An overview of the extratropical storm tracks in CMIP6 historical simulations

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

An Overview of the Extratropical Storm Tracks in CMIP6 Historical Simulations

Matthew D. K. Priestley,[a] Duncan Ackerley,[b] Jennifer L. Catto,[a] Kevin I. Hodges,[c]
Ruth E. McDonald,[b] and Robert W. Lee[c]

[a] *College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom*
[b] *Met Office, Exeter, United Kingdom*
[c] *National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, United Kingdom*

ABSTRACT

The representation of the winter and summer extratropical storm tracks in both hemispheres is evaluated in detail for the available models in phase 6 of the Coupled Model intercomparison Project (CMIP6). The state of the storm tracks from 1979 to 2014 is compared to that in ERA5 using a Lagrangian objective cyclone tracking algorithm. It is found that the main biases present in the previous generation of models (CMIP5) still persist, albeit to a lesser extent. The equatorward bias around the SH is much reduced and there appears to be some improvement in mean biases with the higher-resolution models, such as the zonal tilt of the North Atlantic storm track. Low-resolution models have a tendency to underestimate the frequency of high-intensity cyclones with all models simulating a peak intensity that is too low for cyclones in the SH. Explosively developing cyclones are underestimated across all ocean basins and in both hemispheres. In particular the models struggle to capture the rapid deepening required for these cyclones. For all measures, the CMIP6 models exhibit an overall improvement compared to the previous generation of CMIP5 models. In the NH most improvements can be attributed to increased horizontal resolution, whereas in the SH the impact of resolution is less apparent and any improvements are likely a result of improved model physics.

1. Introduction

Climate models are our primary tool for investigating present and future climate. In this paper, the latest suite of models partaking in phase 6 of the Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016) will be evaluated for their representation of the midlatitude storm tracks and the characteristics of the cyclones within them. Previously, in the Fifth Assessment Report of the IPCC (AR5) it was stated that models can capture the general characteristics of the midlatitude storm tracks (Flato et al. 2013), with models having a more consistent representation for the Northern Hemisphere (NH) than the Southern Hemisphere (SH).

The models partaking in CMIP6 have provided new simulations of the climate system in an attempt to further understand past, present, and future climate variability and change. These new models represent improvements over those used in CMIP5, with the main developments being in model physics and increased resolution (e.g., Wu et al. 2019; Voldoire et al. 2019; Mauritsen et al. 2019; Andrews et al. 2019; Kawai et al. 2019). This study will evaluate this new generation of models and investigate if any of the uncertainties and biases from the previous generation of models have been reduced, if at all. Understanding how well storm tracks and extratropical cyclones are represented in models has considerable implications as extratropical cyclones are the dominant weather type in the midlatitudes and can have significant socioeconomic impacts through their associated extreme precipitation (Hawcroft et al. 2012) and severe winds (Browning 2004). They play a significant role in the

---

[⌖] Denotes content that is immediately available upon publication as open access.

[📎] Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-19-0928.s1.

Corresponding author: M. Priestley, m.priestley@exeter.ac.uk

general circulation by transporting large amounts of heat, moisture, and momentum poleward (Kaspi and Schneider 2013), and assist in maintaining the mean midlatitude westerly flow (Woollings et al. 2010). The midlatitude storm tracks are major features of both the NH (Chang et al. 2002) and SH (Trenberth 1991), particularly in the winter seasons.

The NH generally features two main areas of high synoptic activity, over both the Atlantic and Pacific Ocean basins respectively (Blackmon 1976; Hoskins and Hodges 2002). Conversely, the SH is characterized by a much more zonally symmetric storm track, mainly due to the absence of any significant landmass (Trenberth 1991; Hoskins and Hodges 2005). The storm tracks have generally been identified using two different methods, one being Eulerian, the other Lagrangian. The Eulerian perspective involves the time filtering of meteorological data to isolate variations in the synoptic band, usually taken to be the 2–6-day window (e.g., Blackmon 1976; Hoskins and Valdes 1990; Trenberth 1991). The Lagrangian storm track perspective involves the objective identification and tracking of synoptic features throughout their life cycle and generally results in a more detailed view of the storm tracks. The Lagrangian perspective has a further benefit of allowing storm frequency and intensity to be separated, which is not possible using the Eulerian method as the two properties are conflated. There have been numerous Lagrangian methods developed for this type of analysis (e.g., Murray and Simmonds 1991; Hodges 1994; Sinclair 1994; Wernli and Schwierz 2006), which generally use either mean sea level pressure (MSLP) or low-level relative vorticity at a high temporal resolution (typically 6 hourly) for generating the cyclone tracks.

In AR5 it was stated that there was low confidence in the magnitude of regional storm track changes and their impact on the regional surface climate (Christensen et al. 2013). These projections were largely made using the models in the phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012). There was a generally robust signal in the SH for a poleward shift of the storm track (Bengtsson et al. 2006; Kidston and Gerber 2010; Chang et al. 2012), whereas in the NH the pattern was less robust. Previous modeling studies have demonstrated evidence for a poleward shift in the Pacific sector (Catto et al. 2011), an extension into Europe from the North Atlantic storm track (Zappa et al. 2013b), and reduced activity over the Mediterranean (Nissen et al. 2014; Zappa et al. 2015). Almost all of the projections are marred by considerable intermodel spread and uncertainty (Harvey et al. 2012; Chang et al. 2012; Zappa et al. 2013b), which makes any estimations as to how impacts from extratropical cyclones will change under climate change very difficult (e.g., Chang et al. 2015; Osburn et al. 2018).

The historical representation of the storm tracks in the previous CMIP5 ensemble were rigorously studied, particularly those in the NH (e.g., Eichler et al. 2013; Zappa et al. 2013a; Colle et al. 2013; Lehmann et al. 2014; Lee 2015), with less attention being given to the SH (Chang et al. 2012; Chang 2017). In the evaluations of the CMIP5 storm tracks it was found that the North Atlantic winter storm track [December– February (DJF)] was generally too zonal, and extended too far into Europe compared to reanalyses (Zappa et al. 2013a; Colle et al. 2013), with this being linked to the underrepresentation of blocking (Zappa et al. 2014). The spatial pattern of the Pacific storm track is generally well represented (Yang et al. 2018), with some indications of a slight equatorward bias in the central and western North Pacific in DJF (Chang et al. 2012). Further to this, the CMIP5 models underestimate the total number of cyclones in both hemispheres, with this being most apparent in the NH, and also in each hemisphere's summer season (Lee 2015). There has been evidence for some reduction in biases from CMIP3 to CMIP5 (Chang et al. 2013; Zappa et al. 2013a), and it remains to be seen if there has been further improvement in biases in the CMIP6 ensemble.

In the SH winter season [June–August (JJA)] there is both a zonal and equatorward bias of the storm tracks (Kidston and Gerber 2010; Chang et al. 2012), with these errors possibly being linked to biases in cloud shortwave radiative processes (Ceppi et al. 2012), low-level orographic drag (Pithan et al. 2016), or the localized representation of blocking (Patterson et al. 2019). Atmospheric model resolution, and that of the coupled ocean, has also been shown to be of importance for model representation of the storm tracks in the NH (Woollings et al. 2010; Lee et al. 2018; Small et al. 2019). Higher-resolution ocean–atmosphere coupling appears to help somewhat with the zonal bias in the North Atlantic storm track, with this possibly being linked to the better representation of mesoscale moist processes within cyclones (Zappa et al. 2013a; Willison et al. 2013; Tamarin and Kaspi 2017).

Some of the most intense cyclones are those that rapidly deepen and are known as bomb cyclones and can be associated with both strong winds and extreme precipitation (Sanders and Gyakum 1980). They are most commonly located over the warm western boundary currents of the Atlantic and Pacific during the cold season in DJF (Sanders and Gyakum 1980; Roebber 1984) due to both the strong moisture availability and meridional temperature gradients commonly found in these locations, which results in enhanced baroclinicity and available potential energy. In the SH they are also commonly located near strong SST gradients (Lim and Simmonds 2002), but they are identified in all ocean basins with maxima over the Indian Ocean sector and

south of Australia (Reale et al. 2019). Seiler and Zwiers (2016) showed that the CMIP5 models robustly underestimate the frequency of bomb cyclones in the NH by approximately one-third. They also demonstrated that all models could capture the spatial pattern/location of bomb cyclones and their seasonality, with models that had the lowest fraction of bomb cyclones being the ones that underestimated the deepening rate by the largest amount.

It is hoped that advances in modeling capability in the CMIP6 models, in terms of both resolution and physics, will help reduce the biases in the extratropical storm tracks discussed above. In this study, the current state of the storm track representation in the new CMIP6 historical ensemble is quantified for both the NH and the (less studied) SH in both the winter and the summer seasons. The focus will be on all identified cyclones and the more intense bomb cyclones, with results compared and contrasted with previous results from CMIP5 and the latest global reanalysis products.

The specific aims of this paper are the following:

- To document the cyclone genesis and track density characteristics of the CMIP6 ensemble.
- To identify any improvements in the CMIP6 ensemble of models to the previous CMIP5 ensemble.
- To quantify the representation of cyclone intensity and growth rates for all cyclones and bomb cyclones in the CMIP6 and CMIP5 ensembles.
- To investigate whether increased resolution may be playing a key role in any model improvement of the simulated storm tracks, cyclone intensity and bomb cyclones.

By presenting results for both hemispheres and the solstice seasons, the overall aims of this paper are to document any improvements in storm track representation, and to provide a reference for those wishing to use these models to look at, or evaluate, any future projections or changes of the storm tracks in CMIP6. The biases in these models will be presented; however, for brevity, the physical processes that may be responsible for those biases will be discussed in a follow-up study.

The rest of this paper is structured as follows: in section 2 the data and methods used are discussed, and in section 3 the storm track representation is discussed and biases are highlighted. Finally, in section 4 the main conclusions and results are discussed.

2. Datasets, cyclone tracking, and bomb cyclone identification

a. CMIP6 models

In this study only the historical runs of CMIP6 models will be considered. These historical runs cover the period 1850–2014 and are fully coupled to a dynamical ocean and interactive sea ice, with evolving atmospheric forcings that are closely based on observed forcings (Eyring et al. 2016). The historical runs are mainly created for the purpose of model and climate system evaluation. The 20 CMIP6 models that are used in this study are listed in Table 1. The models vary largely in their setup and span a range of different atmospheric and oceanic resolutions. Most models have an atmospheric grid spacing of 100–200 km, with the lowest-resolution model having a horizontal grid spacing of over 300 km and the highest approximately 100 km. The models also vary in their vertical resolution, with some having a well-resolved stratosphere (top above 1 hPa). For consistency with the reanalysis products used for verification, only the years 1979–2014 are considered. Many of the modeling centers have created an ensemble of simulations of the historical period; however, so as to not disproportionately weight any models against each other only the first ensemble member of each model is taken (i.e., variant label r1i1p1f1). Due to the required temporal resolution of the cyclone identification and tracking algorithm used (discussed later), only models that have so far provided u and v (zonal and meridional winds, respectively) on pressure levels (850 hPa) every 6 h are used, as well as mean sea level pressure (MSLP) at the same temporal resolution. For the CMIP6 models and reanalyses the focus will be on just the winter and summer seasons in both the NH and SH.

To separate the models into higher- and lower-resolution groups, a "nominal resolution" is quoted that characterizes the models into common reference values regardless of their grid design (see Taylor et al. 2017). In this paper, two nominal resolution groups are used:

1) Nominal resolution of 100 km, where the mean maximum distance between grid points is <160 km.
2) Nominal resolution of 250 km, where the mean maximum distance between grid points is <360 km.

The use of the terms "high" or "higher" resolution in the text below refers to the 100-km nominal resolution CMIP6 models, of which there are 10 used in this study, and "low" or "lower" resolution refer to the CMIP6 models with the 250-km nominal resolution, for which there are also 10 models used herein. The nominal resolution of each model is included in Table 1.

b. CMIP5 models

To document the progression of the CMIP6 models from CMIP5, some comparison will also be shown by comparing the multimodel means. A full list of the CMIP5 models used to construct the multimodel mean can be found in Table S1 in the online supplemental

TABLE 1. List of CMIP6 models that have been used in this study. Columns 3 and 4 indicate the horizontal and vertical resolution of the atmospheric component of the model. Any spectral models are first stated by their truncation type and number. "T" stands for triangular truncation; "TL" stands for triangular truncation with linear Gaussian grid. The models with "C" refer to a cubed-sphere finite volume model, with the following number being the number of grid cells along the edge of each cube face. Models with "N" refer to the total number of two-gridpoint waves that can be represented in the zonal direction. Following any grid specification is the dimensions of the model output on a Gaussian longitude × latitude grid. The resolution stated in kilometers is the stated nominal resolution of the atmospheric component of the model from Taylor et al. (2017). (Expansions of acronyms are available online at http://www.ametsoc.org/PubsAcronymList.)

| | | Atmospheric resolution | |
Model name	Institution	Horizontal	Vertical
ACCESS-CM2	CSIRO-ARCCSS; Commonwealth Scientific and Industrial Research Organization, Australian Research Council Centre of Excellence for Climate System Science, Australia	N96; 192 × 144; 250 km	85 levels to 85 km
ACCESS-ESM1.5	CSIRO; Commonwealth Scientific and Industrial Research Organization, Australia	N96; 192 × 144; 250 km	85 levels to 85 km
BCC-CSM2-MR	BCC; Beijing Climate Center, China	T206; 320 × 160; 100 km	46 levels to 1.46 hPa
CNRM-CM6-1-HR	CNRM-CERFACS, Center National de Recherches Meteorologiques, center Européen de Recherche et de Formation Avancée en Calcul Scientifique, France	T359; 720 × 360; 100 km	91 levels to 78.4 km
EC-Earth3	EC-Earth Consortium	TL255; 512 × 256; 100 km	91 levels to 0.01 hPa
EC-Earth3-Veg	EC-Earth Consortium	TL255; 512 × 256; 100 km	91 levels to 0.01 hPa
GFDL CM4	NOAA-GFDL; National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, United States	C96; 360 × 180; 100 km	33 levels to 1 hPa
GISS-E2-1-G	NASA-GISS; Goddard Institute for Space Studies, United States	144 × 90; 250 km	40 levels to 0.1 hPa
HadGEM3-GC3.1-LL	MOHC; Met Office Hadley Centre, United Kingdom	N96; 192 × 144; 250 km	85 levels to 85 km
HadGEM3-GC3.1-MM	MOHC; Met Office Hadley Centre, United Kingdom	N216; 432 × 324; 100 km	85 levels to 85 km
IPSL-CM6A-LR	IPSL; Institut Pierre Simon Laplace, France	N96; 144 × 143; 250 km	79 levels to 40 km
MIROC6	MIROC; MIROC Consortium (JAMSTEC, AORI, NIES, R-CCS), Japan	T85; 256 × 128; 250 km	81 levels to 0.004 hPa
MPI-ESM1-2-HAM	HAMMOZ Consortium	T63; 192 × 96; 250 km	95 levels to 0.01 hPa
MPI-ESM1-2-HR	MPI-M, DWD, DKRZ; Max Planck Institute for Meteorology, Deutscher Wetterdienst, Deutsches Klimarechenzentrum, Germany	T127; 384 × 192; 100 km	95 levels to 0.01 hPa
MPI-ESM1-2-LR	MPI-M, AWI; Max Planck Institute for Meteorology, Alfred Wegener Institute, Germany	T63; 192 × 96; 250 km	47 levels to 0.01 hPa
MRI-ESM2-0	MRI; Meteorological Research Institute, Japan	TL159; 320 × 160; 100 km	80 levels to 0.01 hPa
NESM3	NUIST; Nanjing University of Information Science and Technology, China	T63; 192 × 96; 250 km	47 levels to 1 hPa
NorESM2-LM	NCC; NorESM Climate Modeling Consortium, Norway	144 × 90; 250 km	32 levels to 3 hPa
NorESM2-MM	NCC; NorESM Climate Modeling Consortium, Norway	288 × 192; 100 km	32 levels to 3 hPa
SAM0-UNICON	SNU; Seoul National University, Republic of Korea	288 × 192; 100 km	30 levels to ≈2 hPa

material. As the CMIP5 ensemble is larger than that of the current CMIP6, a subset of the CMIP5 models, denoted the like-for-like models (C5-C6 Like4Like), are also considered. The like-for-like CMIP5 models are those from the same modeling centers as the available CMIP6 models, which are highlighted in bold in Table S1. The C5-C6 Like4Like subset provides a fairer comparison between both model generations. All of the cyclone tracks of the CMIP5 models were produced as part of Lee (2015).

c. Reanalyses

Reanalysis products are commonly used to evaluate climate models, particularly the extratropical storm tracks (Hodges et al. 2011) due to their coherent spatial and temporal documentation of the state of the atmosphere and climate system. Reanalyses are produced by running historical, observationally constrained general circulation models (GCMs) in order to produce the best possible estimate of the atmospheric state over a given period of time (typically the last few decades; see https://reanalyses.org/ for more details and products). As reanalysis estimates based on different models can often differ slightly (Hodges et al. 2011), three reanalysis products from different centers are considered in this study:

1) European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 (Hersbach and Dee 2016).
2) NASA's Global Modeling and Assimilation Offices Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2017).
3) The Japan Meteorological Agency's Japanese 55-year Reanalysis (JRA-55; Kobayashi et al. 2015).

All three products currently provide data up to the end of 2018 and extend back to various start points (January 1958 for JRA-55, January 1979 for ERA5, and January 1980 for MERRA-2). All products provide Earth system data at 6-hourly time intervals. JRA-55 is produced at T319 resolution (available at $1.25° \times 1.25°$) with 60 vertical atmospheric levels up to 0.1 hPa; MERRA-2 is produced at $1/2°$ lat $\times 5/8°$ lon resolution with 72 vertical atmospheric levels to up to 0.01 hPa; ERA5 is produced at T639 ($0.28°$ $0.28°$) with 137 vertical atmospheric levels up to 0.01 hPa. Reanalyses have been shown to have similar storm track features, with deficiencies in older reanalyses being reduced through improved data assimilation methods that extract more information content from the observations, and resolution improvements (Hodges et al. 2011). Throughout this study ERA5 will be used as the main comparison reanalysis due to its higher horizontal and vertical resolution; however, at all stages, reference will also be made to storm track differences relative to the other reanalyses or a multi-reanalysis mean.

d. Feature tracking

The method of Hodges (1994, 1995, 1999) is used to objectively identify and track cyclones and is applied in the same way as described in Hoskins and Hodges (2002). This method uses 850-hPa relative vorticity (ξ_{850}) for the feature tracking. The vorticity is preferred as the tracking and identification variable as it is less influenced by the large-scale background state, is not an extrapolated field, and focuses on the smaller spatial scales (Hoskins and Hodges 2002). Using vorticity also allows for features to be identified earlier in their life cycle as occasionally a cyclone may have a distinct vorticity feature before a local pressure minimum develops. Before any identification takes place the ξ_{850} field is spectrally truncated to T42 and all wavenumbers less than or equal to 5 are removed to eliminate the planetary scales. This spectral filtering ensures that the synoptic scales are retained and also ensures that, regardless of input data, all tracking is done at a consistent resolution.

Cyclones are initially identified by searching for the grid point extrema that exceed $1 \times 10^{-5} s^{-1}$ (multiplied by -1 in the SH) and then refined using a B-spline interpolation and steepest ascent maximization. Cyclones are then grouped into tracks using a nearest neighbor approach. These are then refined by the minimization of a cost function for track smoothness subject to adaptive constraints. Tracks are filtered to retain those that persist for at least 48 h and travel more than 1000 km to focus on long-lived, mobile storms. Sensitivity analysis to the track life cycle criteria have been performed previously (Jung et al. 2012). MSLP values are assigned to tracks following Bengtsson et al. (2009) using the B-spline interpolation and minimization technique within a 5° radius from the cyclone center to identify the minimum MSLP value. Storm track statistics are calculated from the individual cyclone tracks using spherical nonparametric estimators to produce a number of different fields such as cyclone track, genesis, and lysis density, as well as mean intensity, growth/decay rates, and propagation speed (Hodges 1996). The cyclone track density statistics are provided in units of number density per month per unit area where the unit area is equivalent to a 5° spherical cap ($\approx 10^6 km^2$). Genesis densities are defined as the number density of cyclones per month per unit area when only the first identified time step of a cyclone is considered. Cyclone intensities are defined as the peak value of ξ_{850} (at T42 resolution) across the cyclone life cycle. This method has been widely applied for both reanalyses and climate models (e.g., Bengtsson et al. 2006, 2009; Catto et al. 2011; Zappa et al. 2013a; Tamarin and Kaspi 2017; Priestley et al. 2018) and is robust to the choice of input data (Hodges et al. 2011).

e. Bomb cyclones

Bomb (or explosively developing) cyclones are identified from our set of tracks following the widely used method introduced by Sanders and Gyakum (1980). Bomb cyclones are identified if they intensify by at least 1 bergeron (b) in 24 h, where a bergeron is defined as

$$b = \frac{\Delta p_{24h}}{24h} \frac{\sin(60°)}{|\sin(\varphi)|}. \qquad (1)$$

The cyclone MSLP change across 24 h (Δp_{24h}) is scaled by the average latitude of the cyclone over the 24-h period [$|\sin(\varphi)|$]. As bomb cyclones are most common in the winter seasons (Seiler and Zwiers 2016; Reale et al. 2019), the identification and analysis of bombs is only performed for those occurring in the winter season of each hemisphere (i.e., DJF in the NH and JJA in the SH).

3. Results

a. Storm tracks in the reanalyses

Before presenting an evaluation of the models, a brief description of the storm tracks in ERA5 and how they compare to the other reanalyses is given here as an indication of verification data uncertainty. The cyclone track (shading) and genesis (dashed contour lines) densities for the NH and SH solstice seasons are plotted in Fig. 1.

In the NH there are two very clear regions of high track densities in both winter (Fig. 1a) and summer (Fig. 1b), which are separated by orographic features:

- Region 1: From the high topography in East Asia (i.e., the Tibetan Plateau and the Altai–Sayan–Stonovoy range) into the western North Pacific (within the region denoted by the magenta line in Figs. 1a and 1b), and
- Region 2: From the lee of the Rocky Mountains in North America, across the North Atlantic into Scandinavia and northern Russia (within the red outline in Figs. 1a and 1b).

It is clear that the genesis and track densities in winter are higher in both domains (and throughout the hemisphere) than in the summer season (cf. Figs. 1a,b). Moreover, the main storm tracks are also displaced more equatorward in the winter than the summer, which is particularly visible in both the eastern North Atlantic and the central North Pacific. The Mediterranean storm track is also more active in winter than summer. The features described above for the NH are consistent with corresponding assessments of other reanalyses by Hodges et al. (2011) and Hoskins and Hodges (2002, 2019).

In the SH, the storm track is a more continuous feature than in the NH due to the reduced presence of land and topography acting as a barrier to cyclone propagation (see Figs. 1c,d). In summer (Fig. 1c) the storm track is annular in structure with the highest track densities between 50°–70°S.

In the winter (Fig. 1d) the symmetric, annular pattern is less evident as each genesis region downstream of the Andes is farther poleward than the previous one and the cyclones also propagate toward higher latitudes in general. Therefore, the poleward preference for cyclogenesis and propagation results in a spiral-like pattern of track density toward Antarctica (as documented previously; Hoskins and Hodges 2005). There is also another local track density maximum in the southern Pacific, which extends from the east coast of Australia to South America that is associated with the subtropical jet (see Inatsu and Hoskins 2006; Hoskins and Hodges 2005; Hodges et al. 2011).

The structure and strength of the storm track may depend on the choice of reanalysis used, and therefore the cyclone track density statistics for JRA-55 and MERRA2 have also been produced (Fig. S1). The main spatial structure of the storm tracks and the overall cyclone frequency are consistent across the reanalyses, particularly in the NH and also for the summer seasons in both hemispheres. The average difference in the number of tracks for MERRA2 (and JRA-55) relative to ERA5[1] in the NH is −2.1% (−2.2%) during DJF and +1% (−0.1%) during JJA. In the SH, there are fewer cyclones in both the summer and winter seasons for MERRA2 and JRA-55 relative to ERA5. The average difference in number of tracks in MERRA2 (and JRA-55) relative to ERA5 in the SH is −0.9% (−4%) during DJF and −1.7% (−1%) during JJA. The larger differences between reanalyses in the SH have been documented previously (Hodges et al. 2011); however, the differences presented here are smaller than those previously estimated. These smaller differences between reanalysis products are likely to be driven by further improved horizontal resolution, forecasting capabilities, and observation assimilation in both hemispheres compared to those analyzed in Hodges et al. (2011). Despite the disparities noted above, the general structure of the storm tracks is consistent across reanalyses and any would be suitable for evaluating the CMIP6 models. For brevity, only anomalies relative to ERA5 will be shown and discussed in detail for spatial cyclone statistics (e.g., see Fig. 3). Furthermore, ERA5 is primarily chosen due to its superior horizontal resolution over JRA-55 and MERRA2.

b. Representation in the CMIP6 and CMIP5 ensembles

1) NORTHERN HEMISPHERE WINTER (DJF)

To provide a broader evaluation of cyclogenesis rates in the reanalyses and models, statistics of the number of

---

[1] Calculated as the difference in the mean MERRA2 and JRA-55 number of cyclones forming per season poleward of 30°N/S relative to ERA5.
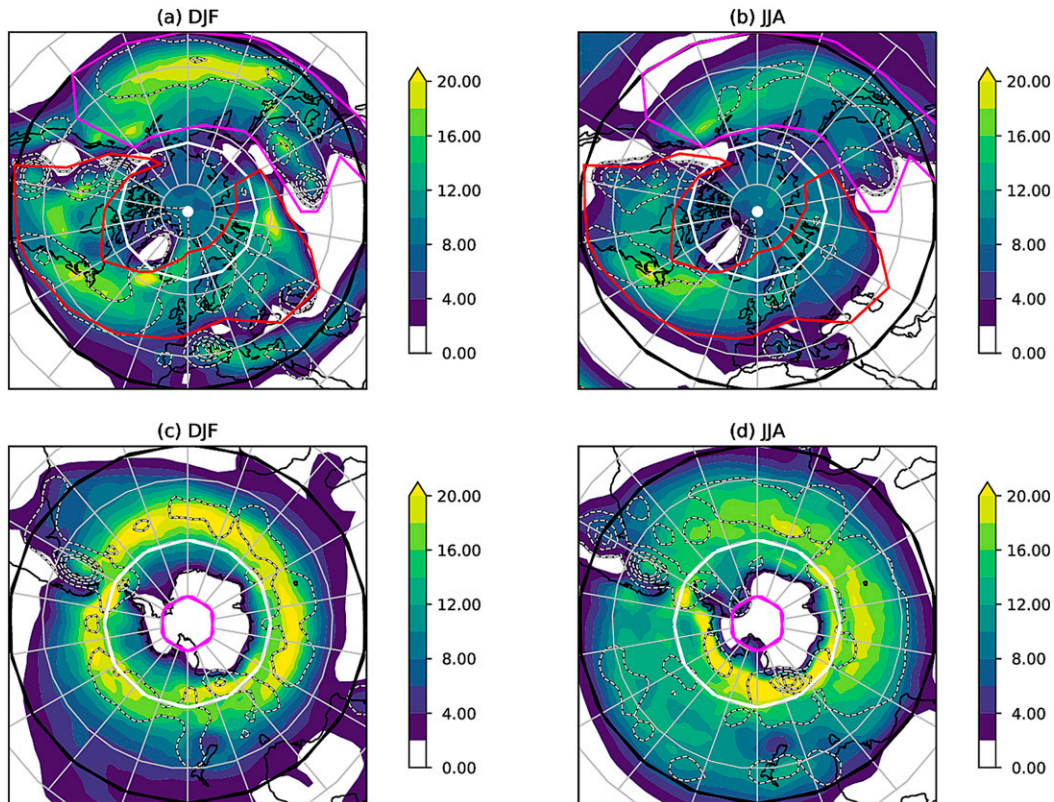
FIG. 1. Track (shading) and genesis (dashed contours) densities from ERA5 for the winter and summer seasons in both the NH and SH: (a) NH DJF, (b) NH JJA, (c) SH DJF, and (d) SH JJA. Units are number of cyclones per month per 5° spherical cap. Genesis density contours are plotted, in steps of 1, from 1 to 4 cyclones per month per 5° spherical cap. In (a) and (b) the black line is at 30°N, the white line at 65°N, the magenta polygon is for region 1 (see text in section 3a), and the red polygon for region 2 (also see text in section 3a). For (c) and (d) the black line is at 30°S, the white line is at 60°S, and the magenta line is at 80°S.

cyclones forming (i.e., cyclogenesis) within four regions of the NH are plotted in Figs. 2a(i)–(iv). The regions correspond to those plotted in Figs. 1a and 1b and specifically are as follows:

1) Poleward of 30°N: A measure of the hemispheric extratropical cyclone activity (poleward of the black circle, Figs. 1a,b).
2) Poleward of 65°N: A measure of Arctic cyclone activity (poleward of the white circle, Figs. 1a,b).
3) Region 1 from section 3a: A measure of cyclone activity from East Asia to the northeast Pacific Ocean (within the magenta polygon, Figs. 1a and 1b; also denoted ASPAC).
4) Region 2 from section 3a: A measure of cyclone activity from lee genesis in the Rockies, across the North Atlantic, and into Scandinavia/Siberia (within the red polygon, Figs. 1a and 1b; also denoted AMATSI).

Region 1 and region 2 extend beyond the North Atlantic and North Pacific storm tracks to capture the continuum

of storm generation and decay that occurs between the natural topographical barriers in East Asia and western North America [as discussed in Hoskins and Hodges (2002)]. Finer regional details within these four domains above will be discussed later in this section.

The cyclogenesis rates for the combined reanalyses are plotted as the red box in Fig. 2a with the median values also given in Table 2. The yellow bar is the median, the notches on the boxes are the 5%–95% confidence intervals on the median, the solid black horizontal lines denote the range of those intervals, and the black dashed lines denote the upper and lower quartiles [i.e., the interquartile range (IQR)]. The genesis rates for the full CMIP6 ensemble combined (ALL_CMIP6; teal), high-resolution ensemble (CMIP6_NR_100; blue), and the low-resolution ensemble (CMIP6_NR_250; cyan) are shown. To both identify model improvements and highlight unresolved issues, it is important to compare the results of the CMIP6 ensemble against CMIP5 and so the CMIP6-CMIP5 like-for-like model ensemble
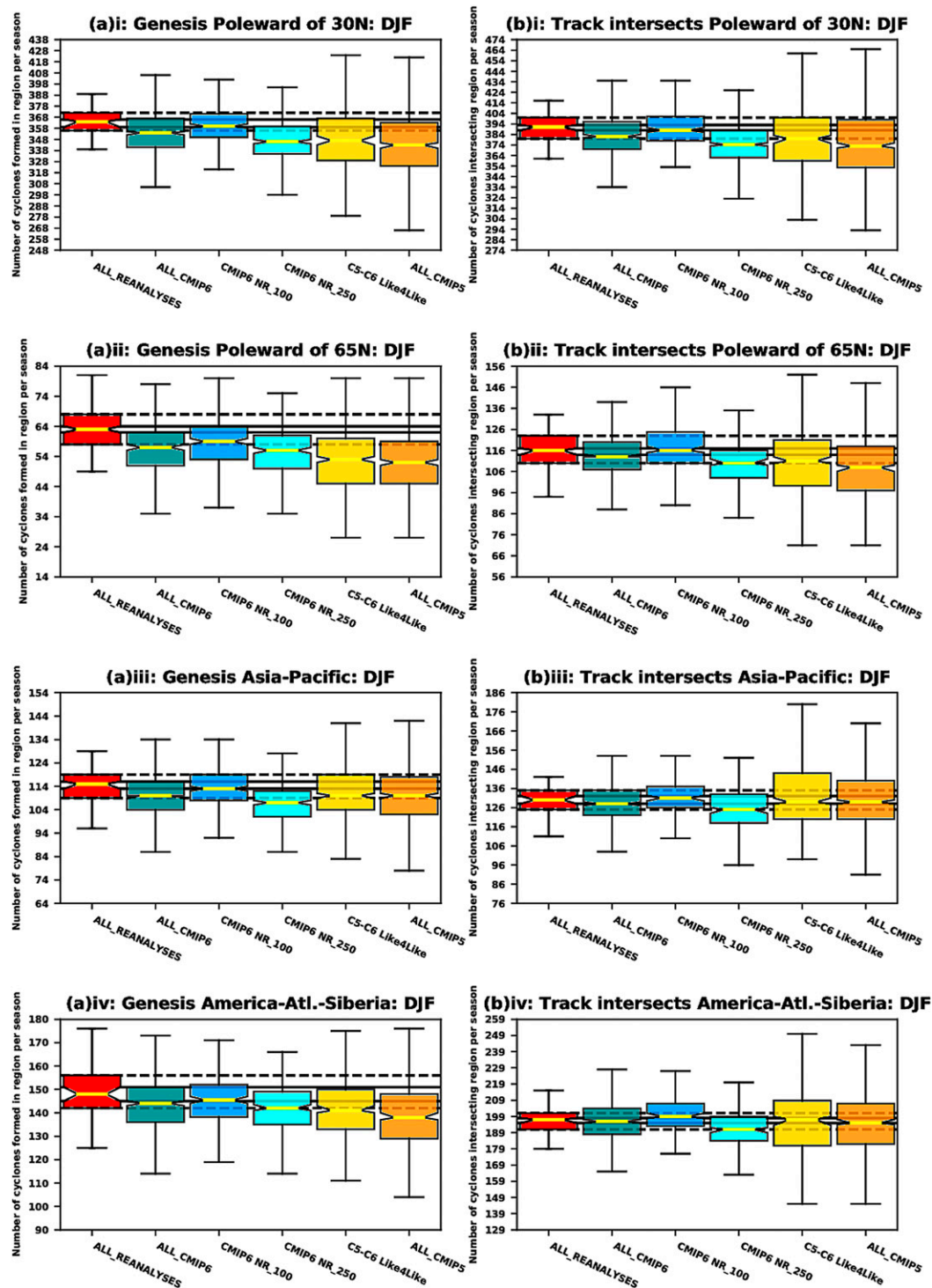
FIG. 2. Boxplots of (a) the number of cyclones forming within or (b) tracking partially/completely through different geographical domains of the NH for DJF. The regions are (i) poleward of 30°N, (ii) poleward of 65°N, (iii) the Asia–Pacific sector, and (iv) the American–Atlantic–Siberian sector. Results are shown for all reanalyses combined (ALL_REANALYSES; red), all CMIP6 models combined (ALL_CMIP6; teal), high-resolution CMIP6 models (CMIP6_NR_100; blue), low-resolution CMIP6 models (CMIP6_NR_250; cyan), CMIP5/CMIP6 like-for-like models (C5-C6 Like4Like; gold boxes), and all CMIP5 models combined (ALL_CMIP5; orange). The yellow lines in the boxes are the median, and the boxes extend to the 25th and 75th percentiles. Whiskers extend to the 1.5 times the interquartile range (IQR). Notches on the boxes represent the 5%–95% confidence range on the median, based on 10 000 bootstrap resamples. The black solid lines extend horizontally from the notches and the black dashed lines extend from the upper and lower quartiles on the reanalysis box in each panel. Units for all boxes and all panels are cyclones per season.

TABLE 2. DJF median genesis (rows 3–6) and regional track intersection (rows 8–11) totals for the NH from reanalyses, and each grouping of CMIP6 and CMIP5 models. The numbers in the parentheses are the differences relative to the reanalyses (%). Bold numbers indicate that the multimodel ensemble median is significantly different from the reanalyses, italicized values indicate that the CMIP6 median is significantly different from the C5-C6 Like4Like models' median, and two asterisks (**) denote that the CMIP6 multimodel median is significantly different from the full CMIP5 multimodel median (significance is achieved for $p \leq 0.05$).

Region	Reanalyses	ALL_CMIP6	CMIP6 NR_100	CMIP6_NR_250	C5-C6 Like4Like	ALL_CMIP5
Genesis						
Poleward 30°N	364	*354 (−2.74)***	360 (−1.10)	**346 (−4.95)**	**347 (−4.67)**	**343 (−5.77)**
Poleward 65°N	63	*57 (−9.52)***	**59 (−6.35)**	**56 (−11.11)**	**53 (−15.87)**	**52 (−17.46)**
ASPAC	115	**110 (−4.35)**	113 (−1.74)	**107 (−6.96)**	**110 (−4.35)**	**110 (−4.35)**
AMATSI	148	*144 (−0.51)***	146 (−1.35)	**142 (−4.05)**	**141 (−4.73)**	**138 (−6.76)**
Tracks						
Poleward 30°N	391	**382 (−2.30)***	388 (−0.76)	**375 (−4.09)**	**380 (−2.81)**	**373 (−4.60)**
Poleward 65°N	116	**113 (−2.59)***	116 (0.00)	**110 (−5.17)**	**111 (−4.31)**	**108 (−6.90)**
ASPAC	131	**128 (−2.29)***	131 (0.00)	**125 (−4.58)**	129 (−1.53)	129 (−1.53)
AMATSI	197	196 (−0.51)	**199 (+1.02)**	**191 (−3.05)**	197 (0.00)	195 (−1.01)

(C5-C6 Like4Like; gold) and the full CMIP5 ensemble (ALL_CMIP5; orange) are also plotted in Fig. 2. Boxplots for the individual CMIP6 models and the combined reanalyses are also given for reference in the supplemental material. As with the reanalyses, the median genesis rates for each model grouping are given in Table 2 where Mood's median test[2] is applied to each relative to the reanalyses and between ALL_CMIP6 and both C5-C6 Like4Like and ALL_CMIP5.

For the NH (poleward of 30°N; Fig. 2a), there are typically 364 (Table 2) cyclones generated during DJF for the 1979/80 to 2013/14 period in the reanalyses. There is a significant ($p \leq 0.05$) underestimation for ALL_CMIP6, which is primarily due to the poorer performance of the lower-resolution models [CMIP6_NR_250; see Fig. 2a(i) and Table 2]. There is clear improvement poleward of 30°N in CMIP6 relative to CMIP5 as the median genesis rates for C5-C6 Like4Like and ALL_CMIP5 lie below the lower quartile of (and are significantly lower than; Table 2) the reanalyses and the CMIP6 median (Fig. 2a).

The CMIP6 models (statistically significant) underestimate cyclogenesis rates at high latitudes (>65°N) in DJF where model medians lie on or outside the reanalyses' lower quartile, which is also visible for the lower-resolution CMIP6 models [Fig. 2a(ii) and Table 2]. Nonetheless, the median for the higher-resolution models [blue box, Fig. 2a(ii)] lies within the reanalyses' interquartile range and suggests that the representation of cyclogenesis at high latitudes improves with increased resolution, although the number is significantly lower than the reanalyses (Table 2). Moreover, the C5-C6 Like4Like and ALL_CMIP5 median genesis rates lie below the CMIP6 median and the lower quartile of the

reanalyses (both significant; Table 2), which indicates an improvement for CMIP6. Similarly, the higher-resolution CMIP6 models perform better on average than the lower-resolution models across the Asia–Pacific region during DJF [cf. blue and cyan boxes in Fig. 2a(iii) and Table 2] with no improvement relative to CMIP5 for the low-resolution models. For the broad American–Atlantic–Siberian domain, there is a clear improvement for CMIP6, regardless of resolution, over CMIP5 [Fig. 2a(iv) and Table 2]; however, the genesis rate for ALL_CMIP6 is significantly lower than the reanalyses' estimate [Fig. 2a(iv) and Table 2].

The genesis rates plotted in Fig. 2a do not provide information on the overall cyclone activity in each domain described above as cyclones may form outside these regions and propagate into them. Therefore, the number of cyclone tracks intersecting each subdomain (i.e., those generated within plus those generated outside that move into the domain) in Figs. 1a and 1b are given in Figs. 2b(i)–(iv) and Table 2. The median number of tracks intersecting the NH-wide domain (poleward of 30°N) during DJF in the reanalyses is 391. Given that the DJF genesis rate median is 364 (see above), this indicates that 93% of cyclones that propagate poleward of 30°N also originate there. A similar proportion of genesis to total tracks crossing into the region poleward of 30°N is also seen for ALL_CMIP6 (93%). For tracks intersecting the >30°N domain, the CMIP6 models outperform CMIP5 [Fig. 2b(i) and Table 2]. The CMIP6_NR_100 median is statistically indistinguishable from the reanalyses' median whereas the median for CMIP6_NR_250 is significantly lower [Fig. 2b(i) and Table 2].

At high latitudes (>65°N) in DJF [Fig. 2b(ii) and Table 2], the median genesis rates are lower than the track intersection numbers for the reanalyses (only 54% of cyclones in the Arctic are formed there). The median numbers of tracks entering the Arctic are significantly

---

[2] See https://www.statisticshowto.datasciencecentral.com/moods-median-test/ for details.

too low for CMIP6_NR_250 and both CMIP5 model groups, which is consistent with their relatively low genesis rates [Figs. 2a(ii) and b(ii) and Table 2]; however, the ratios of genesis to total tracks in domain are also low in these three groups (≈50%) relative to the reanalyses. Interestingly, the median number of tracks intersecting poleward 65°N in CMIP6_NR_100 group is equal to that of the reanalyses [Fig. 2b(ii) and Table 2] whereas the median genesis rates lie within the lower quartile of the reanalysis estimates [Fig. 2a(ii) and Table 2]. Therefore, with reduced genesis and a consistent number of tracks, it may be that there are slightly too many cyclones propagating into the Arctic in the CMIP6_NR_100 models relative to the reanalyses. The CMIP6_NR_100 models, by comparison, represent the Asia–Pacific region well for both genesis and track intersecting but less so in the American–Atlantic–Siberian domain [Figs. 2b(iii) and (iv) and Table 2].

The discrepancy between genesis-in-domain and track-domain-intersection numbers described above may be due to the regions of cyclogenesis being displaced slightly outside the specified domains in the models compared to the reanalyses. Therefore, the focus now is on the finer regional details of the track densities simulated by the models.

For NH winter (DJF), the highest track densities are found over the North Atlantic and North Pacific with other regions of relatively high track densities over North America, the Mediterranean/Middle East, Scandinavia/northern Russia, and East Asia in the CMIP6 multimodel mean (Fig. 3a). The difference between the CMIP6 multimodel mean and ERA5 is plotted in Fig. 3b (shading) with stippling denoting high model consensus (more than 80% model agreement) on the sign of the differences. Over the eastern and central North Pacific, cyclones track too far equatorward (relative to ERA5) as indicated by the positive and negative anomalies separated by the 40°N latitude band (Fig. 3b), with this being a more robust pattern farther east. The eastern North Pacific negative biases are less prominent and less expansive in the higher-resolution models (Fig. 3c) than at lower resolution (Fig. 3d); however, the positive bias to the south is larger at higher resolution. For the North Atlantic, there is high consensus for positive track density biases running southwest to east-northeast from 30°N, 60°W into Europe with negative biases poleward of that band (Fig. 3b), which indicates the storm track is too zonal relative to ERA5 and cyclones propagate too far into Europe. Unlike the North Pacific, the North Atlantic biases are similar in both magnitude and structure for both the high-resolution models and the low-resolution models. Nonetheless, there is a larger underrepresentation on the northern flank of the storm track, and therefore a stronger dipole anomaly, for the

low-resolution models (cf. Figs. 3c,d). There are also large negative track density biases over the Mediterranean/Middle East, North America, northern Russia, and northeast Asia, which are apparent regardless of resolution relative to ERA5 (Figs. 3b–d). With the underrepresentation in track frequency over the Mediterranean, there is also an increase in tracks to the north (east of the Alps). As Mediterranean cyclones are commonly driven through interactions with orography, the biases surrounding the Mediterranean in Figs. 3b–d could point to errors in interaction of the mean flow from the North Atlantic with the Alps, resulting in these track density anomalies.

Biases with high model consensus that are visible in both the CMIP6 and CMIP5 occur over North America, the North Atlantic to Europe, northern Russia, the Mediterranean to the Middle East, and northeast Asia (cf. Figs. 3b,e). There is evidence of the biases being reduced slightly in the regions described above for CMIP6 relative to CMIP5 (Fig. 3f), particularly across the northeastern North Atlantic and western Europe, where a majority of models demonstrate a reduction in the zonal bias and eastward extension. The reduced extension into Europe is likely a result of improved blocking frequencies in CMIP6 compared to CMIP5 (Schiemann et al. 2020); however, the models still underestimate blocking relative to the reanalyses, which reflects the storm track biases. Schiemann et al. (2020) also noted an improvement in blocking representation with resolution up to 25 km, yet our results show no marked improvement in storm track biases over northwestern Europe with higher resolution among the CMIP6 models. Therefore, resolution higher than 100 km may be required for significant reduction in the storm track biases for the European region. The reduction in the zonal bias across the North Atlantic from CMIP5 to CMIP6 is partly related to a reduction in genesis latitude biases in the western North Atlantic sector (30°–60°N, 40°–90°W; not shown). The median genesis latitude of CMIP5 models had an equatorward bias of 0.33° relative to ERA5, whereas the CMIP6 median bias is too poleward by 0.23°. Despite the improvement in genesis latitude, a zonal bias relative to ERA5 still remains, which is likely due to deficiencies in the poleward propagation of cyclones (e.g., Tamarin and Kaspi 2016, 2017). There is also little improvement in the representation of the North Pacific storm track, with there actually being more of an equatorward bias in the east of the basin in CMIP6 than CMIP5 (cf. Figs. 3b,e,f), which is robust across the models. In other regions of the North Pacific, the model consensus is weak as to the improvement compared to CMIP5. Both CMIP5 and CMIP6 models exhibit an equatorward bias in the
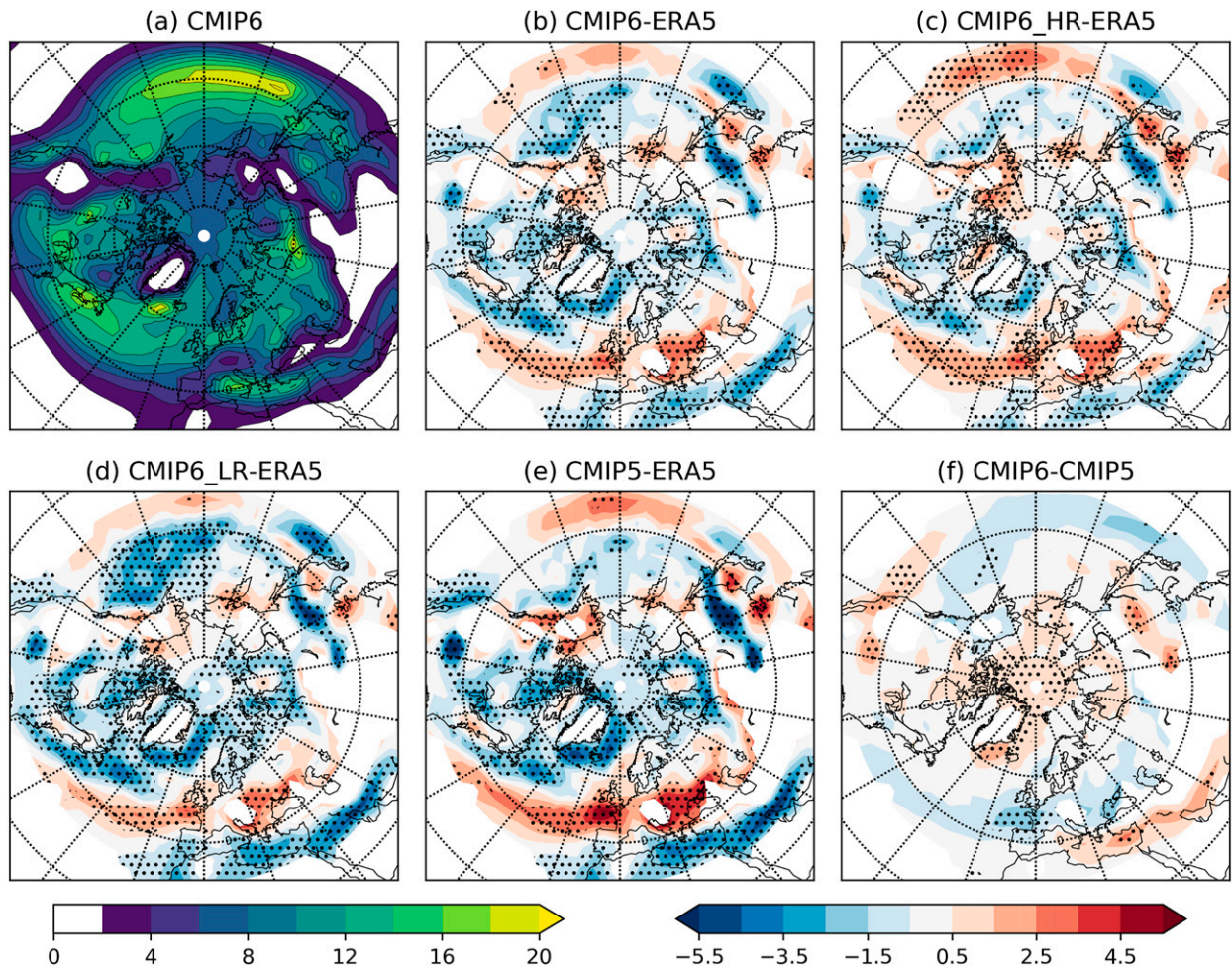
FIG. 3. (a) The CMIP6 multimodel mean track density for DJF in the NH. (b) The CMIP6 track density anomaly relative to ERA5. (c) CMIP6 high-resolution models anomaly. (d) CMIP6 low-resolution models anomaly. (e) CMIP5 multimodel mean anomalies from ERA5. (f) The difference between the CMIP6 and CMIP5 multimodel means. All of the CMIP5 models are included in the multimodel mean. Units are number of cyclones per month per 5° spherical cap. Stippling indicates where 80% of the models agree on the sign of the error. Latitudes are plotted at 20°, 40°, 60°, and 80°N. Longitudes are plotted every 20° (including 0°).

average latitude of cyclogenesis over the western North Pacific (30°–60°N, 140°–180°E; not shown) with the median bias being 0.55° and 0.37°, respectively. The slightly reduced equatorward genesis latitude bias of the CMIP6 models is likely contributing to the decreased anomalies on the equatorward flank of the western and central North Pacific storm track (Fig. 3f).

2) NORTHERN HEMISPHERE SUMMER (JJA)

The median number of cyclones generated in summer is lower than for the winter (277 and 364, respectively) poleward of 30°N for the reanalyses, which is also true for the ALL_CMIP6 group (261 in summer vs 354 in winter). None of the medians for the model ensembles plotted in Fig. 4a(i) lie above the lower quartile of the reanalyses and all the differences are significant

(Table 3). Therefore, there is a general lack of cyclogenesis in the NH summer for the CMIP6 models, which is consistent with the CMIP5 estimations, albeit with a slight improvement in number for the higher-resolution models [Fig. 4a(i) and Table 3]. The underestimation is also a consistent feature in each of the subdomains for the NH [Figs. 4a(ii)–(iv) and Table 3], with the largest negative bias in the Asia–Pacific domain [Fig. 4a(iii)] where there are approximately 10%–20% fewer storms generated for all plotted model combinations relative to the reanalyses (Table 3). Again, all model groups have significantly lower cyclogenesis relative to the reanalyses in each subdomain (Table 3).

Regarding tracks intersecting poleward of 30°N [Fig. 4b(i)], the median for the reanalyses is 302 and, consistent with the genesis results described above, all
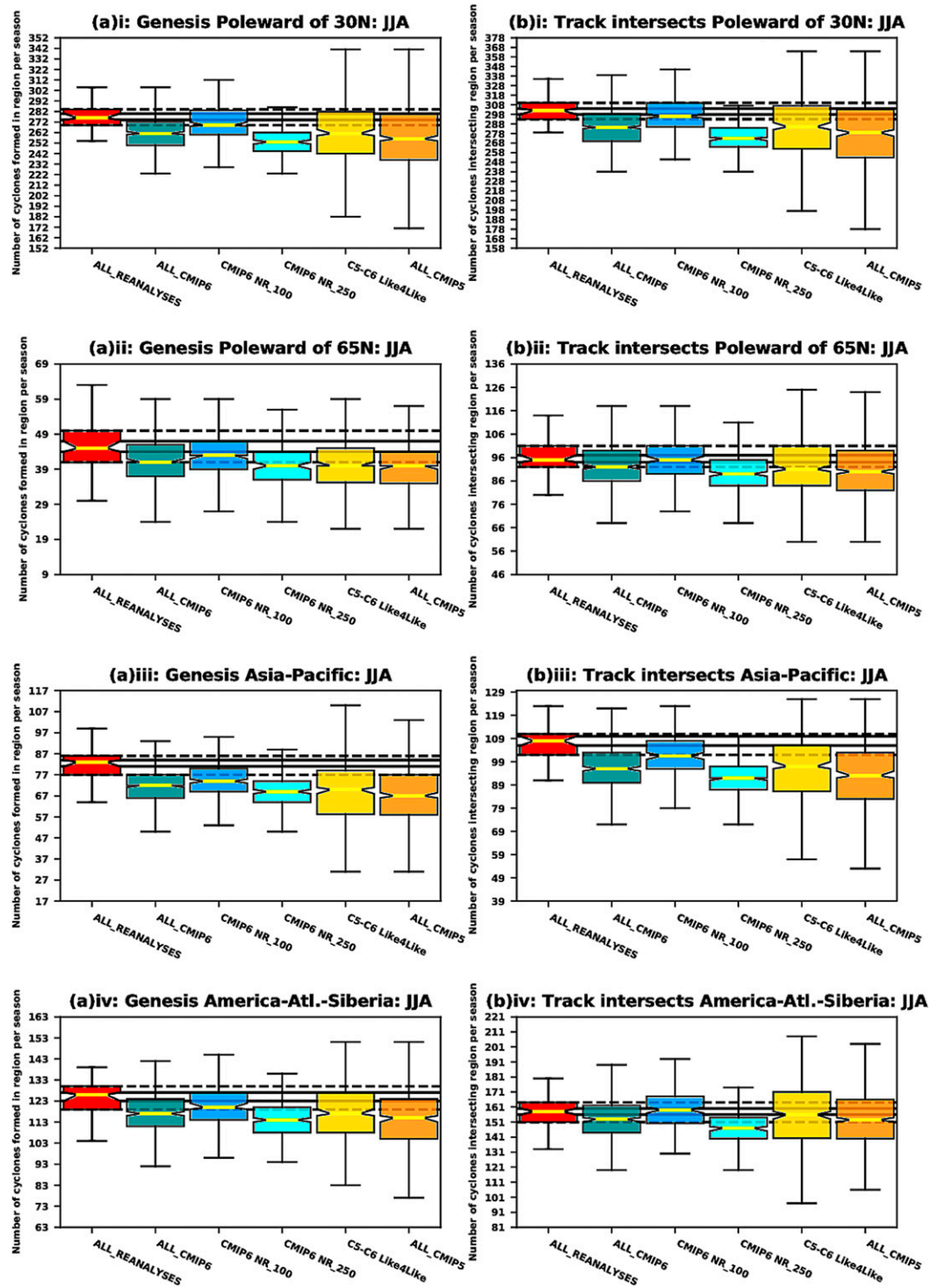
FIG. 4. Boxplots of (a) the number of cyclones forming within or (b) tracking partially/completely through different geographical domains of the NH for JJA. The regions plotted, style of boxplots, and black solid/dashed lines are all the same as in Fig. 2. Units for all boxes and all panels are cyclones per season.

TABLE 3. As in Table 2, but for NH JJA median genesis and track intersection totals.

Region	Reanalyses	CMIP6	CMIP6_NR_100	CMIP6_NR_250	C5-C6 Like4Like	CMIP5
Genesis						
Poleward 30°N	277	**261 (−5.78)****	**269 (−2.89)**	**253 (−8.66)**	**261 (−5.78)**	**256 (−7.58)**
Poleward 65°N	45	**41 (−8–89)****	**43 (−4.44)**	**40 (−11.11)**	**40 (−11.11)**	**40 (−11.11)**
ASPAC	83	*72 (−13.25)***	**74 (−10.84)**	**69 (−16.87)**	**70 (−15.66)**	**67 (−19.28)**
AMATSI	126	**117 (−7.14)****	**120 (−4.76)**	**114 (−9.52)**	**117 (−7.14)**	**115 (−8.73)**
Tracks						
Poleward 30°N	302	**285 (−5.63)****	**296 (−1.99)**	**273 (−9.60)**	**285 (−5.63)**	**279 (−7.62)**
Poleward 65°N	95	**92 (−3.16)**	95 (0.00)	**89 (−6.32)**	**91 (−4.21)**	**90 (−5.26)**
ASPAC	108	**96 (−11.11)****	**102 (−5.56)**	**92 (−14.18)**	**97 (−10.19)**	**93 (−13.89)**
AMATSI	158	*153 (−3.16)*	159 (+0.63)	**147 (−6.96)**	156 (−1.24)	**153 (−3.16)**

medians for the different combinations of CMIP6 and CMIP5 models lie significantly below the reanalyses (Table 3). For the high-latitude domain (>65°N), all model groups apart from the CMIP6_NR_100 have track intersection numbers that are significantly lower than the reanalyses (Table 3). The number of tracks crossing into/within the Asia–Pacific [Fig. 4b(iii)] region (as with cyclogenesis numbers) are significantly lower in all multimodel combinations relative to the reanalyses [Figs. 4a(iii) and b(iii) and Table 3]. It is noteworthy that the higher-resolution CMIP6 model ensemble has the highest median genesis and track intersection numbers for all regions in JJA, which suggests that increasing the resolution reduces some of the error but not all. Finally, for the American–Atlantic–Siberian domain, there are significantly fewer track intersections for all model combinations apart from the CMIP6_NR_100 and C5-C6 Like4Like models [Fig. 4b(iv) and Table 3].

Focusing on the regional detail in summer (JJA), the highest track densities are over northeast Asia, the North Pacific, northeast North America, the North Atlantic, and the northern Russian coast (Fig. 5a). There is a clear underestimation in the cyclone track density from East Asia, over Japan, and into the northeast North Pacific in the CMIP6 multimodel mean (Fig. 5b), which has high model consensus and is a robust feature across all of the CMIP6 models regardless of the horizontal resolution (cf. Figs. 5c,d). There are also smaller negative track density biases across the North Atlantic, although this underrepresentation does appear to be larger and more robust in the lower-resolution models (Figs. 5b–d). There are also positive track density biases over northern Russia, northeast Asia, and central North America (Fig. 5b), which are larger at higher resolution (cf. Figs. 5c,d) and may indicate there are generally more cyclones in those models. It is important to note that the pattern of biases in CMIP6 (and, in particular, around the North Pacific region) still persist from CMIP5 (cf. Figs. 5b,e) and the magnitudes of these

biases are only slightly reduced, with this being most evident in the west of the two main ocean basins and to the south of Alaska (i.e., small positive differences in Fig. 5f). The slight equatorward shift of tracks in the western North Atlantic (Fig. 5f) is associated with a small reduction in the poleward genesis bias from CMIP5 to CMIP6 (0.71° and 0.42° poleward respectively). The poleward genesis latitude bias in the North Pacific sector is not improved from CMIP5 to CMIP6, indicating that this is a robust and persistent bias that has not been improved with model development.

3) SOUTHERN HEMISPHERE SUMMER (DJF)

As with the NH, a broad overview of cyclone formation and track numbers is undertaken regionally for the SH. Due to the more continuous structure of the storm track (see Figs. 1c,d), the hemisphere is simply split into three zonal bands (also highlighted in Figs. 1c and 1d):

1) From 30° to 80°S: A measure of the hemispheric extratropical cyclone activity, which cuts out the highest peaks of Antarctica that lie above the 850-hPa level (poleward of the black circle in Figs. 1c and 1d).
2) From 30° to 60°S: A measure of low-to-midlatitude extratropical cyclone activity (between the black and white circles in Figs. 1c and 1d).
3) From 60° to 80°S: A measure of high-latitude cyclone activity (between the white and magenta circles in Figs. 1c and 1d).

As with the NH, finer geographical details of the SH storm track density biases are discussed later in this section.

Between 30° and 80°S, the median number of cyclones that form during SH summer is 269 for the reanalyses, which is significantly underestimated (Table 4) by all of the different model combinations plotted in Fig. 6a(i). For the 30°–80°S region the CMIP6 models do show a significant improvement compared to CMIP5 as a whole but not relative to the like-for-like group (Table 4).
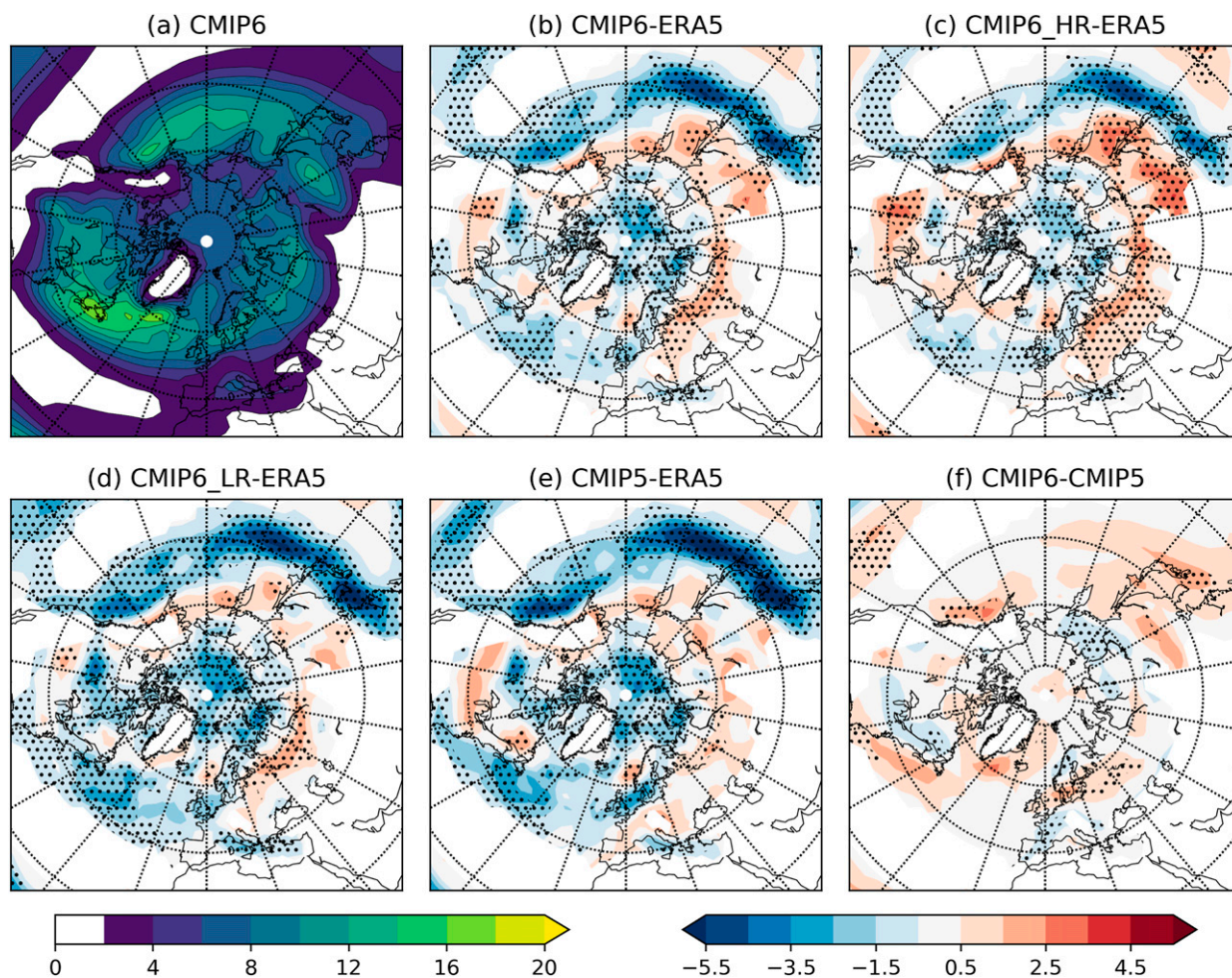
FIG. 5. As in Fig. 3, but for NH JJA.

Looking at the SH subdomains [Figs. 6a(ii) and (iii)], the underestimation in cyclogenesis for the CMIP6 models is primarily from the 30° to 60°S band [Fig. 6a(ii) and Table 4] where the CMIP5 model groups appear to slightly outperform those of CMIP6. Between 60° and 80°S, however, the CMIP6 models compare better with the reanalyses (albeit still too low; Table 4) and outperform the CMIP5 models [Fig. 6a(iii)].

When considering the number of tracks intersecting each SH domain, the differences between the models and the reanalyses mirror the genesis results described above. The number of tracks passing through/within the 30°–80°S domain are consistently and significantly lower in each plotted multimodel ensemble relative to the reanalyses [Fig. 6b(i) and Table 4]. Furthermore, there is no clear improvement from the CMIP5 to the CMIP6 model groups (Table 4). In the 30°–60°S band, there is a clear (and significant; see Table 4) underestimation of the track numbers across the different ensembles and no

improvement from CMIP5 to CMIP6 [Fig. 6b(ii) and Table 4]. The number of cyclone tracks intersecting the 60°–80°S domain is higher in the CMIP6 groups relative to CMIP5 (i.e., improvement; see Table 4) and also higher for the higher-resolution CMIP6 models than the low resolution ones; however, the numbers are consistently lower than the reanalyses regardless of resolution [Fig. 6b(iii) and Table 4].

Having reviewed the broad characteristics of the formation and track numbers for the three SH domains, the focus now moves to the regional detail. For DJF, the storm track, while annular in shape around the hemisphere, contains a region of higher track densities from South America to approximately 120°E along 50°S (Fig. 7a). For the CMIP6 multimodel mean, the storm track biases are minimal with little consensus (Fig. 7b), which indicates that the models (on average) are capturing the main structure and amplitude of the SH storm track well. However, there are indications of an

TABLE 4. DJF median genesis (rows 3–6) and regional track intersection (rows 8–11) totals for the SH from reanalyses, and each grouping of CMIP6 and CMIP5 models. The numbers in the parentheses are the differences relative to the reanalyses (%). Bold numbers indicate that the multimodel ensemble median is significantly different from the reanalyses, italicized values indicate that the CMIP6 median is significantly different from the C5-C6 Like4Like models' median, and two asterisks (**) denote that the CMIP6 multimodel median is significantly different from the full CMIP5 multimodel median (significance is achieved for $p \leq 0.05$).

Region	Reanalyses	CMIP6	CMIP6_NR_100	CMIP6_NR_250	C5-C6 Like4Like	CMIP5
Genesis						
30°–80°S	269	**256** (−4.83)**	**259** (−3.72)	**253** (−5.95)	**254** (−5.58)	**252** (−6.32)
30°–60°S	218	*209 (−4.13)***	**211** (−3.21)	**207** (−5.05)	**214** (−1.83)	**212** (−2.75)
60°–80°S	51	*47 (−7.84)***	47 (−7.84)	46 (−9.80)	**37** (−27.45)	**39** (−23.53)
Tracks						
30°–80°S	310	**294** (−5.16)	**299** (−3.55)	**289** (−6.77)	**296** (−4.52)	**293** (−5.48)
30°–60°S	286	*269 (−5.94)***	**275** (−3.85)	**263** (−8.04)	**277** (−3.15)	**273** (−4.55)
60°–80°S	151	*140 (−7.28)***	**143** (−5.30)	**137** (−9.27)	**121** (−19.87)	**124** (−17.88)

equatorward bias in the Indian and Pacific Ocean sectors, and particularly to the south of New Zealand (positive/negative dipole anomalies relative to ERA5) (Figs. 7b–d). There are also positive biases in the vicinity of the southern tip of South America (Figs. 7b–d), which may indicate problems with the representation of orography or in the way the mean flow interacts with it. There are also large negative anomalies to the east of South America, which is particularly robust in the low-resolution models (Fig. 7d). The improvement in the representation of the number of high-latitude cyclones noted in Fig. 6b(iii) is clear when comparing the CMIP6 and CMIP5 ensembles. There is a clear poleward shift in the cyclone tracks that is robust across the CMIP6 models, and a reduction of the hemispheric equatorward bias that was seen in the CMIP5 models (Figs. 7e,f). This poleward shift in the track density is partnered with a large poleward shift in the median genesis latitude of all cyclones (30°–60°S; not shown) with the large equatorward bias of CMIP5 models (0.69° equatorward) being almost eradicated in the CMIP6 ensemble (0.03° equatorward). Therefore, even though the CMIP6 models appear to be no better than their CMIP5 counterparts in Table 4, in particular between 30° and 60°S, it is clear that there has been an improvement in the overall representation of the SH storm tracks in CMIP6.

4) SOUTHERN HEMISPHERE WINTER (JJA)

As with the SH summer, the broad characteristics of formation and track numbers are assessed and (as with the NH) the amount of cyclogenesis is higher during the winter than the summer in the reanalyses (354 vs 269, respectively; see Table 5). The CMIP6 multimodel ensemble median lies close to the reanalysis estimate [Fig. 8a(i)] but is still significantly lower (Table 5) as are the CMIP6_NR_250 and both CMIP5 groups with only the CMIP6_NR_100 group comparable with the reanalyses [Fig. 8a(i) and Table 5]. Both the CMIP6 and CMIP5 model groups lie within ±2% of the reanalyses'

estimate in the 30°–60°S region for genesis [Fig. 8a(ii)] whereas the CMIP6 models perform better than CMIP5 at higher-latitude cyclogenesis [Fig. 8a(iii)].

The differences in the number of tracks intersecting with each subdomain of the SH mirror those of the genesis results in JJA (as also seen for the SH in DJF). For the hemisphere-wide (30°–80°S) and lower-latitude (30°–60°S) domains, the number of intersecting cyclones lie within ±2.8% of the reanalyses' estimate for both the CMIP6 and CMIP5 groups [Figs. 8b(i) and (ii) and Table 5]. For the higher latitudes (60°–80°S), all CMIP6 model groups have higher numbers of cyclone tracks passing through the domain than the CMIP5 groups, on average; however, the numbers are significantly lower than those of the reanalyses in all groups [Fig. 8b(iii) and Table 5].

Looking at a finer spatial scale for the SH winter, the highest track densities are over the south Indian Ocean and along the Antarctic coast (between 100°E and the Antarctic Peninsula), with a secondary maximum in the South Pacific near 40°S (Fig. 9a). In the CMIP6 multimodel mean, the largest biases are over the southern Indian and Pacific Oceans and also to the south of Australia (Fig. 9b) at the eastern end of the local track density maximum (see Fig. 9a). To the south of Australia there are large positive anomalies with negative anomalies located immediately to the south. Interestingly, the biases to the south of Australia are larger in the high-resolution models than the low-resolution models. This structure indicates that the local track density maximum is displaced too equatorward and likely too zonally oriented, given this is a region of large poleward movement of the cyclones (Hoskins and Hodges 2005). There are also large positive anomalies to the southeast of South Africa, which may arise due to the incorrect interaction of the mean flow with the topography in this region, which model simulations have been shown to be sensitive to (Inatsu and Hoskins 2004). It appears that increasing the
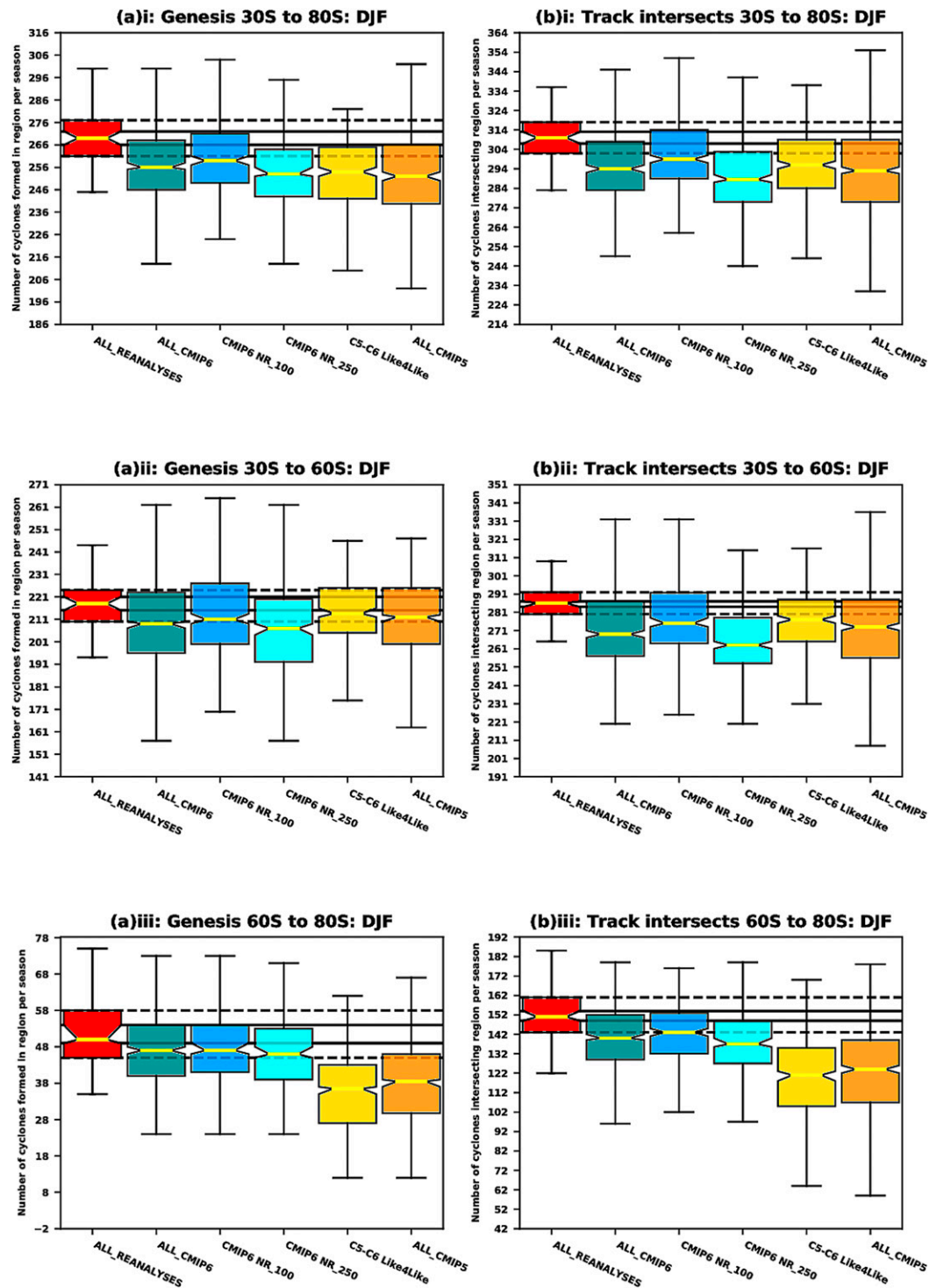
FIG. 6. Boxplots of (a) the number of cyclones forming within or (b) tracking partially/completely through different geographical domains of the SH for DJF. The regions are (i) between the 30° and 80°S band, (ii) between the 30° and 60°S band, and (iii) between the 60° and 80°S band. The style of boxplots and black solid/dashed lines are all as in Fig. 2. Units for all boxes and all panels are cyclones per season.
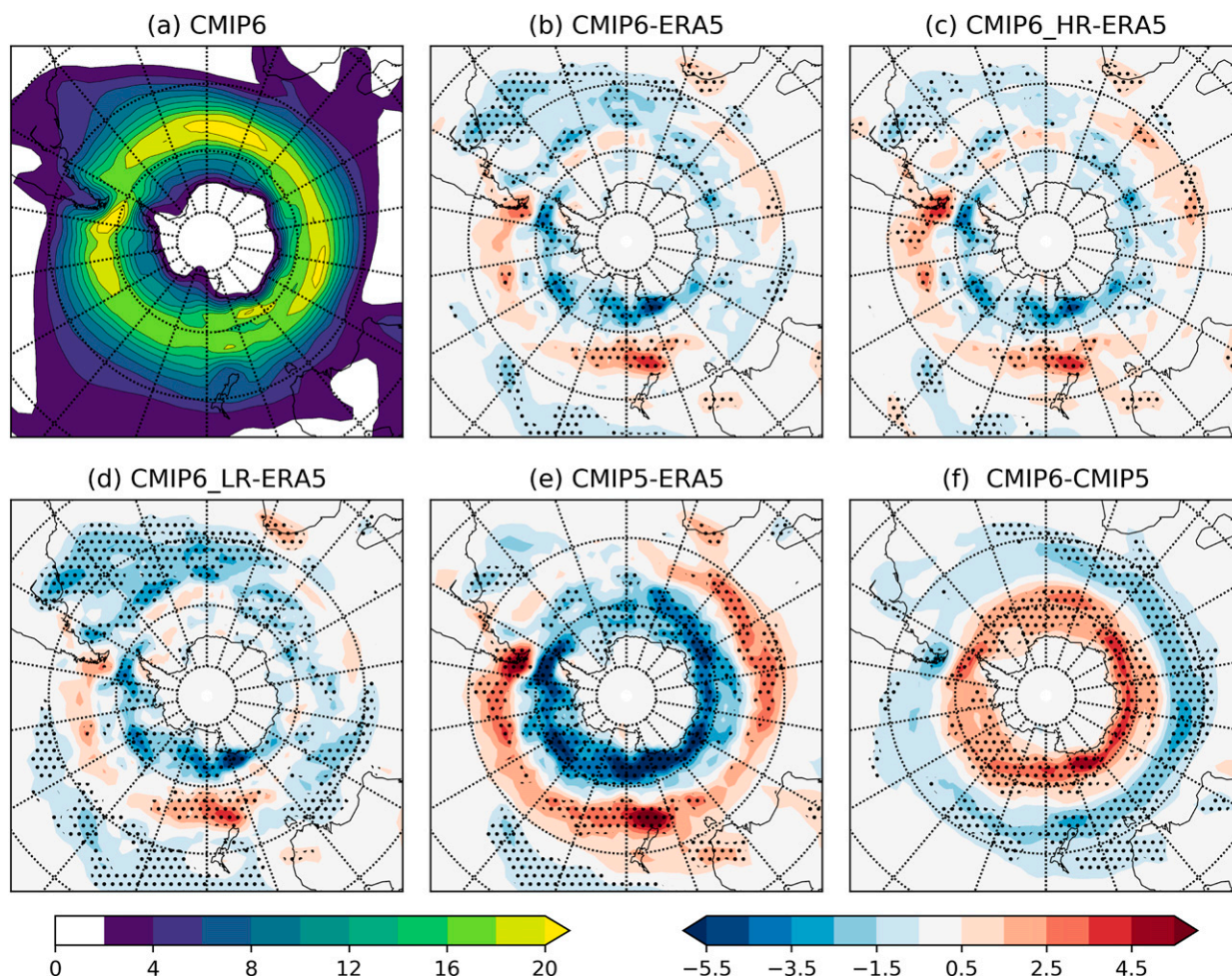
FIG. 7. (a) The CMIP6 multimodel mean track density for DJF in the SH. (b) The CMIP6 track density anomaly relative to ERA5. (c) CMIP6 high-resolution models anomaly. (d) CMIP6 low-resolution models anomaly. (e) CMIP5 multimodel mean anomalies from ERA5. (f) The difference between the CMIP6 and CMIP5 multimodel means. Units are number of cyclones per month per 5° spherical cap. Stippling indicates where 80% of the models agree on the sign of the error. Latitudes are plotted at 20°, 40°, 60° and 80°S. Longitudes are plotted every 20° (including 0°).

resolution has minimal impact with respect to the winter storm track in the SH as the pattern of the anomalies (i.e., magnitude and location) and consensus on the biases is similar in Figs. 9b–d. As in Figs. 7e–f the broadscale equatorward bias of the CMIP5 models is less evident in the CMIP6 models (Fig. 9f) with an increase in track density poleward of 60°S around all of the Antarctic coastline, although this is less clear than in DJF, as was noted from Figs. 7b(iii) and 8b(iii). Consistent with the track density shift poleward, there is also a shift in the

TABLE 5. As in Table 4, but for SH JJA median genesis and track intersection totals.

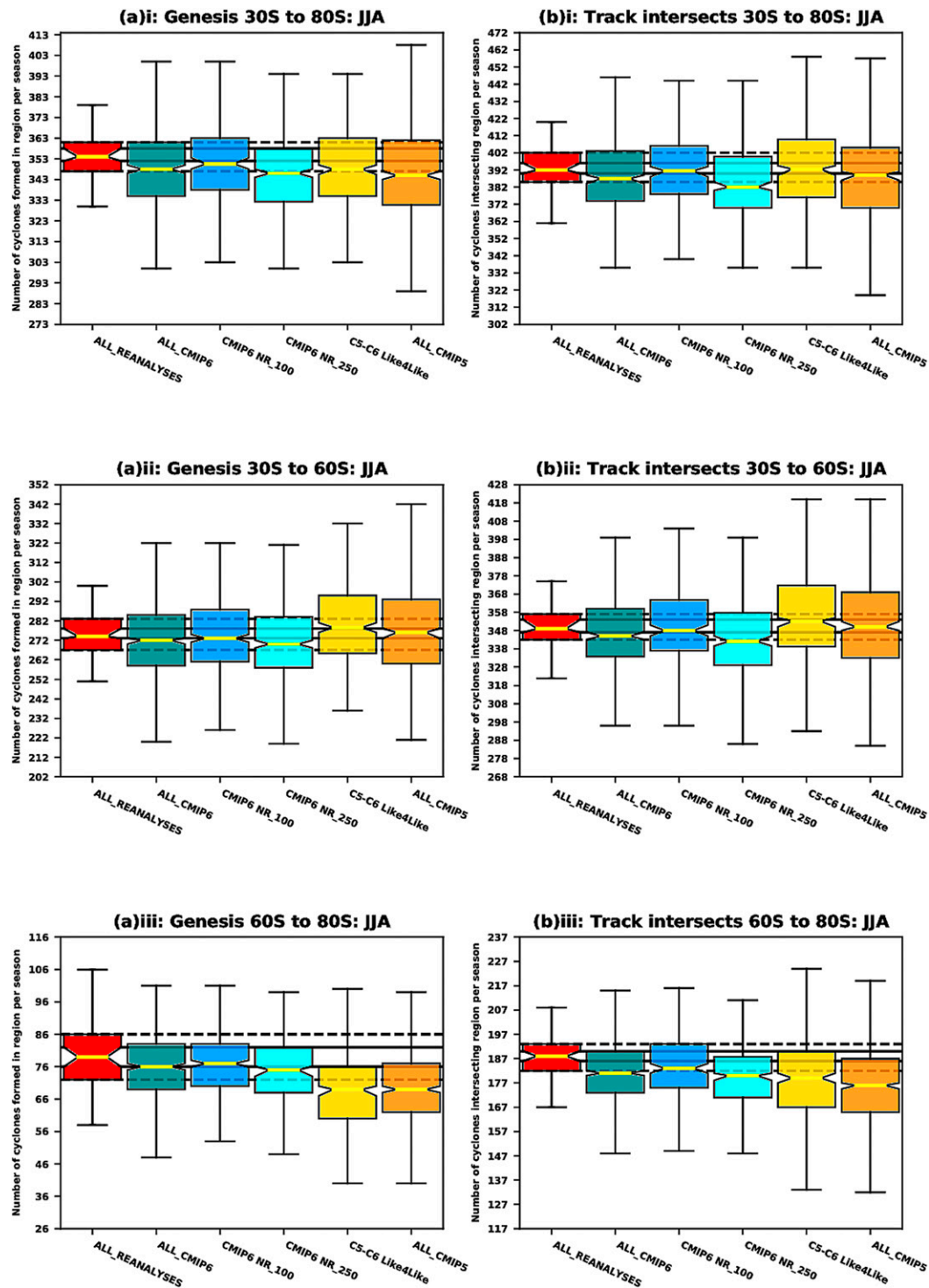| Region	Reanalyses	CMIP6	CMIP6_NR_100	CMIP6_NR_250	C5-C6 Like4Like	CMIP5
Genesis						
30°–80°S	354	**348 (−1.69)**	351 (−0.85)	**346 (−2.26)**	**348 (−1.69)**	**345 (−2.54)**
30°–60°S	275	*272 (−1.09)***	273 (−0.73)	**270 (−1.82)**	279 (+1.45)	276 (+0.36)
60°–80°S	79	*76 (−3.80)***	77 (−2.53)	**75 (−5.06)**	69 (−12.66)	69 (−12.66)
Tracks						
30°–80°S	393	*387 (−1.53)*	392 (−0.25)	**382 (−2.80)**	393 (0.00)	389 (−1.02)
30°–60°S	349	*345 (−1.15)*	348 (−0.29)	**342 (−2.01)**	353 (+1.15)	350 (+0.29)
60°–80°S	188	**181 (−3.72)***	**183 (−2.66)**	**180 (−4.26)**	179 (−4.79)	176 (−6.38)

FIG. 8. Boxplots of the number of cyclones (a) forming within or (b) tracking partially/completely through different geographical domains of the SH for JJA. The regions plotted, style of boxplots, and black solid/dashed lines are all as in Fig. 2. Units for all boxes and all panels are cyclones per season.
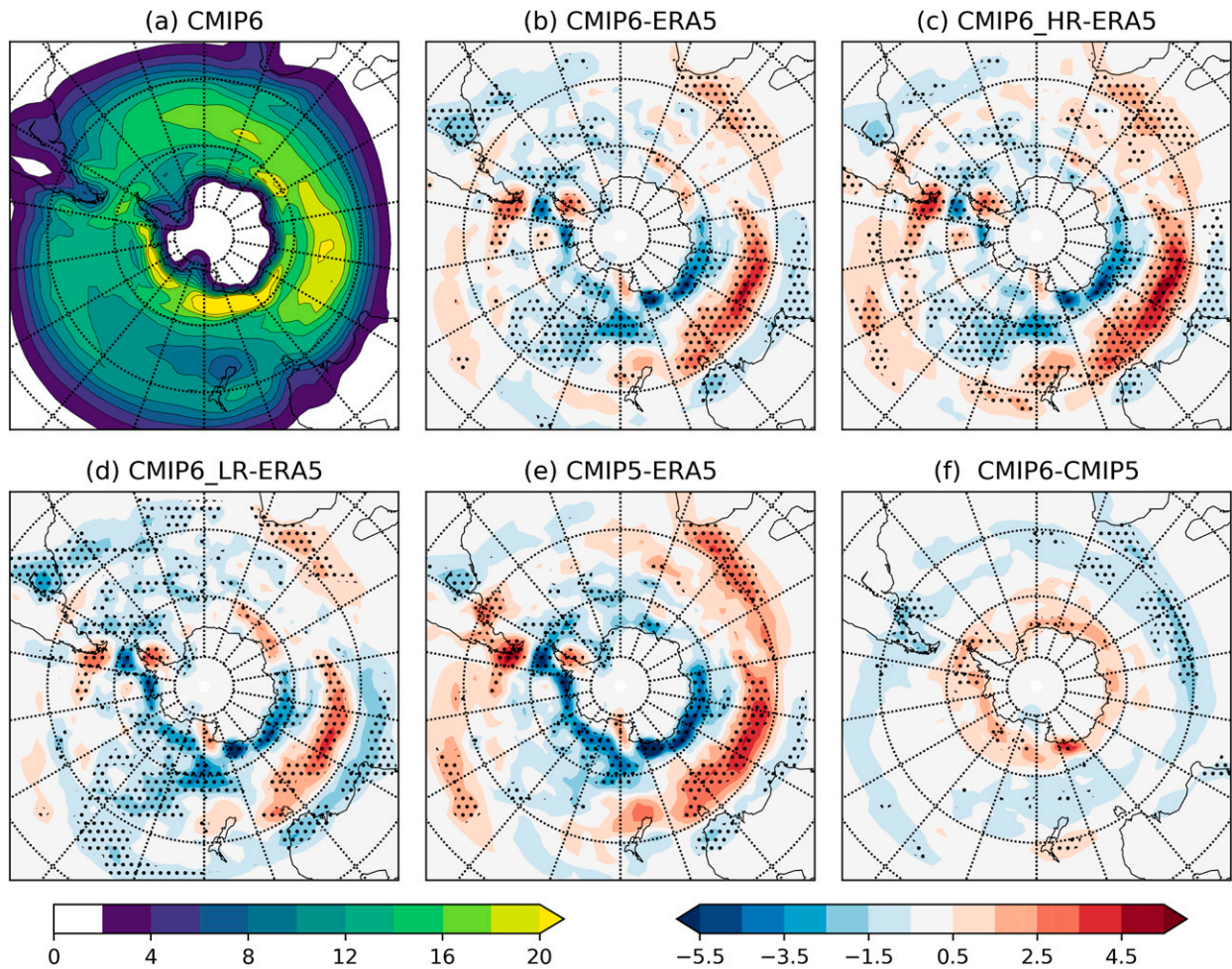
FIG. 9. As in Fig. 7, but for SH JJA.

median genesis latitude, although to a lesser extent than in DJF, with CMIP6 models having a reduced equatorward bias (0.3°) than the CMIP5 models (0.82°).

c. Intensities

The intensities of extratropical cyclones were robustly underestimated in CMIP5 models, particularly for the most intense cyclones (Zappa et al. 2013a; Chang et al. 2012). However, some evidence suggests that higher-resolution models better represent the peak intensity of these phenomena (Colle et al. 2013; Zappa et al. 2013a). In Fig. 10 the peak intensity (as measured by 850-hPa maximum T42 vorticity) is compared for all of the CMIP6 models.

1) NORTHERN HEMISPHERE

The NH mean and median peak vorticity for ERA5 are 6.02×10^{-5} and $5.53 \times 10^{-5} \, \text{s}^{-1}$, respectively, in DJF and correspond to the positively skewed distribution for

cyclone intensity (Fig. 10a). The shape of the multi-model CMIP6 cyclone intensity distribution is very similar to the reanalysis distribution (the mean is $5.85 \times 10^{-5} \, \text{s}^{-1}$ and the median is $5.4 \times 10^{-5} \, \text{s}^{-1}$) and is also positively skewed. The similarity in all the distributions across resolutions and model groups is also consistent with the similar medians, which is likely due to the common resolution used for the feature identification and tracking. Nonetheless, model horizontal resolution still appears to play an important role in the representation of cyclone intensity in CMIP6. In NH winter (and all other seasons) it appears that the higher-resolution models perform better than the low-resolution model set, particularly for the most intense cyclones. In winter, the low-resolution models have a tendency to have slightly too many cyclones at the lower end of the distribution (where the frequencies peak) and too few at the upper end, particularly above a vorticity threshold of $10 \times 10^{-5} \, \text{s}^{-1}$ (Fig. 10a). This underestimation is
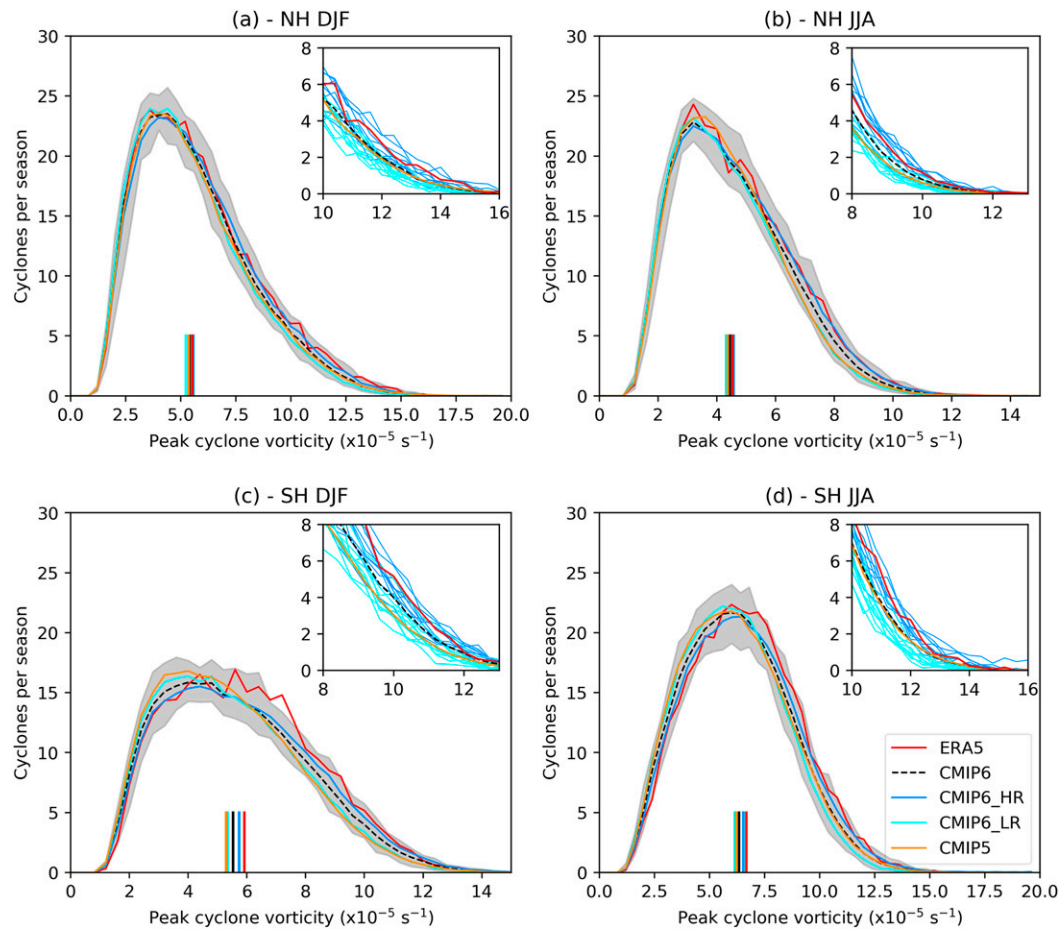
FIG. 10. Distributions of the magnitude of peak cyclone intensity as measured by cyclone T42 vorticity: (a) NH DJF, (b) NH JJA, (c) SH DJF, and (d) SH JJA. The gray shaded region represents the 5th–95th percentile of the CMIP6 models with the black dashed line being the multimodel mean. The cyan and dark blue lines are the means of the low-resolution and high-resolution CMIP6 models, respectively. The orange line shows the mean from the CMIP5 multimodel mean. The red line shows the results from ERA5. The colored columns from the x axis indicate the median of the distribution of each of the model sets. Column colors are the same as those from the distributions. Bin widths are $0.4 \times 10^{-5}\,\mathrm{s}^{-1}$.

also robust to the choice of reanalysis (not shown). Compared to the previous generation of CMIP5 models there have been only minor improvements in the frequency distribution of the peak intensity of cyclones in CMIP6 (the CMIP5 mean is $5.82 \times 10^{-5}\,\mathrm{s}^{-1}$ and the median is $5.35 \times 10^{-5}\,\mathrm{s}^{-1}$). There is, however, a slight improvement in the number of mid-to-strong intensity cyclones in the CMIP6 multimodel ensemble relative to CMIP5. This improvement comes from the improvements in horizontal resolution, as the mean distribution of the high-resolution CMIP6 models is well matched to that of ERA5 (see Fig. 10a inset). Despite the above differences, each model distribution still lies within the ensemble spread of the CMIP6 models. However, the CMIP6 models do have a reduced ensemble spread compared to the total CMIP5 ensemble by an average of

30% across the entire intensity distribution. This difference can be as low as 18% when considering an equal number of (randomly selected) CMIP5 and CMIP6 models (not shown). This demonstrates that the CMIP6 models offer improved confidence of a more accurate peak cyclone intensity compared to CMIP5. This is a feature of the solstice seasons in both hemispheres.

In the NH summer (JJA; Fig. 10b), the mean and median cyclone intensities are lower in ERA5 (4.84×10^{-5} and $4.54 \times 10^{-5}\,\mathrm{s}^{-1}$, respectively) than in winter, which is also simulated by the CMIP6 multimodel ensemble (mean of $4.72 \times 10^{-5}\,\mathrm{s}^{-1}$ and median of $4.45 \times 10^{-5}\,\mathrm{s}^{-1}$). In addition, all model groups and resolutions have very similarly shaped distributions with comparable medians, as is also the case for NH DJF (Fig. 10a). As in the winter, the higher-resolution CMIP6 models

perform better than the lower-resolution models (cf. blue and cyan lines in Fig. 10b). The high-resolution model mean distribution matches the ERA5 distribution very well above $6 \times 10^{-5}\,\mathrm{s}^{-1}$, but there are too few cyclones identified in the low-resolution models for the same intensity range (Fig. 10b inset). For CMIP6 relative to CMIP5, there has been a more marked improvement in JJA than in DJF for the number of mid-to-high intensity cyclones. The CMIP5 distribution is clearly lower than ERA5 in the $6–10 \times 10^{-5}\,\mathrm{s}^{-1}$ range, with the CMIP6 multimodel intensities lying closer to the ERA5 distribution (cf. the gold, black dashed, and red lines in Fig. 10b). The low-resolution CMIP6 models are comparable to the CMIP5 ensemble for the number of mid-to-high intensity cyclones identified. Nevertheless, both demonstrate a large underestimation in the frequency of mid- to high-intensity cyclones compared to the high-resolution models and the reanalysis.

2) SOUTHERN HEMISPHERE

In the SH summer (DJF; Fig. 10c), the magnitudes of the mean and median peak cyclone intensities[3] are 6.09×10^{-5} and $5.91 \times 10^{-5}\,\mathrm{s}^{-1}$, respectively for ERA5. The multimodel peak intensities for CMIP6 are lower than ERA5 for both the mean ($5.82 \times 10^{-5}\,\mathrm{s}^{-1}$) and the median ($5.54 \times 10^{-5}\,\mathrm{s}^{-1}$). Furthermore, there are larger difference in the mean/median between higher- and lower-resolution models (see the blue and cyan lines in Fig. 10c) than there are for both seasons in the NH, indicating there may be more of a resolution sensitivity to the representation in peak cyclone intensity across the CMIP6 ensemble relative to ERA5 in the SH. For higher intensities (above $8 \times 10^{-5}\,\mathrm{s}^{-1}$), the higher-resolution models are much closer to the ERA5 frequencies than the lower-resolution models (see inset of Fig. 10c), with the low-resolution models overestimating frequencies of low-intensity cyclones. Furthermore, the mean ($5.58 \times 10^{-5}\,\mathrm{s}^{-1}$) and median ($5.29 \times 10^{-5}\,\mathrm{s}^{-1}$) multimodel cyclone intensities in CMIP5 are lower than those of CMIP6 (also plotted in Fig. 10c), and there are fewer high-intensity cyclones (above $8 \times 10^{-5}\,\mathrm{s}^{-1}$) for CMIP5 relative to CMIP6. Despite this, the CMIP5 distribution still lies within the CMIP6 ensemble spread. The shape of the distribution, and the medians, of the CMIP5 and low-resolution CMIP6 models are closely matched and are shifted toward lower intensities compared to ERA5. This further suggests that peak intensity is commonly underestimated in the SH for models with a lower resolution. Therefore, there may be systematic

deficiencies in the intensification processes, which may be related to the better representation of high-latitude cyclones and genesis rates in high-resolution models [Fig. 6b(iii)]. This resolution-dependent intensity bias is not the case in the NH where frequency of peak cyclone intensity is well estimated (i.e., comparable medians between CMIP6 and ERA5) and the main underestimations are in the number of high-intensity cyclones. Despite the similarity of the CMIP5 and low-resolution CMIP6 models, there is an improvement for CMIP6 compared to the CMIP5 ensemble in the peak intensity bias, with this coming from the high-resolution models.

The mean ($6.78 \times 10^{-5}\,\mathrm{s}^{-1}$) and median ($6.67 \times 10^{-5}\,\mathrm{s}^{-1}$) cyclone intensities are higher in the SH winter (JJA) than the summer in ERA5 (also cf. Figs. 10c and 10d), as is the case for the NH winter. As with the SH summer (DJF), the mean and median (6.49×10^{-5} and $6.35 \times 10^{-5}\,\mathrm{s}^{-1}$, respectively) cyclone intensities in the CMIP6 multimodel ensemble for SH JJA are lower than ERA5, and the median and mean are closer to the reanalyses in the higher-resolution models than the lower (cf. blue and cyan lines in Fig. 10d). As in SH DJF, and unlike cyclones in the NH, there is a larger difference in the medians, and a shift of the distributions toward lower values for the CMIP6 models compared to ERA5, with a larger shift for the lower-resolution models, further suggesting that higher resolution is required to accurately simulate cyclone peak intensity. Despite this, the difference in the medians is less in SH JJA than SH DJF, suggesting summer intensification mechanisms are less well represented by the models. For the CMIP5 models, the mean peak intensity is lower ($6.38 \times 10^{-5}\,\mathrm{s}^{-1}$) than for CMIP6, with the median ($6.21 \times 10^{-5}\,\mathrm{s}^{-1}$) also being underestimated compared to CMIP6 (also see Fig. 10d). Interestingly, the frequency of cyclones above $10 \times 10^{-5}\,\mathrm{s}^{-1}$ is higher for the CMIP5 models than the low-resolution CMIP6 models (inset in Fig. 10d). The CMIP6 higher-resolution models do nonetheless outperform the CMIP5 models for the high-intensity cyclones (inset in Fig. 10d) and, as in the other seasons/hemispheres, compare better with the reanalysis.

3) INTENSITY SUMMARY

Despite the improvements in the models it is apparent that most CMIP6 models still underestimate the cyclone intensification processes, particularly for the highest-intensity cyclones, and most notably in models with a lower horizontal resolution. It is also notable that in the SH lower resolution leads to an underestimation of peak intensity for cyclones of all intensities compared to the reanalysis estimates. There may therefore be specific intensification processes in the SH that are not fully captured by the models. For the CMIP6 ensemble there

---

[3] All SH cyclone vorticity values have been multiplied by -1 to make them comparable to values obtained from the NH.
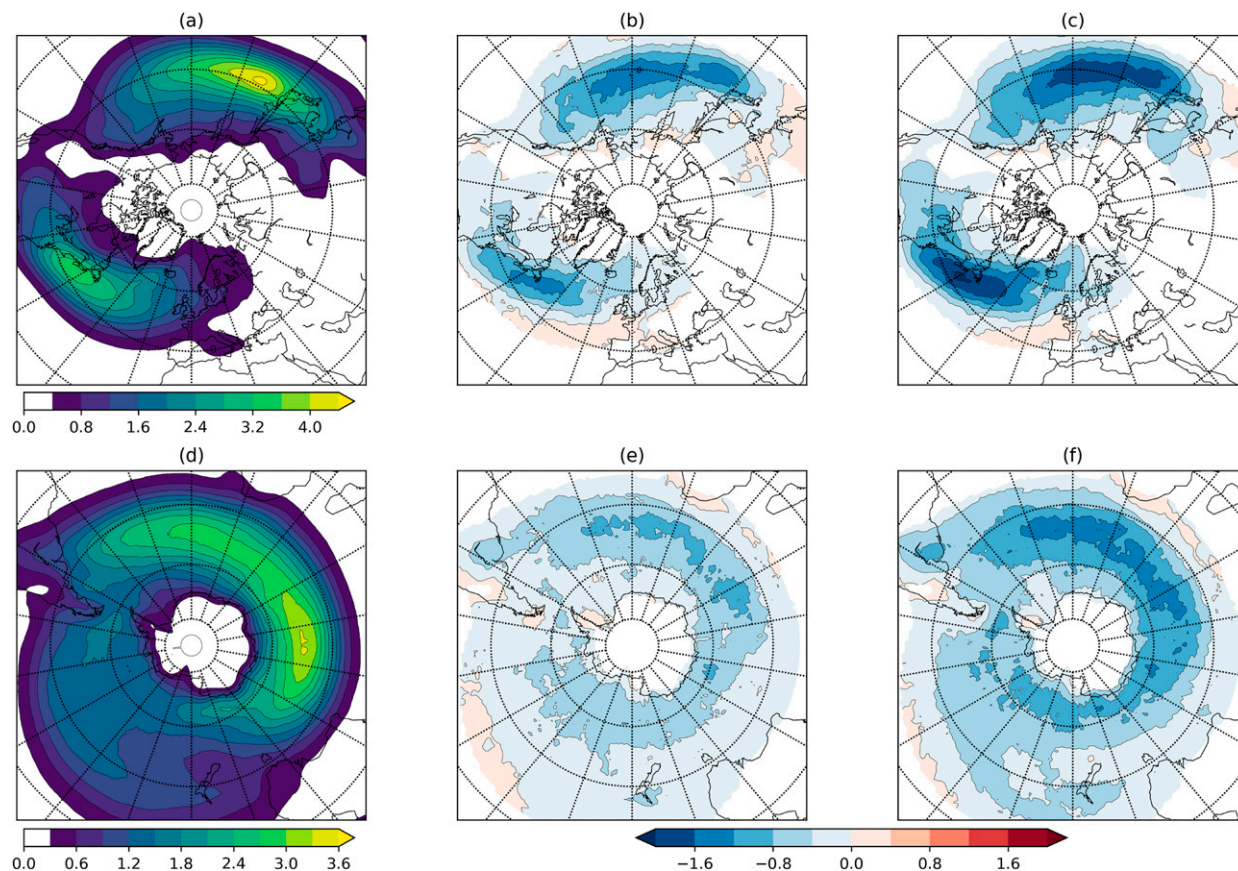
FIG. 11. Track densities of bomb cyclones in the (a) NH DJF and (d) SH JJA season for the CMIP6 multimodel mean. (b),(e) Anomalies of (a) and (d), respectively, relative to the ERA5 climatology. (c),(f) The anomalies of the CMIP5 multimodel mean from the ERA5 climatology. Units are number of cyclones per month per 5° spherical cap.

are improvements in the median peak cyclone intensities compared to the CMIP5 ensemble. Furthermore, estimations of cyclone intensity are less variable for the models in CMIP6 compared to CMIP5, evident from a reduced ensemble spread at all intensities.

d. Bomb cyclones

The most rapidly intensifying cyclones are bomb cyclones and are defined as those that have an intensification rate of at least 1 bergeron (Sanders and Gyakum 1980). Analysis of the CMIP5 models by Seiler and Zwiers (2016) illustrated that the models could represent the spatial pattern of the occurrence frequency of bomb cyclone locations, but not the magnitude of the frequencies in these locations, with lower-resolution models tending to have larger biases than the higher-resolution models. A similar analysis of our CMIP6 ensemble is performed and biases in frequency and intensity are outlined and discussed.

Figures 11a and 11d illustrate where bomb cyclones are most commonly located during DJF for the NH and

JJA for the SH, in the CMIP6 multimodel mean. For NH DJF (Fig. 11a) bomb cyclones are primarily located in the western reaches of the two main ocean basins, with these locations being strongly linked to both the Gulf Stream and the Kuroshio Currents (Seiler and Zwiers 2016; Reale et al. 2019). The bomb cyclones are also collocated with the main storm track regions (i.e., within highest track densities in Figs. 3a and 1a), with the North Atlantic tracks exhibiting more of a southwest to northeast tilt than their North Pacific counterparts. For SH JJA, the tracks of bomb cyclones are also located within the main storm track (i.e., where the highest track densities are in Fig. 9a and 1d) with high track densities over the South Atlantic Ocean and south Indian Ocean (Fig. 11d). There is also a weak, local maximum in bomb cyclone track densities between 160°W and the South American coastline, which is likely to be associated with systems developing in association with the subtropical jet.

In terms of the track density of cyclones compared to ERA5, there are fewer bombs in the CMIP6 ensemble for both hemispheres in the winter seasons (Figs. 11b,e).

In the NH the underestimation is primarily over the western and central regions of the two ocean basins, where the maximum of the bomb cyclone track densities occurs (Fig. 11b). On average, there are 0.46 fewer cyclones per month in the Pacific sector than ERA5 and 0.41 fewer cyclones per month in the North Atlantic sector. As a percentage bias for CMIP6 relative to ERA5, the peak underestimation of the bomb cyclone track densities is between 35%–40% in the central North Pacific and approximately 30% in the central and western North Atlantic. Taking the area average relative to ERA5, the CMIP6 bomb cyclone frequency is 17% lower over the North Pacific (120°E–120°W) and 15% lower over the North Atlantic (80°W–0°) sectors. This represents an improvement of the CMIP6 models relative to CMIP5 models, as shown in Fig. 11c, where negative track density anomalies are present in the same locations as in the CMIP6 models. Underestimations in the CMIP5 models can be up to 2 cyclones per month in both basins and by 27% in the North Pacific and 31% in the North Atlantic sectors [this is consistent with previous underestimations of 22% in the North Pacific and 31% in the North Atlantic by Seiler and Zwiers (2016)].

In the SH, the CMIP6 multimodel mean bomb track densities are lower than ERA5 in all ocean basin sectors (Fig. 11e), as in the NH. On average, there are approximately 0.3 cyclones per month fewer around the entire hemisphere with a peak underestimation of approximately 1 cyclone per month in all ocean sectors for CMIP6 relative to ERA5. The biases in the SH are closely collocated with the highest overall cyclone track densities (see Figs. 9a and 1d) and the pattern of negative bomb cyclone track density biases spirals toward Antarctica across the southern Atlantic and Indian Ocean basins (Fig. 11e). There are also lower bomb cyclone track densities in the South Pacific extending poleward from New Zealand (CMIP6 relative to ERA5; Fig. 11e). Taking the area average differences relative to ERA5, the CMIP6 bomb cyclone frequency is 18% lower for the entire SH, 17% lower over the southern Atlantic sector (60°W–30°E), 17% lower over the Indian Ocean sector (30°–120°E), and 18% lower over the southern Pacific sector (140°E–60°W). As with the NH there are improvements in the SH for bomb cyclones in CMIP6 compared to CMIP5. In the CMIP5 mean (Fig. 11f) the peak underestimation is by approximately 1.5 cyclones per month in the South Atlantic and south Indian Ocean sectors. The percentage underestimations for CMIP5 relative to ERA5 are by 31% for the entire SH, 34% over the southern Atlantic sector, 28% over the Indian Ocean sector, and 31% over the southern Pacific sector.

Biases in bomb intensity are compared in a similar way to those of all cyclones presented in Fig. 10. The

intensity measures used are peak T42 vorticity, minimum MSLP, and maximum 24-h deepening rate (measured in bergerons). The bomb cyclones identified are part of the upper end of the intensity distributions presented in the inset of Fig. 10. For both NH DJF (Figs. 12a–c) and SH JJA (Figs. 12d–f) reanalyses, multimodel means (CMIP5 and CMIP6) and the individual model groups all exhibit a similar shaped distribution regardless of the intensity metric, indicating the models perform well at representing peak intensity of bombs. Nevertheless, it is interesting to note that there are different frequencies of bombs for each group of models compared to the reanalyses. This is most clearly seen in the distributions of peak vorticity (Figs. 12a,d) whereby the reanalyses have the highest number of bombs, then the CMIP6 models, and the least in the CMIP5 models. The same underestimation of both CMIP5 and CMIP6 relative to ERA5 (and CMIP5 relative to CMIP6) is visible for all intensity measures and in both hemispheres/seasons (Fig. 12). Across all intensity measures and in both hemispheres, the higher-resolution CMIP6 models also have bomb frequencies that more closely match the reanalyses. Furthermore, the lower-resolution models tend to have lower bomb frequencies, which are usually lower than the CMIP5 multimodel mean (cf. orange and cyan lines in Fig. 12). The difference in frequency of the intensities between CMIP6 and the reanalyses is consistent with the underestimation of the bomb cyclone track densities in Fig. 11.

It is clear that there has been some improvement in representing bomb cyclones with the newer generation of CMIP6 models compared to CMIP5, particularly in those with the highest horizontal atmospheric resolution; however, there are still too few relative to the reanalyses. It is worth noting that when all cyclones were considered (Fig. 10), the differences in the frequency of high-intensity cyclones was not too dissimilar for CMIP5, CMIP6, and the different reanalyses (particularly for the NH); however, the frequencies of bomb cyclone intensities in Fig. 12 contradict this. Therefore, both CMIP5 and CMIP6 models are capable of simulating the peak vorticity of cyclones at a range of intensities (with CMIP5 being slightly deficient compared to CMIP6), but do not perform well at capturing the rapid intensification mechanisms of some of these high-intensity cyclones. This clearly points to a specific deficiency of the models in capturing the explosive development and is likely the main reason why the number of bombs is underrepresented compared to the reanalyses. It is interesting that there has clearly been some progress in this area from CMIP5 to CMIP6, and that higher-resolution models perform better than lower-resolution models.
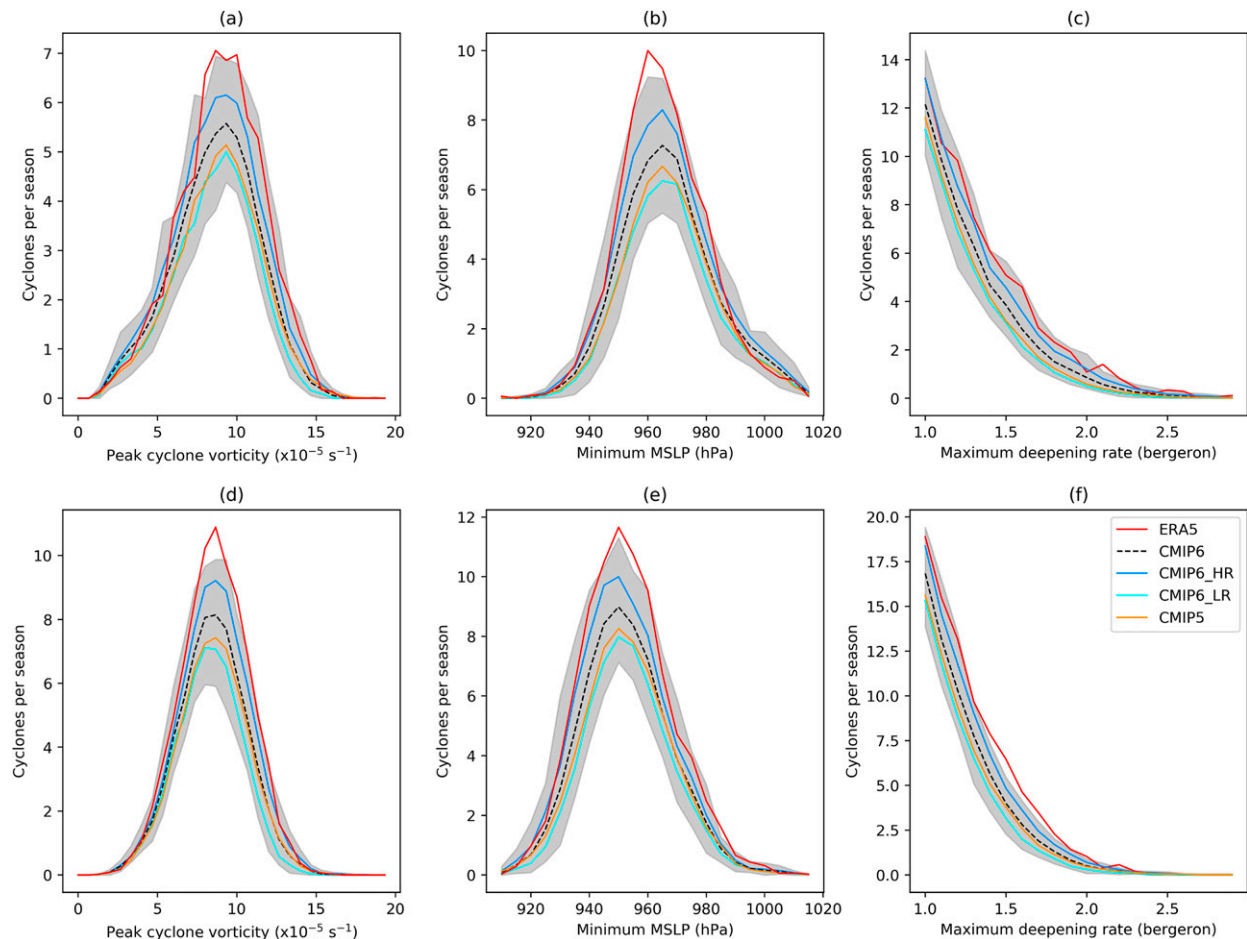
FIG. 12. Intensity distributions of identified bomb cyclones for (top) NH DJF and (bottom) SH JJA. Intensity metrics are (a),(d) the magnitude of peak T42 vorticity at 850 hPa, (b),(e) maximum cyclone MSLP, and (c),(f) maximum deepening rate in bergerons. The gray shaded region represents the 5th–95th percentile of the CMIP6 models with the black dashed line being the multimodel mean. The cyan and dark blue lines are the means of the low-resolution and high-resolution CMIP6 models, respectively. The orange line shows the mean from the CMIP5 multimodel mean. The red line represents the results from ERA5. Note that the bin widths in (a) and (d) are $0.67 \times 10^{-5}\,\text{s}^{-1}$ and different from those in Fig. 10.

Nonetheless, further development, as well as possible increases in resolution, is required in order to capture the frequency of bombs compared to the reanalysis.

4. Discussion and conclusions

In this study an evaluation of the CMIP6 models in terms of their representation of the extratropical storm tracks has been presented. The main biases of the storm track in both the NH and SH in DJF and JJA were discussed, as well as the representation of cyclone peak intensity, and the frequency and intensity distribution biases of explosively developing bomb cyclones. The main results of this work are summarized below.

- In NH winter, cyclogenesis rates are lower in the CMIP6 ensemble relative to reanalyses but higher

than CMIP5 poleward of 30°N and also for three other NH subdomains. The higher-resolution CMIP6 models perform better than those with lower resolution (Fig. 2).
- Biases that were present in CMIP5 are also seen in the CMIP6 models in the NH winter, such as an equatorward bias in the eastern North Pacific storm track and a too zonal storm track in the North Atlantic that extends too far into western Europe, relative to ERA5 (Fig. 3).
- In NH summer, there is a clear lack of cyclogenesis (and general cyclone activity) poleward of 30°N and in all subdomains, with this being particularly evident in the Asia–Pacific region. There is little structural difference in the track density biases for the higher-resolution CMIP6 models relative to the lower-resolution models, despite improvements in total cyclone numbers (Figs. 4 and 5).

- In SH summer, there is a general underrepresentation of cyclogenesis and also the number of cyclones tracking through the 30°–80°S domain in CMIP6, regardless of model resolution. This is primarily due to errors from 30°–60°S and is consistent with CMIP5. There is a large improvement in the 60°–80°S band in CMIP6 relative to CMIP5 with more cyclones farther poleward (Figs. 6 and 7).
- In SH winter, cyclogenesis and track numbers compare well for all model ensembles for the 30°–80°S and 30°–60°S bands relative to the reanalyses; however, the CMIP6 models perform better than CMIP5 in the highest latitude band. In all model ensembles there is a robust positive track density bias to the south of Australia where cyclone tracks are too zonal and do not propagate toward Antarctica (Figs. 8 and 9).
- For cyclone intensity in the NH, there is evidence that the higher-resolution models in CMIP6 outperform the low-resolution CMIP6 and the multimodel mean of the CMIP5 models, particularly with respect to the higher-intensity cyclones. In the NH median peak intensity is well represented in both seasons (Fig. 10).
- In the SH, the high-resolution models have a better frequency distribution of the peak intensity than the low-resolution models. Yet the median peak intensity is underrepresented in the CMIP6 ensemble, with this being worse for low-resolution models and the CMIP5 ensemble. The underrepresentation is also worse in summer rather than winter (Fig. 10).
- Area averages of bomb cyclone frequencies are lower across the main NH and SH storm tracks in CMIP6 relative to ERA5; however, frequencies across all ocean basins, and in both hemispheres, are higher than those of CMIP5 (Fig. 11).
- Higher-resolution models have higher bomb frequencies than low-resolution models. The CMIP6 models struggle to capture the rapid deepening associated with bomb cyclones, despite broadly capturing the correct peak intensities of all cyclones, indicating specific model deficiencies related to rapid intensification rates (Fig. 12).

The CMIP6 models have been shown to be broadly consistent with the reanalyses with regard to the number and frequency of cyclones tracking through specific geographic regions. There is also a general improvement in performance for the CMIP6 models compared to the CMIP5 ensemble. In the NH a reduction of the magnitude of the biases is seen in CMIP6, but the spatial pattern of the biases has changed little from CMIP5. However, in the SH there is a reduction of the overall spatial bias and a large poleward shift in the tracks that largely eliminates the large equatorward bias previously seen in the CMIP5

models. In the NH, resolution appears to play a large role in improving the representation of cyclone track and genesis locations (regardless of season), yet in the SH the increases in resolution within the CMIP6 ensemble do not seem to have such an impact. Despite this, the CMIP6 ensemble still performs better than the CMIP5 ensemble in the SH, particularly with regard to the large equatorward bias around the entire hemisphere.

Our results demonstrate that improving horizontal resolution has positive impacts in the NH, these improvements may be associated with improved mean-flow interaction with orography (Pithan et al. 2016), improved air–sea coupling (Woollings et al. 2010; Lee et al. 2018; Small et al. 2019), or better representation of cyclone moist processes (Willison et al. 2013). In the SH, where the impact of resolution is less apparent, perhaps model physics plays the largest role in the improvements seen from CMIP6 to CMIP5. In CMIP5, shortwave cloud biases were linked to the large equatorward biases in the eddy-driven jet (Ceppi et al. 2012). Recent studies, using a subset of the CMIP6 models, have shown a reduction in shortwave cloud forcing biases in the SH (Kawai et al. 2017, 2019; Voldoire et al. 2019), combined with an overall reduction in low cloud cover in the SH extratropics (Zelinka et al. 2020). Such improvements may have contributed to the poleward shift of the storm track and reduction of the large equatorward bias seen in the CMIP5 models through a modification of the surface temperature gradients (Ceppi et al. 2012).

The connection between horizontal atmospheric resolution and latent heat release may be the reason for the reduction in zonal biases in the storm track that are seen in the NH, most notably in the North Atlantic sector in DJF. Tamarin and Kaspi (2017) discussed how an increase in latent heat release tended to cause cyclones to propagate farther poleward through enhancing the strength of PV anomalies at upper levels. It is likely there are deficiencies in this process in the CMIP6 ensemble, particularly in the North Atlantic, as our results have shown that despite improvements in genesis latitude a zonal bias in track density still remains. The poleward propagation may be better resolved at higher resolutions and explain some improvement in the North Atlantic zonal bias from CMIP5 to CMIP6 and in the high-resolution CMIP6 models compared to the low-resolution ones. The continued presence of the bias, however, indicates that there may need to be further increases in atmospheric resolution or other elements of the model physics. The impact of resolution could further be tested through analysis of historical simulations as part of the HighResMIP project (Haarsma et al. 2016), which will run models with nominal atmospheric resolutions of 25 and 50 km. If the latent heat release within cyclones is better represented in higher-resolution models, this could help

explain the increase in the number of bomb cyclones seen in the higher-resolution models of in our results, as previous studies have shown latent heat release to be important for the rapid deepening of these cyclones (e.g., Hirata et al. 2019).

Numerous biases in the NH CMIP5 storm tracks were shown to be associated with biases in large-scale blocking (Zappa et al. 2014). These blocking biases are associated with the extension of the storm track into western Europe and also the underrepresentation in the Mediterranean. Schiemann et al. (2020) have shown improvements in blocking in the CMIP6 models relative to CMIP5, yet an underestimation relative to the reanalyses still exists. As Mediterranean cyclones require interaction of the mean flow with the Alpine orography (and an underrepresentation of Mediterranean cyclones also still exists), it is likely that any significant improvement in the representation of Mediterranean cyclones will require further improvement in the representation of blocking. These biases are also likely a driver for the increased number of cyclones to the east of the Alps. With a more zonal flow across the mountains, cyclogenetic processes will likely be happening to the east of the mountains, and not over the Gulf of Genoa, as would be expected.

The impact of ocean resolution and coupling has not been explored in this study. Lee (2015) discussed how the equatorward biases in the CMIP5 storm tracks were reduced in AMIP simulations, yet did not improve the major zonal biases or biases in the intensities of the cyclones. However, the resolution of the ocean component of coupled models has been shown to have a positive impact on the representation of the storm tracks through improved atmosphere–ocean coupling (Woollings et al. 2010; Lee et al. 2018; Small et al. 2019). The models utilized as part of this analysis have a range of nominal ocean resolutions from 25 to 100 km and could have varying impacts on the atmospheric circulation. The next step would be to assess AMIP simulations and also fixed SST and high-resolution coupled SST simulations as part of the HighResMIP project (Haarsma et al. 2016) to further assess the impact of ocean resolution and coupling.

Despite reducing biases through increased resolution in the CMIP6 ensemble relative to reanalyses (i.e., North Atlantic zonal bias in DJF) and some significant improvements since CMIP5 (i.e., equatorward bias in SH DJF), there are some features that persist in CMIP6 from CMIP5. The two clearest features are the underestimation of the number of tracks over eastern Asia and the northwestern North Pacific in JJA, as well as the persistent overestimation/zonal nature of tracks to the south of Australia in JJA. These persistent anomalies that have not seen significant robust improvements

require further investigation to isolate the specific model deficiencies leading to these biases.

There are several caveats to the results presented in this study. The most significant is the use of a single tracking scheme (Hodges 1994, 1995, 1999) that focuses on cyclones in the Lagrangian framework. An intercomparison with other methods, whether or not they are Lagrangian feature tracking schemes, or Eulerian filtered methods would be of interest. Initial results from Harvey et al. (2020), using Eulerian methods, show biases in the North Atlantic sector in DJF that are consistent with our findings, with a reduction of the zonal bias compared to CMIP5 estimations. Studies such as those of Neu et al. (2013) and Reale et al. (2019) have shown cyclone identification and tracking methods to be consistent, particularly for well-developed, intense cyclones. Furthermore, only one measure for the intensity of cyclones has been used in this study (T42 relative vorticity), so results may be sensitive to the choice of parameter. Despite this Chang (2017) showed similar distributions and future changes of cyclones based upon a number of different intensity metrics, indicating results may be insensitive to this choice.

This study has evaluated the current state of the representation of the storm tracks in the latest generation of GCMs that are part of CMIP6. A follow-up study will further investigate the main drivers and large-scale features associated with these storm track biases. This study also acts as a basis for further assessments of the future changes and impacts of midlatitude cyclones. Previous studies by Chang et al. (2012) indicated that models with large equatorward biases have larger future climate responses (e.g., a larger poleward shift in the SH) and therefore it will be of interest to see if the CMIP6 models (which have slightly reduced equatorward biases, particularly in the SH) follow the same pattern and have similar projections. Further to this the recent study from Baker et al. (2019) indicated that increasing the atmospheric resolution of a model resulted in a larger increase in the number of cyclones impacting western Europe under future climate conditions. With the CMIP6 models used in this study tending to have a higher horizontal resolution than the previously assessed CMIP5 ensemble it will be interesting to note if any projections follow the same pattern across multiple geographic regions in both hemispheres. Finally, initial estimations have shown that the equilibrium climate sensitivity of the CMIP6 models is higher than the CMIP5 models (e.g., Wu et al. 2019; Andrews et al. 2019; Voldoire et al. 2019; Gettelman et al. 2019) and this may have an impact on the magnitude of any changes to the general circulation of the midlatitudes and the

cyclones that form there under different future forcing scenarios.

REFERENCES

Andrews, T., and Coauthors, 2019: Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3.1 and UKESM1. *J. Adv. Model. Earth Syst.*, **11**, 4377–4394, https://doi.org/10.1029/2019ms001866.

Baker, A. J., and Coauthors, 2019: Enhanced climate change response of wintertime North Atlantic circulation, cyclonic activity, and precipitation in a 25-km-resolution global atmospheric model. *J. Climate*, **32**, 7763–7781, https://doi.org/10.1175/JCLI-D-19-0054.1.

Bengtsson, L., K. I. Hodges, and E. Roeckner, 2006: Storm tracks and climate change. *J. Climate*, **19**, 3518–3543, https://doi.org/10.1175/JCLI3815.1.

——, ——, and N. Keenlyside, 2009: Will extratropical storms intensify in a warmer climate? *J. Climate*, **22**, 2276–2301, https://doi.org/10.1175/2008JCLI2678.1.

Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.*, **33**, 1607–1623, https://doi.org/10.1175/1520-0469(1976)033<1607:ACSSOT>2.0.CO;2.

Browning, K. A., 2004: The sting at the end of the tail: Damaging winds associated with extratropical cyclones. *Quart. J. Roy. Meteor. Soc.*, **130**, 375–399, https://doi.org/10.1256/qj.02.143.

Catto, J. L., L. C. Shaffrey, and K. I. Hodges, 2011: Northern Hemisphere extratropical cyclones in a warming climate in the HiGEM high-resolution climate model. *J. Climate*, **24**, 5336–5352, https://doi.org/10.1175/2011JCLI4181.1.

Ceppi, P., Y.-T. Hwang, D. M. W. Frierson, and D. L. Hartmann, 2012: Southern Hemisphere jet latitude biases in CMIP5 models linked to shortwave cloud forcing. *Geophys. Res. Lett.*, **39**, L19708, https://doi.org/10.1029/2012GL053115.

Chang, E. K. M., 2017: Projected significant increase in the number of extreme extratropical cyclones in the Southern Hemisphere. *J. Climate*, **30**, 4915–4935, https://doi.org/10.1175/JCLI-D-16-0553.1.

——, S. Lee, and K. L. Swanson, 2002: Storm track dynamics. *J. Climate*, **15**, 2163–2183, https://doi.org/10.1175/1520-0442(2002)015<02163:STD>2.0.CO;2.

——, Y. Guo, and X. Xia, 2012: CMIP5 multimodel ensemble projection of storm track change under global warming. *J. Geophys. Res.*, **117**, D23118, https://doi.org/10.1029/2012JD018578.

——, ——, ——, and M. Zheng, 2013: Storm-track activity in IPCC AR4/CMIP3 model simulations. *J. Climate*, **26**, 246–260, https://doi.org/10.1175/JCLI-D-11-00707.1.

——, C. Zheng, P. Lanigan, A. M. W. Yau, and J. D. Neelin, 2015: Significant modulation of variability and projected change in California winter precipitation by extratropical cyclone activity. *Geophys. Res. Lett.*, **42**, 5983–5991, https://doi.org/10.1002/2015GL064424.

Christensen, J., and Coauthors, 2013: Climate phenomena and their relevance for future regional climate change. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1217–1308.

Colle, B. A., Z. Zhang, K. A. Lombardo, E. Chang, P. Liu, and M. Zhang, 2013: Historical evaluation and future prediction of eastern North American and western Atlantic extratropical cyclones in the CMIP5 models during the cool season. *J. Climate*, **26**, 6882–6903, https://doi.org/10.1175/JCLI-D-12-00498.1.

Eichler, T. P., N. Gaggini, and Z. Pan, 2013: Impacts of global warming on Northern Hemisphere winter storm tracks in the CMIP5 model suite. *J. Geophys. Res. Atmos.*, **118**, 3919–3932, https://doi.org/10.1002/JGRD.50286.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Gettelman, A., and Coauthors, 2019: High climate sensitivity in the Community Earth System Model version 2 (CESM2). *Geophys. Res. Lett.*, **46**, 8329–8337, https://doi.org/10.1029/2019GL083978.

Haarsma, R. J., and Coauthors, 2016: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geosci. Model Dev.*, **9**, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016.

Harvey, B. J., L. C. Shaffrey, T. J. Woollings, G. Zappa, and K. I. Hodges, 2012: How large are projected 21st century storm track changes? *Geophys. Res. Lett.*, **39**, L18707, https://doi.org/10.1029/2012GL052873.

——, P. Cook, L. Shaffrey, and R. Schiemann, 2020: The response of the Northern Hemisphere storm tracks and jetstreams to climate change in the CMIP3, CMIP5, and CMIP6 climate models. *EGU General Assembly 2020*, EGU2020-19711, https://doi.org/10.5194/egusphere-egu2020-19711.

Hawcroft, M. K., L. C. Shaffrey, K. I. Hodges, and H. F. Dacre, 2012: How much Northern Hemisphere precipitation is associated with extratropical cyclones? *Geophys. Res. Lett.*, **39**, L24809, https://doi.org/10.1029/2012GL053866.

Hersbach, H., and D. Dee, 2016: ERA5 reanalysis is in production. *ECMWF Newsletter*, No. 147, ECMWF, Reading, United Kingdom, 7, http://www.ecmwf.int/sites/default/files/elibrary/2016/16299-newsletter-no147-spring-2016.pdf.

Hirata, H., R. Kawamura, M. Nonaka, and K. Tsuboki, 2019: Significant impact of heat supply from the Gulf Stream on a "Superbomb" cyclone in January 2018. *Geophys. Res. Lett.*, **46**, 7718–7725, https://doi.org/10.1029/2019GL082995.

Hodges, K. I., 1994: A general method for tracking analysis and its application to meteorological data. *Mon. Wea. Rev.*, **122**, 2573–2586, https://doi.org/10.1175/1520-0493(1994)122<2573:AGMFTA>2.0.CO;2.

——, 1995: Feature tracking on the unit sphere. *Mon. Wea. Rev.*, **123**, 3458–3465, https://doi.org/10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2.

——, 1996: Spherical nonparametric estimators applied to the UGAMP model integration for AMIP. *Mon. Wea. Rev.*, **124**, 2914–2932, https://doi.org/10.1175/1520-0493(1996)124<2914:SNEATT>2.0.CO;2.

——, 1999: Adaptive constraints for feature tracking. *Mon. Wea. Rev.*, **127**, 1362–1373, https://doi.org/10.1175/1520-0493(1999)127<1362:ACFFT>2.0.CO;2.

——, R. W. Lee, and L. Bengtsson, 2011: A comparison of extratropical cyclones in recent reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25. *J. Climate*, **24**, 4888–4906, https://doi.org/10.1175/2011JCLI4097.1.

Hoskins, B. J., and P. J. Valdes, 1990: On the existence of storm-tracks. *J. Atmos. Sci.*, **47**, 1854–1864, https://doi.org/10.1175/1520-0469(1990)047<1854:OTEOST>2.0.CO;2.

——, and K. I. Hodges, 2002: New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**, 1041–1061, https://doi.org/10.1175/1520-0469(2002)059<1041:NPOTNH>2.0.CO;2.

——, and ——, 2005: A new perspective on Southern Hemisphere storm tracks. *J. Climate*, **18**, 4108–4129, https://doi.org/10.1175/JCLI3570.1.

——, and ——, 2019: The annual cycle of Northern Hemisphere storm tracks. Part I: Seasons. *J. Climate*, **32**, 1743–1760, https://doi.org/10.1175/JCLI-D-17-0870.1.

Inatsu, M., and B. J. Hoskins, 2004: The zonal asymmetry of the Southern Hemisphere winter storm track. *J. Climate*, **17**, 4882–4892, https://doi.org/10.1175/JCLI-3232.1.

——, and ——, 2006: The seasonal and wintertime interannual variability of the split jet and the storm-track activity minimum near New Zealand. *J. Meteor. Soc. Japan*, **84**, 433–445, https://doi.org/10.2151/jmsj.84.433.

Jung, T., and Coauthors, 2012: High-resolution global climate simulations with the ECMWF model in project Athena: Experimental design, model climate, and seasonal forecast skill. *J. Climate*, **25**, 3155–3172, https://doi.org/10.1175/JCLI-D-11-00265.1.

Kaspi, Y., and T. Schneider, 2013: The role of stationary eddies in shaping midlatitude storm tracks. *J. Atmos. Sci.*, **70**, 2596–2613, https://doi.org/10.1175/JAS-D-12-082.1.

Kawai, H., S. Yukimoto, T. Koshiro, N. Oshima, T. Tanaka, and H. Yoshimura, 2017: Improved representation of clouds in climate model MRI-ESM2. CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling/WMO 47, 707 pp.

——, ——, ——, ——, ——, ——, and R. Nagasawa, 2019: Significant improvement of cloud representation in the global climate model MRI-ESM2. *Geosci. Model Dev.*, **12**, 2875–2897, https://doi.org/10.5194/gmd-12-2875-2019.

Kidston, J., and E. P. Gerber, 2010: Intermodel variability of the poleward shift of the austral jet stream in the CMIP3 integrations linked to biases in 20th century climatology. *Geophys. Res. Lett.*, **37**, L09708, https://doi.org/10.1029/2010GL042873.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, https://doi.org/10.2151/jmsj.2015-001.

Lee, R. W., 2015: Storm track biases and changes in a warming climate from an extratropical cyclone perspective using CMIP5. Ph.D. thesis, University of Reading, 411 pp.

——, T. J. Woollings, B. J. Hoskins, K. D. Williams, C. H. O'Reilly, and G. Masato, 2018: Impact of Gulf Stream SST biases on the global atmospheric circulation. *Climate Dyn.*, **51**, 3369–3387, https://doi.org/10.1007/s00382-018-4083-9.

Lehmann, J., D. Coumou, K. Frieler, A. V. Eliseev, and A. Levermann, 2014: Future changes in extratropical storm tracks and baroclinicity under climate change. *Environ. Res. Lett.*, **9**, 084002, https://doi.org/10.1088/1748-9326/9/8/084002.

Lim, E.-P., and I. Simmonds, 2002: Explosive cyclone development in the Southern Hemisphere and a comparison with Northern Hemisphere events. *Mon. Wea. Rev.*, **130**, 2188–2209, https://doi.org/10.1175/1520-0493(2002)130<2188:ECDITS>2.0.CO;2.

Mauritsen, T., and Coauthors, 2019: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *J. Adv. Model. Earth Syst.*, **11**, 998–1038, https://doi.org/10.1029/2018MS001400.

Murray, R. J., and I. Simmonds, 1991: A numerical scheme for tracking cyclone centres from digital data. Part I: Development and operation of the scheme. *Aust. Meteor. Mag.*, **39**, 155–166.

Neu, U., and Coauthors, 2013: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bull. Amer. Meteor. Soc.*, **94**, 529–547, https://doi.org/10.1175/BAMS-D-11-00154.1.

Nissen, K. M., G. C. Leckebusch, J. G. Pinto, and U. Ulbrich, 2014: Mediterranean cyclones and windstorms in a changing climate. *Reg. Environ. Change*, **14**, 1873–1890, https://doi.org/10.1007/s10113-012-0400-8.

Osburn, L., K. Keay, and J. L. Catto, 2018: Projected change in wintertime precipitation in California using projected changes in extratropical cyclone activity. *J. Climate*, **31**, 3451–3466, https://doi.org/10.1175/JCLI-D-17-0556.1.

Patterson, M., T. Bracegirdle, and T. Woollings, 2019: Southern Hemisphere atmospheric blocking in CMIP5 and future changes in the Australia–New Zealand sector. *Geophys. Res. Lett.*, **46**, 9281–9290, https://doi.org/10.1029/2019GL083264.

Pithan, F., T. G. Shepherd, G. Zappa, and I. Sandu, 2016: Climate model biases in jet streams, blocking and storm tracks resulting from missing orographic drag. *Geophys. Res. Lett.*, **43**, 7231–7240, https://doi.org/10.1002/2016GL069551.

Priestley, M. D. K., H. F. Dacre, L. C. Shaffrey, K. I. Hodges, and J. G. Pinto, 2018: The role of serial European windstorm clustering for extreme seasonal losses as determined from multi-centennial simulations of high-resolution global climate model data. *Nat. Hazards Earth Syst. Sci.*, **18**, 2991–3006, https://doi.org/10.5194/nhess-18-2991-2018.

Reale, M., M. L. Liberato, P. Lionello, J. G. Pinto, S. Salon, and S. Ulbrich, 2019: A global climatology of explosive cyclones using a multi-tracking approach. *Tellus*, **71A**, 1611340, https://doi.org/10.1080/16000870.2019.1611340.

Roebber, P. J., 1984: Statistical analysis and updated climatology of explosive cyclones. *Mon. Wea. Rev.*, **112**, 1577–1589, https://doi.org/10.1175/1520-0493(1984)112<1577:SAAUCO>2.0.CO;2.

Sanders, F., and J. R. Gyakum, 1980: Synoptic-dynamic climatology of the "Bomb". *Mon. Wea. Rev.*, **108**, 1589–1606, https://doi.org/10.1175/1520-0493(1980)108<1589:SDCOT>2.0.CO;2.

Schiemann, R., and Coauthors, 2020: The representation of Northern Hemisphere blocking in current global climate models. *Wea. Climate Dyn. Discuss.*, https://doi.org/10.5194/wcd-2019-19, in press.

Seiler, C., and F. W. Zwiers, 2016: How well do CMIP5 climate models reproduce explosive cyclones in the extratropics of the Northern Hemisphere? *Climate Dyn.*, **46**, 1241–1256, https://doi.org/10.1007/s00382-015-2642-x.

Sinclair, M. R., 1994: An objective cyclone climatology for the Southern Hemisphere. *Mon. Wea. Rev.*, **122**, 2239–2256, https://doi.org/10.1175/1520-0493(1994)122<2239:AOCCFT>2.0.CO;2.

Small, R. J., R. Msadek, Y.-O. Kwon, J. F. Booth, and C. Zarzycki, 2019: Atmosphere surface storm track response to resolved ocean mesoscale in two sets of global climate model experiments. *Climate Dyn.*, **52**, 2067–2089, https://doi.org/10.1007/s00382-018-4237-9.

Tamarin, T., and Y. Kaspi, 2016: The poleward motion of extratropical cyclones from a potential vorticity tendency analysis. *J. Atmos. Sci.*, **73**, 1687–1707, https://doi.org/10.1175/JAS-D-15-0168.1.

——, and ——, 2017: Mechanisms controlling the downstream poleward deflection of midlatitude storm tracks. *J. Atmos. Sci.*, **74**, 553–572, https://doi.org/10.1175/JAS-D-16-0122.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

——, and Coauthors, 2017: CMIP6 global attributes, DRS, filenames, directory structure, and CV's. Tech. Rep. v6.2.6, Program for Climate Model Diagnosis and Intercomparison, 29 pp., http://goo.gl/v1drZl.

Trenberth, K. E., 1991: Storm tracks in the Southern Hemisphere. *J. Atmos. Sci.*, **48**, 2159–2178, https://doi.org/10.1175/1520-0469(1991)048<2159:STITSH>2.0.CO;2.

Voldoire, A., and Coauthors, 2019: Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *J. Adv. Model. Earth Syst.*, **11**, 2177–2213, https://doi.org/10.1029/2019MS001683.

Wernli, H., and C. Schwierz, 2006: Surface cyclones in the ERA-40 dataset (1958–2001). Part I: Novel identification method and global climatology. *J. Atmos. Sci.*, **63**, 2486–2507, https://doi.org/10.1175/JAS3766.1.

Willison, J., W. A. Robinson, and G. M. Lackmann, 2013: The importance of resolving mesoscale latent heating in the North Atlantic storm track. *J. Atmos. Sci.*, **70**, 2234–2250, https://doi.org/10.1175/JAS-D-12-0226.1.

Woollings, T., B. Hoskins, M. Blackburn, D. Hassell, and K. Hodges, 2010: Storm track sensitivity to sea surface temperature resolution in a regional atmosphere model. *Climate Dyn.*, **35**, 341–353, https://doi.org/10.1007/s00382-009-0554-3.

Wu, T., and Coauthors, 2019: The Beijing Climate Center Climate System Model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geosci. Model Dev.*, **12**, 1573–1600, https://doi.org/10.5194/gmd-12-1573-2019.

Yang, M., X. Li, R. Zuo, X. Chen, and L. Wang, 2018: Climatology and interannual variability of winter North Pacific storm track in CMIP5 models. *Atmosphere*, **9**, 79, https://doi.org/10.3390/atmos9030079.

Zappa, G., L. C. Shaffrey, and K. I. Hodges, 2013a: The ability of CMIP5 models to simulate North Atlantic extratropical cyclones. *J. Climate*, **26**, 5379–5396, https://doi.org/10.1175/JCLI-D-12-00501.1.

——, ——, ——, P. G. Sansom, and D. B. Stephenson, 2013b: A multimodel assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. *J. Climate*, **26**, 5846–5862, https://doi.org/10.1175/JCLI-D-12-00573.1.

——, G. Masato, L. Shaffrey, T. Woollings, and K. Hodges, 2014: Linking Northern Hemisphere blocking and storm track biases in the CMIP5 climate models. *Geophys. Res. Lett.*, **41**, 135–139, https://doi.org/10.1002/2013GL058480.

——, M. K. Hawcroft, L. Shaffrey, E. Black, and D. J. Brayshaw, 2015: Extratropical cyclones and the projected decline of winter Mediterranean precipitation in the CMIP5 models. *Climate Dyn.*, **45**, 1727–1738, https://doi.org/10.1007/s00382-014-2426-8.

Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.