity.

The observed seasonal cycle of NTCs in the NI basin has two peaks: one in May–June and a larger one in September–December. The relatively quiescent period during July–August is associated with the Indian summer monsoon. These two peaks are reproduced by all models except GFDL CM2.0, but with slight variations in phase. The MPI-ECHAM5 model, however, has a larger peak for the May–June period than the September–December period.

### d. Interannual variability

For completeness, we also examine the interannual variability of model-detected and observed TCs. Because the interannual variability of climate models is not expected to be phase locked with observations, we use the coefficient of variability (CV) to assess the interannual variability of model-detected TCs instead of

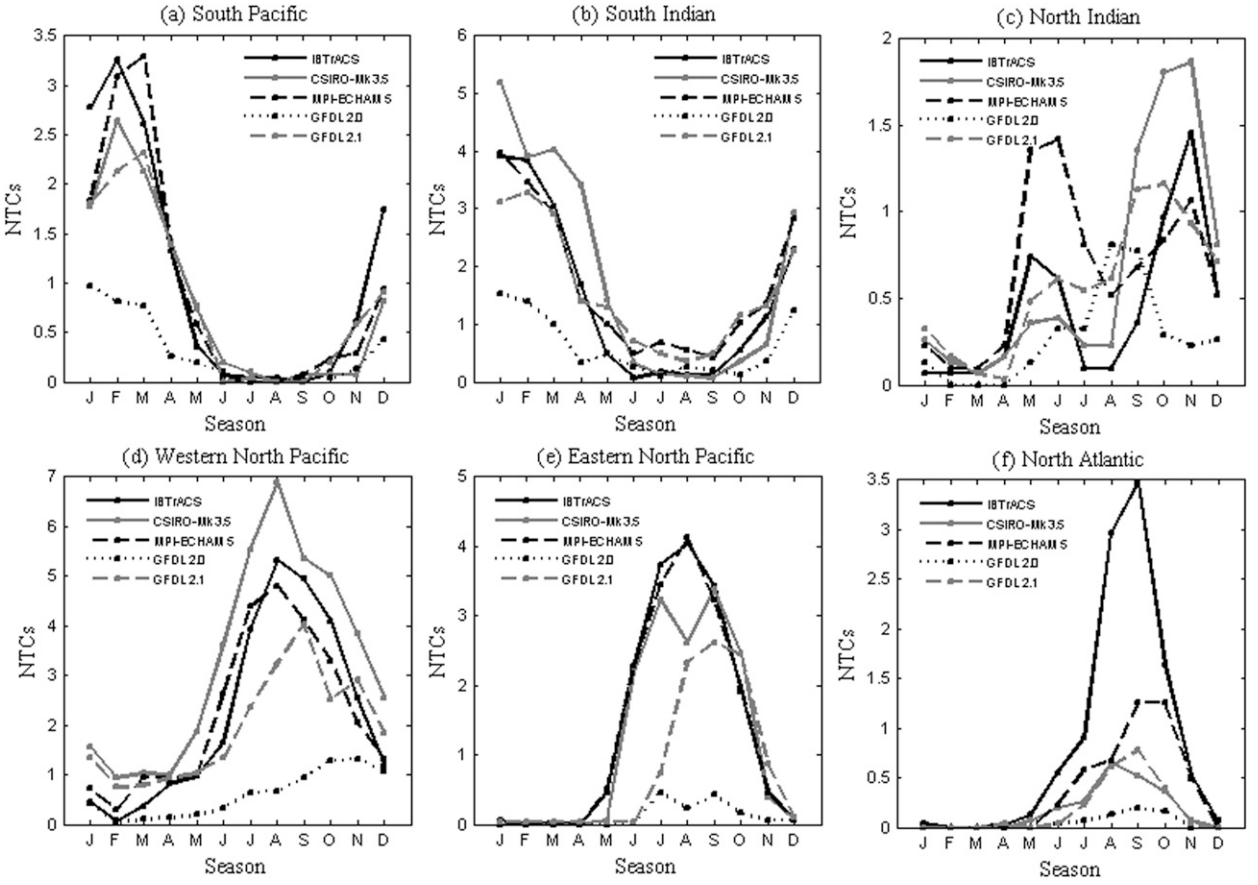


FIG. 5. Seasonal cycle of the number of observed and model-detected TCs for each TC basin.

correlation coefficients. The CV is considered an appropriate measure of dispersion when comparing data with different mean values. It is defined as the ratio of the standard deviation to the mean. To better compare with observations, each model CV is normalized by the observed CV. Thus, the closer the normalized CV is to unity, the better the model representation of interannual variability.

The normalized CVs of the CSIRO-Mk3.5, MPI-ECHAM5, and GFDL CM2.1 models (Table 3) are within 50% of unity for all TC basins except for the NA basin, where underdetections are known to be substantial. The greatest deviation from unity in all basins occurs for the GFDL CM2.0 model. With the exception of that model, the normalized CVs suggest the models capture interannual TC variability reasonably well for most TC basins. While a physical explanation is beyond the scope of this paper, it is likely that variability of ENSO in the models and its well-documented link to TC variability (e.g., Camargo et al. 2010) may explain some of the differences in interannual variability, as the models examined here are capable of generating ENSO-like features (Guilyardi et al. 2009).

### *e. Comparison with other detectors*

It is difficult to make performance comparisons with other techniques because there are very few published studies of TC detections in CMIP3 models. In the few studies of which we are aware, the performance assessments are not particularly comprehensive and tend to be limited to global maps of TC formation density (Walsh et al. 2010, 2013). Conclusions drawn on differences between models or detection schemes are mostly qualitative. With the exception of Fig. 3 of Walsh et al. (2010), most images are too small for easy comparison with our results. Figure 3 of Walsh et al. (2010) contains TC detection frequency using two different detection schemes for one model (CSIRO-Mk3.5) for the period January–March. However, a later paper (Walsh et al. 2013) notes that the results from one of the detection schemes were in error. The figure depicting the results from the other detection scheme (based on Camargo and Zebiak 2002) is reproduced here in Fig. 6 with an equivalent plot from the OWZP detector. The plots show the 20-yr climatology of a 3-month period (January–March) for the CSIRO-Mk3.5 model. In general, the two

TABLE 3. Normalized CV for TC frequency in each basin and model. The model CV is normalized by the observed CV. Values greater (less) than one represent greater (less) variability than observed.

| | SP | SI | NI | WNP | ENP | NA |
|---|---|---|---|---|---|---|
| CSIRO-Mk3.5 | 1.12 | 0.67 | 0.70 | 0.82 | 0.75 | 2.69 |
| MPI-ECHAM5 | 0.88 | 0.82 | 0.60 | 0.64 | 0.60 | 1.45 |
| GFDL CM2.0 | 2.54 | 1.71 | 0.83 | 1.70 | 2.22 | 4.37 |
| GFDL CM2.1 | 0.96 | 0.91 | 0.72 | 0.98 | 0.93 | 2.61 |

TC distributions look quite similar. The main difference appears to be a broader and more southward distribution of Southern Hemisphere formation locations detected by the OWZP scheme, and perhaps a slightly greater eastward distribution in the SP basin. It is possible that the Camargo and Zebiak (2002) scheme identifies circulations earlier than the OWZP scheme, the latter requiring all thresholds and conditions to be satisfied for three consecutive 0000 UTC time periods.

Although it is difficult to make precise comparisons with Walsh et al.'s (2013) graphical results, it is apparent that the OWZP detections are slightly higher in the NA basin than the CSIRO detector for the MPI-ECHAM5 and CSIRO-Mk3.5 models. While the detections from both methods (Fig. 7) depict a TC frequency much lower than observed, the comparison does appear to illustrate that the OWZP detector perhaps has a slightly greater ability to detect NA TCs.

## 5. Projected changes in TC frequency

The main focus of this paper is to demonstrate the performance of the OWZP TC detector in a selection of CMIP3 climate models. Thus, it is not our intention to make any conclusive statements about changing TC frequency in a future climate. However, given that the OWZP detections in three of the CMIP3 climate models appear to well reproduce the observed TC climatology, it is of interest to apply the detector to at least one future climate scenario for each of the three models. In anticipation that future changes in TC frequency will be small compared to the projection uncertainty, we chose the highest emissions scenario (A2; Nakicenovic et al. 2000) and the greatest time range (100 yr) available. The three models used are CSIRO-Mk3.5, MPI-ECHAM5, and GFDL CM2.1, and the two time periods analyzed are 1981–2000 and 2081–2100. In the following discussion, all references to the CMIP3 model results refer to the model–OWZP–detector combination.

A decrease in TC frequency is projected in all three models for the late twenty-first century, both globally and hemispherically (Figs. 8a–c). Error bars have been added to give some indication of the uncertainty of the mean annual TC numbers (95% significance level using the bootstrap sampling method). However, it is important to remember that this represents only the uncertainty that the sample mean is representative of a hypothetical larger sample. In no way does it represent uncertainties in a model's ability to generate realistic TC climatologies or the detector's ability to accurately detect model TCs. At present we do not have a method for calculating or estimating these uncertainties and can thus only consider them qualitatively. Of the three models, only the CSIRO-Mk3.5 and the MPI-ECHAM5 projected decreases are statistically significant at the 95% significance level. Globally, the percentage decrease

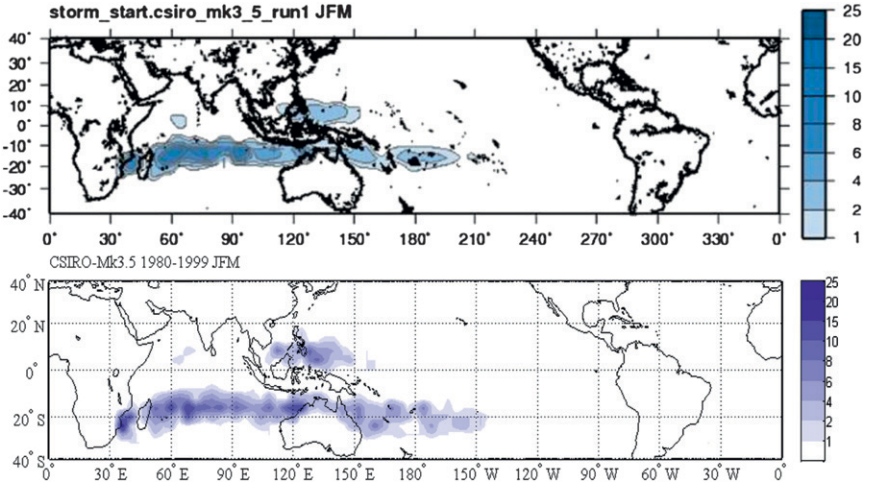


FIG. 6. TC detection distributions for the months January–March from 1980 to 1999. (top) Basin-dependent scheme of Camargo and Zebiak (2002), reproduced from Fig. 3 of Walsh et al. (2010). (bottom) OWZP scheme.

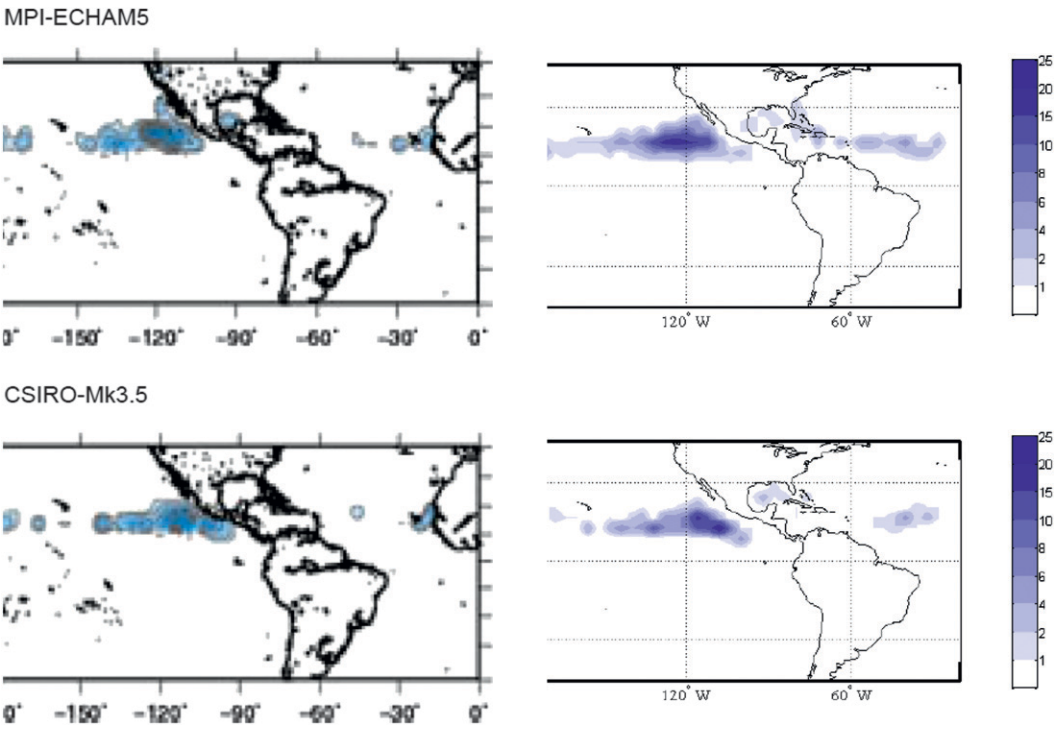


FIG. 7. TC frequency plots for the months July–September from 1980 to 1999, for the ENP and NA basins, in (top) MPI-ECHAM5 and (bottom) CSIRO-Mk3.5 using (left) the Walsh et al. (2007) detector and (right) the OWZP detector. The left panels are reproduced from Walsh et al. (2013).

ranges from 6% to 20% for the GFDL CM2.1 and MPI-ECHAM5 models, respectively (Fig. 9). In the Southern Hemisphere the decrease in TC frequency ranges from 4% to 24% for the GFDL CM2.1 and MPI-ECHAM5 models, respectively, and an equivalent decrease of 6% to 19% in the Northern Hemisphere. These decreases are consistent with the recent work of Knutson et al. (2010). While it is unclear whether these changes exceed the true uncertainty of the calculated mean annual TC numbers, some confidence can be gained from the fact the three models are consistent in predicting a decrease in TC frequency.

The same cannot be said for projected TC frequency changes in the individual basins (Figs. 8d–i). In the SP basin, the CSIRO-Mk3.5 model projects an *increase* of about 23%, whereas a similar magnitude *decrease* is projected in the MPI-ECHAM5 model, and there is no change in the GFDL CM2.1 model (Fig. 9). However, none of these changes is significant at the 95% confidence level. As a consequence, it is difficult to draw conclusions regarding changes in TC frequency from these data alone. Similarly, projections include both decreases and increases in the ENP and NA basins (Fig. 9), with only one model in each basin projecting a change that is significant at the 95% confidence level (CSIRO-Mk3.5, ENP basin, and MPI-ECHAM5, NA basin). Again, it could be misleading to argue that these data support any change in TC frequency, especially in the NA basin, where these models have trouble reproducing the current TC climatology. In the other three basins (i.e., SI, NI, and WNP basins), all models project a decrease in TC frequency, although only three of the nine changes are significant at the 95% confidence level [CSIRO-Mk3.5 and MPI-ECHAM5 in the SI basin (Fig. 8e) and MPI-ECHAM5 in the WNP basin (Fig. 8g)].

The above discussion would suggest there is limited value in the projected TC frequency changes from the data presented here. However, despite the uncertainty, the majority of basinwide projections (13 out of 18) show fewer TCs, and all of the six hemispheric and three global projections also show fewer TCs. This qualitative model ensemble result can be useful if provided with sufficient caveats explaining the uncertainty. Additional model ensemble results are provided in Fig. 10, which shows the geographic distribution of the late-twentieth- and the late-twenty-first-century ensemble mean TC formation density (Figs. 10a,b) and the difference between the two (Fig. 10c). The three-model ensemble mean tendency of Fig. 10c demonstrates that reduced TC frequency is widespread across all basins.

Seasonal variation in TC frequency for the late-twenty-first-century simulations are similar to their

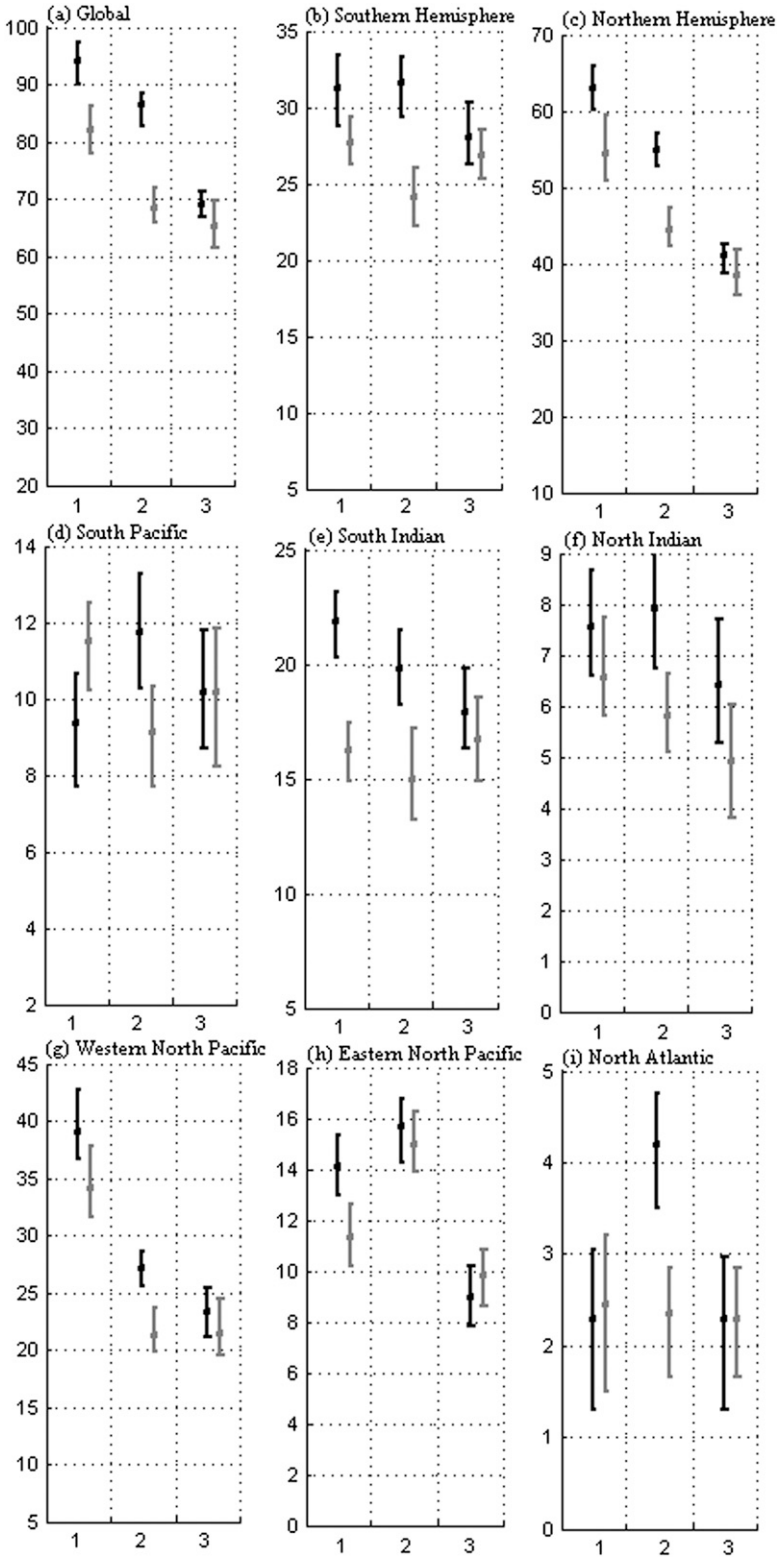


FIG. 8. Mean number of TCs for the late-twentieth-century simulations (black) and late-twenty-first-century simulations (gray), globally, hemispherically, and for the individual TC basins for three models (CSIRO-Mk3.5, MPI-ECHAM5, and GFDL CM2.1). Error bars represent 95% significance level obtained using the bootstrap sampling method.

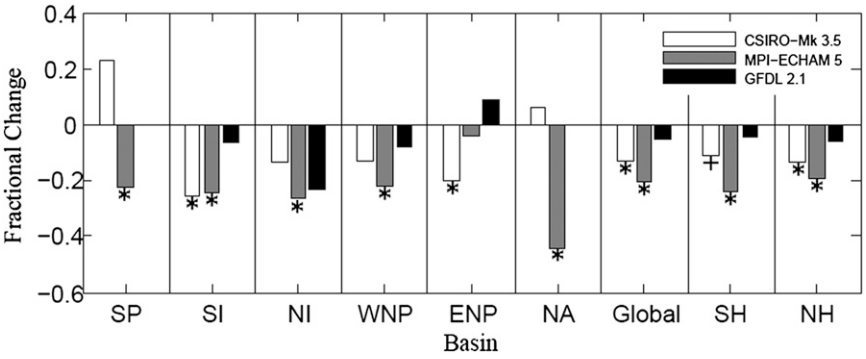


FIG. 9. Fractional change in mean TC frequency between the late twentieth and late twenty-first centuries for the three models. Changes that are significant at 95% and 90% confidence levels are indicated by asterisk and plus symbols, respectively.

twentieth-century counterparts, indicating that no model projected future change in the seasonal cycle (Fig. 11). Peak activity in some cases is slightly reduced in the future climate simulations, consistent with the overall projected decrease. In the SP basin, the ECHAM5 and GFDL CM2.1 models, for example, project slightly lower peak activity in the future climate compared with the current climate. In the WNP basin, reduction in TC numbers for future climate mainly occurs in the late TC season. This is in agreement with a fine-resolution experiment (~20 km) of Murakami et al. (2011).

In general, there is overall TC projection frequency consistency between these results and other finer-resolution studies (e.g., Zhao et al. 2009; Knutson et al. 2010; Murakami and Wang 2010; Murakami et al. 2011). This suggests lower-resolution models, such as those examined here using the OWZP TC detector, are well capable of representing current and future TC climatology.

## 6. Summary

A novel TC detection technique (OWZP; Tory et al. 2013a,b) was applied to a selection of CMIP3 global climate models, and a comprehensive assessment of the combined detector and model performance in reproducing observed TC climatology was documented. Three models (CSIRO-Mk3.5, MPI-ECHAM5, and GFDL CM2.1) produced very reasonable TC formation climatologies in most basins, with the exception being a large and systematic underprediction in the North Atlantic basin, which is consistent with other studies (e.g., Camargo et al. 2005). In terms of TC formations, these three models also showed good seasonal variation (Fig. 5) and interannual variability (Table 3). The fourth model (GFDL CM2.0) suffered from a large, systematic underprediction in all basins [as also found by Walsh et al. (2013)]. As such, it was considered to be of no value for future climate projections of TC formation.

When verifying against observed TC climatology, it is difficult to separate model error from detector error; thus, the performance assessment reflects the combined performance of the model and detector. The fact that other studies using different detectors also identified poor performance in the North Atlantic basin and in the GFDL CM2.0 model suggests the models are responsible for the underperformance rather than the detector. The good performance in the remaining models and basins represents good combined model and

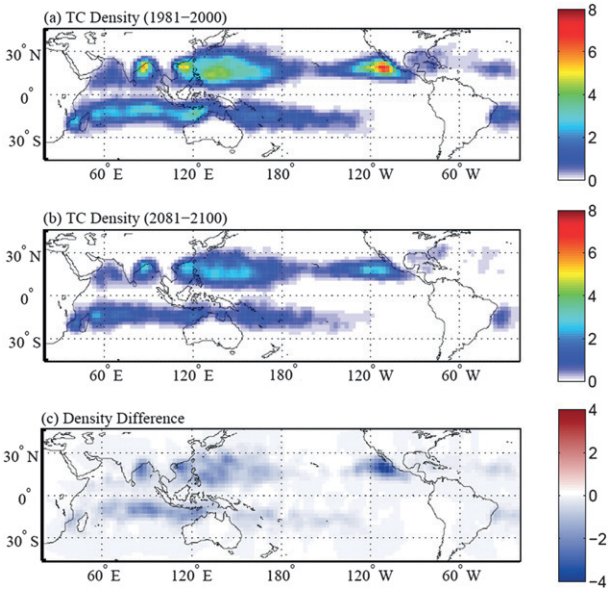


FIG. 10. Anisotropic Gaussian density distribution of ensemble-mean TC genesis positions per year in a 2.5° × 2.5° grid box (a) for the late-twentieth-century simulations and (b) for the late-twenty-first-century simulations. (c) The difference in TC density between (a) and (b).

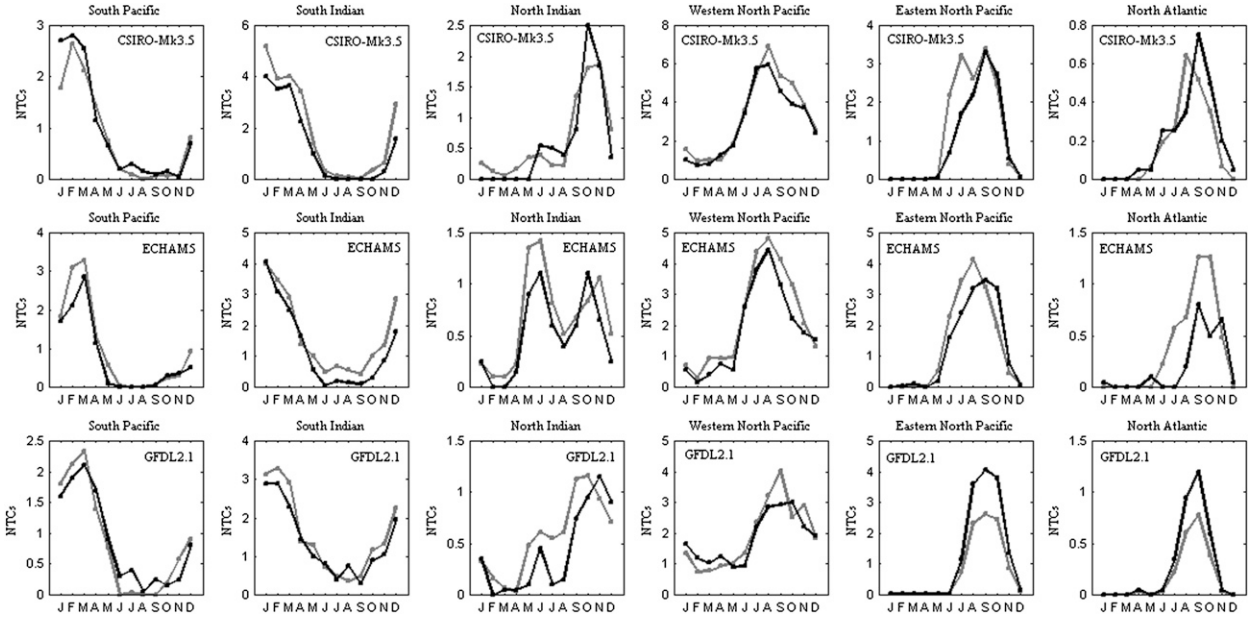


FIG. 11. Seasonal variability in the mean TC frequency for the late-twentieth-century simulations (gray) and the late-twenty-first-century simulations (black) for each basin and model.

detector performance. If the compensation between model and detector error is minimal, then both the model and detector performed well. Although it is not possible to measure any such error compensation, the OWZP was designed to minimize this compensation. It follows that the OWZP TC detection method applied to coarse-resolution models is likely to be particularly robust.

The main focus of this paper was to document the OWZP TC detection performance in CMIP3 models. While a preliminary TC projection analysis was possible using just three models, more models would be required for a comprehensive projection study. The changes we found in TC detection rates between the late twentieth and twenty-first centuries are consistent with higher-resolution studies (e.g., Zhao et al. 2009; Knutson et al. 2010; Murakami and Wang 2010; Murakami et al. 2011). Our results are summarized as follows:

- Averaged across the globe, a decrease in TC detections of 6%–20%, consistent with the 6%–34% reported by Knutson et al. (2010).
- A much larger spread of projections in individual basins, consistent with the 23% increase to 45% decrease reported in Knutson et al. (2010) and other studies.
- No overall change in the seasonal variability of TC formation by the end of the twenty-first century, consistent with other higher-resolution studies (e.g., Murakami and Wang 2010; Murakami et al. 2011).

These projection results are subject to considerable uncertainty, especially as only three models were available. A broader set of possible future outcomes provided by a larger suite of models would offer greater insight into the largely unquantifiable uncertainty. (A more comprehensive study is currently underway using CMIP5 model data.) Additional insight should also be possible by applying a suite of TC detectors. The OWZP TC detector is based on fundamentally different principles to other detectors in the literature (e.g., Camargo and Zebiak 2002; Walsh et al. 2007; Bengtsson et al. 2007; Strachan et al. 2013) and, together with these detectors, would thus be highly valuable for analyzing TC climatology in global climate models.

*Acknowledgments.* We thank Andrew Dowdy and Xingbao Wang for their valuable insight. We acknowledge the Pacific Climate Change and Science Program (PCCSP) project for supporting this work. PCCSP is funded by AusAID, in collaboration with the Department of Climate Change and Energy Efficiency, and delivered by the Bureau of Meteorology and the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

## REFERENCES

Bengtsson, L., H. Bottger, and M. Kanamitsu, 1982: Simulation of hurricane-type vortices in a general circulation model. *Tellus,* **34,** 440–457.

——, M. Botzet, and M. Esch, 1995: Hurricane-type vortices in a general circulation model. *Tellus,* **47A,** 175–196.

——, K. I. Hodges, and M. Esch, 2007: Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analyses. *Tellus,* **59A,** 396–416.

Camargo, S. J., and S. E. Zebiak, 2002: Improving the detection and tracking of tropical storms in atmospheric general circulation models. *Wea. Forecasting,* **17,** 1152–1162.

——, A. G. Barnston, and S. E. Zebiak, 2005: A statistical assessment of tropical cyclone activity in atmospheric general circulation models. *Tellus,* **57A,** 589–604.

——, A. H. Sobel, A. G. Barnston, and J. K. Philip, 2010: The influence of natural climate variability on tropical cyclones, and seasonal forecasts of tropical cyclone activity. *Global Perspectives on Tropical Cyclones: From Science to Mitigation,* J. C. L. Chan and J. D. Kepert, Eds., World Scientific, 325–360.

Chan, J. C. L., 2005: Interannual and interdecadal variations of tropical cyclone activity over the western North Pacific. *Meteor. Atmos. Phys.,* **89,** 143–152.

Chand, S. S., and K. J. E. Walsh, 2009: Tropical cyclone activity in the Fiji region: Spatial patterns and relationship to large-scale circulation. *J. Climate,* **22,** 3877–3893.

Chauvin, F., J.-F. Royer, and M. Déqué, 2006: Response of hurricane-type vortices to global warming as simulated by ARPEGE-Climat at high resolution. *Climate Dyn.,* **27,** 377–399.

Efron, B., and R. Tibshirani, 1991: Statistical data analysis in the computer age. *Science,* **253,** 390–395.

Emanuel, K., R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc.,* **89,** 347–367.

Guilyardi, E., A. Wittenberg, A. Fedorov, M. Collins, C. Wang, A. Capotondi, G. J. van Oldenborgh, and T. Stockdale, 2009: Understanding El Niño in ocean–atmosphere general circulation models. *Bull. Amer. Meteor. Soc.,* **90,** 325–340.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archieve for Climate Stewardship (IBTrACS) unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.,* **91,** 363–376.

Knutson, T. R., and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.,* **3,** 157–163.

Manabe, S., J. L. Holloway, and H. M. Stone, 1970: Tropical circulation in a time-integration of a global model of the atmosphere. *J. Atmos. Sci.,* **27,** 580–613.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: The WCRP CMIP3 multi-model dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.,* **88,** 1383–1394.

Murakami, H., and B. Wang, 2010: Future change of North Atlantic tropical cyclone tracks: Projection by a 20-km-mesh global atmospheric model. *J. Climate,* **23,** 2699–2721.

——, ——, and A. Kitoh, 2011: Future change of western North Pacific typhoons: Projections by a 20-km-mesh global atmospheric model. *J. Climate,* **24,** 1154–1169.

Nakicenovic, N., and Coauthors, 2000: *IPCC Special Report on Emission Scenarios*. Cambridge University Press, 599 pp.

Ramage, C. S., and A. M. Hori, 1981: Meteorological aspects of El Niño. *Mon. Wea. Rev.,* **109,** 1827–1835.

Ramsay, H. A., L. M. Leslie, P. J. Lamb, M. B. Richman, and M. Leplastrier, 2008: Interannual variability of tropical cyclones in the Australian region: Role of large-scale environment. *J. Climate,* **21,** 1083–1103.

Strachan, J., P. L. Vidale, K. Hodges, M. Roberts, and M.-E. Demory, 2013: Investigating global tropical cyclone activity with a hierarchy of AGCMs: The role of model resolution. *J. Climate,* **26,** 133–152.

Thorncroft, C., and K. Hodges, 2001: African easterly wave variability and its relationship to Atlantic tropical cyclone activity. *J. Climate,* **14,** 1166–1179.

Tory, K. J., S. S. Chand, R. A. Dare, and J. L. McBride, 2013a: The development and assessment of a model-, grid- and basin-independent tropical cyclone detection scheme. *J. Climate,* **26,** 5493–5507.

——, R. A. Dare, N. E. Davidson, J. L. McBride, and S. S. Chand, 2013b: The importance of low-deformation vorticity in tropical cyclone formation. *Atmos. Chem. Phys.,* **13,** 2115–2132, doi:10.5194/acp-13-2115-2013.

Vitart, F., J. L. Anderson, and W. F. Stern, 1997: Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations. *J. Climate,* **10,** 745–760.

Walsh, K., M. Fiorino, C. W. Landsea, and K. L. McInnes, 2007: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J. Climate,* **20,** 2307–2314.

——, S. Lavender, H. Murakami, E. Scoccimarro, L.-P. Caron, and M. Ghantous, 2010: The Tropical Cyclone Climate Model Intercomparison Project. *Hurricanes and Climate Change,* Vol. 2, J. B. Elsner et al., Eds., Springer, 1–24.

——, ——, E. Scoccimarro, and H. Murakami, 2013: Resolution dependence of tropical cyclone formation in CMIP3 and finer resolution models. *Climate Dyn.,* **40,** 585–599, doi:10.1007/s00382-012-1298-z.

Wu, G., and N.-C. Lau, 1992: A GCM simulation of the relationship between tropical-storm formation and ENSO. *Mon. Wea. Rev.,* **120,** 958–977.

Zhao, M., M. I. Held, S.-J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, internannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate,* **22,** 6653–6678.