



## Tracking Scheme Dependence of Simulated Tropical Cyclone Response to Idealized Climate Simulations

MICHAEL HORN,\* KEVIN WALSH,\* MING ZHAO,<sup>+</sup> SUZANA J. CAMARGO,<sup>#</sup> ENRICO SCOCCIMARRO,<sup>@</sup>  
 HIROYUKI MURAKAMI,<sup>&</sup> HUI WANG,\*\* ANDREW BALLINGER,<sup>++</sup> ARUN KUMAR,\*\*  
 DANIEL A. SHAEVITZ,<sup>#</sup> JEFFREY A. JONAS,<sup>@</sup> AND KAZUYOSHI OOUCHI<sup>&&</sup>

<sup>\*</sup> School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia

<sup>+</sup> NOAA/GFDL, Princeton, New Jersey

<sup>#</sup> Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

<sup>@</sup> Istituto Nazionale di Geofisica e Vulcanologia, and Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

<sup>&</sup> IPRC/MRI, Tsukuba, Ibaraki, Japan

<sup>\*\*</sup> NOAA/NWS/NCEP/Climate Prediction Center, College Park, Maryland

<sup>++</sup> Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey

<sup>##</sup> Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York

<sup>@@</sup> Center for Climate System Research, Columbia University, and NASA Goddard Institute for Space Studies, New York, New York

<sup>&&</sup> Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

(Manuscript received 6 March 2014, in final form 5 September 2014)

### ABSTRACT

Future tropical cyclone activity is a topic of great scientific and societal interest. In the absence of a climate theory of tropical cyclogenesis, general circulation models are the primary tool available for investigating the issue. However, the identification of tropical cyclones in model data at moderate resolution is complex, and numerous schemes have been developed for their detection.

The influence of different tracking schemes on detected tropical cyclone activity and responses in the Hurricane Working Group experiments is examined herein. These are idealized atmospheric general circulation model experiments aimed at determining and distinguishing the effects of increased sea surface temperature and other increased CO<sub>2</sub> effects on tropical cyclone activity. Two tracking schemes are applied to these data and the tracks provided by each modeling group are analyzed.

The results herein indicate moderate agreement between the different tracking methods, with some models and experiments showing better agreement across schemes than others. When comparing responses between experiments, it is found that much of the disagreement between schemes is due to differences in duration, wind speed, and formation-latitude thresholds. After homogenization in these thresholds, agreement between different tracking methods is improved. However, much disagreement remains, accountable for by more fundamental differences between the tracking schemes. The results indicate that sensitivity testing and selection of objective thresholds are the key factors in obtaining meaningful, reproducible results when tracking tropical cyclones in climate model data at these resolutions, but that more fundamental differences between tracking methods can also have a significant impact on the responses in activity detected.

### 1. Introduction

The nature of possible future changes in tropical cyclone (TC) activity is of great interest not only scientifically, but

to all of society. In the absence of a general climate theory of TC formation, climate models are the primary tool available for investigating the problem. The spatial scales of key TC features such as the eyewall may have suggested that resolutions approaching single kilometers would be necessary to produce TCs in general circulation models (GCMs). However, it is well established now that modern GCMs are capable of producing structures that can be recognized as similar to tropical cyclones

Corresponding author address: Michael Horn, School of Earth Sciences, Corner Swanston and Elgin Streets, University of Melbourne, Parkville, VIC 3010, Australia.  
 E-mail: mjhorn@student.unimelb.edu.au

at resolutions as coarse as 100 km (Knutson et al. 2010). These TC-like vortices are low pressure centers with associated intense winds and a warm core. Because of limited resolution, their spatial extents are larger and intensities lower than observed in real TCs (Walsh et al. 2007). For simplicity, we will refer to these features as TCs throughout.

The identification of these TC features in model output at moderate resolution (i.e., 50–200-km grid spacing) is nontrivial, and numerous different schemes have been developed for the detection and tracking of tropical cyclone-like vortices (e.g., Camargo and Zebiak 2002; Zhao et al. 2009; Walsh et al. 2013; Strachan et al. 2013). These tracking schemes scan model output data and locate points at which certain TC criteria are met. These criteria usually include thresholds in variables such as wind speed and vorticity. The thresholds can be based either on absolute values or on deviations from the mean in that model and/or ocean basin. If thresholds based on absolute values are derived from observations, tracking schemes taking this approach cast clear light on the ability of a model to reproduce a realistic genesis climatology, as the possible tuning of such thresholds is limited. The disadvantage of this approach is that it does not easily allow for the correction of model biases. The latter approach uses relative thresholds that are adjusted model to model or basin to basin. This approach is motivated by the assumption that TCs represent the extreme tails of the distributions in relevant variables, and that the position of TCs in these distributions (in terms of standard deviations from the mean) will remain the same in different models, even if the distributions themselves are substantially different. By design, these schemes produce a fairly realistic present-day climatology in most models (Camargo and Zebiak 2002). One scheme considered here takes this relative approach, with the remainder using absolute thresholds.

Different schemes also differ in the way they join detection points into tracks. Some simply apply the same criteria to all points and then join spatially and temporally adjacent detections into tracks. Camargo and Zebiak (2002) point out that, in some cases, this approach results in unrealistically short tracks. To address this shortcoming, their tracking scheme and other schemes apply some relaxation of detection criteria after an initial detection (Camargo and Zebiak 2002; Walsh et al. 2013). In some cases, this includes reanalysis of time steps preceding a detection with relaxed criteria (Camargo and Zebiak 2002). These differences in tracking may have substantial impacts on the statistics of detected TCs.

At present, there is little uniformity between tracking methods and criteria used in different GCM TC studies.

The use of a 10-m wind speed criterion is a notable exception, where the objective resolution-based thresholds determined by Walsh et al. (2007) have been adopted in a number of studies (e.g., Stowasser et al. 2007; Bengtsson et al. 2007; Zhao et al. 2009; Scoccimarro et al. 2011; Vecchi et al. 2013). However, other studies use resolution-independent thresholds by interpolating all data to a fixed resolution before tracking (Strachan et al. 2013). Thresholds in other variables such as low-level vorticity, sea level pressure (SLP), sea surface temperature (SST), and measures of the warm core such as wind speed and temperature anomalies vary widely among tracking schemes.

Little previous work has directly addressed the potential significance of tracking scheme differences in analyzing responses in TC activity in climate models. However, some previous studies have examined the sensitivities of their detection numbers to threshold values within a single scheme. Li et al. (2013) find little sensitivity to any thresholds except those in genesis location and the strength of the warm core, although it should be noted that their study is in an aquaplanet context. Zhao et al. (2009) find some sensitivity to all thresholds, with an especially strong dependence on the duration threshold. They also find that these sensitivities are much reduced when focusing only on the most intense TCs produced. Generally, it is unclear how the different threshold sensitivities observed within single tracking schemes and experiments in different studies may vary between tracking schemes or across different experiments.

Ideally, all schemes would be sufficiently objective to detect the same or similar TC activity in any GCM data. However, it is known that different schemes give different numbers in individual experiments. For example, Tory et al. (2013) report that eight reliable models from phase 5 of the Coupled Model Intercomparison Project (CMIP5) project decreases in global TC frequency using their unique TC detection method, while Camargo (2013) reports relative increases in projected global TC frequency for a number of CMIP5 models, including some of those analyzed by Tory et al. (2013), based on the use of the detection algorithm of Camargo and Zebiak (2002). Here, we investigate whether differences in responses detected between tracking schemes remain consistent over different experiments, or whether different tracking schemes have the potential to alter the detected response of GCMs to different perturbations. In the process of this analysis, we separate the effects of different thresholds in wind speed, duration, and formation latitude from differences we regard as more fundamental to the different tracking schemes, such as differently functioning warm-core checks and different

methods of combining detection points into tracks. Thresholds in vorticity and the strength of the warm core are included in the latter category, as the strengths used for these are more highly dependent on the details of the tracking process used (in ways that, for example, a duration threshold is not).

Section 2 provides details of the modeling and tracking scheme methods used. Section 3 then presents results comparing tracking scheme performance for present climate and altered climate experiments, and section 4 discusses the relevance of these results to TC GCM research in general. Finally, section 5 provides our conclusions.

## 2. Methods

We analyze results from a suite of idealized altered-climate experiments performed in four different GCMs and tracked using multiple schemes. The experiments were performed as part of the U.S. Climate Variability and Predictability Research Program (CLIVAR) Hurricane Working Group (HWG; <http://www.usclivar.org/working-groups/hurricane>). The HWG experiments are designed to compare the drivers of trends in TC activity in different GCMs. They consist of atmosphere-only runs in a number of GCMs forced for the different experiments as follows:

- (a) 1992 atmospheric gas concentrations and 1985–2001 seasonally varying climatological SSTs and sea ice concentration (SIC);
- (b) As in (a), but with a uniform global + 2K SST anomaly;
- (c) As in (a), but with doubled CO<sub>2</sub> concentration; and
- (d) As in (a), but with a uniform global + 2K SST anomaly and doubled CO<sub>2</sub> concentration.

Full details on the methodology of these experiments can be found in Held and Zhao (2011). In this work, we use data from a subset of the HWG models: the Centro Euro-Mediterraneo per i Cambiamenti Climatici (CMCC)–Istituto Nazionale di Geofisica e Vulcanologia (INGV) ECHAM5 (T159 resolution, ~90-km grid spacing at equator; Roeckner et al. 2003), the National Aeronautic and Space Administration (NASA) Goddard Institute for Space Studies (GISS) model (1° resolution, ~110-km grid spacing at equator; Schmidt et al. 2014), the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) (T126 resolution, ~110-km grid spacing at equator; Saha et al. 2014), and the Meteorological Research Institute Atmospheric General Circulation Model, version 3.2 (MRI AGCM3.2) (TL319 resolution, ~60-km grid spacing at equator; Mizuta et al. 2012).

The experimental design differs slightly for the MRI model, using Atmospheric Model Intercomparison Project (AMIP)-style SSTs instead of seasonal climatologies. Specifically, the MRI model is forced for the different experiments with

- (i) 1979–2003 yearly global mean atmospheric gas concentrations and monthly observed SSTs and SIC;
- (ii) 1979–2003 yearly global mean atmospheric gas concentrations and 2075–99 SSTs and SIC from models from phase 3 of CMIP (CMIP3) using the Intergovernmental Panel on Climate Change (IPCC) A1B scenario;
- (iii) 2075–99 atmospheric gas concentrations from the IPCC A1B scenario and 1979–2003 SSTs and SIC; and
- (iv) Atmospheric gases set to 2075–99 values from the IPCC A1B scenario with 1979–2003 monthly observed SST plus a 1.83-K global anomaly.

While these are different from the climatological experiments used for the other three models, experiments a–d do correspond qualitatively to experiments i–iv. The directions of TC genesis changes resulting, if not the magnitudes, can still be compared meaningfully.

The experiments for the MRI model are 25 years long, 20 years for the GISS model, and 10 years for the CMCC-INGV and NCEP models. These different experiment durations are due to the different model resolutions and amounts of computer time available to the different institutions involved. IBTrACS best-track data for 1980–99 are used to compare model genesis patterns and tracks with the real climate (Knapp et al. 2010).

We apply two different tracking schemes to these data. A modified version of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) tracking scheme (Walsh et al. 2007; Horn et al. 2013) is used across all four models, and the Zhao tracking scheme (Zhao et al. 2009) is used for all but the MRI data, where necessary data were not archived. These two tracking schemes are selected because they are versions of two of the most widely used schemes in TC GCM studies (see e.g., Stowasser et al. 2007; Zhao et al. 2009; Scoccimarro et al. 2011; Held and Zhao, 2011; Murakami et al. 2012; Walsh et al. 2013; Vecchi et al. 2013). We also analyze the tracks provided by each modeling group, which were produced using different tracking schemes depending on the group. Model data used are 6-hourly in all cases.

The modified CSIRO tracking scheme uses the following detection criteria to locate TCs:

- 1) An absolute value of 850-hPa vorticity greater than  $10^{-5} \text{ s}^{-1}$ ;

- 2) A closed pressure minimum within a distance in both the  $x$  and  $y$  directions of 350 km from a point satisfying condition 1 above (distance chosen empirically to give a good geographical association between vorticity maxima and pressure minima). This minimum pressure is taken as the center of the storm;
- 3) A mean wind speed in the region  $700 \text{ km} \times 700 \text{ km}$  square around the center of the storm at 850 hPa greater than at 300 hPa; and
- 4) Maximum 10-m wind speeds exceeding a resolution-dependent value as specified in [Walsh et al. \(2007\)](#).

Detections are allowed only over ocean, based on topography fields degraded to model resolution, unless a previous detection exists within a resolution-dependent distance. These detections are then linked into tracks by associating consecutive detections within  $6^\circ$  of each other (for 6-hourly data). Tracks lasting less than 24 h are excluded. No latitude restriction is imposed, and the TCs are instead partitioned from extratropical storms using the separation in the latitudinal distribution of their genesis points caused by the extratropical ridges in both hemispheres. This is one point of departure from the original CSIRO scheme; another is the removal of a computationally demanding warm core check that was found to be unnecessary at the higher resolutions used in the HWG experiments ([Horn et al. 2013](#)).

The Zhao scheme identifies TCs by locating grid points meeting the following criteria:

- 1) An 850-hPa relative vorticity maximum exceeding  $3.5 \times 10^{-5} \text{ s}^{-1}$  within a  $6^\circ \times 6^\circ$  latitude/longitude box;
- 2) A local minimum of sea level pressure within  $2^\circ$  latitude/longitude from the vorticity maximum; and
- 3) A local maximum anomaly in the temperature averaged between 300 and 500 hPa located within  $2^\circ$  of the SLP minimum. Temperature must be at least  $1^\circ\text{C}$  warmer than the surrounding local mean.

The resulting detections are then combined into trajectories by associating the closest successive (i.e., 6 h separated) detections within 400 km of each other. If there are multiple possibilities, preference is given to storms to the west and poleward of the previous detection. Trajectories lasting less than 3 days are eliminated. Storms are also required to have a maximum surface wind speed greater than  $12 \text{ m s}^{-1}$  during at least 2 days (not necessarily consecutive). Only trajectories beginning within  $50^\circ$  of the equator are considered.

The MRI group tracks use a method based on [Murakami et al. \(2012\)](#). The criteria considered are as follows:

- 1) The maximum relative vorticity at 850 hPa exceeds  $8.0 \times 10^{-5} \text{ s}^{-1}$ .

- 2) The maximum wind speed at 850 hPa exceeds  $13.0 \text{ m s}^{-1}$ .
- 3) There is an evident warm core aloft, with the sum of the temperature deviations at 300, 500, and 700 hPa exceeding 0.8 K. The temperature deviation for each level is computed by subtracting the maximum temperature from the mean temperature over the  $10^\circ \times 10^\circ$  grid box centered nearest to the location of maximum vorticity at 850 hPa.
- 4) The maximum wind speed at 850 hPa is greater than the maximum wind speed at 300 hPa.
- 5) To remove tropical monsoon depressions in the north Indian Ocean (NIO), the radius of maximum mean wind speed must be less than 200 km from the detected storm center. This condition is applied in the NIO only.

The duration of each detected storm must exceed 36 h. When a single TC satisfies all the criteria intermittently, it is considered as multiple TC generation events. To prevent multiple counts of a single TC, a single time-step failure is allowed. These criteria are optimized to produce around 84 TCs per year in the MRI model.

The GISS group tracks use the tracking scheme of [Camargo and Zebiak \(2002\)](#). This is the only tracking scheme used here that does not employ absolute thresholds. The scheme uses model-dependent thresholds based on selecting the tails of the probability distribution functions (PDFs) of relevant variables. Based on analysis of the joint PDFs obtained in the 850-hPa relative vorticity, the 850- to 300-hPa anomalous integrated temperature, and the surface wind speed for observations and GCMs, the following model-dependent criteria are chosen:

- 1) 850-hPa relative vorticity at least twice the standard deviation of the vorticity;
- 2) 850- to 300-hPa anomalous integrated temperature threshold greater than or equal to the standard deviation calculated over only those cases where there is a warm core; and
- 3) Surface wind speed greater than or equal to the global average wind speed (over ocean only) plus the standard deviation in the relevant basin.

The scheme also imposes the following model-independent criteria:

- 1) A local minimum in sea level pressure;
- 2) A positive local temperature anomaly at 850, 700, 500, and 300 hPa;
- 3) A larger local temperature anomaly at 850 hPa than at 300 hPa; and
- 4) Higher mean wind speeds at 850 hPa than at 300 hPa.

The closest successive detections within  $5^\circ$  of each other are then connected into tracks. Tracks of at least

TABLE 1. Tracking schemes (columns) used for each model (rows).

	CSIRO/ Horn	Zhao	Murakami	CSIRO/ CMCC-INGV	Camargo
MRI	X		X		
NCEP	X	X			
CMCC- INGV	X	X		X	
GISS	X	X			X

1.5 days are considered to be TCs. These tracks are then extended forward and backward in time by tracking the vorticity maximum while the absolute value exceeds a relaxed vorticity threshold. This is intended to achieve more realistic track lengths.

For the CMCC-INGV group tracks, storm centers were detected using an alternate version of the CSIRO scheme. These points were then combined into tracks based on spatial and temporal continuity. Tracks lasting less than 18 h were removed. The NCEP group tracks use the Zhao tracking scheme and are identical to those results, and so are not used here.

The models and tracking schemes used are summarized in Table 1. The gaps in coverage of the models by the different tracking schemes are due to the unavailability of data necessary for a scheme in a given model. This is also the reason for the chosen subset of HWG models; others included in the project did not archive sufficient data for at least the CSIRO scheme. Correlations stated in the text are Pearson correlation coefficients. Statistical significance of changes in mean genesis rates between experiments is found by a *t* test of two independent series of annual genesis rates, assuming identical variance.

### 3. Results

#### a. Present climate

Table 2 gives the mean yearly TC numbers for each model, experiment, and scheme. The rate of genesis varies substantially between models and schemes. Compared to the observed present-day climatological mean genesis rate of around 90 TCs per year, GISS shows very low formation when tracked with the CSIRO and Zhao schemes, with much stronger performance in the relative Camargo (“group”) scheme. NCEP shows realistic genesis in the CSIRO scheme, but only around half as much in the Zhao scheme. CMCC-INGV shows realistic present-day genesis rates of between 85 and 91 TCs per year in all three schemes. The MRI model also performs reasonably well with the CSIRO scheme,

TABLE 2. Mean yearly TC detection numbers from each tracking scheme. Bold text indicates an increase of greater than 5% relative to the equivalent present-day experiment and italic text a decrease of greater than 5%. For the MRI model, AMIP refers to the present-day experiment, CO<sub>2</sub>-f SST-p to the future CO<sub>2</sub> and present-day SST experiment, CO<sub>2</sub>-p SST-f to the reverse, and SST-1.83K CO<sub>2</sub>-f to the future CO<sub>2</sub> and SST experiment.

		CSIRO/Horn	Zhao	Group- supplied
GISS	Present day	11.1	16.9	76.4
	Double CO <sub>2</sub>	<b>12.7</b>	<b>18.3</b>	68.35
	SSTp2K	11.3	<i>11.65</i>	<b>87.45</b>
	SSTp2K 2×CO <sub>2</sub>	<b>12.0</b>	<i>12.0</i>	79.4
NCEP	Present day	86.3	46.7	—
	Double CO <sub>2</sub>	83.3	<b>50.0</b>	—
	SSTp2K	82.7	<i>40.7</i>	—
	SSTp2K 2×CO <sub>2</sub>	<i>81.3</i>	39.8	—
CMCC-INGV	Present day	91.0	86.3	85.3
	Double CO <sub>2</sub>	78.5	<i>79.1</i>	76.6
	SSTp2K	<b>93.6</b>	84.3	88.5
	SSTp2K 2×CO <sub>2</sub>	<i>80.5</i>	<i>77.3</i>	<i>80.4</i>
MRI	AMIP	45.0	—	83.0
	CO <sub>2</sub> -f SST-p	<i>42.5</i>	—	<i>78.4</i>
	CO <sub>2</sub> -p SST-f	44.0	—	<i>71.3</i>
	SST-1.83K CO <sub>2</sub> -f	37.5	—	<i>66.8</i>

although the genesis rate is lower than with the group-supplied scheme.

Figures 1 and 2 provide a comparison of the present-day January–March (JFM) and July–September (JAS) genesis densities (genesis per 20 years per 4° square box) generated in the four models as detected by the available tracking schemes. As the relative performance of each model in reproducing real-world genesis densities is not the direct concern of this paper, we will not dwell on the details of this geographic comparison. In most cases, the models produce moderately realistic genesis patterns, subject to the typical shortfalls of the TC genesis distribution in GCMs, especially low North Atlantic genesis rates (Camargo 2013).

Importantly, however, we should note that the different tracking schemes produce less variation in the geographic distribution of genesis within each model than they do in the global mean genesis rates (except for in the GISS model, where the group scheme uses relative thresholds resulting in a substantially different distribution). Generally, the differences between the geographical distributions derived using the different schemes are smaller than the differences observed among the four models, and between the models and the best-track data. The fact that the geographical distributions detected with the different schemes are similar despite differences in the total number of TCs detected indicates that there are no regions where the schemes are substantially more or less likely to differ. Instead, the

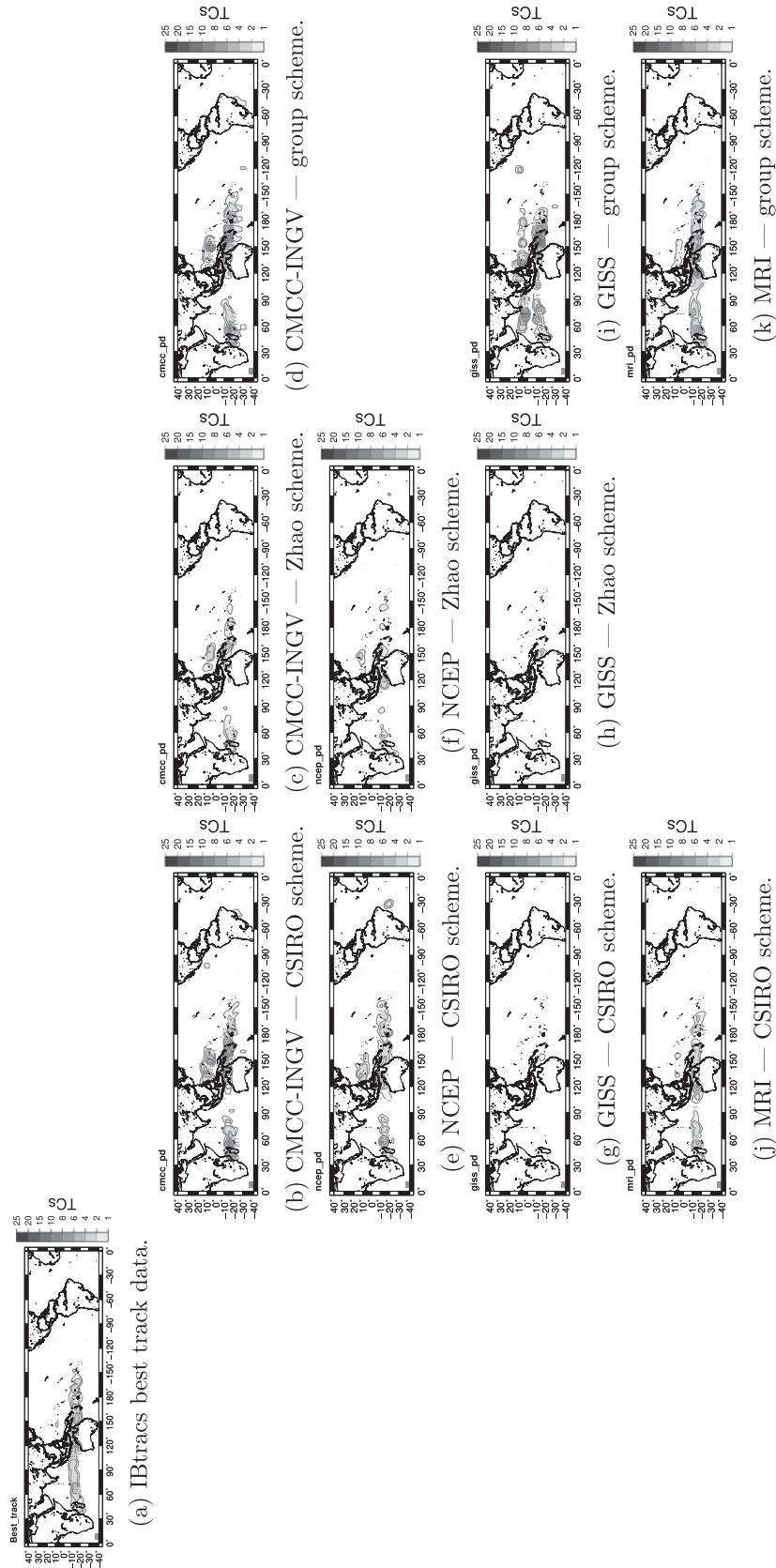


FIG. 1. January–March genesis density per  $4^\circ \times 4^\circ$  box per 20 years. (Note that NCEP and CMCC-INGV experiments are 10 years long; the genesis density shown for those models is double the total from the experiment.)



FIG. 2. As in Fig. 1, but for the July–September period.

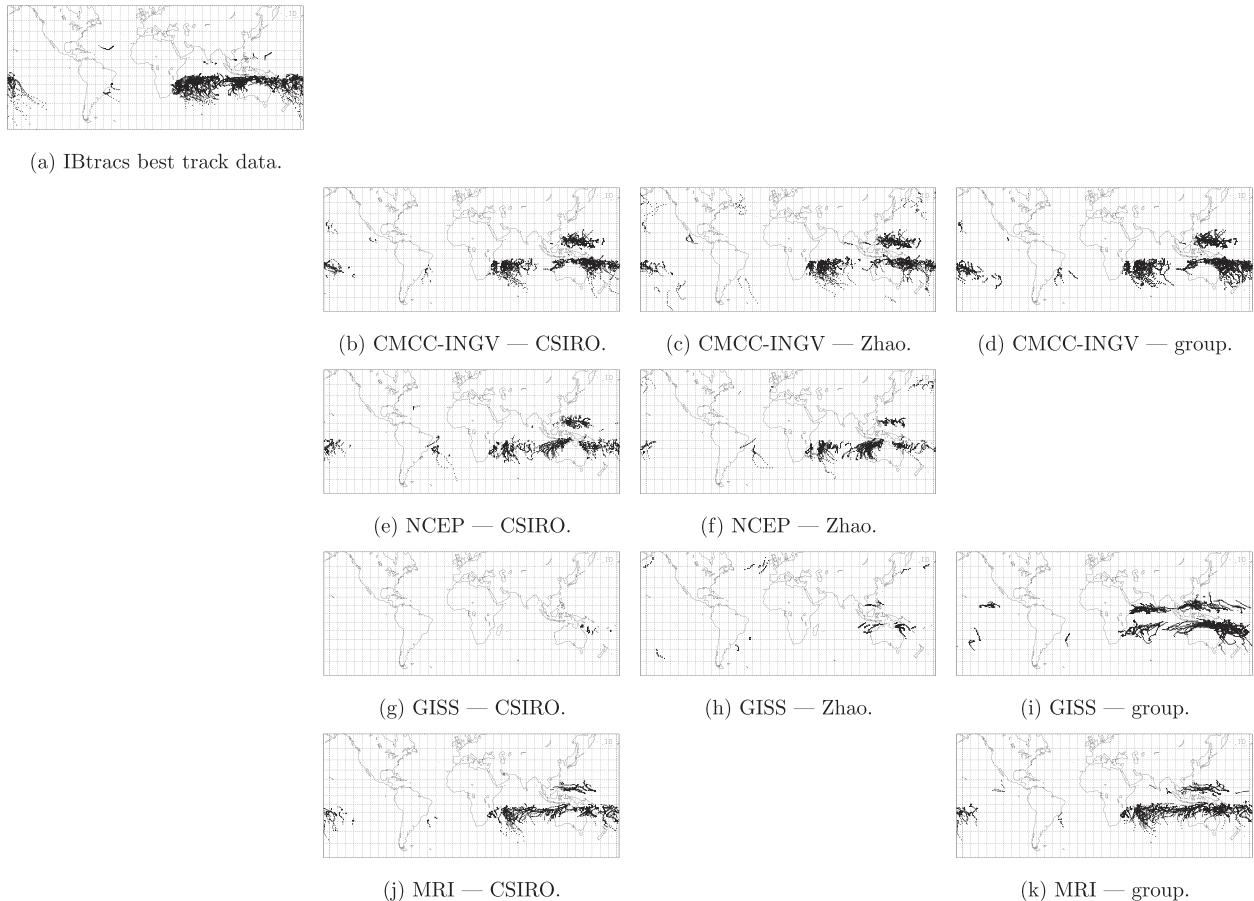


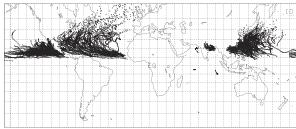
FIG. 3. January–March TC tracks; 10 years of data are used in all cases.

additional detections of those schemes that give higher genesis rates appear to be spread relatively evenly across the geographical distribution in most cases. This is promising, as it suggests—insofar as different ocean basins can be seen as analogous to different climate regimes—that the schemes may not differ substantially in their response to different climates.

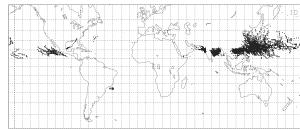
Figures 3 and 4 compare the present-day January–March (JFM) and July–September (JAS) tracks. The Zhao tracks for the CMCC-INGV model show a significant presence of extratropical storms. This is unsurprising, because these storms are not explicitly excluded (if forming equatorward of  $50^{\circ}$ ) in the Zhao scheme as they are in the modified CSIRO scheme. A moderate warm-core check will not exclude all such detections, because, as Walsh et al. (2014) point out, a subset of extratropical storms will evolve by the warm seclusion method proposed by Shapiro and Keyser (1990) and potentially possess a warm core for part of their lifetime. Figure 5 gives the latitudinal distribution of initial detections from the CMCC-INGV and NCEP models in the Zhao scheme. This figure confirms that the Zhao scheme is

effectively excluding extratropical storms in the NCEP model, but not the CMCC-INGV model. This suggests that the warm seclusion process may be more prominent in the CMCC-INGV model, or that tropical and extratropical storms are otherwise less distinguished in this model. The few extratropical detections that do occur in the NCEP model should not have a significant effect on the statistics. In the CMCC-INGV model, the larger numbers of extratropical storms (visible north of  $30^{\circ}\text{N}$  and south of  $30^{\circ}\text{S}$ ) may have some influence on detected TC statistics. Even in this case, however, the influence should be limited, as the extratropical storms occur largely in the winter hemisphere.

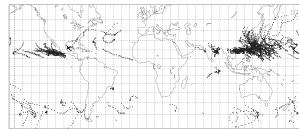
If the extratropical storms are disregarded, the tracks from the different schemes in Figs. 3 and 4 appear broadly similar in most cases. The Zhao tracks are in general more elongated, but few clear differences between the schemes not already described by differences in the detected genesis densities are apparent. Overall, the representation of the geographic pattern and tracks of TC activity in the current climate is reasonably similar in each model across schemes. Between models, the



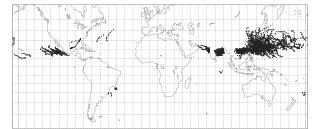
(a) IBTracs best track data.



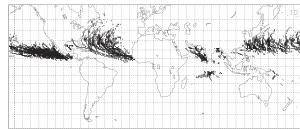
(b) CMCC-INGV — CSIRO.



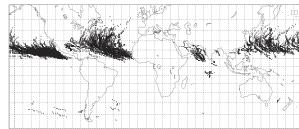
(c) CMCC-INGV — Zhao.



(d) CMCC-INGV — group.



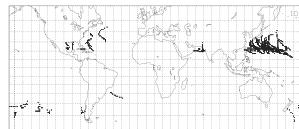
(e) NCEP — CSIRO.



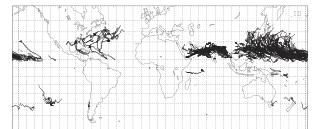
(f) NCEP — Zhao.



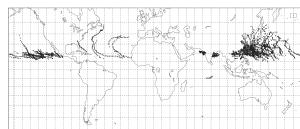
(g) GISS — CSIRO.



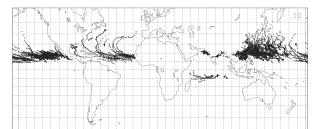
(h) GISS — Zhao.



(i) GISS — group.



(j) MRI — CSIRO.



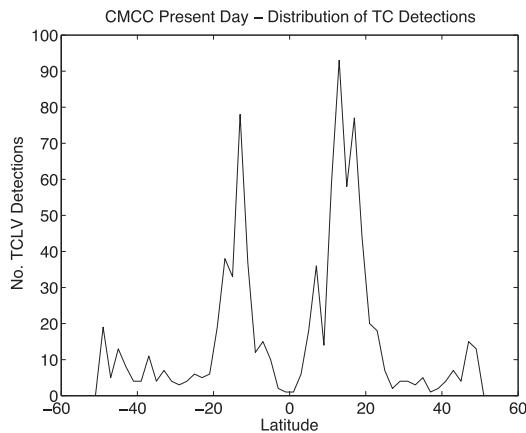
(k) MRI — group.

FIG. 4. July–September TC tracks; 10 years of data are used in all cases.

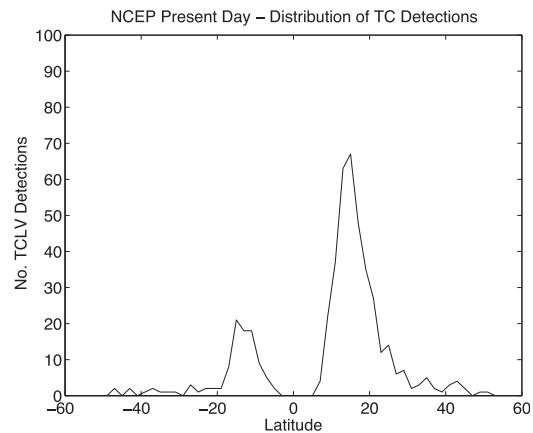
most notable difference is in the Atlantic basin for JAS, where NCEP performs well but all other models show very little genesis. We should also note that, unlike the IBTrACS observations, the model tracks do not show

extratropical transitions, as the tracking schemes used are configured to track only tropical storms.

The similarities in the present-day climatological geographic distributions from the different tracking



(a) CMCC-INGV.



(b) NCEP.

FIG. 5. Latitudinal distribution of initial TC detections in the Zhao scheme for two models.

TABLE 3. Correlations between monthly genesis time series for all models, experiments, and tracking scheme pairs. All correlations are statistically significant.

		CSIRO–Zhao	CSIRO–group	Zhao–group
GISS	Present day	0.501	0.357	0.225
	Double CO <sub>2</sub>	0.626	0.473	0.421
	SSTp2K	0.360	0.314	0.335
	SSTp2K 2×CO <sub>2</sub>	0.487	0.417	0.321
NCEP	Present day	0.688	—	—
	Double CO <sub>2</sub>	0.670	—	—
	SSTp2K	0.807	—	—
	SSTp2K 2×CO <sub>2</sub>	0.712	—	—
CMCC-INGV	Present day	0.447	0.780	0.464
	Double CO <sub>2</sub>	0.457	0.771	0.470
	SSTp2K	0.577	0.648	0.487
	SSTp2K 2×CO <sub>2</sub>	0.386	0.763	0.462
MRI	CO <sub>2</sub> -p SST-p	—	0.667	—
	CO <sub>2</sub> -f SST-p	—	0.602	—
	CO <sub>2</sub> -p SST-f	—	0.724	—
	SST-1.83K CO <sub>2</sub> -f	—	0.713	—

schemes within each model do not necessarily indicate good agreement between the schemes on a storm-by-storm basis. To better compare the results of the schemes in the present climate, we compare monthly time series of genesis between schemes in each model and experiment. The correlations between monthly genesis time series for both the present-day and altered forcing experiments are given in Table 3. All correlations are statistically significant. The highest correlation in the present-day experiments (0.78) is obtained between the CSIRO and the CMCC-INGV group scheme in the CMCC-INGV data. This is unsurprising, as the CMCC-INGV group used a variant of the CSIRO scheme. Correlations are lower between the other tracking scheme pairs for the CMCC-INGV data. The CSIRO–Zhao correlation for the NCEP model is high, and the MRI model data also show reasonable agreement between the two available genesis time series. Correlations for the GISS model are generally smaller, especially between the GISS group scheme and the other schemes. This is to be expected, as the GISS group scheme uses the substantially different relative threshold approach. The low overall TC genesis rate detected with the CSIRO and Zhao schemes in the GISS data may also contribute to the low correlation by allowing small variations in numbers to have disproportionate impact.

Overall, the detected genesis shows the best correspondence between available schemes for the NCEP model, with the CMCC-INGV and MRI models also showing some agreement across tracking schemes. This agreement across spatial and temporal scales suggests that these schemes may also respond similarly to changes

in the climate produced in the idealized future climate experiments for these models.

#### b. Future climate

Figure 6 shows the percentage changes in TC numbers in the three altered climate experiments as detected by each tracking scheme in each model. In the increased SST experiment (Fig. 6a), there is substantial disagreement in the direction and magnitude of the trend, which ranges from a decrease of 30% in the GISS model tracked by the Zhao scheme to an increase of around 15% in the same model when tracked by the group's own scheme. Furthermore, both these responses are statistically significantly to at least the  $p = 0.05$  level. It is clear that the different tracking methods are detecting substantially different trends in this case.

In the doubled CO<sub>2</sub> experiment, there is again a wide range of responses detected. The GISS model tracked with the CSIRO scheme shows the largest positive trend at around 14%, while the CMCC-INGV model tracked with CSIRO shows a decrease of a similar magnitude. There is again substantial variation in these results, although the majority of models and trackers show a decrease with magnitude less than 20%, with all statistically significant responses being decreases.

The combined SST/CO<sub>2</sub> experiment shows the best trend agreement across models and tracking schemes. In all but the GISS model, moderate decreases in TC frequency are indicated. The GISS model shows small increases with the CSIRO and group-specific tracking schemes, and a large decrease with the Zhao scheme. Again, all statistically significant responses are decreases.

It is clear from these results that the responses detected are in many cases dependent on the tracking

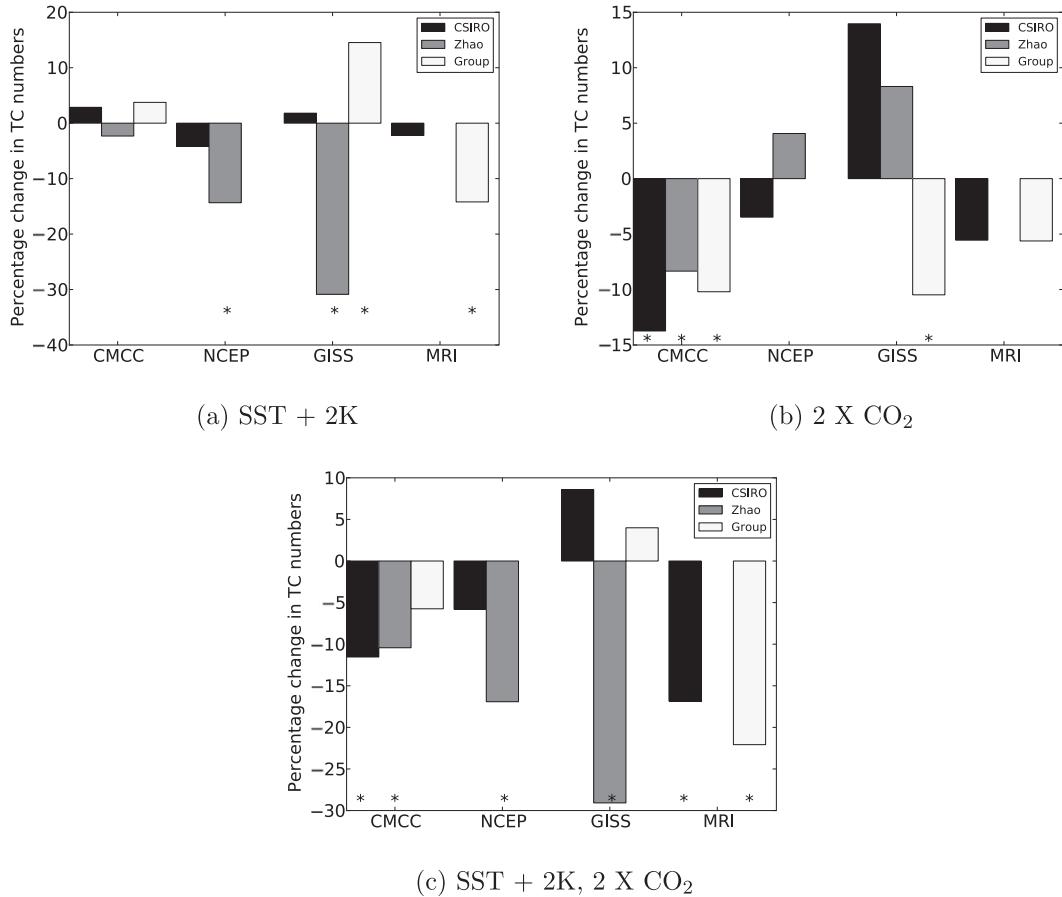


FIG. 6. Percentage change in TC numbers in each model for the three altered climate experiments relative to the present-day experiment, as tracked by the CSIRO, Zhao, and individual group schemes. Asterisks indicate statistical significance to at least the  $p = 0.05$  level.

scheme used. However, this is not the case for all models. In the MRI model, the different tracking schemes agree on the direction of the response in every experiment (although we must keep in mind that only two schemes are applied to this model). In the CMCC-INGV model, the responses in an experiment never differ between different tracking schemes by substantially more than 5%. Conversely, some models show much lower agreement between tracking schemes. The NCEP model shows divergent responses in the doubled CO<sub>2</sub> experiment and substantial differences in the magnitudes of the responses detected in the other experiments. For the GISS model, the schemes disagree over the direction of the trend in every experiment. In the increased SST experiment, the Zhao tracks show a decrease of around 30% while the group-supplied tracks show an increase of around 15%. This wide disparity is likely a function primarily of the low overall TC numbers in the GISS model when tracked with either the CSIRO or Zhao schemes. Low overall genesis could

allow small disagreements over TC numbers to appear as widely divergent responses.

As well as showing better agreement for some models, the different tracking schemes also show better agreement for some experiments than others. The combined SST/CO<sub>2</sub> experiment shows much better agreement between schemes within each model (as well as between models) than is seen in the increased SST experiments, with all tracking schemes in agreement for all but the GISS model.

These two factors suggest that differences in tracking scheme methods and parameters produce different sensitivities to both the differences in storm representation in different models, and the changes in storm activity in altered climate experiments. We will attempt to explain the reasons for these varied sensitivities in the next section.

### c. Threshold sensitivities

Differences in thresholds between the tracking schemes seem likely to account for a substantial proportion of the

TABLE 4. Correlations between monthly genesis time series for all models, experiments, and tracking scheme pairs after homogenization in thresholds for duration, wind speed, and latitude of formation. All correlations are statistically significant. Bold text indicates increased correlation compared to Table 3.

		CSIRO–Zhao	CSIRO–group	Zhao–group
GISS	Present day	<b>0.534</b>	<b>0.382</b>	<b>0.260</b>
	Double CO <sub>2</sub>	0.625	<b>0.508</b>	<b>0.510</b>
	SSTp2K	0.321	<b>0.386</b>	<b>0.423</b>
	SSTp2K 2×CO <sub>2</sub>	<b>0.580</b>	<b>0.557</b>	<b>0.336</b>
NCEP	Present day	<b>0.787</b>	—	—
	Double CO <sub>2</sub>	<b>0.732</b>	—	—
	SSTp2K	0.789	—	—
	SSTp2K 2×CO <sub>2</sub>	0.683	—	—
CMCC-INGV	Present day	<b>0.493</b>	0.767	<b>0.568</b>
	Double CO <sub>2</sub>	<b>0.524</b>	<b>0.829</b>	<b>0.532</b>
	SSTp2K	0.487	<b>0.653</b>	<b>0.513</b>
	SSTp2K 2×CO <sub>2</sub>	<b>0.486</b>	0.757	<b>0.486</b>
MRI	CO <sub>2</sub> -p SST-p	—	0.604	—
	CO <sub>2</sub> -f SST-p	—	0.543	—
	CO <sub>2</sub> -p SST-f	—	0.635	—
	SST-1.83K CO <sub>2</sub> -f	—	0.607	—

disagreement between schemes, with the remainder due to actual differences in the basic functioning of each scheme. The schemes used here differ in their duration criteria, wind speed criteria, and allowed latitudes of formation. We therefore filter the tracks produced by each scheme for each model to remove those storms that do not meet the strictest thresholds of any scheme for that model. For duration, this means removing all storms that last less than 2 days (the Zhao scheme threshold). For latitude, we filter all tracks forming poleward of 30° in order to remove the influence of the different treatment of latitude of formation in the different schemes. This latitude is chosen as it is the strictest cutoff applied at any point by the CSIRO scheme, which uses a variable phenomenon-based latitude cutoff, or by any of the other schemes. For wind speed, we remove those storms with maximum wind speeds below the strictest threshold among the various tracking schemes for that model. It should be noted that these changes do not completely account for the influence of the differing thresholds because of differences in how the thresholds operate in each model. For example, the CSIRO scheme requires its wind speed threshold to be met at every time step, whereas the Zhao scheme requires only that the threshold be exceeded for at least three (not necessarily consecutive) days. Such effects are regarded as more fundamental to the tracking scheme, and so we do not attempt to correct for them here.

If the differences between tracking scheme results are largely due to differing thresholds, then we might expect correcting for these differences to improve the correlations between tracker detection time series. Correlations for all experiments after homogenization in

duration, minimum wind speed, and latitude of formation are shown in Table 4. Generally, correlations are improved. For the present-day experiments, only two cases occur where correlation does not improve. The first is for the CMCC-INGV model with the CSIRO–group tracking scheme pair. In this case, the correlation was already high (0.780) and does not decrease substantially. The second is the MRI model in the CSIRO–group tracking scheme pair, where the reduction in correlation is also small. In the altered climate experiments, similar results are seen, with improved correlations in most cases, especially where correlations are initially low. We should also note that although correlations do improve in most cases, many of these improvements are small. This suggests that differences in the basic functioning of the tracking schemes remain significant in some cases when considering the details of detected activity in a single experiment.

To determine the relative importance of the different thresholds, we can examine the change in correlation with homogenization in each variable individually. The correlations between tracker pairs in each model in the original tracks, with each of duration, latitude, and minimum wind speed homogenized individually, and with all three factors homogenized together, are given for the present-day experiment in Fig. 7. No consistent pattern is evident in the contribution of the individual homogenizations to the combined improvement in correlations. In some cases, one threshold appears to provide all the improvement (e.g., duration for the CSIRO–Zhao pair in the NCEP model), while in others moderate improvements in several thresholds individually combine for a greater improvement in the combined

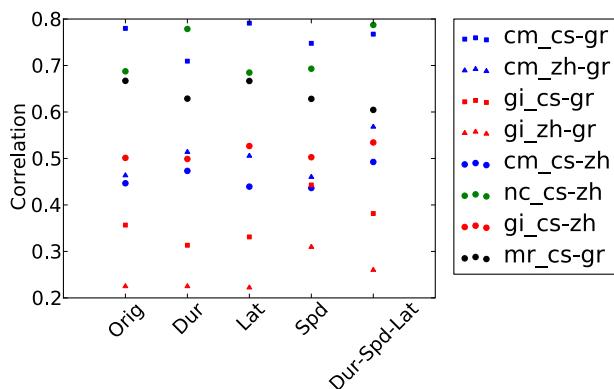


FIG. 7. Correlations between tracking scheme pairs for the present-day experiment in original tracks; with homogenization in each of duration, formation latitude, and wind speed; and with homogenization in all three variables together. Red points are for GISS model data, green for NCEP, blue for CMCC-INGV, and black for MRI. Circles indicate the correlation between CSIRO and Zhao tracks, squares between CSIRO and group-supplied tracks, and triangles between Zhao and group-supplied tracks.

homogenization (e.g., the Zhao–group pair in the CMCC-INGV model). In some cases homogenization in a single variable produces a larger improvement in correlation than is seen when all thresholds are homogenized together. This is seen most clearly in the case of the GISS model for the Zhao–group and CSIRO–group tracking scheme pairs when wind speed alone is homogenized, and also for the CMCC-INGV model with the CSIRO and group schemes when formation latitude alone is homogenized. However, the latter case is characterized by a high correlation in all cases with little to no homogenization improvement in general. The former case may be affected by the high degree of noise resulting from low detection numbers in GISS with both the CSIRO and Zhao schemes. In most cases, the combined homogenization produces the best available correlation. Similar results are observed in the changes in correlations with individual and combined homogenizations in the altered climate experiments (not shown).

Given these moderate improvements in agreement between tracking schemes with threshold homogenization, it is reasonable to expect some improvement in agreement on responses to altered climate forcing also. Figure 8 shows the responses in TC frequency in the idealized future climate experiments after track homogenization in duration, formation latitude, and wind speed. The level of agreement between tracking schemes on the signs of the responses in each model (i.e., increased or decreased TC genesis) is improved. After homogenization, the doubled  $\text{CO}_2$  experiment, which had shown disagreement between trackers in two of the four models, shows agreement between the trackers in

all models. The models themselves disagree, but this is not due to tracking methods. Rather, it appears that the clarification of the responses detected across tracking schemes may uncover fundamental disagreement between models here. However, the HWG simulations are too short to obtain statistical significance for many of these conflicting cases, and further work is required to confirm this. All statistically significant cases, after homogenization, show a reduction in the TC genesis rate. Improvement is also seen for the increased SST experiments, where trackers are brought into agreement for the CMCC-INGV and GISS models. Some reduction in agreement is also seen in this case, however. For the NCEP model, the CSIRO tracks move away from the Zhao results to show a very slight increase in frequency. The MRI model shows very little change, maintaining reasonable agreement between the CSIRO and group results. For the combined experiment, agreement between trackers was already good. Homogenization brings the magnitudes of the projected decreases closer together for the CMCC-INGV, NCEP, and MRI models. It does not bring the noisy GISS data into agreement. Statistical significance is obtained in fewer cases, likely because the homogenization reduces the number of samples in the datasets, making it more difficult to distinguish changes in the mean from statistical noise.

This increased agreement on the sign of responses for the CMCC-INGV, NCEP, and MRI models is associated with some increase in agreement on the magnitudes. The spread of responses (the difference between the largest/most positive response and the smallest/most negative response) is reduced moderately or unchanged with homogenization for most models and experiments. However, the spread does increase with homogenization in some cases—most noticeably, for NCEP in the increased SST experiment and MRI in the increased  $\text{CO}_2$  experiment. Furthermore, although the spread is reduced in many cases, many of these reductions are relatively small. For example, the combined increase experiment in the NCEP model gives a response of between  $-5\%$  and  $-17\%$  before homogenization. After homogenization, this spread reduces only marginally to between  $-5\%$  and  $-15\%$ . The lack of major improvements in agreement on the magnitudes of the responses indicates that some of the improvement in agreement on the signs of the responses is likely to be due to statistical noise rather than genuinely increased agreement. It is clear that homogenization brings some improvement in tracking scheme agreement, but large disparities remain in some cases. These disparities are smallest in the combined increase experiment, where both the CMCC-INGV and MRI models show spreads of less than 5% after

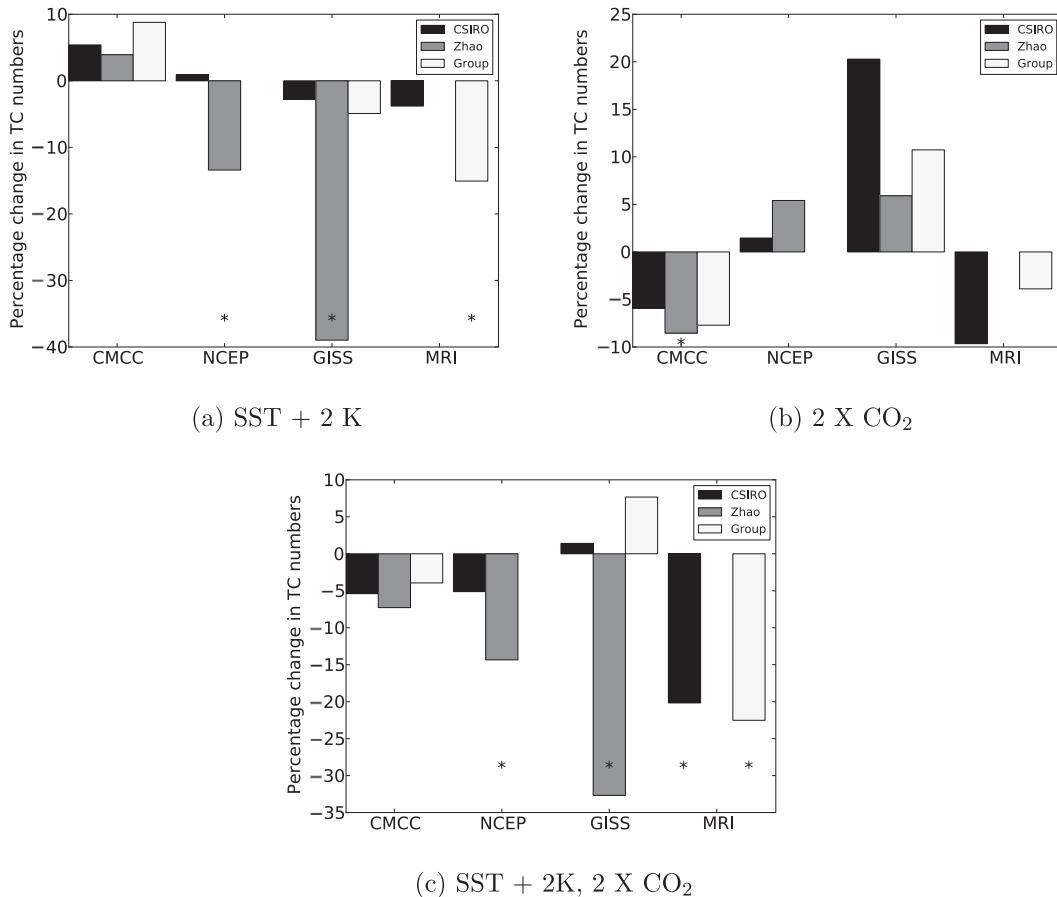


FIG. 8. Percentage change in TC numbers in each model for the three altered climate experiments relative to the present-day experiment, as tracked by the CSIRO, Zhao, and individual group schemes, after homogenization in duration, wind speed, and latitude of formation. Asterisks indicate statistical significance to at least the  $p = 0.05$  level.

homogenization, and NCEP shows a moderate spread of 10%. We have also examined the improvements in response agreement with homogenization in individual thresholds only (not shown), and find that for every experiment, the agreement in responses is greatest with homogenization of all three variables, indicating that the responses show significant sensitivity to the thresholds in all three factors.

Overall, the agreement obtained between the tracking schemes when considering responses to altered forcing is moderate. We will not here attempt to explain the responses seen in the altered climate experiments. Initial analysis can be found in [Held and Zhao \(2011\)](#).

#### 4. Discussion

Before discussing the results, it is worth considering differences between the model runs that could have complicated the relationships between the different tracking schemes. The most obvious possibilities are the

interannually varying (instead of climatological) SSTs used in the MRI model experiments, and the resolutions of the models. The use of interannually varying conditions is likely to alter the temporal and geographic distribution of TCs in the MRI experiments. However, it is not clear that this year-to-year variation should have any influence on the physical characteristics of any individual TC generated in the model. The structure of the individual storms is likely to be unaffected by the choice between climatological or varying conditions, and therefore the relationships observed between the different tracking schemes applied to the data should also be unaffected. The same argument applies to results in the altered climate experiments; while the responses in the MRI model may differ from those in the other models as a result of the different experimental design, there is no reason to expect the relationships between tracking schemes to vary.

When it comes to the effect of varying horizontal resolution, there may be more reason to believe that tracking scheme relationships may vary. As discussed in

the introduction, the characteristics of TCs generated in GCMs at these resolutions are substantially different from those of real-world TCs. It is likely that the spread of resolutions used here results in some spread in the realism of the TCs generated. The more realistic TCs with less ambiguous structures generated by the higher-resolution models are likely to be more consistently detected across different tracking schemes. This effect may account for the greater agreement between tracking schemes observed for the MRI model, which has the highest resolution of the models used here. However, substantial disagreements between schemes persist even in the MRI model. The cause of these disagreements between schemes for moderate-resolution models remains an important issue, even if moving to higher resolutions may alleviate the problem.

It is clear that the choice and setup of a tracking scheme will influence the results obtained for TC activity in moderate-resolution GCM experiments. Different tracking methods will usually not represent exactly the same storms even when they employ similar thresholds and show good general agreement on the total genesis rate. However, when it comes to detecting changes in the genesis rate under altered climate conditions, moderate agreement between tracking schemes can be obtained. When using uniform thresholds, different tracking schemes are likely to produce comparable results when comparing mean genesis rates between experiments. However, even using uniform thresholds, basic tracking scheme differences can still lead to moderate disparities in the numbers of detected storms in some cases.

Our results indicate that the choice of thresholds in duration, wind speed, and latitude of formation is critical. Small shifts in these thresholds can reverse the direction of a trend across experiments. Selection of meaningful thresholds is therefore crucial. Ideally, these thresholds could be chosen objectively to represent equivalent conditions to those defining a tropical cyclone in reality. Walsh et al. (2007) made progress in this direction for wind speeds by degrading real-world data to model resolutions and determining equivalent TC wind speed at these resolutions. Of course, the hurricane/tropical storm boundary is itself arbitrary, and testing the sensitivity of any simulated changes to this threshold remains advisable.

The duration threshold imposed by most tracking schemes differs from the wind speed threshold in that it does not correspond to any equivalent threshold in the assessment of real TCs. However, the duration distribution produced by most tracking schemes in most models differs substantially from the real-world distribution, with a substantial overestimation of the

prevalence of TCs lasting two days or less. This factor necessitates the introduction of a duration criteria. However, the prevalence of short-lived storms also leaves detected storm numbers highly sensitive to the exact duration threshold chosen. As a further result of this, detected future responses are likely to be dominated by changes at the short-lived end of the duration distribution, and changes in duration threshold can lead to reversed responses. Landsea et al. (2010) found that a similar problem was occurring in the Atlantic Hurricane Database (HURDAT) observational dataset for the Atlantic basin, where the observed increasing trend in TC numbers was found to be largely a result of increases in the recording of short-lived storms. Selecting a standard duration criteria to exclude all storms below a certain threshold may not necessarily be the most useful approach. Instead, the best approach may be to consider the entire range of storm durations as a matter of course, and divide into short- and long-lived storms during analysis. This approach would ensure that sensitivities to duration thresholds did not go unexamined.

The other variable we have filtered tracks for here, latitude of formation, does not permit the development of simple objective thresholds. Most tracking schemes simply cut off detections poleward of a certain latitude, but this is entirely subjective. Ideally, a tracking scheme should include no latitude threshold at all, as it would be able to differentiate TCs from extratropical storms dynamically. However, GCMs at moderate resolution do not reproduce these dynamical criteria sufficiently well. Latitude thresholds in some form may therefore be the only option, unless one is interested in detecting only those TCs with the most intense warm cores. The necessity of a latitude threshold for eliminating extratropical detections does not necessarily preclude accounting for shifts in the region of tropical genesis. The modified CSIRO scheme used here attempts this by tracking TC features over all latitudes, then counting those detections equatorward of the minimum in detections in each hemisphere caused by the extratropical ridge as TCs. This allows the region of TC formation to expand with expansion of the Hadley cell.

In general, the treatment of tracking scheme thresholds must be informed by the possibility that increases or decreases in TC frequency may not occur as simple amplifications or suppressions of the existing distribution, but rather through a shift in the distribution. This concept is clearly linked to the emerging consensus that overall TC numbers will decrease but with an increase in the most intense TCs (e.g., Knutson et al. 2010). It is therefore essential to test the sensitivity of any results to thresholds in the tracking scheme. We have established here that these thresholds take precedence over more

basic tracking scheme characteristics in determining variations in detected TC activity. Such sensitivity testing should therefore allow relatively objective representation of the TC activity present in model data. However, we must caution that even with uniform thresholds, our results indicate that it is still possible in some cases to observe substantial disparities between results from different tracking schemes. In some cases, sensitivity testing with multiple schemes may be advisable.

## 5. Conclusions

We have investigated the influence of tracking scheme differences on tropical cyclone (TC) activity and responses to altered climate conditions in the Hurricane Working Group experiments, which represent present climate, increased SST, increased CO<sub>2</sub>, and combined increased SST and CO<sub>2</sub> conditions. Our results indicate that when analyzing the detail of TC activity from a single experiment, basic differences in tracking scheme methods can produce substantially different numbers of TCs from the same model output, even when differences in TC detection thresholds between the schemes are accounted for. When analyzing the responses of TC numbers to the HWG perturbation experiments listed above, responses for each experiment varied moderately between tracking schemes. When differences in detection thresholds between schemes were removed, the signs and magnitudes of the changes in TC genesis became more similar among the tracking schemes used for each model, especially in the combined increase experiment. Some variation between tracking scheme responses can be explained by differences in the thresholds selected, although a substantial portion is also due to basic differences between the schemes. Partly because of the nature of storm distribution in climate models, small shifts in the choice of thresholds in wind speed, duration, or latitude of formation can lead to large changes in response magnitudes and directions. This result highlights the importance of objective threshold selection where possible, and wide-ranging sensitivity testing in other cases. It also indicates that sensitivity testing with multiple tracking schemes may be advisable in some cases.

*Acknowledgments.* The authors thank the U.S. Climate Variability and Predictability Program, the Australian Research Council Centre of Excellence for Climate System Science, and their respective institutions for supporting this work. Suzana Camargo, Dan Shaevitz, and Kevin Walsh are supported by NSF Grant AGS 1143959. Suzana Camargo is supported by NASA Grant NNX09AK34G. Enrico Scoccimarro acknowledges support from the Italian Ministry of Education,

University and Research and the Italian Ministry of Environment, Land and Sea under the GEMINA project. Thanks also to Sally Lavender of CSIRO for help with plotting software. We acknowledge also the support provided by Naomi Henderson, who compiled the HWG model output archive.

## REFERENCES

- Bengtsson, L., K. I. Hodges, M. Esch, N. Keenlyside, L. Kornbluh, J.-J. Luo, and T. Yamagata, 2007: How may tropical cyclones change in a warmer climate? *Tellus*, **59A**, 539–561, doi:10.1111/j.1600-0870.2007.00251.x.
- Camargo, S. J., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate*, **26**, 9880–9902, doi:10.1175/JCLI-D-12-00549.1.
- , and S. E. Zebiak, 2002: Improving the detection and tracking of tropical cyclones in atmospheric general circulation models. *Wea. Forecasting*, **17**, 1152–1162, doi:10.1175/1520-0434(2002)017<1152:ITDATO>2.0.CO;2.
- Held, I. M., and M. Zhao, 2011: The response of tropical cyclone statistics to an increase in CO<sub>2</sub> with fixed sea surface temperatures. *J. Climate*, **24**, 5353–5364, doi:10.1175/JCLI-D-11-00050.1.
- Horn, M., K. Walsh, and A. Ballinger, 2013: Detection of tropical cyclones using a phenomenon-based cyclone tracking scheme. *Proc. U.S. CLIVAR Hurricane Workshop*, Princeton, NJ, Geophysical Fluid Dynamics Laboratory. [Available online at [http://www.usclivar.org/sites/default/files/meetings/Walsh\\_clivar\\_gfdl\\_2013.pdf](http://www.usclivar.org/sites/default/files/meetings/Walsh_clivar_gfdl_2013.pdf).]
- Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc.*, **91**, 363–376, doi:10.1175/2009BAMS2755.1.
- Knutson, T. R., and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.*, **3**, 157–163, doi:10.1038/ngeo779.
- Landsea, C. W., G. A. Vecchi, L. Bengtsson, and T. R. Knutson, 2010: Impact of duration thresholds on Atlantic tropical cyclone counts. *J. Climate*, **23**, 2508–2519, doi:10.1175/2009JCLI3034.1.
- Li, F., W. D. Collins, M. F. Wehner, and L. R. Leung, 2013: Hurricanes in an aquaplanet world: Implications of the impacts of external forcing and model horizontal resolution. *J. Adv. Model. Earth Syst.*, **5**, 134–145, doi:10.1002/jame.20020.
- Mizuta, R., and Coauthors, 2012: Climate simulations using MRI-AGCM3.2 with 20-km grid. *J. Meteor. Soc. Japan*, **90A**, 233–258, doi:10.2151/jmsj.2012-A12.
- Murakami, H., and Coauthors, 2012: Future changes in tropical cyclone activity projected by the new high-resolution MRI-AGCM. *J. Climate*, **25**, 3237–3260, doi:10.1175/JCLI-D-11-00415.1.
- Roeckner, E., and Coauthors, 2003: The atmospheric general circulation model ECHAM5: Part 1: Model description. Max Planck Institute for Meteorology Rep. 349, 127 pp.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.
- Schmidt, G. A., and Coauthors, 2014: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.*, **6**, 141–184, doi:10.1002/2013MS000265.

- Scoccimarro, E., and Coauthors, 2011: Effects of tropical cyclones on ocean heat transport in a high-resolution coupled general circulation model. *J. Climate*, **24**, 4368–4384, doi:10.1175/2011JCLI4104.1.
- Shapiro, M. A., and D. A. Keyser, 1990: Fronts, jet streams, and the tropopause. *Extratropical Cyclones: The Erik Palmén Memorial Volume*, C. Newton and E. Holopainen, Eds., Amer. Meteor. Soc., 167–191.
- Stowasser, M., Y. Wang, and K. Hamilton, 2007: Tropical cyclone changes in the western North Pacific in a global warming scenario. *J. Climate*, **20**, 2378–2396, doi:10.1175/JCLI4126.1.
- 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, doi:10.1175/JCLI-D-12-00012.1.
- Tory, K. J., S. S. Chand, J. L. McBride, H. Ye, and R. A. Dare, 2013: Projected changes in late-twenty-first-century tropical cyclone frequency in 13 coupled climate models from phase 5 of the Coupled Model Intercomparison Project. *J. Climate*, **26**, 9946–9959, doi:10.1175/JCLI-D-13-00010.1.
- Vecchi, G. A., S. Fueglistaler, I. M. Held, T. R. Knutson, and M. Zhao, 2013: Impacts of atmospheric temperature trends on tropical cyclone activity. *J. Climate*, **26**, 3877–3891, doi:10.1175/JCLI-D-12-00503.1.
- 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, doi:10.1175/JCLI4074.1.
- , S. Lavender, 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.
- , F. Giorgi, and E. Coppola, 2014: Mediterranean warm-core cyclones in a warmer world. *Climate Dyn.*, **42**, 1053–1066, doi:10.1007/s00382-013-1723-y.
- Zhao, M., I. M. Held, S.-J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate*, **22**, 6653–6678, doi:10.1175/2009JCLI3049.1.