Drivers of biases in the CMIP6 extratropical storm tracks. Part I: Northern Hemisphere

Article

Accepted Version

Drivers of biases in the CMIP6 extratropical storm tracks. Part 1: Northern Hemisphere

Matthew D. K. Priestley*

College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK

Duncan Ackerley

Met Office, Exeter, UK

Jennifer L. Catto

College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK

Kevin I. Hodges

NCAS, Department of Meteorology, University of Reading, Reading, UK

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

ABSTRACT

The ability of climate models to represent extratropical storm tracks is vital to provide useful projections. In previous work the representation of the extratropical storm tracks in the Northern Hemisphere was found to have improved from the 5th to 6th coupled model intercomparison project. Here we investigate the remaining and persistent biases in models from the 6th phase of the Coupled Model Intercomparison Project (CMIP), by contrasting the atmosphere-only simulations (AMIP6) with the historical coupled simulations (CMIP6). The comparison of AMIP6 and CMIP6 simulations reveal that biases in sea surface temperatures (SSTs) in the coupled simulations across the North Pacific in winter modify the atmospheric temperature gradient, which is associated with an equatorward bias of the storm track. In the North Atlantic, cyclones do not propagate poleward enough in coupled simulations, which is partly driven by cold SSTs to the south of Greenland, decreasing the latent heat fluxes. In summer, excessive heating across central Asia and the Tibetan Plateau reduces the local baroclinicity causing fewer cyclones to form and propagate from eastern China into the North Pacific in both the coupled and atmosphere-only simulations. Several of the biases described in the coupled models are reduced considerably in the atmosphere-only models when the SSTs are prescribed. For example the equatorward bias of the North Pacific storm track is reduced significantly. However, other biases are apparent in both CMIP6 and AMIP6 (e.g. persistent reduction in track density and cyclogenesis over eastern Asia in Summer), which are associated with other processes (e.g. land surface temperatures).

1. Introduction

Climate models utilize mathematical formulations of the laws of motion and thermodynamics to represent the complex interactions between the atmosphere, ocean, land, biosphere, and numerous other aspects of the Earth system. These models routinely have errors in their representation of the extratropical circulation (Iqbal et al. 2018) and in particular the mid-latitude storm tracks (Chang et al. 2013). Recently, data has become available from the 6th generation of the Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016), which provides the current most advanced coupled atmosphere-ocean model datasets from numerous centers around the world.

The CMIP6 coupled models are able to successfully reproduce the two main Northern Hemisphere (NH) storm tracks over the North Pacific and North Atlantic Oceans when compared with reanalyses (Priestley et al. 2020). However, biases in their representation, which have been evident throughout numerous phases of CMIP, still remain. Priestley et al. (2020) showed that the North Pacific storm track is generally too zonal, with minimal improvements in storm track latitude compared to CMIP5 (see also Chang et al. 2012; Harvey et al. 2020). In the North Atlantic there is still a zonal bias of the storm track, albeit reduced compared with CMIP5 (see also Zappa et al. 2013). In the past, improvements have been linked to increases in the model horizontal and vertical resolutions (Colle et al. 2013; Zappa et al. 2013), and this is also evident in the CMIP6 models. For example, models with horizontal atmospheric resolutions of at least 100km show reduced track density biases and better distributions of peak cyclone intensity (Priestley et al. 2020).

The representation of the oceans in coupled models, and specifically the sea surface temperature (SST) can have widespread impacts on the storm tracks, and also the wider atmospheric general circulation. Errors in North Atlantic SSTs and SST gradients can modify the intensity and propagation of cyclones considerably (de Vries et al. 2019), with SST biases also generating large anomalous Rossby wave trains that impact the general circulation (Lee et al. 2018). The representation of SSTs in the region to the south of Greenland has also been shown to have a significant impact on the atmospheric circulation over the North Atlantic (Keeley et al. 2012; Scaife et al. 2011) and is a bias that arises in ocean models independent of the atmospheric forcing (Tsujino et al. 2020). Most coupled models commonly feature a cold bias to the south of Greenland associated with a Gulf stream that does not turn poleward enough (Zhang and Zhao 2015). One way to determine the influence of SST errors is through atmosphere-only (*amip*) experiments with the same models (Gates et al. 1999) in which only the atmospheric and land components of the models are interactive. In these models the SSTs and sea ice concentration are prescribed and based upon observed values, therefore any errors associated with the ocean and its interaction with the atmosphere should be minimized. Models that have been run in an atmosphere-only configuration tend to show an improved representation of cyclones and the North Atlantic circulation (O'Reilly et al. 2017; Keeley et al. 2012) as well as improving the location and frequency of blocking (Scaife et al. 2011; O'Reilly et al. 2016).

Blocking has been shown to be a major influence on the representation of the storm tracks and affects both the North Atlantic and North Pacific storm tracks (Zappa et al. 2014; Booth et al. 2017a). The representation of blocking has improved in CMIP6 relative to CMIP5 (Davini and D'Andrea 2020; Schiemann et al. 2020) with further improvements gained from increasing resolution (Schiemann et al. 2017). Therefore, the representation of the storm tracks may be simulated better in high-resolution CMIP6 models relative to their lower-resolution counterparts due to this better representation of blocking.

Despite improvements in storm track representation from CMIP5 to CMIP6, there are still some considerable biases of note such as an equatorward bias in the North Pacific and a zonal bias of tracks in the North Atlantic (Priestley et al. 2020). Through further examination of coupled and *amip* simulations it may be possible to isolate, and attribute biases to specific deficiencies in either model physics or the accuracy of represented large-scale features. In this study the aim is to identify the possible drivers of the persistent storm track biases, and also to understand why these biases are present. Consequently, the main research questions to be addressed in this study are as follows:

- What impact do SST biases from a fully coupled, dynamical ocean have on the storm tracks in the Northern Hemisphere?
- Can coupled model storm track biases be linked to large-scale, mean-state biases in CMIP6 models?

The paper continues as follows. In section 2 the data and methods used for this work are described. In section 3 the results and findings will be presented. Finally, in section 4 the key points of this work and its implications in the wider scientific context will be discussed.

2. Data and Methods

a. Datasets

1) CMIP6 Models

In this study, models that are part of the CMIP6 DECK experiments are used (Eyring et al. 2016). The *historical* and *amip* model runs are analyzed covering the period from 1979-2014. Analysis focuses on the NH winter and summer seasons, these being the December, January, February

(DJF) and June, July, August (JJA) periods respectively. The models used in this study are listed in Table 1. Data is available from 24 models, which include both coupled atmosphere-ocean model *historical* simulations and atmosphere-only *amip* simulations. The number of models is restricted to those which provide the variables required for cyclone tracking at a 6-hourly temporal resolution. A full explanation of the differences between the experiments can be found in Eyring et al. (2016). In this study the coupled atmosphere-ocean *historical* models will be referred to as the *CMIP6* models, with the *amip* models being referred to as *AMIP6*. For all models only a single ensemble member (*r1i1p1f1* or lowest available) is used in the study.

2) Reanalysis

The ERA5 reanalysis (Hersbach et al. 2020) is employed as the reference for real-world atmospheric variability and is used to compare with the CMIP6 and AMIP6 models used in this study. ERA5 data is available from January 1950, however the period 1979-2014 is used in this study to provide a consistent comparison period. ERA5 data are 0.28°× 0.28° (~31 km) spatial resolution. For ERA5 and the CMIP6 models described above, all analyses are performed on the native model resolution, then data are re-gridded onto a 1°× 1° grid for the purposes of visualization and comparison.

b. Feature Tracking

TRACK code (Hodges 1994, 1999) is used for the objective identification and tracking of extratropical cyclones, as in Priestley et al. (2020). Relative vorticity at 850 hPa is used as the input variable, which allows for a reduced influence of the background state on cyclonic features and focuses on smaller spatial scales. As the model and reanalysis data is provided at different horizontal resolutions, the relative vorticity field is first truncated to T42 resolution with all planetary wavenumbers (5 and below) removed. This ensures tracking and cyclone identification are performed at a common resolution. Cyclones are initially identified prior to tracking as maxima above a threshold of $1\times10^{-5}s^{-1}$ on a polar stereographic projection. To ensure only long-lived, mobile synoptic systems are included in the analysis all analyzed cyclones must travel at least 1000 km and have a lifetime of at least 48 hours.

Cyclone track density is calculated using spherical nonparametric estimators from the individual cyclone tracks (Hodges 1996). In cases where cyclone genesis and lysis latitude are quantified this is taken respectively as the latitude of the first and last timestep that the cyclone is identified. The poleward displacement of cyclones is analyzed for the early part of the lifecycle and is taken as the latitude difference between the 9th and 1st timestep of the cyclone track (i.e. first 48 hours of the lifecycle).

Cyclogenesis rates for two large regions will be considered in the main text. These regions follow on from Priestley et al. (2020) and are described therein (see also their Fig. 1a). The two regions capture the main North Atlantic and North Pacific storm tracks and cyclones must form within their bounds to count toward that region's cyclogenesis rate. Region 1 extends from North America, across the North Atlantic, and into Siberia (also described as the America-Atlantic-Siberia region). Region 2 encompasses from eastern Asia and the Tibetan Plateau eastwards to the far eastern North Pacific (also called the Asia-Pacific region).

c. Temperature Gradients

Temperature gradients are calculated using the potential temperature (θ) on pressure levels. The meridional gradient of θ is used and is calculated by the Iris package (Met Office 2010 - 2013) and gradients are quoted in units of K degree^{-1}. In our calculations the θ gradient is required to be positive and is therefore multiplied by -1 for the NH.

3. Results

a. Cyclone Track Densities and Statistics

The CMIP6 coupled model biases were extensively documented in Priestley et al. (2020). Figures 1a and 1d show the track density biases of the CMIP6 multi-model mean, which are almost indistinguishable from those presented for the 20 model ensemble in Priestley et al. (2020). The biases for the corresponding AMIP6 experiments are shown in figures 1b and 1e, with the differences between AMIP6 and CMIP6 in figures 1c and 1f.

For the NH winter (DJF; Figs. 1a–c) a general poleward displacement of both the North Atlantic and North Pacific storm tracks is observed in the AMIP6 experiments compared to CMIP6 (Fig. 1c). The largest poleward displacement in the AMIP6 storm tracks relative to CMIP6 is seen in the west of both ocean basins, where there is high model agreement (Fig. 1c). This is where observed SST gradients are largest in the mid-latitudes and where the coupled models commonly have large errors, which have an impact on the atmospheric circulation (e.g. Woollings et al. 2010; Lee et al. 2018). Notably, there is also a reduction of the zonal bias in the North Atlantic. This reduction in AMIP6 extends from the Gulf of Mexico towards western Europe along ~40°N. The equatorward storm track bias in the North Pacific is substantially lower in AMIP6 than CMIP6 (compare Figs. 1a,b). Despite these improvements, there is still an underestimation of track density in both the North Atlantic and North Pacific in the AMIP6 models (Fig. 1b). Some of this underestimation is likely a result of track density being a function of the number of tracks, as well as the cyclone path, and that there are too few cyclones generated by models around the NH (Table S2 and Priestley et al. 2020).

There are positive track density biases over western Europe and negative biases over the Mediterranean in both the AMIP6 and CMIP6 runs, with no improvement in the former. The western European-Mediterranean track biases have been shown to be associated with blocking (see Zappa et al. 2014). It is interesting that track density biases in AMIP6 and CMIP6 models are similar in this region as there is evidence of North Atlantic SSTs modulating blocking frequency over Europe, however the strength of this link has been debated (Scaife et al. 2011; O'Reilly et al. 2016; Davini and D'Andrea 2016). Recently, however, the representation of blocking has been shown to be similar in coupled and atmosphere-only models, yet sensitive to changes in ocean resolution (Schiemann et al. 2020).

In the NH summer (JJA; Figs 1d–f) the AMIP6 models feature a similar pattern of biases to CMIP6, but with larger magnitudes. This is particularly notable for the large underestimation of track densities over the North Pacific from eastern Asia (~30°N-40°N, 120°E-160°E) and also the western North Atlantic. As the patterns of the CMIP6 and AMIP6 track densities are similar in JJA and the AMIP6 biases are generally larger in magnitude than CMIP6, it is likely that the presence of coupling (and its associated biases) is having a compensating effect on biases that originate in the atmosphere and land components of the models. The overall number of cyclones simulated in AMIP6 and CMIP6 models is very similar (Table S2), with both simulating significantly fewer than identified in ERA5.

In order to further examine the differences between the AMIP6 and CMIP6 storm tracks, and to understand how the characteristics of the cyclones contribute, statistics of genesis latitude, lysis latitude, and poleward displacement of the cyclones have been generated for the North Atlantic (Fig. 2). The statistics presented in Fig. 2 are for cyclones that form within the core genesis region of the North Atlantic storm track (cyan box in Fig. S1a/e). During DJF (Fig. 2a–c) the CMIP6 model cyclones in the North Atlantic have a median genesis latitude that is ~0.6° further poleward than is observed in the reanalyses (significant, $p<0.05$). Atmosphere-only models tend to have a poleward bias relative to reanalyses (Kodama et al. 2015; Bodman et al. 2020) and AMIP6 models are also further poleward than the CMIP6 models. Despite the poleward genesis bias of the CMIP6 models, the lysis latitude is comparable with the reanalyses (Fig. 2b); however, the AMIP6 models are significantly ($p<0.05$) further poleward than CMIP6 in their lysis by ~0.6°. This is notable as the track density bias in the North Atlantic is zonal/equatorward in nature, indicating that this bias does not result from biases in genesis or lysis location, but instead from the track of the cyclones. Both the CMIP6 and AMIP6 models underestimate the cyclone poleward movement relative to the reanalyses (Fig. 2c). Despite an underestimation relative to ERA5, the AMIP6 models show an improved poleward displacement of cyclones compared to CMIP6, which is consistent with the improvements in track density noted in Fig. 1c. Therefore, the bias in track density in the North Atlantic is to some extent driven by the rate at which cyclones are moving polewards. As the poleward movement bias is lessened in AMIP6 models, errors in either the atmosphere-ocean coupling or absolute SST field are likely responsible for the strong zonal bias.

In JJA, cyclones forming in the North Atlantic generally form significantly too far poleward, similar to DJF (Fig. 2d), with the AMIP6 models simulating cyclones forming further poleward than CMIP6. With regards to the lysis latitude (Fig. 2e) the cyclones in JJA generally also dissipate too far poleward. The poleward genesis bias is a result of too few cyclones forming over the southeastern USA, and too many over the northeastern USA (Fig. S1e). Both the CMIP6 and AMIP6 models have very similar 48 hour latitude changes compared to the reanalyses (Fig. 2f). Therefore it appears that the AMIP6 models have storm tracks that are systematically too far poleward in JJA, yet both CMIP6 and AMIP6 models have a good representation of the tilt of the storm track in the summer season. As CMIP6 biases are minor compared to the reanalyses, it appears that the negative track density biases in the North Atlantic in Fig. 1d–e are mostly a result of an insufficient cyclogenesis rate (Table S3).

In the North Pacific in DJF, cyclogenesis generally occurs slightly too far poleward in both CMIP6 and AMIP6 models compared to the reanalyses (Fig. 3a). Despite the bias in the genesis latitude not being significant between CMIP6 and the reanalyses, the AMIP6 models simulate genesis significantly further poleward than CMIP6, with a median latitude that is above the 75th percentile of the reanalyses. With regards to the lysis latitude, it is too far equatorward in the CMIP6 models and too far poleward in the AMIP6 models. Consequently, CMIP6 cyclones do not propagate far enough poleward compared to AMIP6 or the reanalyses.

In JJA in the North Pacific (Fig. 3d–f) all model groups have a very large and significant poleward bias of the median genesis latitude of at least 1.5°. Both the CMIP6 and AMIP6 model ensembles simulate genesis and lysis that is too poleward relative to the reanalyses (Figs. 3d,e) by at least 1.8°. Despite this, the poleward propagation is well represented, with CMIP6 and AMIP6 medians being indistinguishable from the reanalyses median (Fig. 3f). The poleward bias of cyclones is evident in the track density (Fig. 1d–f) and from the underestimation of genesis density equatorward of 40°N (Fig. S1e–g). Therefore, it appears that the poleward bias is a result of a large underestimation of cyclogenesis (and resultant track density) on the equatorward flank of the storm track in JJA and not of excess cyclogenesis on the poleward flank as may be suggested from Fig. 3d–f.

b. Large-scale biases and their impact on the storm track

In this section the relationships between large-scale model biases and the extratropical storm track biases as described above and in Priestley et al. (2020) are investigated and discussed. Focus will be on evaluating seasonal mean features and differences between the AMIP6 and CMIP6 models.

1) North Pacific - Winter

The main differences between the CMIP6 and AMIP6 simulations are the dynamical ocean and its coupling to the atmosphere. The mean DJF SSTs used in ERA5 are shown in Fig. 4a. Cyclone growth commonly occurs in association with the largest SST gradients, which are shown in Fig. S2. The CMIP6 model SSTs show large errors in the vicinity of the Kuroshio current, with SSTs that are too high on the cold side of the strongest gradient and too low on the warm side (Fig. 4b). In the central North Pacific the SSTs are underestimated by a majority of the CMIP6 models across the entire ocean basin by over 2°C from 150°E-200°E along 30°N. This SST bias is similar, albeit larger in magnitude and extended zonally, compared to that demonstrated in OMIP experiments, which are forced by atmospheric reanalysis (Tsujino et al. 2020).

In addition to the differences in the SST field, there are also differences in the representation of the atmospheric circulation between CMIP6 and AMIP6 models (Fig. 5a–c). As with the storm track density (Fig. 1a–c) there is a robust zonal bias of the zonal wind across the North Pacific in CMIP6 models, particularly east of 180°W, which is directly east of the largest SST anomalies. In the AMIP6 models there is a poleward shift of the zonal wind relative to CMIP6 (Fig. 5c) across the entire North Pacific, and therefore small biases relative to ERA5 (not shown). To quantify if there is any relationship between the SST anomaly and the atmospheric circulation a grid-point regression of both the storm track density and 850 hPa zonal wind against seasonal mean SST bias in the central North Pacific (20°N-40°N, 160°W-200°W) is performed (Fig. 6). This regression is performed across model climatologies of zonal wind, storm track density, and SST. For both the zonal wind (Fig. 6a) and the storm track density (Fig. 6b) a statistically significant dipole pattern is present in the North Pacific and North Atlantic that indicates an equatorward displacement of the jet/storm track when there are larger negative SST anomalies in the central North Pacific. This also suggests that in models when the SST bias is smaller, the storm track has less of an equatorward bias and is likely to be in a similar location to the AMIP6 models' mean position (Fig. 1b, 5c). We also performed the regressions in Fig. 6b with the AMIP6 track densities and no dipole relationship was observed (not shown). This demonstrates the importance of SSTs biases in the large-scale atmospheric circulation of coupled climate models.

One possible way in which SST biases contribute to the shift in the jet and storm track is through the modification of the atmospheric temperature gradient, which is plotted in Fig. 7a for ERA5. Cyclones tend to preferentially form in regions of higher temperature gradients and in the CMIP6 models the strongest gradients are shifted equatorward relative to ERA5 (Fig. 7b). In all models the maximum temperature gradient is located 5-10° equatorward of the maximum storm track density, with the biases also showing this behaviour. Across a majority of the North Pacific there is an equatorward shift of the maximum temperature gradient, which is a result of a cooling of the lower atmosphere directly above the cold SST bias in the central North Pacific (Fig. 4b). In the AMIP6 models, the largest temperature gradient is further poleward than in the CMIP6 models across the entire North Pacific and also North America, with there being no atmospheric cooling from the underlying SST biases (Fig. 7c) and minimal biases relative to ERA5 (Fig. S3). As the SST bias appears even when forced by the observed atmosphere (Tsujino et al. 2020), it is likely that the initial mean-state equatorward bias of the storm track and zonal wind in the CMIP6 models is a result of this forcing. However, with the SST bias being zonally extended in CMIP6 models compared to OMIP experiments (Tsujino et al. 2020), there is likely a feedback from the storm track onto the ocean acting to amplify and extend the cold bias (as in Dacre et al. 2020). As a subsequent poleward shift of the storm track in AMIP6 experiments is seen when forced by SSTs that do not have these inherent biases, it is evident that the coupling to an interactive ocean is the leading driver of the equatorward bias in the storm track.

2) North Atlantic - Winter

As in the North Pacific, there are large SST anomalies associated with the region of largest SST gradients (Gulf Stream) in CMIP6 (Fig. 4b,c). Temperatures are too low on the warm side of the strongest gradient and too high on the cold side, resulting in an SST gradient that is weaker than in ERA5 (Fig. S2).

A large number of the biases in the storm tracks in the North Atlantic region noted in Fig. 1 are further identifiable in the zonal wind at 850 hPa (Fig. 5a–c). There is a zonal bias of the jet over the eastern North Atlantic into western Europe, which is identifiable throughout the depth of the troposphere (not shown). In the AMIP6 models (Fig. 5c) there is a poleward shift of the North Atlantic jet relative to CMIP6 across the entire basin. The strength of the poleward shift in zonal wind from CMIP6 to AMIP6 across the Gulf of Mexico, North America, and the western North Atlantic, is larger than the bias of CMIP6 models relative to ERA5 (compare Figs. 5b,c). Therefore, this poleward shift of the zonal wind in the AMIP6 models is in agreement with the poleward genesis and lysis bias of cyclones in AMIP6 models relative to the reanalyses (Fig. 2a), and the poleward shift in the storm track across North America for AMIP6 relative to CMIP6 (Fig. 1c). Despite a poleward shift of the circulation across the North Atlantic there are minimal improvements in the 850 hPa zonal wind over Europe in the AMIP6 models, as with the track density, relative to CMIP6.

In the CMIP6 models, the latitude of the storm track across North America and the North Atlantic appears to be related to the North Pacific SST biases (Fig. 6). Over North America, cyclone tracks in the CMIP6 models have a tendency to be displaced toward the Gulf of Mexico when the cold SST bias over the central North Pacific is larger (Fig. 6b), with this also being the case for the zonal wind (Fig. 6a) throughout the troposphere (not shown). The shift of the circulation/storm track is associated with an equatorward bias in the largest lower-tropospheric potential temperature gradient (Fig. 7b). Consequently, there is an excess of cyclogenesis/track density in the CMIP6 models (relative to ERA5) over the Gulf of Mexico and Southern USA (20-35°N, 250-270°E; Fig. S1d), and lower track density across the continental USA relative to ERA5. The biases across the Gulf of Mexico are reduced considerably in the AMIP6 models as the circulation, temperature gradient, zonal wind, and track density shift poleward (Fig. 1c, S1d, 5c, 7c). The impact of cyclogenesis biases in this region on the North Atlantic storm track can be tested by isolating all cyclones forming in this anomalous cyclogenesis region over the Gulf of Mexico (Fig. 8a) and removing them from the CMIP6 track density (Fig. 8b).

By removing cyclones that form in this anomalous region over the Gulf of Mexico the pattern of track density in Fig. 8b presents a different picture to that for all cyclones in CMIP6 (Fig. 1a). There is a reduction in the positive track density bias that originates in the Gulf of Mexico that extends to the northeast across Florida and into the western North Atlantic (compare Figs. 8b and 1a). Removing the Gulf of Mexico cyclones from the track density appears to have little impact on track density bias east of 60°W. The removal of Gulf of Mexico cyclones from CMIP6 models (Fig. 8c) also results in a track density pattern that is strikingly similar to the AMIP6 model bias in the western North Atlantic and across the southern USA (compare Fig. 8b and 1b). Therefore, having the correct SST distribution in the North Pacific reduces the equatorward bias of track density in the Gulf of Mexico and western North Atlantic.

In addition to biases surrounding the Gulf Stream there is also a negative SST anomaly to the south of Greenland in the North Atlantic (Fig. 4b). This bias has been identified in numerous modeling studies (e.g. Scaife et al. 2011; Wang et al. 2014) and is associated with atmospheric circulation biases in the northeastern North Atlantic. Situated above this negative SST bias is a large underestimation in the strength of the meridional wind at 700 hPa in CMIP6, relative to ERA5 (Fig. 9b). In the AMIP6 models there is an increase in the meridional wind relative to CMIP6 to the south of Greenland (Fig. 9c), which is directly west of the poleward shift of the zonal wind in the North Atlantic in AMIP6 relative to CMIP6 (Fig. 5c). As low-to-mid level winds are often eddy-driven, these meridional wind anomalies may be a result of, rather than a driver of, the changes in cyclone motion. However, this increased meridional wind to the south of Greenland likely indicates where the increased poleward propagation of cyclones in the first 48 hours of their lifecycle is occurring in AMIP6 models relative to CMIP6 (Fig. 2c) and results in a poleward shift of the circulation downstream (Fig. 5c). To test this hypothesis we have performed linear least squares regression of the SST bias to the south of Greenland against the storm track density and 850 hPa zonal wind (Fig. S4). We find a relationship between the atmospheric circulation variables and the SSTs which confirms that models with colder SSTs to the south of Greenland are associated with a more equatorward storm track density over the eastern North Atlantic, as we observe in Fig. 1a).

To understand how the Greenland SST bias is influencing the atmosphere and reduced cyclone poleward propagation in CMIP6 models, we identify a reduction in the CMIP6 ocean-atmosphere latent heat flux by over 90 W m^{-2} (40°N-50°N, 40°W; Fig. 10b). This reduction in heat flux is a direct result of the reduced temperatures of the ocean, with lower surface temperatures resulting in less heat transfer to the lower atmosphere (consistent with; Kushnir and Held 1996; Keeley et al. 2012). In the AMIP6 models this negative heat flux anomaly is not present (therefore it is a positive anomaly relative to the coupled models; Fig. 10c) and consequently the AMIP6 models have a greater source of energy from the ocean. Studies by Tamarin and Kaspi (2016, 2017) concluded that cyclones with larger latent heat release tended to be more intense and feature stronger poleward movement through modification of upper-level potential vorticity (PV). As there is a reduction in the zonal bias of track density in AMIP6 relative to CMIP6 east of 60°W (and greater poleward propagation in this region; Fig. 2c, 9c), the additional latent heat flux from the ocean may be driving this process.

Despite improvements in the zonal track density bias over the North Atlantic there is still a lack of improvement in storm track density over Europe in AMIP6 models relative to CMIP6 (Fig. 1c). This is thought to be linked to limited improvements in the representation of blocking in atmosphere-only simulations (Schiemann et al. 2020). Block amplitude and onset are commonly linked to the amount of latent/condensational heating within the warm conveyor belt (WCB) of upstream extratropical cyclones (Pfahl et al. 2015; Steinfeld and Pfahl 2019; Steinfeld et al. 2020; Maddison et al. 2020). Precipitation is a good proxy for cyclone latent heating, and despite

AMIP6 models having the correct ocean-atmosphere latent heat flux (Fig. 10c) they simulate less precipitation per day in the North Atlantic than cyclones in ERA5, and more than in CMIP6 models (not shown). We therefore hypothesize that the AMIP6 models have sufficient latent heating to yield an improvement in poleward propagation, but insufficient to drive the condensational heating required to have a downstream impact on block formation over Europe. One way in which this may be improved is through higher atmospheric resolution, which has been shown to improve the rate of diabatic heating within cyclones (Willison et al. 2013).

3) North Pacific - Summer

In JJA the SST anomalies in the CMIP6 models are almost identical to those in DJF (Fig. 4) and therefore will not be explored in as much detail. The negative SST bias across the central North Pacific is a persistent feature of the model mean and due to the lower SSTs, there is a reduced SST gradient along 40°N east of 180°E (as in Fig. S2b–d), which may have an influence on the location of maximum baroclinicity.

The biases in the zonal wind in summer (Fig. 5d–f) are smaller than in the winter, however, the pattern of biases is consistent with the storm track biases, particularly west of 170°E and across eastern Asia (Fig. 5 in Priestley et al. 2020, and Fig. 1d–f). Relative to ERA5, the maximum zonal wind is situated further poleward across the North Pacific in JJA for the CMIP6 models (Fig. 5e). Other notable features are the weaker zonal wind to the southeast of Japan and the poleward shift of the jet across the east of the basin, both of which are features consistent with the track density bias (Fig. 1d). The AMIP6 models are broadly consistent with CMIP6, but feature further weakening of the zonal wind to the south of Japan and a more pronounced poleward shift of the jet (Fig. 5f). The poleward bias of the jet across the west of the basin is consistent with the cyclogenesis latitude biases (Fig. 3d–f), and also the underestimation of cyclogenesis for the lower latitudes of eastern Asia (Fig. S1f–h). As the AMIP6 models represent an amplification of the biases in zonal wind and track density (Figs. 1f, 5f) this suggests that coupling with an interactive ocean may actually be counteracting deficiencies in the atmosphere model.

As in DJF, the simulated gradients of potential temperature appear critical in controlling the biases in track density and zonal wind (Fig. 7d–f). The presence of the persistent cold biases in the central North Pacific acts to decrease the temperature gradient from 40-50°N and increase it from 20-40°N, with this dipole being most prominent east of 180°W (Fig. 7e). As the storm track is situated farther poleward in JJA (Fig. 1), it is influenced by the reduction in temperature gradient on the poleward side of the cold anomaly, with the CMIP6 models demonstrating a reduction in storm track density/zonal wind strength in this location (Fig. 1e, 5e). Furthermore, there is a strengthening of the temperature gradient in the high latitude North Pacific (Fig. 7e) as a result of negative SST biases surrounding the Bering Strait, and warm biases to the south of Alaska, therefore contributing to the increased track density noted in Fig. 1e.

In the AMIP6 models, the modifications of the temperature gradient are reduced and there is a large-scale warming, relative to CMIP6, across the North Pacific centred on 40-50°N (gray contours Fig. 7f). As a result the temperature gradients are further increased at high latitudes, and decreased at low latitudes, contributing to the poleward shift in zonal wind and track density in the AMIP6 models relative to CMIP6 (Fig. 1f, 5f). As in DJF, the poleward shift of temperature gradients in the North Pacific also appears to result in a similar shift over North America, and also extending downstream toward the North Atlantic (Fig. 7f) indicating that the two storm tracks should not necessarily be treated as independent features.

The persistent underestimation of track density across eastern Asia and the western Pacific appears to originate from a reduction in cyclogenesis over the continent. The region of reduced cyclogenesis is situated directly over a region of underestimated temperature gradient across eastern China (Fig. 11b). The temperature gradients are also weaker over southern Japan and the western North Pacific, which are connected to the negative SST bias (Fig. 7e). In the AMIP6 models, the temperature gradients are even weaker than in the CMIP6 models (Fig. 11c), hence the cyclogenesis rate and track density are lower (Fig. S1h, 1f). The northern genesis region is co-located with positive temperature gradient anomalies which is likely to be more conducive to cyclogenesis (Fig. 11a–c).

The reason for these differences in the temperature gradient can be traced back to excess heating occurring over the Tibetan Plateau and northern India (Fig. 11d–f) which increases (decreases) the temperature gradient on the poleward (equatorward) flank of the Tibetan Plateau and across large parts of northern Asia (Fig. 11b–c). This increase in potential temperature, and therefore changes in temperature gradient, are more visible in the AMIP6 simulations. These changes to the temperature gradient lead to changes in the baroclinicity that acts to increase the cyclogenesis for the northern genesis region and reduces cyclogenesis for the southern genesis region (Fig. S1f–g and 11b–c). Furthermore, increasing the temperature gradient across northern Asia explains the positive track density bias across all of northern Eurasia noted in Fig. 5 of Priestley et al. (2020). This temperature gradient shift may also act to increase the strength of the jet farther polewards.

The excess heating of central and southern Asia is associated with excess surface sensible heat flux from the land to the atmosphere over large regions of northern India in all model ensembles (Fig. S5), of which there may be many origins which would need investigating further. The resultant track density underestimation is a robust bias that has been present since the CMIP5 models and is independent of ocean variability as all these features are more evident in the AMIP6 models. Therefore, reducing this positive heating bias is a clear region for further model development and should have a direct impact on the latitude of the summer storm track over the North Pacific.

4) North Atlantic - Summer

In JJA, cyclones are generally situated too far poleward in the North Atlantic, with genesis being 0.5° and 1° too poleward in CMIP6 and AMIP6 models respectively (Fig. 2d). This is a result of genesis rates being underestimated across the southeastern USA, and slightly overestimated over the northeast USA (Fig. S1f–h). The additional poleward bias in AMIP6 relative to CMIP6 comes from an amplification of these biases. Examining the biases in temperature gradient it can be seen how the CMIP6 models have a gradient that is too low across the southeast USA and too high across large parts of eastern and central Canada (Fig. 7e). As temperature gradients play a strong role in atmospheric baroclinicity it is likely that this is a large driver of the genesis (and therefore track density) biases. The temperature gradient bias is a result of temperatures over the Rocky mountains (and downstream) being too high (gray contours Fig. 7e), which may be influenced by an incorrect representation of the orographic features. In the AMIP6 models, the temperature gradient is even higher over Canada and lower over the eastern USA (Fig. 7f) as a result of even further warming across a central band of the USA, peaking over the Rocky mountains.

Interestingly, the pattern of zonal wind biases does not reflect the temperature gradient, or track density biases, in CMIP6 models across the eastern USA (Fig. 5e). The winds are generally biased equatorward in the CMIP6 models, with a poleward shift in the AMIP6 models (Fig. 5f), that may be associated with the shift in the temperature gradient. It was shown in DJF that the poleward shift of the jet and temperature gradient over North America is linked to the SST bias in the North Pacific, therefore the shift observed in JJA in the AMIP6 models (relative to CMIP6) may also be influenced by correcting the distribution of North Pacific SSTs in the AMIP6 models. The poleward shift in temperature gradient across North America does appear to be coherently downstream of the shift in temperature gradient resulting from the differences in the models (Fig. 7f).

Across the North Atlantic there are minimal biases in the zonal wind of the CMIP6 models (Fig. 5e), with a slight reduction in temperature gradient to the south of Greenland (Fig. 7e), which is likely associated with the negative SST bias in the centre of the basin (as in Fig. 4b). In the AMIP6 models a further poleward shift of the zonal wind and temperature gradient is seen east of 50°W (Figs. 5f, 7f). These poleward shifts are consistent with the poleward shift in track density to the south of Greenland and are likely driven by the warming of the lower troposphere from 40°N-50°N (gray contours Fig. 7f) that have a maximum over the region where the negative SST anomaly is found in the CMIP6 models. Therefore, correcting the SST bias alters the temperature distribution of the ocean and atmosphere, modulating the temperature gradient and creating an environment more preferential for cyclone growth and development on its northern flank, as was also observed in the same region during DJF. For improvements in track density representation in models, the long-standing, robust, underestimation of track density in the North Atlantic in JJA may be improved through increasing the temperature gradient across North America, as there are minimal biases in any of the other large-scale fields.

4. Discussion and Conclusions

In this study the large-scale drivers of biases in simulated Northern Hemisphere extratropical storm tracks have been investigated. Comparisons have been made between the coupled models documented in Priestley et al. (2020) and the corresponding atmosphere-only models. For a majority of the major track density biases the forcing of these biases can be traced back to errors in the ocean state and the forcing applied by these persistent errors. Furthermore, there is significant influence from discrepancies in the large-scale temperature gradients and jet structures, as well as in interactions between the land and the atmosphere. The key findings of this work are summarized as:

- A large number of the major storm track biases seen in the CMIP6 models in winter are smaller in the AMIP6 simulations (Fig. 1a–c). There is a reduced equatorward bias in the North Pacific and reduced zonal bias in the North Atlantic. Despite improvements, some biases are still present in the AMIP6 storm tracks, such as a reduction in overall cyclogenesis relative to both CMIP6 and the reanalyses.

- In DJF, the AMIP6 simulations show increased poleward displacement of cyclones for both storm tracks in the early part of their lifecycles, reducing the zonal bias of tracks seen in CMIP6 simulations (Fig. 2a–c and 3a–c).

- The equatorward bias in the North Pacific in the CMIP6 models originates from large negative SST biases (Fig. 4b) which are associated with shifts of the temperature gradient and zonal wind equatorwards (Figs. 4–7b). In the AMIP6 models the SST bias is not present, so there are minimal biases in the latitude of the maximum temperature gradient or zonal wind (Figs. 5c, 7c).

- In the North Atlantic in winter, the too weak poleward displacement of cyclones is associated with a persistent cold anomaly in the North Atlantic to the south of Greenland (Fig. 4b). This SST anomaly reduces the latent heat flux from the ocean to the atmosphere (Fig. 10b) and consequently there is a reduced meridional component to the steering flow (Fig. 9b) and large underestimation of cyclone poleward propagation.

- Over the western North Atlantic in winter the positive track density bias in the CMIP6 models is a result of excess cyclogenesis occurring over the Gulf of Mexico (Fig. 8 and S1a–d). This excess cyclogenesis is a result of the equatorward biased jet extending from the North Pacific combining with a higher temperature gradient to the South of the Rocky mountains and over the Gulf of Mexico, creating an environment favourable for cyclogenesis.

- In summer, both the North Atlantic and North Pacific storm tracks show a poleward shift in the location of the largest track densities in AMIP6 compared to CMIP6, with the major biases in CMIP6 also being visible in the AMIP6 models (Fig. 1d–f) with underestimations in the North Atlantic and across eastern Asia.

- There are minimal biases in the North Atlantic in summer with regards to the poleward displacement or genesis/lysis latitude in the CMIP6 models (Fig. 2d–f). However, in the North Pacific, cyclones are too poleward by up to 2.5° in both genesis and lysis latitude (Fig. 3d–f).

- In the North Pacific in summer both the CMIP6 and AMIP6 models are characterized by an underestimation of track density across eastern Asia and the western North Pacific on the southern flank of the storm track (Fig. 1d–f). This underestimation results from an almost absence of cyclogenesis in the southern of the two genesis regions over eastern Asia (Fig. 11a–c). The lack of cyclogenesis is driven by a reduced temperature gradient in this region from an increase in the surface sensible heat flux (Fig. S5) that contributes to increased heating of central Asia and over the Tibetan Plateau (Fig. 11d–f).

- Track density and cyclogenesis in summer are underestimated for the whole North Atlantic storm track in the CMIP6 and AMIP6 models (Fig. 1d–f and S1f–h). The underestimation is driven by reduced temperature gradients across the southeastern USA (Fig. 7e).

Many of the results summarized in this paper demonstrate that by forcing models with the correct SST and sea ice distribution leads to improvements in the mean state flow and therefore in the seasonal storm track density. This is particularly notable for the AMIP6 simulations in winter where the North Pacific storm track has a reduced equatorward bias and the zonal bias of the North Atlantic storm track is removed. Despite the numerous improvements in the AMIP6 models used in this study, there are still biases that remain in the models that are independent of coupling, or even compensated by the presence of coupling to an interactive ocean. The most striking of these is the general underestimation of cyclogenesis, which is present in both DJF and JJA across both ocean basins (see section 3a). Furthermore, simulated cyclones struggle to travel poleward enough, especially in the North Atlantic in DJF (Fig. 2c). Finally, the entire storm track tends to be too far poleward in the North Pacific in JJA. This is another bias that has long persisted since CMIP5 (see Priestley et al. 2020), with minimal evidence of improvement.

One factor that can influence cyclogenesis is the improvement of large-scale temperature gradients by increasing model resolution. Priestley et al. (2020), Bracegirdle et al. (2021), Zappa et al. (2013), and Baker et al. (2019) have all shown that higher atmospheric resolution models tended to have better cyclogenesis rates or improved jet latitude. Improving resolution may also have other impacts such as strengthening eddy feedbacks (Scaife et al. 2019) and improving the representation of orography and associated wave drag (Pithan et al. 2016; Davini et al. 2021). Increasing ocean resolution can also have a significant impact (Woollings et al. 2010) and can be tested through the HighResMIP project (Haarsma et al. 2016) and the *highresSST-present* experiments which are atmosphere-only simulations but with an ocean horizontal resolution of $\frac{1}{4}^{\circ}$, therefore much higher than in the AMIP6 experiments. Despite improvements in resolution, considerable variability in ocean representation remains across model families (Chassignet et al. 2020) and therefore increasing ocean resolution may not correct all the remaining biases described above.

One finding of this study regards the increased poleward movement of cyclones in the AMIP6 models compared to CMIP6. Tamarin and Kaspi (2016, 2017) found that cyclones that moved farther poleward tended to be of a higher intensity, or intensify rapidly. The cyclones in the AMIP6 simulations do receive additional moisture and heat from the ocean via enhanced latent heat fluxes. Nevertheless, it is likely that the models are incapable of producing the correct amount of condensational heating to resolve the additional intensification seen in the reanalyses (as in Keeley et al. 2012). It will be of interest to see if cyclones in the AMIP6 simulations do achieve higher intensities than cyclones in the CMIP6 simulations, and simulate an increased number of explosive cyclones with improved heat fluxes (e.g. Hirata et al. 2019), or if this is something that can only be improved with further increased horizontal resolution (e.g. Jiaxiang et al. 2020).

One persistent bias that has a large influence on the model storm track, and is not improved in the AMIP6 models, is the warm bias over central Asia and the Tibetan Plateau. Model simulations of surface temperatures over the Tibetan Plateau are generally poor (Su et al. 2013; Zhu and Yang 2020) and have been linked to biases in surface albedo and snow cover (Chen et al. 2017). This bias is the likely reason for the limited improvement in the North Pacific summer storm track structure compared to the CMIP6 models. Hoskins and Hodges (2019) showed that cyclones from the western North Pacific played a key role in aiding cyclogenesis for the eastern North Pacific, therefore any improvements for the west of the basin would first require improvements in cyclogenesis and temperature gradients over eastern Asia.

The large negative SST bias in the North Pacific in the coupled models, which has a large impact on the structure of the North Pacific storm track and a downstream influence on the North Atlantic storm track, is a feature of many ocean models, even when forced from reanalysis data (Tsujino et al. 2020). As this bias still persists in the CMIP6 simulations and has substantial impact on the extratropical circulation, it is clearly something that requires further attention with some studies having demonstrated a connection to the strength of the Atlantic Meridional Overturning Circulation (Wang et al. 2014; Zhang and Zhao 2015). Associated with this bias is the evidence of connectivity between the storm tracks in the two ocean basins and how biases in the North Pacific can have a downstream effect over the North Atlantic. Our results have demonstrated that the long-observed zonal bias of the North Atlantic storm track (e.g. Doblas-Reyes et al. 1998; Ulbrich et al. 2008; Zappa et al. 2013; Chang et al. 2012; Colle et al. 2013; Booth et al. 2017b) is not a bias that has its origins solely in the North Atlantic, but also has influences from the North Pacific.

There are several elements that have not been discussed in this paper with regards to isolating biases in the storm track. Of these, an important issue is variations on small time and space scales. Mesoscale dynamics within cyclones play a critical role in cyclone evolution and development (e.g. Willison et al. 2013) and this is not something considered in our assessment as this work has only focussed on large-scale variations on seasonal timescales. Due to the limited number of variables and temporal resolution of data in CMIP6, in depth analyses on these scales are not possible. Furthermore, CMIP6 models do not possess high enough spatial resolution to resolve the relevant mesoscale processes accurately. Therefore, it is recommended that either more detailed modeling studies are undertaken or increased output is made available from models in future MIPs to further investigate the findings outlined in this study.

Accompanying this study, an analysis by the authors of the drivers of the Southern Hemisphere storm tracks is presented in Part 2 (Priestley et al. 2022). This analysis will focus on similar features and assess the influence of SSTs, the mid-latitude jet, and the large-scale temperature gradients on the structure and variability of the storm tracks and the cyclones within them.

Acknowledgments. M. D. K. Priestley and J. L. Catto are supported by the Natural Environment Research Council (NERC) grant NE/S004645/1. D. Ackerley is supported by the Joint BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). K. I. Hodges was funded by the United Kingdom's Natural Environment Research Council (NERC) as part of the National Centre for Atmospheric Science. The cyclone tracking algorithm TRACK is available from https://gitlab.act.reading.ac.uk/track/track. We thank the ECMWF for their ERA5 reanalysis, which is available from the Copernicus Climate Change Service Climate Data Store (https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset). CMIP6 data is publicly available through the Earth System Grid Federation (https://esgf-node.llnl.gov/projects/cmip6/).

References

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. *Journal of Climate*, **32 (22)**, 7763–7781, doi:10.1175/JCLI-D-19-0054.1.

Bodman, R. W., D. J. Karoly, M. R. Dix, I. N. Harman, J. Srbinovsky, P. B. Dobrohotoff, and C. Mackallah, 2020: Evaluation of CMIP6 AMIP climate simulations with the ACCESS-AM2 model. *Journal of Southern Hemisphere Earth Systems Science*, **70**, 166–179, doi:10.1071/ES19033.

Booth, J. F., E. Dunn-Sigouin, and S. Pfahl, 2017a: The Relationship Between Extratropical Cyclone Steering and Blocking Along the North American East Coast. *Geophysical Research Letters*, **44 (23)**, 11,976–11,984, doi:10.1002/2017GL075941.

Booth, J. F., Y.-O. Kwon, S. Ko, R. J. Small, and R. Msadek, 2017b: Spatial Patterns and Intensity of the Surface Storm Tracks in CMIP5 Models. *Journal of Climate*, **30 (13)**, 4965–4981, doi:10.1175/JCLI-D-16-0228.1.

Bracegirdle, T. J., H. Lu, and J. I. Robson, 2021: Early-winter North Atlantic low-level jet latitude biases in climate models: implications for simulated regional atmosphere-ocean linkages. *Environmental Research Letters*, doi:10.1088/1748-9326/ac417f.

Chang, E. K. M., Y. Guo, and X. Xia, 2012: CMIP5 multimodel ensemble projection of storm track change under global warming. *Journal of Geophysical Research: Atmospheres*, **117 (D23)**, doi:10.1029/2012JD018578.

Chang, E. K. M., Y. Guo, X. Xia, and M. Zheng, 2013: Storm-Track Activity in IPCC AR4/CMIP3 Model Simulations. *Journal of Climate*, **26 (1)**, 246–260, doi:10.1175/JCLI-D-11-00707.1.

Chassignet, E. P., and Coauthors, 2020: Impact of horizontal resolution on global ocean–sea ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2). *Geoscientific Model Development*, **13 (9)**, 4595–4637, doi:10.5194/gmd-13-4595-2020.

Chen, X., Y. Liu, and G. Wu, 2017: Understanding the surface temperature cold bias in CMIP5 AGCMs over the Tibetan Plateau. *Advances in Atmospheric Sciences*, **34**, 1447–1460, doi:10.1007/s00376-017-6326-9.

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. *Journal of Climate*, **26 (18)**, 6882–6903, doi:10.1175/JCLI-D-12-00498.1.

Dacre, H. F., S. A. Josey, and A. L. M. Grant, 2020: Extratropical-cyclone-induced sea surface temperature anomalies in the 2013–2014 winter. *Weather and Climate Dynamics*, **1 (1)**, 27–44, doi:10.5194/wcd-1-27-2020.

Davini, P., and F. D'Andrea, 2020: From CMIP3 to CMIP6: Northern Hemisphere Atmospheric Blocking Simulation in Present and Future Climate. *Journal of Climate*, **33 (23)**, 10 021–10 038, doi:10.1175/JCLI-D-19-0862.1.

Davini, P., and F. D'Andrea, 2016: Northern Hemisphere Atmospheric Blocking Representation in Global Climate Models: Twenty Years of Improvements? *Journal of Climate*, **29 (24)**, 8823–8840, doi:10.1175/JCLI-D-16-0242.1.

Davini, P., F. Fabiano, and I. Sandu, 2021: Orographic resolution driving the improvements associated with horizontal resolution increase in the Northern Hemisphere winter mid-latitudes.

Weather and Climate Dynamics Discussions, 1–25, doi:10.5194/wcd-2021-51.

de Vries, H., S. Scher, R. Haarsma, S. Drijfhout, and A. v. Delden, 2019: How Gulf-Stream SST-fronts influence Atlantic winter storms. *Climate Dynamics*, **52 (9)**, 5899–5909, doi:10.1007/s00382-018-4486-7.

Doblas-Reyes, F. J., M. DéQué, F. Valero, and D. B. Stephenson, 1998: North Atlantic wintertime intraseasonal variability and its sensitivity to GCM horizontal resolution. *Tellus A*, **50 (5)**, 573–595, doi:10.1034/j.1600-0870.1998.t01-4-00002.x.

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. *Geoscientific Model Development*, **9 (5)**, 1937–1958, doi:10.5194/gmd-9-1937-2016.

Gates, W. L., and Coauthors, 1999: An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, **80 (1)**, 29–56, doi:10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2.

Haarsma, R. J., and Coauthors, 2016: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, **9 (11)**, 4185–4208, doi:10.5194/gmd-9-4185-2016.

Harvey, B. J., P. Cook, L. C. Shaffrey, and R. Schiemann, 2020: The Response of the Northern Hemisphere Storm Tracks and Jet Streams to Climate Change in the CMIP3, CMIP5, and CMIP6 Climate Models. *Journal of Geophysical Research: Atmospheres*, e2020JD032701, doi:10.1029/2020JD032701.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049, doi:10.1002/qj.3803.

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. *Geophysical Research Letters*, **46 (13)**, 7718–7725, doi:10.1029/2019GL082995.

Hodges, K. I., 1994: A General Method for Tracking Analysis and Its Application to Meteorological Data. *Monthly Weather Review*, **122 (11)**, 2573–2586, doi:10.1175/1520-0493(1994)122<2573:AGMFTA>2.0.CO;2.

Hodges, K. I., 1996: Spherical Nonparametric Estimators Applied to the UGAMP Model Integration for AMIP. *Monthly Weather Review*, **124 (12)**, 2914–2932, doi:10.1175/1520-0493(1996)124<2914:SNEATT>2.0.CO;2.

Hodges, K. I., 1999: Adaptive Constraints for Feature Tracking. *Monthly Weather Review*, **127 (6)**, 1362–1373, doi:10.1175/1520-0493(1999)127<1362:ACFFT>2.0.CO;2.

Hoskins, B. J., and K. I. Hodges, 2019: The Annual Cycle of Northern Hemisphere Storm Tracks. Part I: Seasons. *Journal of Climate*, **32 (6)**, 1743 – 1760, doi:10.1175/JCLI-D-17-0870.1.

Iqbal, W., W.-N. Leung, and A. Hannachi, 2018: Analysis of the variability of the North Atlantic eddy-driven jet stream in CMIP5. *Climate Dynamics*, **51 (1-2)**, 235–247, doi:10.1007/s00382-017-3917-1.

Jiaxiang, G., and Coauthors, 2020: Influence of model resolution on bomb cyclones revealed by HighResMIP-PRIMAVERA simulations. *Environmental Research Letters*, **15 (8)**, 084 001, doi:10.1088/1748-9326/ab88fa.

Keeley, S. P. E., R. T. Sutton, and L. C. Shaffrey, 2012: The impact of North Atlantic sea surface temperature errors on the simulation of North Atlantic European region climate. *Quarterly Journal of the Royal Meteorological Society*, **138 (668)**, 1774–1783, doi:10.1002/qj.1912.

Kodama, C., and Coauthors, 2015: A 20-Year Climatology of a NICAM AMIP-Type Simulation. *Journal of the Meteorological Society of Japan*, **93 (4)**, 393–424, doi:10.2151/jmsj.2015-024.

Kushnir, Y., and I. M. Held, 1996: Equilibrium Atmospheric Response to North Atlantic SST Anomalies. *Journal of Climate*, **9 (6)**, 1208–1220, doi:10.1175/1520-0442(1996)009<1208:EARTNA>2.0.CO;2.

Lee, R. W., 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 Dynamics*, **51 (9)**, 3369–3387, doi:10.1007/s00382-018-4083-9.

Maddison, J. W., S. L. Gray, O. Martínez-Alvarado, and K. D. Williams, 2020: Impact of model upgrades on diabatic processes in extratropical cyclones and downstream forecast evolution. *Quarterly Journal of the Royal Meteorological Society*, **146 (728)**, 1322–1350, doi:https://doi.org/10.1002/qj.3739.

Met Office, 2010 - 2013: *Iris: A Python library for analysing and visualising meteorological and oceanographic data sets*. Exeter, Devon, v1.2 ed., URL http://scitools.org.uk/.

O'Reilly, C. H., S. Minobe, and A. Kuwano-Yoshida, 2016: The influence of the Gulf Stream on wintertime European blocking. *Climate Dynamics*, **47 (5)**, 1545–1567, doi:10.1007/s00382-015-2919-0.

O'Reilly, C. H., S. Minobe, A. Kuwano-Yoshida, and T. Woollings, 2017: The Gulf Stream influence on wintertime North Atlantic jet variability. *Quarterly Journal of the Royal Meteorological Society*, **143 (702)**, 173–183, doi:10.1002/qj.2907.

Pfahl, S., C. Schwierz, M. Croci-Maspoli, C. M. Grams, and H. Wernli, 2015: Importance of latent heat release in ascending air streams for atmospheric blocking. *Nature Geoscience*, **8**, 610–614, doi:10.1038/ngeo2487.

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. *Geophysical Research Letters*, **43 (13)**, 7231–7240, doi:10.1002/2016GL069551.

Priestley, M. D. K., D. Ackerley, J. L. Catto, and K. I. Hodges, 2022: Drivers of biases in the CMIP6 extratropical storm tracks. Part 2: Southern Hemisphere. *Journal of Climate*, In Review.

Priestley, M. D. K., D. Ackerley, J. L. Catto, K. I. Hodges, R. E. McDonald, and R. W. Lee, 2020: An Overview of the Extratropical Storm Tracks in CMIP6 Historical Simulations. *Journal of Climate*, **33 (15)**, 6315–6343, doi:10.1175/JCLI-D-19-0928.1.

Scaife, A. A., and Coauthors, 2011: Improved Atlantic winter blocking in a climate model. *Geophysical Research Letters*, **38 (23)**, doi:10.1029/2011GL049573.

Scaife, A. A., and Coauthors, 2019: Does increased atmospheric resolution improve seasonal climate predictions? *Atmospheric Science Letters*, **20 (8)**, e922, doi:https://doi.org/10.1002/asl.922.

Schiemann, R., and Coauthors, 2017: The Resolution Sensitivity of Northern Hemisphere Blocking in Four 25-km Atmospheric Global Circulation Models. *Journal of Climate*, **30 (1)**, doi:10.1175/JCLI-D-16-0100.1.

Schiemann, R., and Coauthors, 2020: Northern Hemisphere blocking simulation in current climate models: evaluating progress from the Climate Model Intercomparison Project Phase 5 to 6 and sensitivity to resolution. *Weather and Climate Dynamics*, **1 (1)**, 277–292, doi:10.5194/wcd-1-277-2020.

Steinfeld, D., M. Boettcher, R. Forbes, and S. Pfahl, 2020: The sensitivity of atmospheric blocking to upstream latent heating – numerical experiments. *Weather and Climate Dynamics*, **1 (2)**, 405–426, doi:10.5194/wcd-1-405-2020.

Steinfeld, D., and S. Pfahl, 2019: The role of latent heating in atmospheric blocking dynamics: a global climatology. *Climate Dynamics*, **53**, 6159–6180, doi:10.1007/s00382-019-04919-6.

Su, F., X. Duan, D. Chen, Z. Hao, and L. Cuo, 2013: Evaluation of the Global Climate Models in the CMIP5 over the Tibetan Plateau. *Journal of Climate*, **26 (10)**, 3187–3208, doi:10.1175/JCLI-D-12-00321.1.

Tamarin, T., and Y. Kaspi, 2016: The Poleward Motion of Extratropical Cyclones from a Potential Vorticity Tendency Analysis. *Journal of the Atmospheric Sciences*, **73 (4)**, 1687–1707, doi:10.1175/JAS-D-15-0168.1.

Tamarin, T., and Y. Kaspi, 2017: Mechanisms Controlling the Downstream Poleward Deflection of Midlatitude Storm Tracks. *Journal of the Atmospheric Sciences*, **74 (2)**, 553–572, doi:10.1175/JAS-D-16-0122.1.

Taylor, K. E., 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, http://goo.gl/v1drZl.

Tsujino, H., and Coauthors, 2020: Evaluation of global ocean–sea-ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2). *Geoscientific Model Development*, **13 (8)**, 3643–3708, doi:10.5194/gmd-13-3643-2020.

Ulbrich, U., J. G. Pinto, H. Kupfer, G. C. Leckebusch, T. Spangehl, and M. Reyers, 2008: Changing Northern Hemisphere Storm Tracks in an Ensemble of IPCC Climate Change Simulations. *Journal of Climate*, **21 (8)**, 1669–1679, doi:10.1175/2007JCLI1992.1.

Wang, C., L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso, 2014: A global perspective on CMIP5 climate model biases. *Nature Climate Change*, **4 (3)**, 201–205, doi:10.1038/nclimate2118.

Willison, J., W. A. Robinson, and G. M. Lackmann, 2013: The Importance of Resolving Mesoscale Latent Heating in the North Atlantic Storm Track. *Journal of the Atmospheric Sciences*, **70 (7)**, 2234–2250, doi: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 Dynamics*, **35 (2)**, 341–353, doi:10.1007/s00382-009-0554-3.

Zappa, G., G. Masato, L. Shaffrey, T. Woollings, and K. Hodges, 2014: Linking Northern Hemisphere blocking and storm track biases in the CMIP5 climate models. *Geophysical Research Letters*, **41 (1)**, 135–139, doi:10.1002/2013GL058480.

Zappa, G., L. C. Shaffrey, and K. I. Hodges, 2013: The Ability of CMIP5 Models to Simulate North Atlantic Extratropical Cyclones. *Journal of Climate*, **26 (15)**, 5379–5396, doi:10.1175/JCLI-D-12-00501.1.

Zhang, L., and C. Zhao, 2015: Processes and mechanisms for the model SST biases in the North Atlantic and North Pacific: A link with the Atlantic meridional overturning circulation. *Journal of Advances in Modeling Earth Systems*, **7 (2)**, 739–758, doi:10.1002/2014MS000415.

Zhu, Y.-Y., and S. Yang, 2020: Evaluation of CMIP6 for historical temperature and precipitation over the Tibetan Plateau and its comparison with CMIP5. *Advances in Climate Change Research*, **11 (3)**, 239–251, doi:10.1016/j.accre.2020.08.001.

LIST OF TABLES

Model Name	Institution	Atmospheric Resolution	
		Horizontal	Vertical
ACCESS-CM2	CSIRO-ARCCSS; Commonwealth Scientific and Industrial Research Organisation, Australian Research Council Centre of Excellence for Climate System Science, Australia	N96; 192×144; 250km	85 levels to 85 km
ACCESS-ESM1-5	CSIRO; Commonwealth Scientific and Industrial Research Organisation, Australia	N96; 192×144; 250km	85 levels to 85 km
BCC-CSM2-MR	BCC; Beijing Climate Center, China	T206; 320×160; 100km	46 levels to 1.46 hPa
CMCC-CM2-HR4	CMCC; Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	288×192; 100km	26 levels to ~2 hPa
CMCC-CM2-SR5	CMCC; Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	288×192; 100km	30 levels to ~2 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; 100km	91 levels to 78.4km
EC-Earth3	EC-Earth-Consortium	TL255; 512×256; 100km	91 levels to 0.01 hPa
EC-Earth3-Veg	EC-Earth-Consortium	TL255; 512×256; 100km	91 levels to 0.01 hPa
GFDL-CM4	NOAA-GFDL; National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, USA	C96; 360×180; 100km	33 levels to 1 hPa
HadGEM3-GC3.1-LL	MOHC; Met Office Hadley Centre, UK	N96; 192×144; 250km	85 levels to 85 km
HadGEM3-GC3.1-MM	MOHC; Met Office Hadley Centre, UK	N216; 432×324; 100km	85 levels to 85 km
IPSL-CM6A-LR	IPSL; Institut Pierre Simon Laplace, France	N96; 144×143; 250km	79 levels to 40 km
KACE-1-0-G	NIMS-KMA; National Institute of Meteorological Sciences/Korea Meteorological Administration, Republic of Korea	N96; 192×144; 250km	85 levels to 85 km
KIOST-ESM	KIOST; Korea Institute of Ocean Science and Technology, Republic of Korea	C48; 192×96; 250km	32 levels to 2 hPa
MIROC-ES2L	MIROC; MIROC Consortium (JAMSTEC, AORI, NIES, R-CCS), Japan	T42; 128×64; 500km	40 levels to 3 hPa
MIROC6	MIROC; MIROC Consortium (JAMSTEC, AORI, NIES, R-CCS), Japan	T85; 256×128; 250km	81 levels to 0.004 hPa
MPI-ESM1-2-HR	MPI-M, DWD, DKRZ; Max Planck Institute for Meteorology, Deutscher Wetterdienst, Deutsches Klimarechenzentrum, Germany	T127; 384×192; 100km	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; 250km	47 levels to 0.01 hPa
MRI-ESM2-0	MRI; Meteorological Research Institute, Japan	TL159; 320×160; 100km	80 levels to 0.01 hPa
NESM3	NUIST; Nanjing University of Information Science and Technology, China	T63; 192×96; 250km	47 levels to 1 hPa
NorESM2-LM	NCC; NorESM Climate Modelling Consortium, Norway	144×90; 250km	32 levels to 3 hPa
SAM0-UNICON	SNU; Seoul National University, Republic of Korea	288×192; 100km	30 levels to ≈2 hPa
TaiESM1	AS-RCEC; Research Center for Environmental Changes, Academia Sinica, Taiwan	288×192; 100km	30 levels to ≈2 hPa
UKESM1-0-LL	UKESM Consortium (MOHC, NERC, NIMS-KMA, NIWA)	N96; 192×144; 250km	85 levels to 85 km

TABLE 1. List of CMIP6/AMIP6 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' refers to a cubed-sphere finite volumes 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 2 grid point waves that can be represented in the zonal direction. Following any grid specification is the dimensions of the model output on a gaussian longitude x latitude grid. The resolution stated in kilometres is the stated nominal resolution of the atmospheric component of the model from Taylor et al. (2017).

LIST OF FIGURES

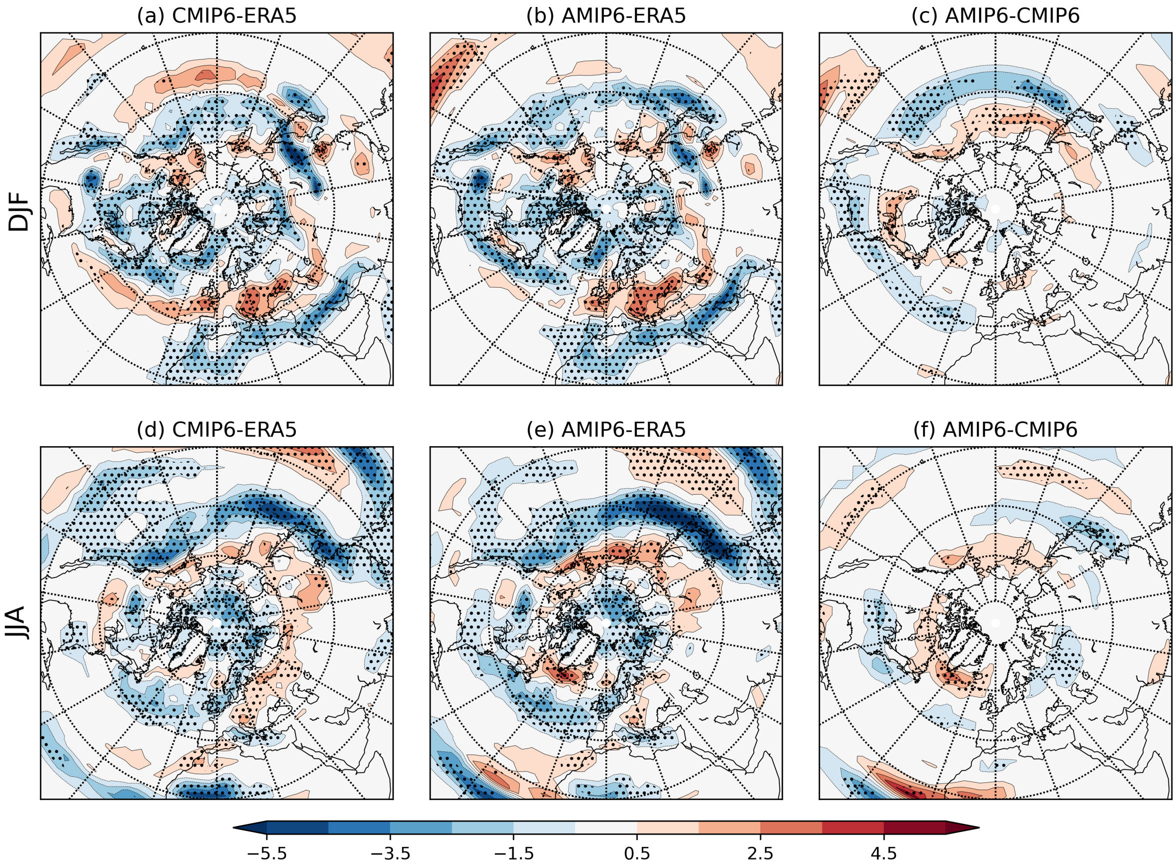


FIG. 1. Track densities of CMIP6 model ensembles for (a-c) DJF and (d-f) JJA from 1979/80 to 2013/14. Differences are shown relative to ERA5 for the (a,d) historical coupled models and (b,e) corresponding AMIP6 runs. AMIP6-CMIP6 is shown in (c,f). Units are number of cyclones per 5° spherical cap per month. Stippling indicates where more than 80% of models agree on the sign of the error. Only models with both a *historical* and *amip* simulation are shown (see Table 1).

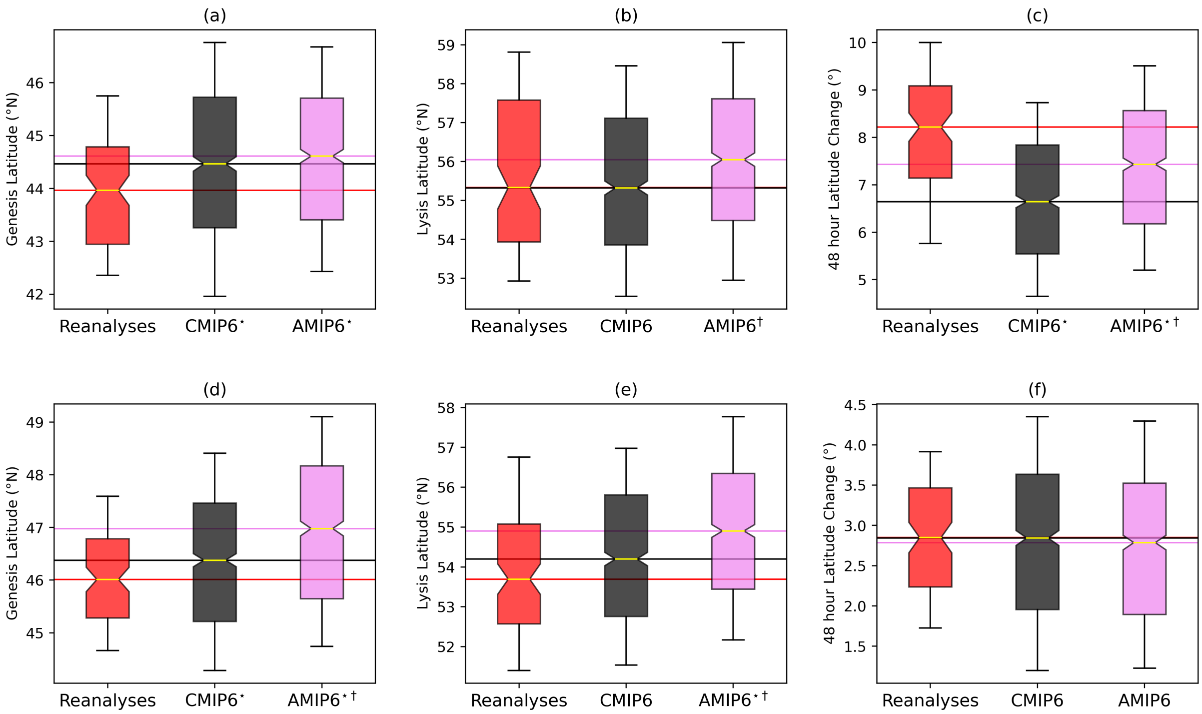


FIG. 2. Boxplots of annual mean cyclogenesis latitude (a,d), cyclolysis latitude (b,e), and cyclone 48-hour latitude change (c,f) for the core cyclogenesis region of the North Atlantic in DJF (a–c) and JJA (d–f). The core cyclogenesis regions for the North Atlantic is the cyan region in Fig. S1a. Horizontal coloured lines indicate the median value for each model distribution. Boxes extend to the 25th and 75th percentile respectively with yellow lines indicating the distribution median. Notches around the median show the uncertainty estimate based on 10,000 random samples and whiskers extend to the 10th and 90th percentiles. In the labels ⋆ indicates where the model group is significantly different from the reanalyses and † indicates where AMIP6 and CMIP6 are significantly different. Significance tests performed using a Mood's Median test and quoted at the 5% level.

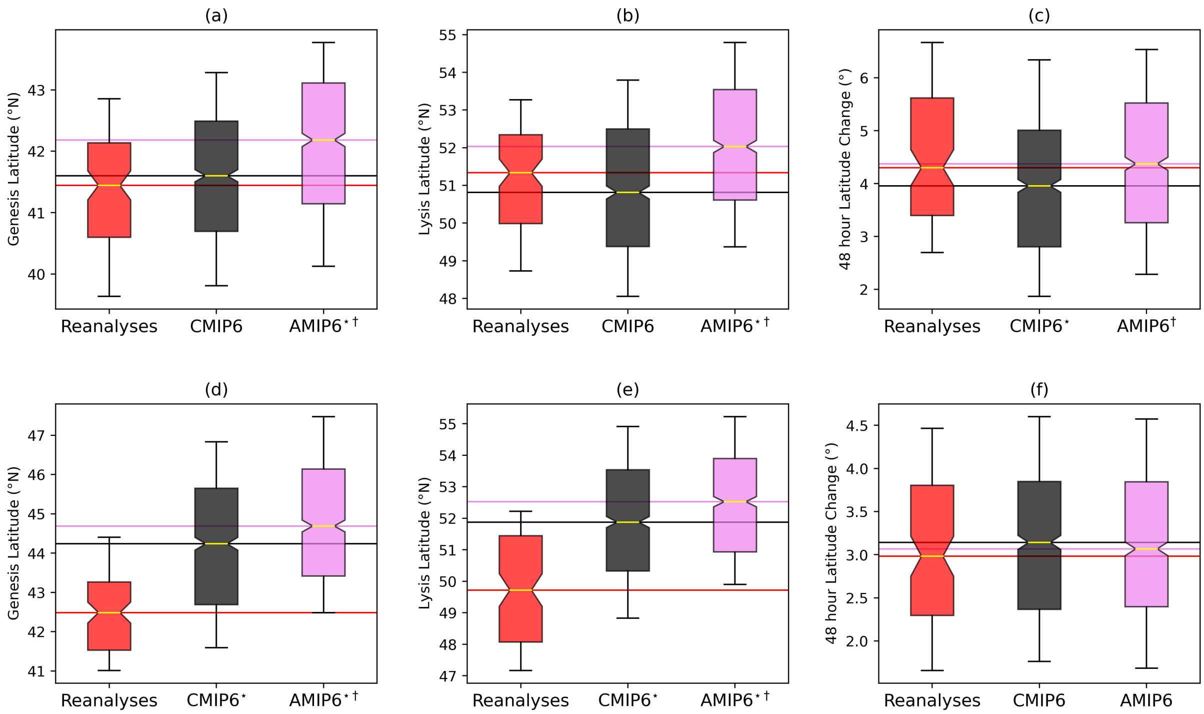


FIG. 3. As Fig. 2 but for genesis occurring in the core North Pacific region. This regions is encapsulated by the red box in Fig. S1a.

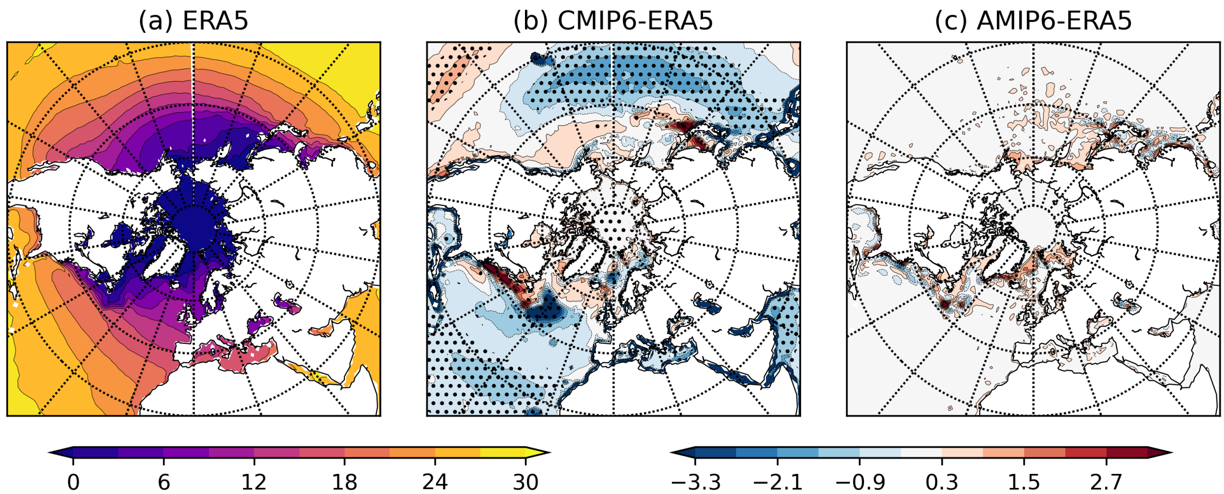
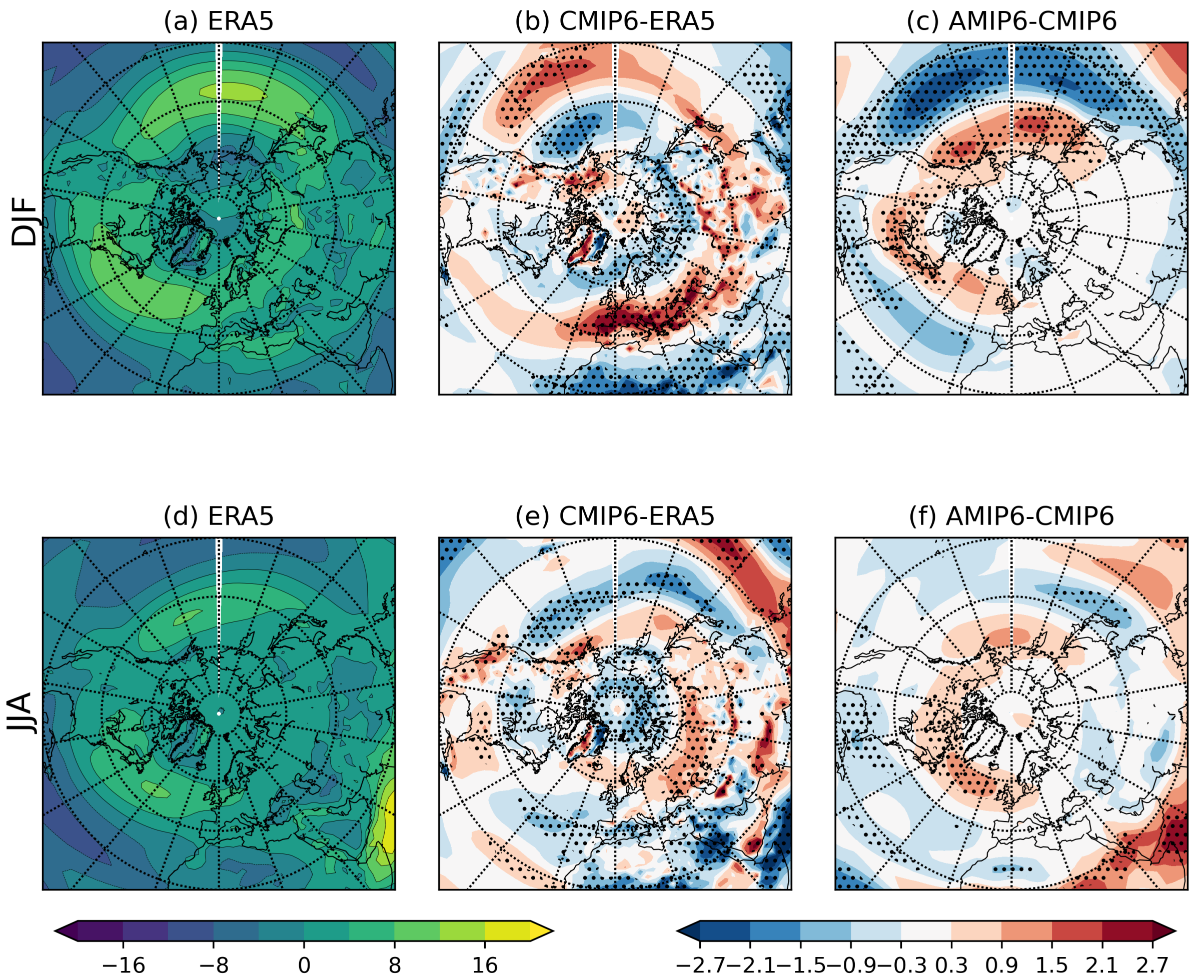


FIG. 4. DJF averaged sea surface temperature (SST) for (a) ERA5, (b) CMIP6-ERA5, and (c) AMIP6-ERA5. Units are °C. Stippling in (b) indicates where there is 80% model agreement on the sign of the error.



896 FIG. 5. Seasonal mean zonal (u) wind at 850 hPa for (a,d) ERA5, (b,e) CMIP6-ERA5, and (c,f) AMIP6-CMIP6 for (a–c) DJF and (d–f) JJA. Units are m s^{-1}. Panel stippling indicates where there is 80% model agreement on the sign of the error.

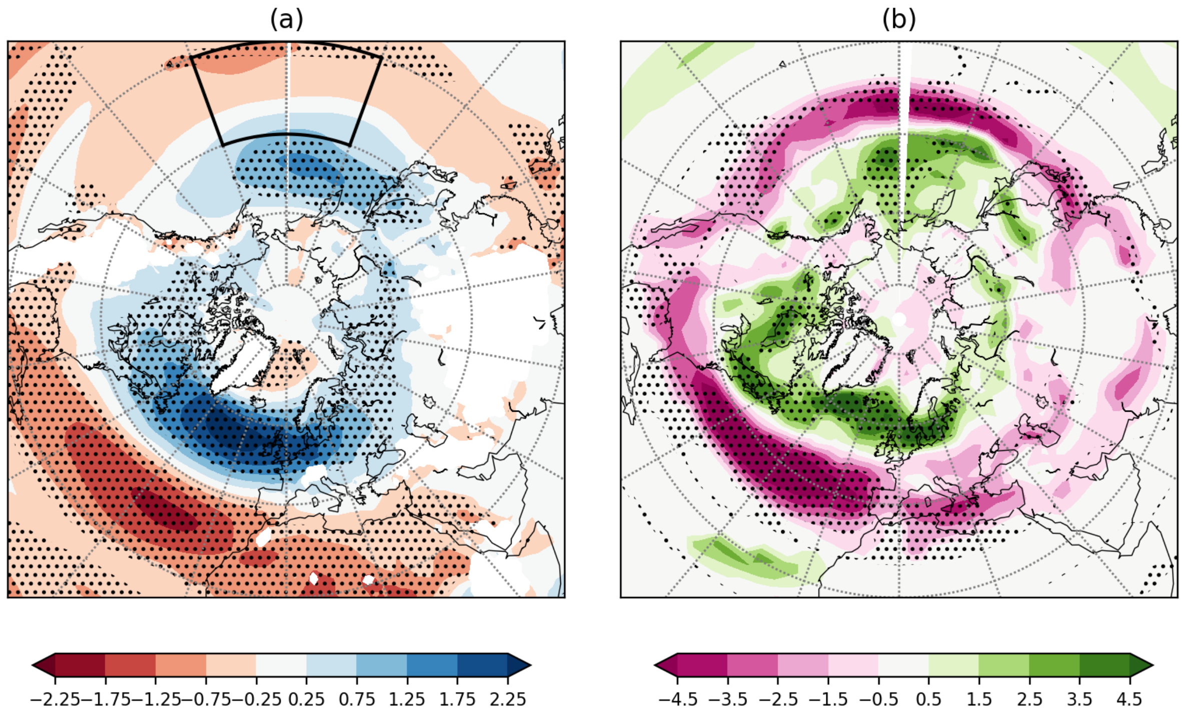
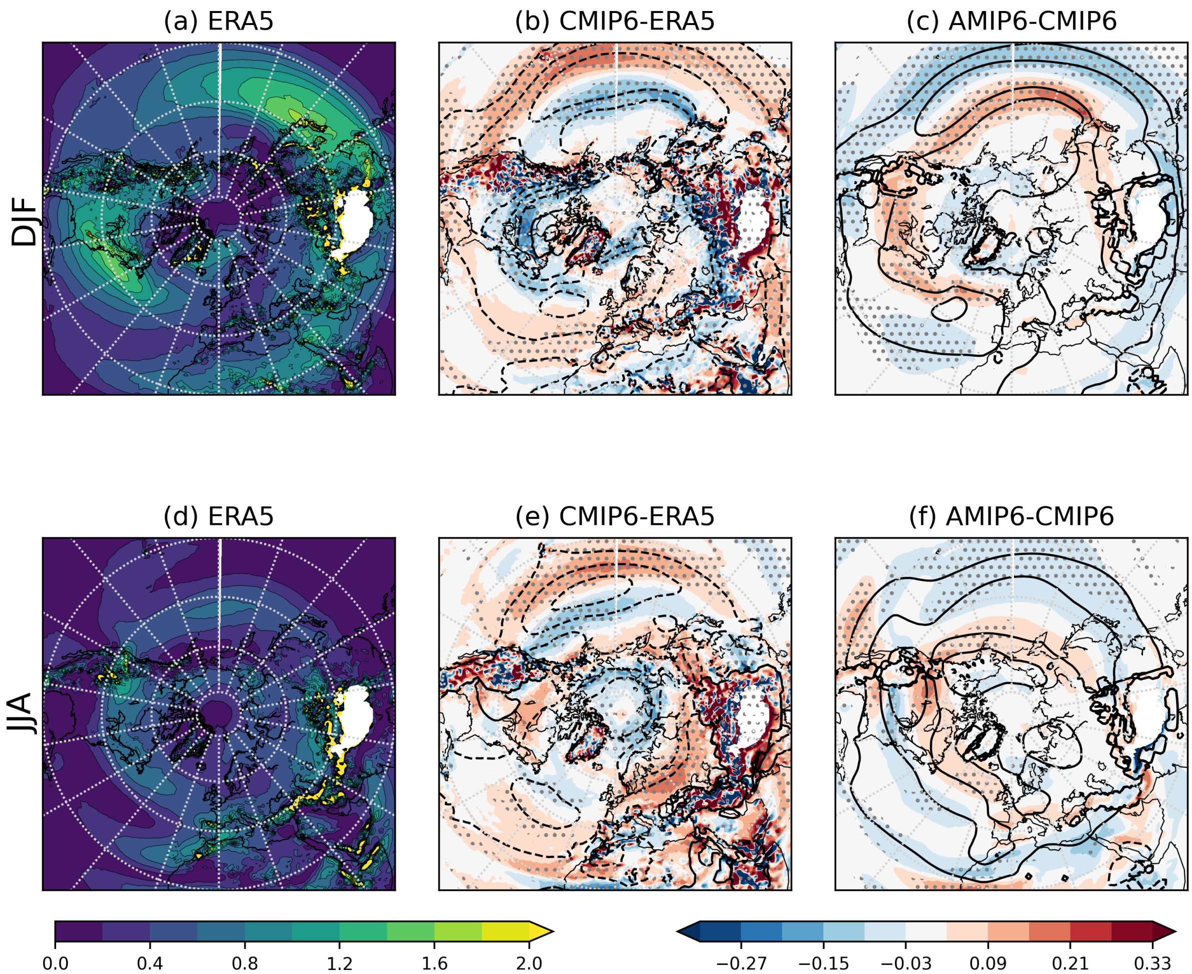


FIG. 6. Linear least-squares regression slope maps of DJF seasonal mean (a) 850 hPa zonal wind and (b) storm track density, against area averaged SST from 20°N-40°N, 160°W-200°W. Regression is performed across all model climatologies. Stippling indicates where regressions are significant at the 5% level. The black box in (a) indicates the region of SSTs used in the regression calculations. Units are (a) m s^{-1} K^{-1} and (b) cyclones per month K^{-1}.



904 FIG. 7. Seasonal mean potential temperature gradient in the lower troposphere (700-850 hPa average, colored shading) for (a,d) ERA5, (b,e) CMIP6-ERA5, and (c,f) AMIP6-CMIP6 for (a–c) DJF and (d–f) JJA. Units are K degree^{-1}. The gray contours show the difference in the absolute potential temperature field on each respective panel. Contour intervals are ± 1 and 2 K with solid (dashed) contours indicating positive (negative) values. Panel stippling indicates where there is 80% model agreement on the sign of the error.

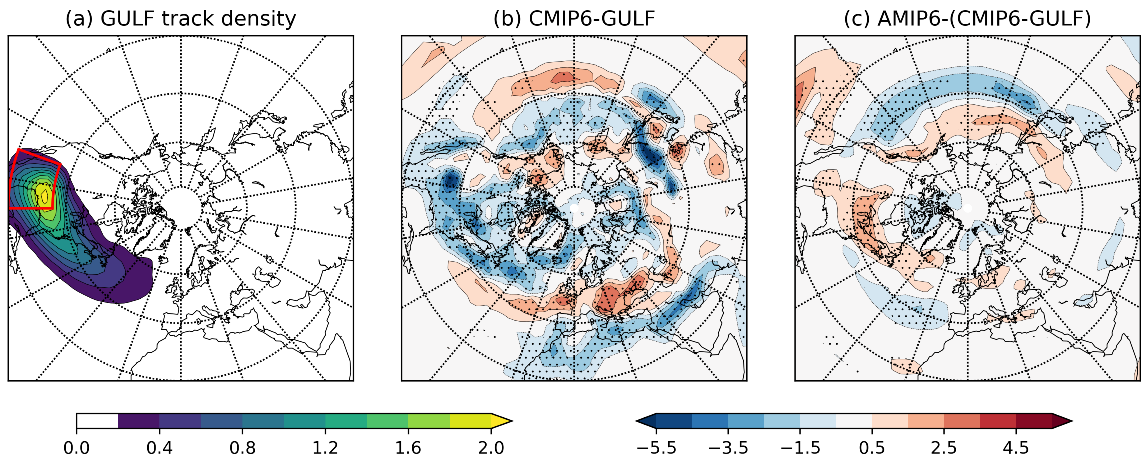


FIG. 8. (a) Track density of all cyclones forming within red box region (20-35°N, 250-270°E) of the CMIP6 models. (b) Track density bias of CMIP6 models without Gulf of Mexico cyclones relative to ERA5. (c) AMIP6-CMIP6 (with no Gulf of Mexico cyclones).

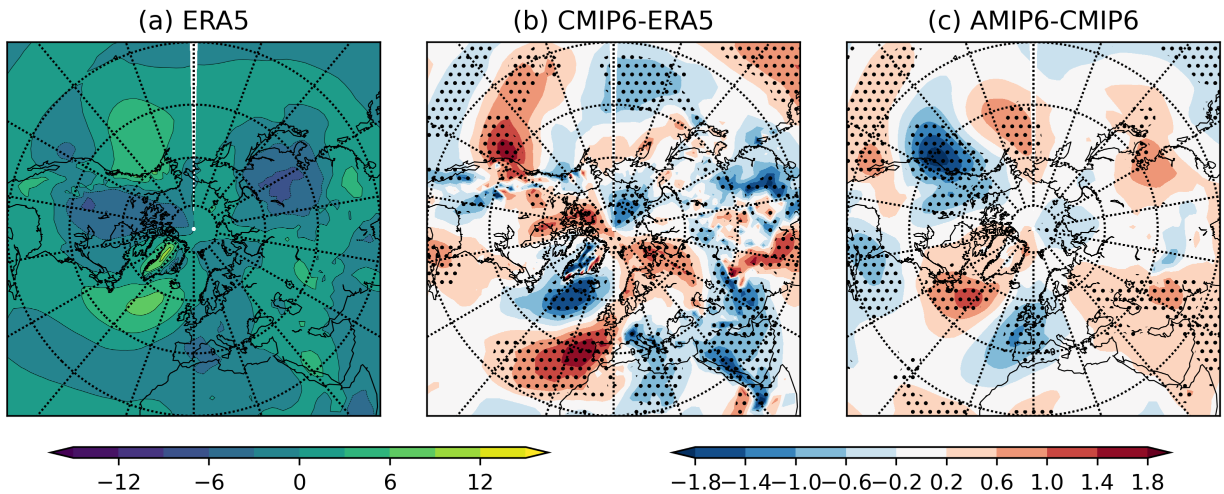
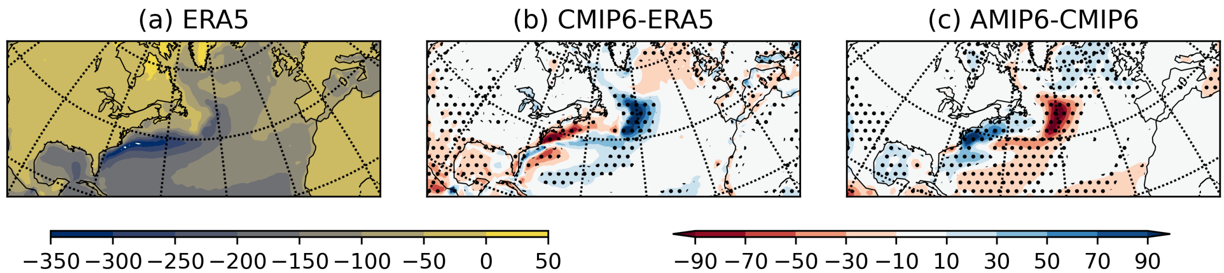


FIG. 9. As Figure 5 but for the DJF meridional wind at 700 hPa.



912 FIG. 10. DJF seasonal mean surface to atmosphere latent heat flux for (a) ERA5, (b) CMIP6-ERA5, and (c) AMIP6-CMIP6. Units are W m^{-2}. Stippling in (b) and (c) indicates where there is 80% model agreement on the sign of the error.

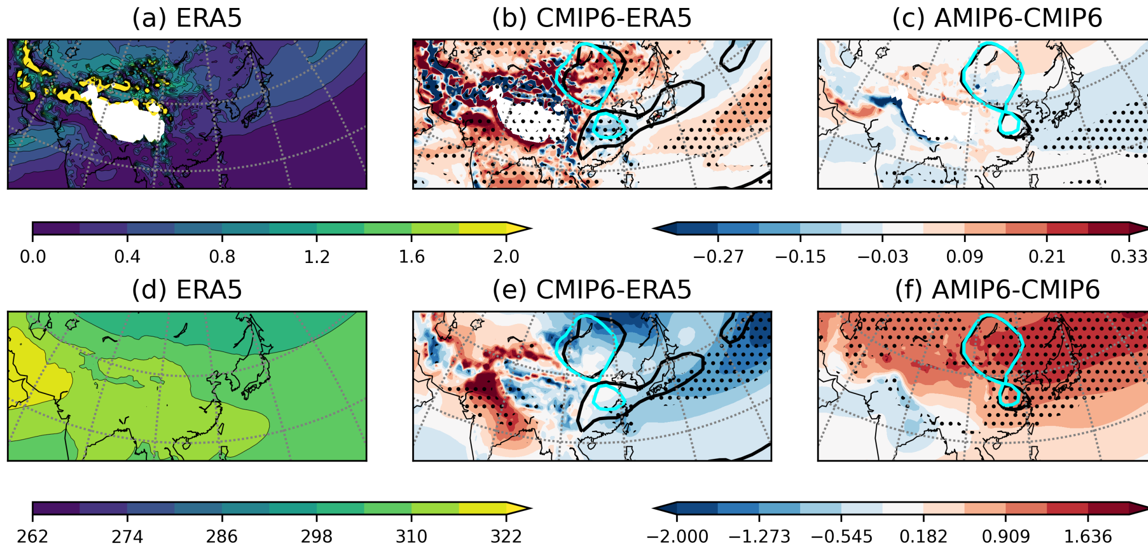


FIG. 11. JJA mean potential temperature gradient (a–c) and absolute potential temperature (d–f) in the lower troposphere (700-850 hPa average) across eastern Asia for (a,d) ERA5, (b,e) CMIP6-ERA5, and (c,f) AMIP6-CMIP6. Units are (a–c) K degree $^{-1}$ and (d–f) K. Black and cyan contours indicate regions of genesis density greater than 1 cyclone per month. The black genesis contour represents the reference dataset (right of panel title) and the cyan genesis contour represents the difference dataset (left of panel title).