Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multi-model ensemble

Article

Accepted Version

To link to this article DOI: http://dx.doi.org/10.1175/jcli-d-19-0639.1

Publisher: American Meteorological Society

Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multi-model ensemble

Malcolm John Roberts[1], Joanne Camp[1], Jon Seddon[1], Pier Luigi Vidale[2], Kevin Hodges[2], Benoit Vanniere[2], Jenny Mecking[3], Rein Haarsma[4], Alessio Bellucci[5], Enrico Scoccimarro[5], Louis-Philippe Caron[6], Fabrice Chauvin[7], Laurent Terray[8], Sophie Valcke[8], Marie-Pierre Moine[8], Dian Putrasahan[9], Christopher Roberts[10], Retish Senan[10], Colin Zarzycki[11], Paul Ullrich[12]

1 Met Office, Exeter EX1 3PB, U.K.

2 National Centre for Atmospheric Science (NCAS), University of Reading, Reading, U.K.

3 University of Southampton, Southampton, U.K. (now at National Oceanography Centre, Southampton, U. K.)

4 Koninklijk Nederlands Meteorologisch Instituut (KNMI), De Bilt, The Netherlands

5 Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Bologna, Italy

6 Barcelona Supercomputing Center – Centro Nacional de Supercomputación (BSC), Barcelona, Spain

Early Online Release: This preliminary version has been accepted for publication in *Journal of Climate*, may be fully cited, and has been assigned DOI 10.1175/JCLI-D-19-0639.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

7 Centre National de Recherches Météorologiques - Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique (CNRM-CERFACS), Toulouse, France

8 CECI, Université de Toulouse, CERFACS/CNRS, Toulouse, France

9 Max Planck Gesellschaft zur Foerderung der Wissenschaften E.V. (MPI-M), Hamburg, Germany

10 European Centre for Medium Range Weather Forecasting (ECMWF), Reading, U.K.

11 Penn State University, Pennsylvania, USA

12 University of California, Davis, Davis, California, USA

Corresponding author: Malcolm John Roberts; email: malcolm.roberts@metoffice.gov.uk

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Abstract

A multi-model, multi-resolution set of simulations over the period 1950-2014 using a common forcing protocol from CMIP6 HighResMIP have been completed by six modelling groups. Analysis of tropical cyclone performance using two different tracking algorithms suggests that enhanced resolution towards 25 km typically leads to more frequent and stronger tropical cyclones, together with improvements in spatial distribution and storm structure. Both of these factors reduce typical GCM biases seen at lower resolution.

Using single ensemble members of each model, there is little evidence of systematic improvement in interannual variability in either storm frequency or Accumulated Cyclone Energy compared to observations when resolution is increased. Changes in the relationships between large-scale drivers of climate variability and tropical cyclone variability in the Atlantic are also not robust to model resolution.

However using a larger ensemble of simulations (of up to 14 members) with one model at different resolutions does show evidence of increased skill at higher resolution. The ensemble mean correlation of Atlantic interannual tropical cyclone variability increases from ~0.5 to ~0.65 when resolution increases from 250 km to 100 km. In the North West Pacific the skill keeps increasing with 50 km resolution to 0.7. These calculations also suggest that more than six members are required to adequately distinguish the impact of resolution within the forced signal from the weather noise.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

1. Introduction

Tropical cyclone impacts globally are important for life and economies, being the largest driver of losses among natural hazards (Landsea, 2000; Aon Benfield, 2018). They also contribute significantly to regional seasonal rainfall totals (Jiang et al. 2010; Scoccimarro et al. 2014; Guo et al. 2017; Franco-Diaz et al. 2019) and hence form an important part of the mean climate. In order to achieve improved forecasts, risk assessment and projections of future changes of tropical cyclones, better understanding of the drivers of interannual variability, and hence potential future changes in frequency or intensity, are key. Such understanding can only come from a combination of observations and modelling.

Previous assessments of tropical cyclone performance within global multi-model simulation comparisons have been hampered by a variety of factors (Camargo and Wing, 2016). Use of models from the Coupled Model Intercomparison Projects (CMIP3 and CMIP5; Walsh et al. 2013; Camargo et al. 2013) typically implies that model grid spacing is greatly restricted, typically to coarser than 100 km, and often considerably coarser, when effective resolution determined from the kinetic energy spectrum is considered (Klaver et al. 2019). This has consequences for both the model mean state and tropical cyclone characteristics. Specific projects such as the Tropical Cyclone-Model Intercomparison Project (TC-MIP; Walsh et al. 2011) and the US Climate and Ocean: Variability, Predictability and Change (CLIVAR) Hurricane Working Group (Walsh et al. 2014) have investigated higher resolutions, but the simulations (and tracking algorithms) were not designed to be uniform and hence the results can be difficult to interpret (Camargo et al. 2013; Shaevitz et al. 2014; Nakamura et al. 2017). There is also a need for multiple ensemble members in

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

order to separate the forced signal from the weather noise (e.g. Zhao et al. 2009; Roberts et al. 2015; Mei et al. 2019).

There have also been many studies of the impact of horizontal resolution on tropical cyclones (Zhao et al. 2009; Manganello et al. 2012; Wehner et al. 2014; Kodama et al. 2015; Murakami et al. 2015; Roberts et al. 2015; Yoshida et al. 2017; Chauvin et al. 2019). These mainly used individual climate models, but due to differences in experimental design, tracking algorithm, model parameters and other factors it can be difficult to understand how generally applicable the results are likely to be for other models.

The CMIP6 High Resolution Model Intercomparison Project (HighResMIP; Haarsma et al. 2016), in a new experimental design for CMIP6 (Eyring et al, 2016), that provides a common protocol for a multi-model, multi-resolution ensemble. Some aspects of the simulation have been deliberately simplified (for example aerosol effects are imposed via specified optical properties), so that a comparison of model performance is made more manageable. This protocol extends the period of atmosphere-only simulations to 1950-2014 (compared to the standard CMIP6 period of 1979-2014; Eyring et al. 2016), in order to assess a longer period of variability and drivers of change and increase the tropical cyclone (TC) sample sizes for climatology.

The European Union Horizon 2020 project PRIMAVERA has six different contributing global atmospheric models, each run using the HighResMIP protocol at both a standard CMIP6-type resolution (typically 100 km) and at a significantly higher resolution (towards 25 km), to investigate the impact this has on the simulation of climate variability and extremes, including tropical cyclones. It is a unique opportunity

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

to understand the robustness of such changes across a range of models and resolutions. Two tracking algorithms—TRACK (Hodges et al. 2017) and TempestExtremes (Ullrich and Zarzycki 2017; Zarzycki and Ullrich 2017)—have been applied uniformily across all models and reanalyses to provide an indication in the uncertainties in the TC identification.

The key science questions addressed in this study are:

1. Are there robust impacts of higher resolution on explicit tropical cyclone simulation across the multi-model ensemble using different tracking algorithms?
2. What are the possible processes responsible for any changes with resolution?
3. How many ensemble members are needed to assess the skill in the interannual variability of tropical cyclones?

In section 2 we describe the models, forcing and reanalysis datasets used in this study, together with the tracking algorithms and other datasets. In section 3 we describe our multi-model, multi-resolution assessment of tropical cyclone performance, both as a global overview and then with focus on the North Atlantic. Here we also describe the impact of a larger ensemble size and the impact on skill for interannual variability. In section 4 we discuss the implications of our results and future work.

2. Model description, forcing, datasets and tracking algorithms

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Six PRIMAVERA modelling groups have configured global models at (at least) two horizontal resolutions and completed the Tier 1 CMIP6 HighResMIP atmosphere-only simulations (Haarsma et al. 2016) for 1950-2014. The models and resolutions are detailed in Table 1, including the ratio of the lower to higher grid spacing at the equator (Table 2). The effective resolution of the models (relating to the kinetic energy spectra) is described in Klaver et al. (2019) and is also included. Further HighResMIP experiments (Tier 2 coupled simulations and Tier 3 future projections) have also been completed, but the analysis of these is outside the scope of this work.

Detailed documentation on all models can be found in the following references, and is briefly summarised in Appendix A: ECMWF-IFS, Roberts et al. (2018); CMCC-CM2, Cherchi et al. (2019); CNRM-CM6, Voldoire et al. (2019); MPI-ESM1-2, Gutjahr et al. (2019); EC-Earth3P, Haarsma et al. (2019); HadGEM3-GC3.1, Vidale et al. (in prep) and Roberts et al. (2019a). The HighResMIP protocol recommends minimal changes in model parameters between low and high resolution simulations in order that differences caused by resolution alone are emphasised. Table 3 describes all the model parameters that are explicitly changed with resolution.

The inclusion of stochastic physics schemes, which attempt to represent the dynamical aspects of sub-grid scale processes, is becoming common for weather and seasonal forecasting (Palmer et al. 2009; MacLachlan et al. 2015; Walters et al. 2019), and is now being included in some global climate models (Batté and Doblas-Reyes, 2015; Walters et al. 2019). Amongst the models used in this study, only the HadGEM3-GC31 and ECMWF-IFS contain such schemes. The influence of these schemes is designed to automatically decrease as model resolution becomes finer

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

(i.e. by self-tuning rather than explicit parameter change, Sanchez et al. 2016), and hence needs to be considered when assessing “model resolution” impacts. Stochastic schemes have been shown to increase tropical cyclone mean frequency by up to 30% at some resolutions in multiple models (e.g. Met Office and ECMWF models; Vidale et al., in prep), at least partly via moistening the tropical environment in the regions where the TCs have genesis (Watson et al. 2017).

All the models use an atmospheric initial condition at 1950 from the ECMWF Reanalysis of the 20th Century (ERA-20C; Poli et al. 2016). Components of the land surface with longer memory (such as soil temperature and moisture) are initialised differently by each group – however, since the focus here is on the later 1979-2014 period of the simulations, this should have minimal impact on the results.

a. Forcing

The HighResMIP experimental design has been followed for the forcing datasets (Haarsma et al. 2016), including using simplified aerosol optical properties apart from one model (see below). These optical properties are a combination of a model constant background natural aerosol (typically diagnosed from a pre-industrially-forced simulation), together with time-varying volcanic and anthropogenic aerosol from the Max Planck Institute Aerosol Climatology version 2 (MACv2-SP; Stevens et al. 2015) scheme. The latter uses sulphate aerosol patterns to scale the aerosol forcing magnitude over time. Note that this forcing by design excludes natural aerosol (including dust) variability and hence the simulations do not explicitly account for any variability driven by such forcing (Reed et al. 2019), apart from that which is integrated in the SST forcing itself. The exception to this is the CNRM-CM6-1 model, which uses its own aerosol scheme (Voldoire et al. 2019; Chauvin et al. 2019). A

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

comparison of performance between MACv2-SP and prognostic aerosol is included in Vidale et al. (in prep).

The sea surface temperature (SST) and sea-ice forcings used in the HighResMIP protocol are based on the daily, ¼ degree Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST.2.2.0; Kennedy et al. 2017) dataset, with area-weighted regridding used to map this to each model grid. Mean differences between this dataset and the standard monthly Program for CLimate Model Diagnosis and Intercomparison (PCMDI) SST used in Atmospheric Model Intercomparison Project (AMIPII; Taylor et al. 2000) are shown in Vidale et al. (in prep). The CMIP6 (Eyring et al. 2016) historic, time-varying forcings for solar (Matthes et al. 2017), ozone concentration (Hegglin et al. 2016) and greenhouse gases (GHG) (Meinshausen and Vogel 2016) are used. The land surface properties and land use remain constant, representative of the year 2000 using a repeating seasonal cycle.

b. Datasets

(1) Reanalyses

The following reanalysis datasets are used: the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-analysis project (ERA-Interim; Dee et al., 2011; 1979-2014); Fifth Generation ECMWF Reanalysis (ERA5; Copernicus Climate Change Service, 2017; 1979-2014); NASA Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2; Gelaro et al. 2017; 1980-2014); National Center for Atmospheric Research - Climate Forecast System Reanalysis (NCAR-CFSR; Saha et al. 2014; 1979-2014); Japanese 55-year Reanalysis (JRA55; Kobayashi et al. 2015; 1959-2014). An overview of the

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

properties of these reanalysis datasets is given in Table 4. Tropical cyclones in these datasets (apart from ERA5) have been compared in Hodges et al. (2017) and Murakami et al. (2014b).

(2) Observations

Observed tropical cyclone tracks for the North Atlantic and Eastern Pacific basins are obtained from the National Oceanic and Atmospheric Administration (NOAA) National Hurricane Center's best- track Hurricane Database (HURDAT2 (Jan 2018 version); Landsea and Franklin, 2013). Observed tropical cyclone data for all remaining basins are obtained from the US Navy's Joint Typhoon Warning Centre (JTWC) best-track database (Chu et al., 2002). We define an observed tropical cyclone as having a 1-min maximum sustained wind speed of 34 kt (17.5 m s^{-1}) or higher, to give a globally-uniform criteria, and we exclude subtropical storms (SS) from observations when they have SS as their officially designated maximum classification. We use these datasets in preference to IBTrACS (Knapp et al. 2010) for the consistency of 1-min averaging periods for all TCs around the world.

(3) Models

Model simulation output can be obtained via the Earth System Grid Federation (ESGF) nodes from the following: Roberts (HadGEM3-GC31; 2017a, 2017b, 2017c), Roberts et al. (ECMWF-IFS; 2017a, 2017b), Voldoire (CNRM-CM6-1; 2017, 2018), Scoccimarro et al. (CMCC-CM2-(V)HR4; 2017a, 2017b), EC-Earth Consortium (EC-Earth3P; 2018a, 2018b), von Storch et al. (MPI-ESM1-2; 2017, 2019). The storm tracks derived from these datasets and analysed here are available from Roberts (2019b, 2019c).

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

c. Analysis information

The analysis presented here focuses on the 1979-2014 period due to both the satellite observations providing a more homogeneous observational reference dataset, and the availability of multiple reanalysis datasets for validation.

The Accumulated Cyclone Energy (ACE) index (Bell et al. 2000) is an integrated measure of tropical cyclone activity, and is calculated for model and observed tropical cyclones using the same method as Camp et al. (2015). For observed tropical cyclones, ACE is the sum of the square of the maximum sustained 10 m wind speed every 6 hours whilst the cyclone is at least tropical storm strength (34 kts; 17.5 m s^{-1}). For model and reanalysis tropical cyclones, the wind speeds are lower than observed (Williams et al. 2015), and therefore the wind speed threshold is removed entirely, and instead we calculate ACE throughout the lifetime of the storm during its warm core phase using winds at 925 hPa to better compare the seasonal cycle and interannual variability with observations (henceforth ACE_{925}), as in Camp et al. (2015). The ACE metric has been found to be a more robust measure for interannual variability than simple storm counts (e.g. Villarini and Vecchi, 2012; Scoccimarro et al. 2018), partly because it may reduce the impact of observational methods and short-lived storms (Landsea 2010).

In general, models at the resolutions shown here are not able to represent very intense wind speeds (see Davis (2018) for theoretical/numerical limits), but are more able to generate strong minima in surface pressure (Manganello et al. 2012). Hence in order to better stratify the model storms by intensity, we use a surface pressure

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

scale for the model intensity, rather than wind speed (Caron and Jones 2012; Roberts et al. 2015). The categories are defined in Table 5.

d. Tracking algorithms (trackers)

The tropical cyclones are diagnosed from models and reanalyses using two feature tracking algorithms (henceforth trackers): TRACK (Hodges et al. 2017) and TempestExtremes (Ullrich and Zarzycki 2017; Zarzycki and Ullrich 2017). These are described in detail in Appendix B, and briefly summarised here. TRACK is based on tracking vorticity features on a common T63 spectral grid with criteria for warm-core and lifetime. TempestExtremes tracks features using sea level pressure on the model grid, with criteria for warm-core and lifetime. Models and reanalyses are all tracked in the same way with the same parameters - for both trackers, the parameter choices are primarily derived from comparing tracked reanalysis datasets and observations (Hodges et al. 2017; Zarzycki et al. 2017), although with differing emphasis (Appendix B). One notable difference between the application of the trackers is the dependence on the model grid - TRACK transforms each model output to a common T63 grid for tracking, while TempestExtremes operates on the native model grid. No wind speed thresholds are applied to either tracker. A more detailed comparison between several trackers to better understand the cause of the differences, including using application of classification schemes to the systems (McTaggart-Cowan et al. 2013; Yanase et al. 2014), is ongoing (Roberts et al. in prep).

We chose to use two trackers in order to obtain complementary viewpoints of model performance. We expect results to depend on the details of each trackers' criteria, as is found in other feature tracking comparisons, for example Horn et al. (2014) for

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

TCs, Neu et al. (2013) for extra-tropical cyclones and Shields et al. (2018) for atmospheric rivers. In cases where both trackers broadly agree, we can be more confident that our conclusions are not dependent on tracker details.

3. Results

a. Global TC activity and track density

Realistic simulation of the frequency and spatial distribution of tracks of tropical cyclones is an important prerequisite for understanding the risk of landfall and climate impacts, as well as for potential changes in regional mean precipitation.

A simple initial assessment of TC frequency from models, reanalyses and observations is shown in Figs. 1,2, illustrating the total number of storms in the northern and southern hemispheres (NH, SH) and the distribution in each NH ocean basin. It is informative to show this using two different trackers since there are several aspects that might be misinterpreted when just a single tracker is used. With TRACK (Fig. 1) there is a distinct increase in TC frequency with resolution for HadGEM3-GC31, CMCC-CM2-(V)HR4 and EC-Earth3P models, while all models and reanalyses typically have a smaller asymmetry of NH:SH TCs than is seen in the observations. The proportions of storms in each ocean basin agree reasonably well with observations, though for most models the relative frequency in the North Atlantic is less than observed while in the North Indian it is more. The overall NH TC frequency for the high resolution models typically approaches or exceeds that observed.

Using TempestExtremes (Fig. 2) a somewhat different picture emerges compared to the above. Now there are only two models (HadGEM3-GC31 and CNRM-CM6-1)

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

which have NH frequencies approaching or exceeding the observed. There is now a more systematic increase in TC frequency with resolution, and the hemispheric asymmetry is more consistent with that observed.

Several conclusions can be drawn from this simple comparison of models and trackers. Great care is needed when interpreting absolute TC frequency from a single tracker, since this will depend on many factors, including the tracker criteria and analysis grid. Features such as the hemispheric asymmetry could lead to the conclusion that the models produce too many SH TCs, but at least in part this seems to depend on how such storms are initially characterised (by vorticity or sea level pressure); observational issues could also contribute to the difference between models and observations, for example because SH tropical depressions and sub-tropical cyclones are not included in Best Track data whereas they are in the NH (Strachan et al. 2013; Hodges et al. 2017).

Evaluation of the models' ability to simulate the spatial distribution of tropical cyclone tracks globally is shown in Fig. 3. This shows track density derived from TRACK and observations, defined by the mean number of tracks per month through a 4° cap at each point during May-November in the NH and November-May in the SH on a common grid. For each pair of plots, the bias in the higher resolution model is shown first, followed by the difference between higher and lower resolution model.

Key aspects include:

- Most models show a reduction in the negative density bias in the North Atlantic, North Western and Eastern Pacific when resolution is increased;

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

- Many models have an excess of activity in the Southern Hemisphere, including in the South Atlantic, which is enhanced at higher resolution, as discussed above;
- There is a common negative bias in the Western Pacific which would indicate a lack of simulated TCs making landfall in the Philippines and Southern China;
- Two models (HadGEM3-GC31 and CMCC-CM2-(V)HR4, both grid point models) show a larger change with resolution, including: a positive bias near the equator extending across the Pacific which is enhanced at higher resolution, and larger positive biases extending into the mid-latitudes;
- The MPI-ESM1-2 model has very few TCs in any basin.

Results from TempestExtremes (not shown) have similar biases to Fig. 3, with slightly larger negative biases in the tropics and reduced positive biases in the extra-tropics, consistent with the lower frequencies shown in Figs. 1, 2. The resolution differences are also similar, enhanced in HadGEM3-GC31 and CNRM-CM6-1 where the lower resolution has fewer TCs, and hence the key aspects are common to both trackers apart from the Southern Hemisphere activity.

The models tend to fall into groups of responses. The HadGEM3-GC31 and CMCC-CM2-(V)HR4 models show similar biases and differences with resolution, as do the EC-Earth3P and ECMWF models. The latter is probably unsurprising given the common basis of their dynamical cores, while the former are the only grid point models.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

A summary of the impact of horizontal resolution on the TC spatial distribution is shown in Fig. 4, using the warm core segments of the cyclone tracks only. The multi-model ensemble mean resolution difference (top) and Root Mean Square Error (RMSE) difference compared to the observed track density (bottom) are shown for both TRACK and TempestExtremes. Both trackers have very consistent increases in track density with higher horizontal resolution, and this leads to decreases in RMSE of more than 50% in the North Atlantic, Eastern and North Western Pacific and the Southern Indian and Australian regions (blue regions in Fig. 4 (c,d).

There is a slight southwards shift of activity in the Eastern Pacific at higher resolution with the TRACK tracker, which causes a larger error, and the positive error towards the mid-latitudes is more evident when using TRACK than TempestExtremes, consistent with the longer tracks as seen in the track densities in Fig. 3.

In summary, enhanced horizontal resolution generally reduces some typical TC biases found in CMIP-class models, and the relative improvements are consistent across two trackers. Biases remain in the southern sector of the North Western Pacific at high resolution which will impact on TC landfall statistics there. The North Atlantic remains a challenging region to simulate (Camargo et al. 2013), perhaps partly due to low rates of intensification (see later and Manganello et al. 2012) as well as sensitivity to model physics (Bruyère et al. 2017; Chauvin et al. 2019), though the low biases are generally improved at higher resolution. Ongoing work suggests that one reason for increased TC frequency in all basins with higher horizontal resolution is a higher conversion rate of pre-TC “seeds” into TCs (Vecchi et al. 2019).

b. Tropical cyclone intensity

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Many recent studies have indicated that although changes in aspects of future tropical cyclone climatology are uncertain, it is likely that strong storms could become stronger due to increased energy availability (in the form of increasing SSTs and column water vapour; Walsh et al. 2016). Elsner et al. (2008) suggest there is already evidence for this in the historic record, while Kossin et al. (2014) suggest an observed poleward shift to the latitude of maximum intensity, though the uniformity of the observational record is questionable (Barcikowska et al. 2012; Ren 2011). However, modelling such changes is challenging for multi-decadal global climate simulations, in which the horizontal resolution is such that few models can simulate strong (Cat4/5) hurricanes, particularly in terms of surface wind speeds (Murakami et al. 2012; Murakami et al. 2015; Wehner et al. 2014). Without this capability, drawing conclusions on changing intensities determined by wind speed is somewhat questionable, and hence here we focus on minimum surface pressure instead.

Figs. 5 shows the intensity scatter and best fit (maximum 10 m wind speed vs minimum MSLP at peak storm intensity) for models, reanalyses and observations, for the North Atlantic, North Western and Eastern Pacific basins respectively. In each basin there is a systematic shift of the model intensities to higher values as resolution is increased (moving from dashed to solid lines) which is as expected; all the models struggle to achieve storm intensities much greater than Cat 2-3 using 10 m wind speeds apart from the CNRM-CM6-1-HR model. This model is an outlier, matching observations extremely closely in the Atlantic and somewhat overestimating them in the NW Pacific.

Such strong wind speeds are beyond the expected capability of the resolved dynamics of a model at this resolution according to Davis (2018). The TC intensities

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

in CNRM-CM6-1-HR are also quite different from the previous CNRM-CM5-1 model (Voldoire et al. 2012). Understanding how this model is able to generate such strong TCs is the subject of an ongoing study (Chauvin et al., 2019; Chauvin et al. in prep), but preliminary results suggest that the new CBR turbulence scheme (Cuxart et al. 2000) and the coefficients therein play an important role in enhancing the TC strength via convection. This could be viewed as either a parameterisation of an unresolved process, or as an outcome of parameter choices and hence perhaps as the right result for the wrong reason.

The models are able to capture the difference in storm intensities in each basin, with more frequent stronger storms in the NW Pacific and North Atlantic and typically weaker storms in the Eastern Pacific. It is also evident here that the reanalyses also struggle to sample the more intense TC activity.

It should be noted that TC intensity is artificially higher in these SST-forced simulations, and it has been shown that interaction with the ocean (i.e. the TC-ocean negative feedback) plays a pivotal role in reducing it (Zarzycki 2016; Scoccimarro et al. 2017). Hence coupled model simulations are likely to produce weaker TCs.

In order to examine where the TCs have their peak intensity, Fig. 6 shows the joint pdf of the mean sea level pressure (MSLP) and latitude of tropical cyclones at peak intensity for all the models, reanalyses using TRACK, and observations. The observations indicate that the TCs at their peak tend to be found at latitudes between 10-30°N with some weaker storms found further north. The low resolution models cannot capture very low MSLP and hence the MSLP distribution with latitude is more uniform or even with a peak at higher latitudes. This likely reflects lower growth rates and also that at mid-latitudes the model resolution becomes more suitable for the

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

scale of the dynamics. In some of the higher resolution models the low latitude "bulge" is more consistent with the observations, although they still have too much activity at higher latitudes. The equivalent TempestExtremes figure (not shown) is broadly similar, though the density of storms at higher latitudes is reduced due to the shorter tracks.

In attempting to understand the behaviour of model storm intensity further, Fig. 7(a,b) shows normalised pdfs of winds at both 925 hPa and 10 m from each TC at peak storm intensity for Northern Hemisphere storms. The CMCC-CM2-VHR4 and CNRM-CM6-1 HR models have maximum 925 hPa winds reaching around 80 ms^{-1} (Fig. 7a), while most of the other HR models achieve around 65 ms^{-1}. For 10 m winds, the CNRM-CM6-1 HR model has wind speeds in excess of 60 ms^{-1}, while CMCC-CM2-VHR4 reaches 55 ms^{-1} and other models more typically 40 ms^{-1}. The equivalent figure for TempestExtremes is very similar.

This would indicate that, in order for a model to attain Cat4-5 10 m wind speeds, it both requires high winds at 925 hPa, and for that momentum to be efficiently exchanged with the near surface via the boundary layer. More detailed process-level analysis will be required to understand whether this is a well-modelled physical process improvement (perhaps relating to boundary layer, convection or surface drag schemes), or whether they are an indication of marginally resolving grid-scale features.

To illustrate that the storms produced in the models do indeed reflect the observed tropical cyclone structure, Fig. 8 shows composite structures of the 10 m tangential wind speeds and MSLP from the low and high resolution model groups and reanalyses, stratified in columns by intensity based on minimum surface pressure.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

The structures are broadly consistent across models, with the core becoming smaller and more intense at higher resolution as expected. The CNRM-CM6-1 HR and CMCC-CM2-VHR4 models have a larger proportion of storms contributing to the composites at the highest intensity, consistent with the results described above, and hence the more robust composites. Note that for some models and categories, the sample of storms can become very small.

In summary, the higher resolution models are able to produce more intense TCs in terms of 10m wind speed and surface pressure. Only the CNRM-CM6-1-HR model is able to simulate above Cat3 10 m wind speeds, and hence these models do not have the capabilities of some other models at around 25 km resolution (Murakami et al. 2012; Murakami et al. 2015; Wehner et al. 2014).

c. North Atlantic mean frequency and seasonal cycle

The May-November mean tropical cyclone frequency in the North Atlantic from models and reanalyses using TRACK and TempestExtremes, and observations, over 1979-2014 (using the longer 1950-2014 period for the models shows only minor differences), is shown in Table 6, together with a breakdown to intensity classes (as measured by minimum SLP during storm lifetime). Common features include:

- The frequencies and standard deviations are mostly reduced using TempestExtremes compared to TRACK, as seen previously, and this is mainly due to a reduction in the weaker storms;

- All models (apart from HadGEM3-GC31-MM) have standard deviations which are lower than observations and reanalyses; this has implications when

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

considering climate risks from interannual-decadal tropical cyclone variability, and is sensitive to tracker;

- All the higher resolution models have an increase in storms at higher intensities, with CMCC-CM2-VHR4 and CNRM-CM6-1-HR beginning to reflect similar distributions to the observations and surpassing reanalyses in this respect;
- The CNRM-CM6-1 model has a high frequency even at low resolution using TRACK with little change between resolutions, but many of these are weak storms, and with TempestExtremes the CNRM-CM6-1-LR has much lower frequency;
- Apart from MPI-ESM1-2, all the higher resolution models have mean TRACK TC frequency within the standard deviation of the observations (and the range as represented by the reanalyses datasets).

As seen previously, the use of TempestExtremes tends to considerably reduce the numbers of storms found, with the largest differences found in the weaker storm categories. Appendix B discusses potential reasons why the trackers may act in this way. There is some evidence that the difference between trackers reduces at higher resolution, which is an expected result given that higher resolution simulates stronger storms and tracker variability is dominated by weak, short-lived systems (Zarzycki and Ullrich, 2017). The particular reasons for why some storms are detected by one tracker and not another are outside the scope of this study but remain a target for future work.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

The seasonal cycle of ACE and frequency for the North Atlantic is shown in Fig. 9 for all models and reanalyses (using TRACK and ACE_{925}) and observations over 1979-2014. The peak in activity in observations is between August-September, and the ECMWF-IFS, CNRM-CM6-1 and EC-Earth3P models mirror this well. HadGEM3-GC31 and CMCC-CM2-(V)HR4 have a slightly delayed peak in September-October, and also have too much activity early in the season, which is also true of the frequency distribution. The timing of peak activity does not seem to change with model resolution for either frequency or ACE_{925}. For most models the seasonal cycle based on TempestExtremes (not shown) scales the frequency and ACE_{925} consistent with earlier results, but for HadGEM3-GC31-HM the phase error above almost disappears, which perhaps suggests that the late-season activity with TRACK is due to weaker storms.

d. Interannual variability and ensemble size

Future projections of the frequency and variability of tropical cyclones strongly depend on how the forcing environment (e.g. global and local drivers such as SST, ENSO, humidity) will change in the future (Zhao and Held, 2012; Murakami et al. 2012; Roberts et al. 2013; Sun et al. 2017). However, our confidence in model projections of future variability is increased if we can show that past performance agrees well with observations, and particularly if models have similar dependencies to both global and regional drivers as are observed. In this section we examine the importance of ensemble size and model resolution to the skill in interannual variability.

Previous studies have shown, in individual models, that higher model resolution with small ensemble sizes (Zhao et al. 2014; Roberts et al. 2016) and larger ensemble

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

sizes at one resolution (Yoshida et al. 2017; Mei et al. 2019) are both important to capture skill in interannual variability of TCs as compared to observations. The larger ensemble sizes mean that the TC internal variability (weather noise) can be averaged out to give increasing correlation with observations (Mei et al. 2019).

In the present study the ensemble size is generally small (1-3 members) across the multi-model dataset, however for the HadGEM3-GC31 model this has been enhanced. A total of 14 members have been produced for the period 1979-2014, at both LM and MM resolutions (nominally 250 km, 100 km resolution, as part of the H2020 Blue-Action project (http://blueaction.eu), together with five members at 50km resolution. A stochastic perturbation is applied to the initial conditions to generate the ensemble. Fig. 10 shows the correlation of each set of combinations of (non-independent) *n* ensemble members within the whole ensemble for 1979-2014 for both frequency and ACE_{925} in the North Atlantic, NW Pacific and E Pacific using TRACK (sold lines) and TempestExtremes (dashed lines); the box indicates the inter-quartile range, while the whiskers show the range of the data, and the lines join the mean correlation achieved for each ensemble size. The significance levels at 95% and 99% are also indicated, based on 36 years of data.

For ACE_{925} and frequency (apart from the NW Pacific), the 100 km model has higher correlation than the 250 km model in all three basins using all ensemble members. It seems that at least 6-8 members selected from this ensemble size are needed for the correlations at these two resolutions to become distinct (as measured by non-overlapping inter-quartile ranges). The 100 km ensemble mean correlation for frequency and ACE_{925} in the North Atlantic seem to asymptote at around 0.75 and 0.70 respectively, which for example compares to a range of correlation between

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

0.4-0.85 using particular combinations of three member ensembles. Note that the combinations are not independent, hence the reduction in range for larger ensemble sizes. Since the 50 km model only has five ensemble members it is difficult to compare this to the lower resolutions, but there are indications that there is potentially extra ACE_{925} skill in this model in the NW Pacific, in contrast to little or no improvement in hindcast skill in a coupled seasonal forecast model with similar resolutions (Scaife et al. 2019).

The correlations shown in Fig. 10 using TRACK and TempestExtremes become more similar as resolution is increased, and indeed mostly overlay each other at HM resolution. This could indicate that: (1) as resolution increases, the tracker details become less important and a more common set of TCs is detected; (2) the influence of the weaker TCs on the interannual variability signal reduces as resolution increases. For the North Atlantic, Fig. 10 also shows that ACE is a more robust measure of variability (e.g. Villarini and Vecchi, 2012; Scoccimarro et al. 2018), since the LM curves are closer together in Fig. 10b compared to Fig. 10a. This reflects the much smaller number of TCs detected by TempestExtremes and hence the weaker signal in terms of variability detected with that tracker using frequency alone, but the more integral ACE measure combining frequency, intensity and lifetime is able to better sample the variability.

Mei et al. (2019) suggest that an ensemble of 20 members should be sufficient to skilfully simulate hurricane frequency in the North Atlantic (as opposed to tropical cyclone frequency shown here). Fig. 10 suggests that more than 10 members are required to fully distinguish the skill at different model resolutions for the tropical cyclones used here, and that such an ensemble size represents most of the skill in

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

the system (noting that some ensemble members can reach skills of over 0.8 here, perhaps indicating where the curve might asymptote to given enough members).

Since our ensemble size is much smaller in most models used here, can we say anything robust about variability and multi-model resolution? Fig. 11 shows the running 30 year correlation over the 1950-2014 period against observations for the North Atlantic, where each timeseries has been detrended over the whole period. There is little clear signal that the higher resolution models obtain an improved correlation for this period using one ensemble member. It is notable that nearly all correlations improve over time, perhaps indicating that:

- The models are better in periods of increased activity and/or can detect trends in activity;
- Uncertainty in the SST forcing further into the past, and the methods used in HadISST.2.2.0.0 (Kennedy et al. 2016) to reconstruct the daily, ¼ degree dataset;
- Uncertainty in the tropical cyclone frequency and ACE variability before the global satellite era due to changes in observations and procedures;

The thicker lines in Fig. 11 show model ensemble means (of up to 3 members) where available, and these typically increase the correlation compared to using only one member. However, for two models the lower resolution ensemble (thick dashed lines) has a greater correlation than the high resolution ensemble (thick solid lines), suggesting either that three members is insufficient to show an improvement with

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

resolution (consistent with Fig. 10), or else that other models could have a different resolution dependence than that shown in Fig. 10.

Table 7 shows the correlation of interannual variability with observations over the period 1979-2014 for one ensemble member for each model-resolution, for both tropical cyclone frequency and ACE_{925}. For reanalyses it is clear that the ACE_{925} correlation is more robust and consistent than frequency (as shown in Villarini and Vecchi, 2012, and Fig. 10(a,b)) and hence we focus on ACE. The models with an ensemble (of size 3 and above) have significant correlations about 0.5, while for the models with only one member only CNRM-CM6-1 at both resolutions nears 0.5.

The correlation of the TC interannual variability against selected individual drivers is shown in Table 8 for models and reanalyses. While it is difficult to assess the correlations with only one ensemble member, the models with at least 3 members have ensemble mean correlations that are consistent with the range seen in the reanalyses. Hence there is no reason to believe that the simulated TC variability has drivers different from the observations. The range of correlations using only one member may be simply indicative of internal variability, or else reflect that different models have TC genesis in different regions of the North Atlantic - different drivers influence particular regions, so if cyclogenesis is shifted (for example equatorwards or westwards) then these correlations will differ from the observed.

e. Impact of mean state in the Atlantic

Simple relationships between simulated mean state, model bias and TC climatology are generally difficult to establish (e.g. Camargo et al. 2013; Murakami et al, 2014a;

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Tang and Camargo, 2014; Kim et al. 2018) and are often model dependent. Here we briefly examine whether the models show any gross biases in key parameters known to be important for TC performance.

The mean 850-250 hPa wind shear over the June-October period for 1979-2014 is shown in Fig. 12 for models and reanalyses. Each model tends to have its own pattern of shear, and there seems little systematic change with resolution. The CNRM-CM6-1 model has the weakest shear across the North Atlantic, which is consistent with their large number of TCs produced at both resolutions using TRACK. The HadGEM3-GC31 model has its minimum shear further south than observed, and this may be linked with the low latitude of the African Easterly Jet (AEJ) in that model (Fig. 13). The MPI-ESM1-2 and ECMWF-IFS models have slightly higher shear (in the Eastern Atlantic) at higher resolution. The shear over West Africa and the Eastern Atlantic is too high in CMCC-CM2-(V)HR4.

In general the latitudes of the AEJ (Fig. 13a) are consistent with the shear, with several models (MPI-ESM1-2 for example) having the mean jet somewhat further north than indicated by the reanalyses, while HadGEM3-GC31-LM is too far to the south. Some previous work (Patricola et al. 2018) has suggested that African Easterly Waves (AEWs) play little role in setting North Atlantic tropical cyclone numbers, while Thorncroft and Hodges (2001) and Roberts et al. (2016) showed some relationship with TC variability at higher resolutions for storms with genesis in the eastern Atlantic. The mean number of African Easterly Waves (AEWs) is shown in Fig. 13b, and the maximum vorticity of these waves in Fig. 13c, calculated following the Bain et al. (2014) simple Hovmöller algorithm calculated on a common grid. There is little evident resolution sensitivity in mean AEW number, and no

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

obvious relation with each model having its own character. All the models are within the range of the reanalyses. There is a more systematic increase in the vorticity of the AEWs with model resolution and perhaps this helps to improve the storm distribution in the eastern Atlantic (Fig. 3) by enabling earlier genesis.

4. Conclusions

The CMIP6 HighResMIP experimental design enables a more systematic assessment of the role of horizontal resolution in the simulation of global tropical cyclones over the period 1950-2014 across multiple models. The results from six modelling groups within the European PRIMAVERA project have been analysed in this work, with resolutions spanning from around 200 km to 25 km. There are several seemingly consistent changes when resolution is increased:

- Increased tropical cyclone frequency and seasonal ACE index in the North Atlantic
- Improved capability to represent the spectrum of tropical cyclone intensities
- Improved distribution of tropical cyclone tracks (and genesis regions)

These conclusions seem to be robust to (at least two) different trackers used in this study, TRACK and TempestExtremes. These improvements are consistent with previous studies using multi-decadal simulations of individual climate models at similar 25 km resolutions (e.g. Zhao et al. 2008; Caron et al. 2011; Murakami et al. 2012; Wehner et al. 2014; Murakami et al. 2015; Roberts et al. 2016).

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Correlations of interannual ACE variability with observations seem to be more robust than using simple storm frequency, but there is no obvious relationship between increased resolution and improved correlation using only one ensemble member.

Using the HadGEM3-GC31 model and several resolutions with an ensemble of 14 members does indicate that increasing resolution from 200 km to 100 km improves model skill for North Atlantic interannual variability. In this case, at 100 km resolution the ensemble mean correlation tends towards ~0.75/0.7 (frequency/ACE), with a sub-sample of ensemble size of 6-8 suggestive of being sufficient to be a robust measure. Hence for this simulation protocol and model, we can explain ~50% of the variance in observed tropical cyclone interannual ACE variability. In the NW Pacific, there is evidence that 50 km resolution offers a further increase in skill.

Future work is needed to discover what factors are missing that could allow more of the variance to be explained. This may lie within the HighResMIP protocol itself (which, for example, excludes interannual variations in natural aerosol, and uses one specific set of SST-sea ice forcing datasets), or could lie with the models themselves (via model bias, lack of key processes, requirement for even higher resolution or limitations in physics such as convection schemes).

Further investigation of the CNRM-CM6-1 model is required to understand how it is ableto achieve such outstanding surface wind speeds compared to all other models, which allows this model to represent the full tropical cyclone intensity spectrum. The other models in this study are not able to simulate above Cat3 intensities as measured by 10 m wind speeds. Davis (2018) suggest that somewhat higher intensities should be possible in theory at 25 km resolution, and indeed other models have shown such capability (e.g. Wehner et al. 2014; Murakami et al. 2015).

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Use of the CMIP6 HighResMIP coupled model simulations can be used to further assess drivers of variability and intensity when the atmosphere and ocean are able to fully interchange fluxes. This configuration may also be useful to understand likely future changes in tropical cyclone characteristics, and is addressed in Roberts et al. (2019d).

Additional assessment of different tracking trackers is needed to better understand their strengths, weaknesses and sources of difference but this needs to be done fairly with some well constrained criteria for evaluation. Using multiple trackers is also likely to be important when assessing future climate simulations, which also form a part of the HighResMIP experimental design.

APPENDIX A

Brief model descriptions.

Brief descriptions of the different models used in this study are included here, in particular aspects that are relevant to tropical cyclones. A summary of the model components is shown in Table 1, and all the parameter changes between model resolutions are shown in Table 3.

The standard HadGEM3-GC31 model configuration is described in Williams et al. (2018), with the atmosphere configuration (GA7.1) further described by Walters et al. (2019) and the HighResMIP configuration in Vidale et al (in prep) and Roberts et al. (2019). The dynamical core uses a semi-implicit semi-Lagrangian formulation to

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

solve the non-hydrostatic, fully-compressible deep-atmosphere equations of motion (Wood et al., 2014) on a regular latitude-longitude grid, with 85 levels with a top at 85 km. This model has been used to generate a larger ensemble size (of up to 14 members) to examine the robustness of some results. Each resolution has at least three ensemble members over 1950-2014. In addition, over the 1979-2014 period, stochastic perturbation of the initial conditions is used and 10 additional members are produced for LM and MM models, and two more members for HM.

The ECMWF-IFS model used for HighResMIP is documented in Roberts et al. (2018) and references therein. The atmospheric component of the Integrated Forecasting System (IFS cyc43r1) model is based on a hydrostatic, semi-Lagrangian, semi-implicit dynamical core with computations alternated between spectral and reduced Gaussian grid-point representations each time step. The vertical discretization is based on a hybrid sigma-pressure coordinate, with 91 levels in the vertical, with top at 0.01 hPa. Additional ensemble members have been generated by random perturbations to the initial stochastic perturbed parameterized tendencies (SPPT) scheme.

The EC-Earth3P model is documented in Haarsma et al. (2019, in prep). The atmospheric component of the Integrated Forecasting System (IFS cyc36r4) model is based on a hydrostatic, semi-Lagrangian, semi-implicit dynamical core. The vertical discretization is based on a hybrid sigma-pressure coordinate, with 91 levels in the vertical, with top at 0.01 hPa.

The MPI-ESM1-2 model is documented in Gutjahr et al (2019) and references therein. The atmospheric submodel of MPI-ESM1.2 is ECHAM6.3, with a dynamical core based on a vorticity and divergence form of the primitive equations, solved

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

using a spectral-transform method. The vertical discretization uses a hybrid sigma-pressure coordinate system with 95 vertical levels with a top at 0.01 hPa.

The CNRM-CM6-1 model is documented in Voldoire et al. (2019) for CMIP6 DECK experiments. It is based on four main components for atmosphere, surface and ocean and sea ice. The atmospheric component is based on the spectral atmospheric model ARPEGE-Climat version 6.3. There are 91 vertical levels following a hybrid σ pressure discretization with 15 levels in the boundary layer. Since the previous version of the model, changes have been introduced in the parameterizations and mainly concern the convection (Piriou et al. 2007, Gueremy et al. 2011), microphysics (Lopez 2002) and turbulence (Cuxart et al. 2000). The surface component SURFEX (Masson et al. 2013) includes 3 surface types: ocean, land and lakes.

A general description of CMCC-CM2 models family used in CMIP6 can be found in Cherchi et al. (2019). In the present study, the CMCC-CM2-(V)HR4 configuration is used, specifically developed for HighResMIP. This model differs from the standard resolution CMCC-CM2 configuration (CMCC-CM2-SR5; Cherchi et al., 2019) in that it makes use of the Community Atmosphere Model vn4 (CAM4; Neale et al., 2010), in alternative to CAM5. This choice allowed a substantial reduction of computational costs, especially beneficial for the high-resolution (CMCC-CM2-VHR4) experiments, and it made possible the implementation of the MACv2-SP "simple plume" scheme for the anthropogenic aerosols (Stevens et al., 2017), following the HighResMIP protocol. Specific aspects concerning the CMCC-CM2-(V)HR4 ability in reproducing the characteristics of TCs in the West North Pacific are documented in Scoccimarro et al. 2019.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

APPENDIX B

Brief tracking algorithm (tracker) descriptions

Brief descriptions of the two trackers used to find tropical cyclones within the model simulations are included here, for TRACK (Hodges et al. 2017), and TempestExtremes (Ullrich and Zarzycki, 2017; Zarzycki and Ullrich 2017). There are no changes in the trackers used between models and resolutions. Note that the variables used are on the Analysis grid (Table 2) for each model.

TRACK uses relative vorticity as the feature-tracking variable. The vorticity over 850, 700, 600 hPa is averaged on the analysis grid, and then spectrally filtered to a common T63 grid using triangular truncation to retain wavenumbers 6-63. The tracking proceeds by identifying the off-grid vorticity maxima, by applying a maximization scheme (Hodges 1995), if they exceed a value of $5x10^{-6}$ s^{-1} in each time frame (SH scaled by -1). These are initially linked together using a nearest-neighbor approach and then refined by minimizing a cost function for track smoothness, subject to adaptive constraints on displacement distance and track smoothness (Hodges 1999). Only tracks that last at least 2 days (eight time steps) are retained for further analysis. Identification criteria post tracking are used to isolate warm-core tropical cyclones: 1) T63 relative vorticity at 850 hPa must attain a threshold of $6x10^{-5}$ s^{-1}; 2) the difference in vorticity between 850 and 250 hPa (at T63 resolution) must be greater than $6x10^{-5}$ s^{-1} to provide evidence of a warm core; 3) the T63 vorticity centre must exist at each level (850, 700, 600, 500, 250 hPa) for a coherent vertical structure; 4) 1-3 must be jointly attained for at least four

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

consecutive timesteps (one day) and only apply over the oceans; 5) tracks must start between 30°S-30°N.

TempestExtremes uses sea level pressure (SLP) as its feature-tracking variable on the native analysis grid. Candidates are initially identified by minima in SLP, and a closed contour criteria is applied, requiring an increase in SLP of at least 2 hPa within 5.5° of the candidate node. A decrease in geopotential height difference (250 - 500 hPa) of 6 m within 6.5° of the candidate within 1° of the candidate with maximum geopotential height. Candidates are then stitched in time to form paths, with a maximum distance between candidates of 8°, consisting of at least ten candidates per path and with a maximum gap size of three (number of time steps where no identification occurred). For at least ten timesteps the underlying topographic height must be at most 1500 m, and for at least four timesteps it must be at most 10 m, and the storm must form between 10-40°. The storm must also travel at least 8°.

The TRACK configuration is tuned to capture roughly the number of tropical storms including possibly tropical depressions and sub-tropical storms found in observations, primarily using the ECMWF operational analyses (Bengtsson et al. 2007). The TempestExtremes configuration was developed by performing a sensitivity analysis and optimizing against high-resolution reanalysis products as described in Zarzycki and Ullrich (2017). It has attempted to keep the false alarm rate to acceptable levels, which may have the effect of reducing the detection of weaker storms.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Acknowledgements

MR, JS, PLV, KH, BV, RH, AB, ES, LPC, LT, CR, RS, DP acknowledge funding from the PRIMAVERA project, funded by the European Union's Horizon 2020 programme under Grant Agreement no. 641727. JM acknowledges funding from the Blue-Action project, funded by the European Union's Horizon 2020 programme under Grant Agreement no. 727852. MR and JC acknowledge support from the UK-China Research & Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund. Funding for PU and CZ to support use of the TempestExtremes suite was provided under NASA award NNX16AG62G and the US Department of Energy Office of Science award number DE-SC0016605.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Many thanks to the editor and three anonymous reviewers for their comments which have greatly strengthened this manuscript.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

References

Aon Benfield, 2018: Weather, Climate & Catastrophe Insight — 2017 Annual Report. Global Economic Losses. Aon Benfield UCL Hazard Research Centre. 56pp.

Bain, C. L., Williams, K. D., Milton, S. F. and Heming, J. T., 2014: Objective tracking of African Easterly Waves in Met Office models. *Q.J.R. Meteorol. Soc*., **140**, 47-57, https://doi.org/10.1002/qj.2110

Barcikowska, M., F. Feser, and H. von Storch, 2012: Usability of Best Track Data in Climate Statistics in the Western North Pacific. *Mon. Wea. Rev*., **140**, 2818–2830, https://doi.org/10.1175/MWR-D-11-00175.1

Batté, L. and F.J. Doblas-Reyes, 2015: Stochastic atmospheric perturbations in the EC-Earth3 global coupled model: impact of SPPT on seasonal forecast quality. *Clim. Dyn*., **45**, 3419-3439, https://doi.org/10.1007/s00382-015-2548-7.

Bell, G. D., and Coauthors, 2000: Climate assessment for 1999. *Bull. Amer. Meteor. Soc*., **81** (6), s1-s50, https://doi.org/10.1175/1520-0477(2000)81\%5Bs1:caf\%5D2.0.co;2

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

Bruyère, C. L., and Coauthors, 2017: Impact of Climate Change on Gulf of Mexico Hurricanes. NCAR Technical Note NCAR/TN-535+STR, 165 pp, https://doi.org/10.5065/D6RN36J3.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Camargo, S. J., 2013: Tropical cyclones in high-resolution climate models. *U.S. CLIVAR Variations*, Vol. 11, No. 3, 4-11.

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

Camargo, S. J. and A. A. Wing, 2016: Tropical cyclones in climate models. *WIREs Clim. Change*, **7**, 211-237, https://doi.org/10.1002/wcc.373

Camp, J., M. Roberts, C. MacLachlan, E. Wallace, L. Hermanson, A. Brookshaw, A. Arribas, A. A. Scaife, 2015: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. *Q.J.R. Meteorol. Soc*., **141** (691), 2206-2219, https://doi.org/10.1002/qj.2516

Caron, L-P, C. G. Jones and K. Winger, 2011: Impact of resolution and downscaling technique in simulating recent Atlantic tropical cyclone activity. *Clim. Dyn*., **5**, 869-892.

Caron, L.-P. and C. G. Jones, 2012: Understanding and simulating the link between African Easterly Waves and Atlantic Tropical Cyclones using a Regional Climate Model: The role of domain size and lateral boundary conditions. *Clim. Dyn*., **39**, 113-135, https://doi.org/10.1007/s00382-011-1160-8.

Chauvin, F., and Coauthors, 2019: Future changes in Atlantic hurricanes with the rotated-stretched ARPEGE-Climat at very high resolution. *Clim. Dyn*., https://doi.org/10.1007/s00382-019-05040-4.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Cherchi, A., and Coauthors, 2019: Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. *J. Adv. Model. Earth Syst*., **11**, https://doi.org/10.1029/2018MS001369.

Chu, J. H., C. R. Sampson, A. S. Levine, E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone Best-Tracks, 1945-2000. Tech. rep., naval Research Laboratory Tech. Rep. NRL/MR/7540-02-16.

Copernicus Climate Change Service (C3S), 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), date of access:15.03.2019. https://cds.climate.copernicus.eu/cdsapp#!/home

Cuxart J, P. Bougeault, J. L. Redelsperger, 2000: A turbulence scheme allowing for mesoscale and large-eddy simulations. *Q J R Meteorol. Soc*., **126**, 1-30, https://doi.org/10.1002/qj.49712656202

Daloz A.S., and Coauthors, 2015: Cluster analysis of explicitly and downscaled simulated North Atlantic tropical cyclone tracks. *J. Climate*., **28**, 1333–1361, https://doi.org/10.1175/JCLI-D-13-00646.1

Davis, C. A., 2018: Resolving Tropical Cyclone Intensity in Models. *Geophys. Res. Lett.*, ***45***(4), 2082–2087, https://doi.org/10.1002/2017GL076966.

Dee, D. P., and Co-authors, 2011: The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc*., **137** (656), 553-597, https://doi.org/10.1002/qj.828

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

EC-Earth Consortium (EC-Earth), 2018a: *EC-Earth-Consortium EC-Earth3P model output prepared for CMIP6 HighResMIP*. Earth System Grid Federation. http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.HighResMIP.EC-Earth-Consortium.EC-Earth3P

EC-Earth Consortium (EC-Earth), 2018b: *EC-Earth-Consortium EC-Earth3P-HR model output prepared for CMIP6 HighResMIP*. Earth System Grid Federation. http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.HighResMIP.EC-Earth-Consortium.EC-Earth3P-HR

Elsner, J. B., J. P. Kossin, and T. H. Jagger, 2008: The increasing intensity of the strongest tropical cyclones. *Nature*, ***455***(7209), 92–95, https://doi.org/10.1038/nature07234

Eyring, V., S. Bony, G. A. Meehl, C. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2015: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation. *Geosci. Model Dev.*, **9**(5), 1937-1958, https://doi.org/10.5194/gmd-9-1937-2016.

Franco-Diaz, A., N. P. Klingaman, P. L. Vidale, L. Guo, and M.-E. Demory, 2019: The contribution of tropical cyclones to the atmospheric branch of Middle America's hydrological cycle using observed and reanalysis tracks. *Clim. Dyn*., https://doi.org/10.1007/s00382-019-04920-z.

Gelaro, R., and Coauthors, 2017: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Climate*, **30**(14), 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Guérémy J.-F., 2011: A continuous buoyancy based convection scheme: One- and three-dimensional validation. *Tellus A: Dyn. Meteorol. Oceanogr*., 63, 687-706, https://doi.org/10.1111/j.1600-0870.2011.00521.x

Guo, L., N.P. Klingaman, P.L. Vidale, A.G. Turner, M. Demory, and A. Cobb, 2017: Contribution of Tropical Cyclones to Atmospheric Moisture Transport and Rainfall over East Asia. *J. Climate,* **30**, 3853–3865, https://doi.org/10.1175/JCLI-D-16-0308.1

Gutjahr, O., D. Putrasahan, K. Lohmann, J. H. Jungclaus, J.-S. von Storch, N. Brüggemann, H. Haak, and A. Stössel, 2019: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP), *Geosci. Model Dev*., **12**, 3241-3281, https://doi.org/10.5194/gmd-12-3241-2019.

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

Haarsma, R, and Coauthors, 2019: HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR. Description, model performance, data handling and validation. *Geosci. Model Dev.*, submitted.

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

Horn, M., K. Walsh, M. Zhao, S.J. Camargo, E. Scoccimarro, H. Murakami, H. Wang, A. Ballinger, A. Kumar, D.A. Shaevitz, J.A. Jonas, and K. Oouchi, 2014: Tracking

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Scheme Dependence of Simulated Tropical Cyclone Response to Idealized Climate Simulations. *J. Clim.*, **27**, 9197–9213, https://doi.org/10.1175/JCLI-D-14-00200.1

Jiang, H. and E.J. Zipser, 2010: Contribution of Tropical Cyclones to the Global Precipitation from Eight Seasons of TRMM Data: Regional, Seasonal, and Interannual Variations. *J. Climate,* **23**, 1526–1543, https://doi.org/10.1175/2009JCLI3303.1.

Kennedy, J., H. Titchner, N. Rayner, M. Roberts, 2017: *input4MIPs.MOHC.SSTsAndSeaIce.HighResMIP.MOHC-HadISST-2-2-0-0-0*. Version *20170505*.Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1221

Kim, D., Y. Moon, S.J. Camargo, A.A. Wing, A.H. Sobel, H. Murakami, G.A. Vecchi, M. Zhao, and E. Page, 2018: Process-Oriented Diagnosis of Tropical Cyclones in High-Resolution GCMs. *J. Climate,* **31**, 1685–1702, https://doi.org/10.1175/JCLI-D-17-0269.1

Klaver, R., R. Haarsma, P.L. Vidale, W. Hazeleger, 2019: Effective resolution in high resolution global atmospheric models for climate studies. *Atmos*. *Sci*. *Lett*., in press.

Knapp, K.R., M.C. Kruk, D.H. Levinson, H.J. Diamond, and C.J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc*., **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1

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.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Kodama, C., and Coauthors, 2015: A 20-Year climatology of a NICAM AMIP-type simulation. *J. Met. Soc. Japan*, **93**(4), 393–424. https://doi.org/10.2151/jmsj.2015-024.

Kossin, J. P., Emanuel, K. A., & Vecchi, G. A., 2014: The poleward migration of the location of tropical cyclone maximum intensity. *Nature*, **509**(7500), 349–352. https://doi.org/10.1038/nature13278.

Landsea, C. W., J. L. Franölin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev*. **141** (10), 3576-3592, https://doi.org/10.1175/mwr-d-12-00254.1

Landsea, C.W., G.A. Vecchi, L. Bengtsson, and T.R. Knutson, 2010: Impact of Duration Thresholds on Atlantic Tropical Cyclone Counts. *J. Clim.*, **23**, 2508–2519, https://doi.org/10.1175/2009JCLI3034.1

Landsea, C. W., 2000: El Niño-Southern Oscillation and the seasonal predictability of tropical cyclones. In El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts, edited by H. F. Diaz and V. Markgraf. pp.149-181

Lopez P, 2002: Implementation and validation of a new prognostic large-scale cloud and precipitation scheme for climate and data-assimilation purposes. *Q. J. R. Meteorol. Soc.,* **128***,* 229-257*,* https://doi.org/ 10.1256/00359000260498879.

MacLachlan, C., and Coauthors, 2014: Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Q. J. R. Meteorol. Soc*., **141**, https://doi.org/10.1002/qj.2396, 2014.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Manganello, J.V., and Coauthors, 2012: Tropical Cyclone Climatology in a 10-km Global Atmospheric GCM: Toward Weather-Resolving Climate Modeling. *J. Climate*, **25**, 3867–3893, https://doi.org/10.1175/JCLI-D-11-00346.1

Masson V., and Coauthors, 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geosci. Model Dev*., **6**, 929-960, https://doi.org/10.5194/gmd-6-929-2013

Mei, W., Y. Kamae, S. Xie, and K. Yoshida, 2019: Variability and Predictability of North Atlantic Hurricane Frequency in a Large Ensemble of High-Resolution Atmospheric Simulations. *J. Climate*, **32**, 3153–3167, https://doi.org/10.1175/JCLI-D-18-0554.1

Murakami H., R. Mizuta, E. Shindo, 2012: Future changes in tropical cyclone activity project by multi-physics and multi-SST ensemble experiments using 60km mesh MRI-AGCM. *Clim. Dyn*., 39, 2569– 2584, https://doi.org/10.1007/s00382-011-1223-x

Murakami, H., P. Hsu, O. Arakawa, and T. Li, 2014a: Influence of Model Biases on Projected Future Changes in Tropical Cyclone Frequency of Occurrence. *J. Climate,* **27**, 2159–2181, https://doi.org/10.1175/JCLI-D-13-00436.1

Murakami, H., 2014b: Tropical cyclones in reanalysis data sets. *Geophys. Res. Lett.*, **41**, 2133–2141, doi:10.1002/2014GL059519.

Murakami, H., G.A. Vecchi, S. Underwood, T.L. Delworth, A.T. Wittenberg, W.G. Anderson, J. Chen, R.G. Gudgel, L.M. Harris, S. Lin, and F. Zeng, 2015: Simulation and Prediction of Category 4 and 5 Hurricanes in the High-Resolution GFDL HiFLOR Coupled Climate Model. *J. Clim.*, **28**, 9058–9079, https://doi.org/10.1175/JCLI-D-15-0216.1.

Nakamura, J., Coauthors, 2017: Western North Pacific tropical cyclone model tracks in present and future climates. *J. Geophys. Res. Atmos.*, **122**, 9721–9744, https://doi.org/10.1002/2017JD027007.

Neale, R. B. and Coauthors 2010: Description of the NCAR Community Atmosphere Model (CAM4.0). NCAR/TN-485+STR, NCAR Technical Note

Palmer, T. N., R. Buizza, F. J. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer: Stochastic parametrization and model uncertainty, Tech. Rep. 1, ECMWF RD Technical Memorandum, ECMWF, Reading, UK, 2009.

Patricola, C. M., R. Saravanan, and P. Chang, 2018: The response of Atlantic tropical cyclones to suppression of African easterly waves. *Geophys. Res. Lett*., **45**, 471– 479. https://doi.org/10.1002/2017GL076081

Piriou J.-M., and Coauthors, 2007: An Approach for Convective Parameterization with Memory: Separating Microphysics and Transport in Grid-Scale Equations. *J. Atmos. Sci.*, **64**:4127–4139, https://doi.org/ 10.1175/2007JAS2144.1

Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**(11), 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1

Reed, K. A., J. T. Bacmeister, J. J. A. Huff, X. Wu, S. C. Bates and N. A. Rosenbloom, 2019: Exploring the impact of dust on North Atlantic hurricanes in a high-resolution climate model. *Geophys. Res. Lett.*, **46**, 1105– 1112, https://doi.org/10.1029/2018GL080642.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Ren, F., J. Liang, G. Wu, W. Dong, and X. Yang, 2011: Reliability Analysis of Climate Change of Tropical Cyclone Activity over the Western North Pacific. *J. Climate*, **24**, 5887–5898, https://doi.org/10.1175/2011JCLI3996.1

Roberts, C. D., R. Senan, F. Molteni, S. Boussetta, M. Mayer and S. Keeley, 2018: Climate model configurations of the ECMWF Integrated Forecast System (ECMWF-IFS cycle 43r1) for HighResMIP. *Geosci. Model Dev.*, **11**, 3681-3712, https://doi.org/10.5194/gmd-11-3681-2018.

Roberts, C. D., R. Senan, F. Molteni, S. Boussetta, S. Keeley, 2017a: *ECMWF ECMWF-IFS-LR model output prepared for CMIP6 HighResMIP*. Version *20170915*.Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2463

Roberts, C. D., R. Senan, F. Molteni, S. Boussetta, S. Keeley, 2017b: *ECMWF ECMWF-IFS-HR model output prepared for CMIP6 HighResMIP*. Version *20170915*.Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2461

Roberts, M. J., Coauthors, 2013: Sensitivity of tropical cyclone simulation to SST forcing. *U.S. CLIVAR Variations*, Vol. 11, No. 3, 12-17.

Roberts, M. J., Coauthors, 2015: Tropical cyclones in the UPSCALE ensemble of high-resolution global climate models. *J. Climate*, **28**(2), 574–596, https://doi.org/10.1175/JCLI-D-14-00131.1

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Roberts, M., 2017a: *MOHC HadGEM3-GC31-LM model output prepared for CMIP6 HighResMIP*. Version *20170906*.Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1321

Roberts, M., 2017b: *MOHC HadGEM3-GC31-MM model output prepared for CMIP6 HighResMIP*. Version *20180818*.Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1902

Roberts, M., 2017c: *MOHC HadGEM3-GC31-HM model output prepared for CMIP6 HighResMIP*. Version *20170831*.Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.446

Roberts, M. J., and Coauthors, 2019a: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geosci. Model Dev., in press*, https://doi.org/10.5194/gmd-2019-148.

Roberts, M., 2019b: CMIP6 HighResMIP: Tropical storm tracks as calculated by the TRACK algorithm. Centre for Environmental Data Analysis, *2019*. http://catalogue.ceda.ac.uk/uuid/0b42715a7a804290afa9b7e31f5d7753

Roberts, M., 2019c: CMIP6 HighResMIP: Tropical storm tracks as calculated by the TempestExtremes algorithm. Centre for Environmental Data Analysis, 2019. http://catalogue.ceda.ac.uk/uuid/438268b75fed4f27988dc02f8a1d756d

Roberts, M. J., and Coauthors, 2019d: Projected future changes in tropical cyclones using the CMIP6 HighResMIP multi-model ensemble. *Geophys. Res. Lett*., in prep.

Saha, S. and Coauthors, 2014: The NCEP climate forecast system version 2. *J. Climate*, **27**(6), 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Sanchez, C., K. D. Williams and M. Collins, 2016: Improved stochastic physics schemes for global weather and climate models, *Q. J. R. Meteorol. Soc*., **142**, 147-159, https://doi.org/10.1002/qj.2640.

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

Scoccimarro, E., S. Gualdi, G. Villarini, G.A. Vecchi, M. Zhao, K. Walsh, and A. Navarra, 2014: Intense Precipitation Events Associated with Landfalling Tropical Cyclones in Response to a Warmer Climate and Increased CO_2. *J. Climate,* **27**, 4642–4654, https://doi.org/10.1175/JCLI-D-14-00065.1

Scoccimarro E., A. Bellucci, A. Storto, S. Gualdi, S. Masina, and A. Navarra, 2018: Remote sub-surface ocean temperature as a predictor of Atlantic hurricane activity. *PNAS*, **115** (45), 11460-11464, https://doi.org/10.1073/pnas.1810755115.

Scoccimarro E., P.G. Fogli, K. Reed, S. Gualdi, S.Masina, A. Navarra, 2017: Tropical cyclone interaction with the ocean: the role of high frequency (sub-daily) coupled processes. *J. Climate*, **30**, 145–162, https://doi.org/10.1175/JCLI-D-16-0292.1

Scoccimarro, E., A. Bellucci, D. Peano, 2017a: *CMCC CMCC-CM2-HR4 model output prepared for CMIP6 HighResMIP*. Version *YYYYMMDD*[1].Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1359

Scoccimarro, E., A. Bellucci, D. Peano, 2017b: *CMCC CMCC-CM2-VHR4 model output prepared for CMIP6 HighResMIP*. Version *YYYYMMDD*[1].Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1367

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Scoccimarro, E., S. Gualdi, A. Bellucci, D. Peano, A. Cherchi, A. Navarra, 2019: The typhoon-induced drying of the Maritime Continent. *PNAS*, under review.

Shaevitz, D. A., and Coauthors, 2014: Characteristics of tropical cyclones in high-resolution models in the present climate, *J. Adv. Model. Earth Syst.*, **6**, 1154– 1172, https://doi.org/10.1002/2014MS000372.

Stevens, B., and Coauthors, 2017: MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6. *Geosci. Model Dev.*, **10**(1), 433–452. https://doi.org/10.5194/gmd-10-433-2017.

Strachan, J., P. L. Vidale, K. Hodges, M. Roberts and M. E. Demory, 2013: Investigating global tropical cyclone activity with a hierarchy of AGCMs: The role of model resolution. *J. Climate*, **26**(1), 133–152, https://doi.org/10.1175/JCLI-D-12-00012.1

Sun, Y., Z. Zhong, T. Li, L. Yi, S. J. Camargo, Y. Hu, K. Liu, H. Chen, Q. Liao, and J. Shi, 2017: Impact of ocean warming on tropical cyclone track over the western north pacific: A numerical investigation based on two case studies, *J. Geophys. Res. Atmos.*, 122, 8617– 8630, https://doi.org/10.1002/2017JD026959.

Taylor, K.E., D. Williamson and F. Zwiers, 2000: The sea surface temperature and sea ice concentration boundary conditions for AMIP II simulations. In PCMDI Report 60, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, 25 pp

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Tang, B., and S. J. Camargo, 2014: Environmental control of tropical cyclones in CMIP5: A ventilation perspective, *J. Adv. Model. Earth Syst.*, 6, 115– 128, https://doi.org/10.1002/2013MS000294.

Thorncroft, C. and K. Hodges, 2001: African Easterly Wave Variability and Its Relationship to Atlantic Tropical Cyclone Activity. *J. Climate*, **14**, 1166–1179, https://doi.org/10.1175/1520-0442(2001)014<1166:AEWVAI>2.0.CO;2.

Ullrich, P. A., and C. M. Zarzycki, 2017: TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geosci. Model Dev.*, **10**(3), 1069–1090. https://doi.org/10.5194/gmd-10-1069-2017.

Vannière, B., P. L. Vidale, M.-E. Demory, R. Schiemann, M. J. Roberts, C. D. Roberts, M. Matsueda, L. Terray, T. Koenigk, R. Senan, 2018: Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution. *Clim. Dyn*., **52**, 6817–6846, https://doi.org/10.1007/s00382-018-4547-y.

Vecchi, G.A., and Coauthors, 2019: Tropical cyclone sensitivities to CO2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. *Clim. Dyn*., **53**, 5999–6033, https://doi.org/10.1007/s00382-019-04913-y.

Villarini, G., and G. A. Vecchi, 2013: Multiseason lead forecast of the north atlantic power dissipation index (PDI) and accumulated cyclone energy (ACE). *J. Climate*, **26**(11), 3631–3643, https://doi.org/10.1175/JCLI-D-12-00448.1

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

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Voldoire, A., 2018: *CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 HighResMIP*. Earth System Grid Federation. http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.HighResMIP.CNRM-CERFACS.CNRM-CM6-1

Voldoire, A., 2017: *CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 HighResMIP*. Earth System Grid Federation. http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.HighResMIP.CNRM-CERFACS.CNRM-CM6-1-HR

Voldoire, A., and Coauthors, 2013: The CNRM-CM5.1 global climate model: description and basic evaluation. *Clim. Dyn*., **40,** 2091-2121, https://doi.org/10.1007/s00382-011-1259-y

von Storch, J.-S., and Coauthors, 2017: *MPI-M MPI-ESM1.2-HR model output prepared for CMIP6 HighResMIP*. Earth System Grid Federation. http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.HighResMIP.MPI-M.MPI-ESM1-2-HR

von Storch, J.-S., and Coauthors, 2019: *MPI-M MPI-ESM1.2-XR model output prepared for CMIP6 HighResMIP*. Earth System Grid Federation. http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.HighResMIP.MPI-M.MPI-ESM1-2-XR

Walsh, K., S. Lavender, H. Murakami, E. Scoccimarro, L.-P. Caron and M. Ghantous, 2011: The Tropical Cyclone Climate Model Intercomparison Project. Hurricanes and Climate Change (Volume 2), Springer, 1-24.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Walsh, K., S. Lavender, E. Scoccimarro, H. Murakami, 2013: Resolution dependence of tropical cyclone formation in CMIP3 and finer resolution models. *Clim. Dyn*., **40**, 585. https://doi.org/10.1007/s00382-012-1298-z

Walsh, K.J., and Coauthors, 2015: Hurricanes and Climate: The U.S. CLIVAR Working Group on Hurricanes. *Bull. Amer. Meteor. Soc.,* **96**, 1440, https://doi.org/10.1175/BAMS-D-15-00232.1

Walsh, K. J., and Coauthors, 2016: Tropical cyclones and climate change. *WIREs Clim. Change*, **7**, 65-89, https://doi.org/10.1002/wcc.371.

Walters, D., and Coauthors, 2019: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci. Model Dev*., **12**, 1909-1963, https://doi.org/10.5194/gmd-12-1909-2019.

Watson, P. A. G., Berner, J., Corti, S., Davini, P., von Hardenberg, J., Sanchez, C., Weisheimer, A., and Palmer, T. N., 2017: The impact of stochastic physics on tropical rainfall variability in global climate models on daily to weekly time scales. *J. Geophys. Res. Atmos*., **122**, 5738– 5762, doi:10.1002/2016JD026386.

Wehner, M. F., K. A. Reed, F. Li, Prabhat, J. Bacmeister, C.-T. Chen, C. Paciorek, P. J. Gleckler, K. R. Sperber, W. D. Collins, A. Gettelman, and C. Jablonowski, 2014: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *J. Adv. Model. Earth Syst*., **6**, 980–997, doi:10.1002/2013MS000276.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Williams, K. D., and Coauthors, 2015: The Met Office Global Coupled Model 2.0 (GC2) configuration. *Geosci. Model Dev.*, **8** (5), 1509-1524, https://doi.org/10.5194/gmd-8-1509-2015

Wood, N., and Coauthors, 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Q. J. R. Meteorol. Soc*., **140**, 1505–1520, https://doi.org/10.1002/qj.2235.

Yoshida, K., Sugi, M., Mizuta, R., Murakami, H., and Ishii, M., 2017: Future changes in tropical cyclone activity in high-resolution large-ensemble simulations. *Geophys. Res. Lett*., **44**, 9910– 9917. https://doi.org/10.1002/2017GL075058.

Zarzycki, C. M., 2016: Tropical cyclone intensity errors associated with lack of two-way ocean coupling in high-resolution global simulations. *J. Clim.*, **29**(23), 8589-8610. https://doi.org/10.1175/JCLI-D-16-0273.1.

Zarzycki, C. M., and P. A. Ullrich, 2017: Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophys. Res. Lett.*, **44**(2), 1141–1149, https://doi.org/10.1002/2016GL071606.

Zhao, M., I.M. Held, S. Lin, and G.A. Vecchi, 2009: Simulations of Global Hurricane Climatology, Interannual Variability, and Response to Global Warming Using a 50-km Resolution GCM. *J. Climate,* **22**, 6653–6678, https://doi.org/10.1175/2009JCLI3049.1

Zhao, M. and I.M. Held, 2012: TC-Permitting GCM Simulations of Hurricane Frequency Response to Sea Surface Temperature Anomalies Projected for the Late-

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Twenty-First Century. *J. Climate,* **25**, 2995–3009, https://doi.org/10.1175/JCLI-D-11-00313.1

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Tables

Institution	MOHC, UREAD, NERC	EC-Earth KNMI, SHMI, BSC, CNR	CERFACS	MPI-M	CMCC	ECMWF
Model name	HadGEM3-GC31	EC-Earth3P	CNRM-CM6-1	MPI-ESM1-2	CMCC-CM2-(V)HR4	ECMWF-IFS
Resolution names	LM, MM, HM	LR, HR	LR, HR	HR, XR	HR4, VHR4	LR, HR
Model atmosphere	MetUM	IFS cyc36r4	ARPEGE6.3	ECHAM6.3	CAM4	IFS cyc43r1
Atmos dynamical scheme (grid)	Grid point (SISL, lat-lon)	Spectral (linear, reduced Gaussian)	Spectral (linear, reduced Gaussian)	Spectral (triangular, Gaussian)	Grid point (finite volume, lat-lon)	Spectral (cubic octohedral, reduced Gaussian)
Atmos grid name	N96, N216, N512	Tl255, Tl511	Tl127, Tl359	T127, T255	1°x1°, 0.25°x0.25°	Tco199, Tco399
Atmos mesh spacing (0N), km	208, 93, 39	78, 39	156, 55	100, 52	100, 28	50, 25
Atmos mesh spacing (50N), km	135, 60, 25	71, 36	142, 50	67, 34	64, 18	50, 25
Atmos nominal res (CMIP6)	250, 100, 50	100, 50	250, 50	100, 50	100, 25	50, 25
Atmos model levels (top)	85 (85 km)	91 (0.01 hPa)	91 (78.4 km)	95 (0.01 hPa)	26 (2 hPa)	91 (0.01 hPa)

Table 1: Summary of models and their properties as used in PRIMAVERA project to complete the CMIP6 HighResMIP *highresSST-present* experiments. SISL = semi-implicit, semi-Lagrangian.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

LR-MR-HR / Model	HadGEM3-GC31 **LM, (MM), HM**	EC-Earth3P **LR, HR**	CNRM-CM6-1 **LR, HR**	MPI-ESM1-2 **HR, XR**	CMCC-CM2-(V)HR4 **HR4, VHR4**	ECMWF-IFS **LR, HR**
Lbox	217, (96.7), 40.8	107, 54.2	207, 75.3	134, 66.9	153, 38.2	123, 62.8
Effective resolution (LR, (MR), HR)	590, (330), 135	375, 165	625, 230	605, 190	490, 150	290, 125
Resolution ratio (low/high) using Lbox (Eff resol)	5.32 (4.37)	1.98 (2.2)	2.75 (2.71)	2.0 (3.18)	4.0 (3.2)	1.95 (2.32)
Analysis grid	Native	Regridded 0.7x0.7, 0.35x0.35	Regridded 1.4x1.4, 0.5x0.5	Native	Native	Regridded 1x1, 0.5x0.5

Table 2: Information about model resolutions as used in this study. The effective resolution is taken from Klaver et al. (2019) and derived from examining model kinetic energy spectra, as is the Lbox value (calculated as a weighted grid box distance). Ratio of the low and high model resolution, calculated from both Lbox and the effective resolution. The analysis grid is the grid of the data as published on ESGF and as used for this analysis.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Model	Timestep (min)	Parameter changes (reason)	Parameter values by resolution (low to high)
HadGEM3-GC31 LM, MM, HM	20, 15, 10	USSP launch factor (QBO period)	1.3, (1.2), 1.2
EC-Earth3P LR, HR	45, 15	No changes	
CNRM-CM6-1 LR, HR	15, 15	No changes	
MPI-ESM1-2 HR, XR	3.3, 1.5	Horizontal diffusion damping term (stability)	1.5, 0.5
CMCC-CM2 HR4, VHR4	30, 15	No changes	
ECMWF-IFS LR, HR	30, 20	Autoconversion threshold for rain over ocean RCLCRIT_SEA (net surface energy balance)	2.5×10^{-4}, 2.0×10^{-4}

Table 3: Summary of parameter differences between horizontal resolutions of the PRIMAVERA models used in HighResMIP *highresSST-present* simulations.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Reanalysis	ERA-Interim	MERRA2	JRA55	NCEP-CFSR	ERA5
Model grid (resolution)	TL255 (80 km)	Cubed sphere (50 km)	TL319 (55 km)	T382 (38 km)	TL1279 (31 km)
Assimilation	4D-Var	3D-Var GSI+IAU	4D-Var	3D-Var GSI	4D-Var
Atmos model levels (top)	L60 (0.1 hPa)	L72 (0.01 hPa)	L60 (0.1 hPa)	L64 (0.26 hPa)	L137 (0.01 hPa)
Analysis grid	480x241	576x361	288x145	720x361	1440x720

Table 4: Properties of the reanalysis datasets used in this study. Abbreviations: 4D-Var, 4D variational data assimilation; 3D-Var, 3D variational data assimilation; TL255, triangular truncation 255, with linear grid (approximate horizontal grid spacing in parentheses); L60 60 vertical levels; GSI, Grid-point Statistical Interpolation; IAU, Incremental Analysis Update. Analysis grid is the grid on which the tracking is performed.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Category (CatP)	MSLP range	Official intensity using 1 min sustained wind speed (ms^{-1})
0	>= 994	18-32
1	980 <= x < 994	33-42
2	965 <= x < 980	43-49
3	945 <= x < 965	50-58
4	920 <= x < 945	58-70
5	860 <= x < 920	>70

Table 5: The storm intensity categories (CatPx) as measured by mean sea level pressure (MSLP) ranges as used in this work, together with the official Saffir-Simpson 1 minute sustained wind speed classification.

Model/	Resol	Mean,std	%	%	%	%	%	%

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

mean freq	(nominal, km)	TRACK (*Tempest*)	TS	Cat1P	Cat2P	Cat3P	Cat4P	Cat5P
HadGEM 3-GC31	250	8.5, 2.7	84	12	3	0	0	0
		(1.9, 1.0)	*64*	*28*	*8*	*0*	*0*	*0*
	100	15.1, 4.6	72	21	5	2	0	0
		(9.8, 2.8)	*60*	*31*	*7*	*2*	*0*	*0*
	50	14.8, 3.3	57	24	13	6	0	0
		(16.0, 3.6)	*50*	*28*	*16*	*5*	*0*	*0*
EC-Earth	100	3.3, 2.2	84	12	3	2	0	0
		(0.7, 0.8)	*77*	*16*	*9*	*0*	*0*	*0*
	50	6.0, 3.2	81	6	6	7	0	0
		(2.3, 2.1)	*65*	*14*	*10*	*11*	*0*	*0*
CNRM-CERFACS	250	14.7, 3.5	91	7	2	0	0	0
		(2.9, 2.0)	*80*	*15*	*4*	*1*	*0*	*0*
	50	15.0, 3.1	60	16	12	9	3	0
		(12.6, 3.4)	*42*	*26*	*15*	*13*	*4*	*0*
MPI	100	2.9, 2.7	92	3	2	2	1	0
		(0.6, 0.7)	*86*	*14*	*0*	*0*	*0*	*0*
	50	2.6, 1.6	85	5	3	7	0	0
		(0.7, 1.0)	*87*	*8*	*4*	*0*	*0*	*0*
CMCC	100	3.4, 1.8	75	13	11	1	0	0
		(NA)						
	25	9.4, 3.0	49	21	12	13	5	0
		(NA)						

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

ECMWF	50	7.9, 3.3	78	14	6	2	0	0
		(4.3, 2.5)	*68*	*21*	*8*	*3*	*1*	*0*
	25	10.0, 3.2	69	14	9	7	1	0
		(7.4, 3.2)	*57*	*19*	*15*	*8*	*1*	*0*
Reanalyses	ERA-I	8.7, 3.3	73	16	8	3	0	0
		(5.2, 3.0)	*66*	*24*	*10*	*1*	*0*	*0*
	CFSR	15.5, 4.3	85	10	4	1	0	0
		(7.2, 3.5)	*70*	*22*	*7*	*1*	*0*	*0*
	MERRA2	12.0, 4.9	69	16	13	2	0	0
		(4.7, 2.0)	*60*	*21*	*17*	*2*	*0*	*0*
	JRA55	13.6, 4.0	76	15	8	1	0	0
		(6.0, 3.14)	*60*	*25*	*14*	*1*	*0*	*0*
	ERA5	10.9, 4.1	63	15	12	9	1	0
		(7.0, 3.5)	*46*	*24*	*17*	*11*	*1*	*0*
Obs		11.3 (4.7)	43	23	10	9	10	3

Table 6: Mean tropical cyclone frequency in the North Atlantic basin during May-November 1979-2014. Mean (std) indicates the mean frequency (standard deviation) of storms of all strengths, TS (tropical storm) and Cat 1P-5P show the percentage of this mean value that lies within these pressure-based categories. The mean and std are shown for both TRACK and TempestExtremes (in italics) respectively, where available.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Model	Resol	Frequency corr (all, >= Cat1P)	ACE corr (all, >= Cat1P)	ACE corr (1950-2014)	ACE corr (ensemble mean)
HadGEM3-GC31	LM	**0.48, 0.46**	0.26, 0.26	0.23	**0.54** (14)
	MM	**0.68, 0.59**	**0.46, 0.45**	**0.35**	**0.68** (14)
	HM	0.32**, 0.37**	**0.50, 0.48**	0.29	**0.56** (5)
ECMWF	LR	**0.52, 0.46**	**0.42, 0.40**	0.27	**0.52** (3)
	HR	**0.41,** 0.25	0.30, 0.26	0.34	**0.50** (3)
EC-Earth	LR	0.33, 0.13	0.27, 0.23	0.24	**0.44** (2)
	HR	**0.34**, 0.26	0.28, 0.28	0.25	0.33 (3)
CNRM-CERFACS	LR	**0.5**, **0.4**	**0.49, 0.46**	**0.45**	
	HR	0.26, 0.13	**0.48, 0.45**	0.35	
CMCC	LR	**0.54, 0.45**	0.31, 0.29	0.24	
	HR	**0.51, 0.47**	**0.37, 0.35**	0.30	
MPI-M	LR	**0.33**, 0.12	**0.34**, 0.31	0.26	
	HR	**0.52, 0.43**	**0.38, 0.37**	0.16	
Reanalyses	ERA-I	**0.78, 0.73**	**0.86, 0.85**		
	CFSR	0.32**, 0.35**	**0.86, 0.85**		
	MERRA2	**0.78, 0.66**	**0.87, 0.85**		
	ERA5	**0.83, 0.72**	**0.91, 0.9**		
	JRA55	**0.68, 0.70**	**0.82, 0.82**	**0.82** (1957-2014)	

Table 7: Correlations of Atlantic tropical cyclone interannual variability frequency and ACE_{925} from TRACK against observations, during May-November 1979-2014.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Correlations shown (a,b) are against observed all storms (tropical storm intensity and above), and against observed hurricanes only (>=CatP1). Correlations of ensemble means are shown where available, with the ensemble size as indicated in brackets.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Model/variate correlation	Nino3.4 ACE member 1 (ensemble mean)	AMO member 1 (ensemble mean)	AMM member 1 (ensemble mean)
HadGEM3-GC31			
LM	-0.3 (**-0.55**)	0.28 (0.37)	**0.4** (**0.56**)
MM	**-0.45** (**-0.55**)	0.29 (**0.53**)	**0.38** (**0.70**)
HM	-0.25 (**-0.41**)	**0.41** (**0.41**)	**0.58** (**0.62**)
ECMWF			
LR	-0.26 (**-0.46**)	0.23 (0.34)	**0.43** (**0.56**)
HR	**-0.51** (-0.40)	0.22 (0.37)	0.27 (**0.48**)
EC-Earth			
LR	-0.18 (-0.28)	0.19 (0.32)	0.23 (**0.43**)
HR	-0.03 (-0.19)	0.35 (0.28)	0.35 (0.34)
CNRM			
LR	-0.22	0.27	0.31
HR	-0.27	0.15	0.34
CMCC			
LR	-0.15	0.10	0.26
HR	**-0.41**	**0.41**	**0.42**
MPI			
LR	**-0.40**	0.10	0.25
HR	-0.10	**0.40**	**0.40**

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

ERAI	**-0.42**	**0.56**	**0.64**
MERRA2	**-0.41**	**0.63**	**0.74**
CFSR	**-0.49**	**0.45**	**0.58**
JRA55	**-0.44**	0.39	**0.55**
ERA5	**-0.42**	**0.56**	**0.65**

Table 8: Correlations of the Atlantic tropical cyclone interannual ACE_{925} variability from TRACK for the North Atlantic (May-Nov, 1979-2014) with some potential drivers of that variability (Nino3.4 index, AMO, AMM) for each model-resolution. The ensemble mean correlations (where available) are shown in brackets, ensemble size as in Table 7.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Figure caption list

Fig. 1: Tropical cyclone frequency (mean storms per year during May-November in Northern Hemisphere, and October-May for the Southern Hemisphere, 1979-2014) from models, reanalyses and observations, as diagnosed using the TRACK algorithm. The donut chart is divided into ocean basins, the totals in the centre are (NH, SH) mean storms per year. The thickness of the donut is scaled to the total NH TC observed frequency (i.e. donuts thicker than in panel r indicate more NH TCs while thinner indicate fewer NH TCs.).

Fig. 2: As Fig. 1 but using the TempestExtremes algorithm. Note that the required diagnostics are not available for the CMCC-CM2-(V)HR models.

Fig. 3: Model tropical cyclone track density (storm transits per month per 4 degree cap): for each pair of models, the bias for the higher resolution model, and the difference between higher and lower resolution models, are shown respectively, compared to observations (last plot). The period used is 1979-2014. Note the two reanalyses products (ERA-Interim, MERRA2).

Fig. 4: (a), (b) Ensemble mean of the track density difference between pairs of high and low resolution models using TRACK and TempestExtremes respectively; (c), (d) Ensemble mean of the track density RMSE difference between pairs of high and low resolution models using TRACK and TempestExtremes respectively.

Fig. 5: Scatter plot of the 10 m wind speed vs minimum MSLP of (a) North Atlantic, (b) North Western Pacific and (c) Eastern Pacific tropical cyclones at the peak of 925 hPa wind speed. Each model is indicated (in pairs of lower and higher resolution, dashed and solid lines respectively), together with best-fit curves to all storms

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

(indicated by symbols). Reanalyses from ERA-Interim MERRA2 and JRA55 (in gray), and observations, are also included.

Fig. 6: Joint pdf of the normalised frequency of the MSLP and latitude at peak storm intensity from models, reanalyses and observations for all Northern Hemisphere tropical cyclones over 1979-2014.

Fig. 7: Normalised probability density function of wind speeds at (a) 925 hPa (vmax) and (b) 10 m, taken at the lifetime peak of the tropical cyclone intensity, for models, reanalyses and observations for Northern Hemisphere storms. Dashed lines show the low resolution models and solid lines are high resolution.

Fig. 8: Composite storm structures from (a) lower and (b) higher resolution models, together with ERA-I, JRA55, CFSR and MERRA2 reanalyses, stratified by minimum surface pressure at peak storm intensity. Colour indicates the surface pressure, and contours the tangential velocity at 925 hPa. The dashed contour is 20 ms^{-1} and the solid contours are at 40, 60 ms^{-1}. The numbers on the right are the total number of tropical cyclones over the period, of which the percentage inset indicates how many occur for each category.

Fig. 9: Mean seasonal cycle of tropical cyclone ACE and frequency in the North Atlantic for models and reanalyses (using TRACK) and observations. In each subplot, the gray bars represent the observed monthly mean ACE over the 1950-2014 period, with the solid lines representing the modelled ACE_{925}. The dashed lines show the TC frequency for observations (black) and models. The red line is the lower resolution and the blue line is the higher resolution for each model or reanalysis.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Fig. 10: Correlation of model tropical cyclone frequency (left column) and ACE_{925} (right column) for the North Atlantic (NA), NW Pacific (WP) and NE Pacific (EP) over 1979-2014 against observations for ensembles of HadGEM3-GC31 simulations (a total of 14 members at both MM (100 km) resolution and LM (250 km), and 5 members at HM (50 km) resolution). For each combination of n ensemble members (x axis), a box and whiskers are plotted (the box showing the lower to upper quartile range, with a line at the median, while the whiskers show the range of the data). The mean correlations for each n ensemble member correlation are joined up by the line. The solid lines are for TRACK and the dashed lines for TempestExtremes. The solid and dashed black lines are approximations of the 95% and 99% confidence levels (assuming each of the 36 years are independent samples).

Fig. 11: Correlation of TRACK ACE_{925} from models and reanalyses for North Atlantic tropical cyclone variability against observed ACE as a function of time, using a moving 30 year period centred on the year shown. The dashed lines are for lower resolution, and solid lines for higher resolution models and reanalyses. The -ENS lines are for up to 3 member ensemble means from the available models.

Fig. 12: Wind shear between 850 and 250 hPa for models and reanalyses. Mean over July-October 1980-2013. The dashed line shows 10 ms^{-1}, and the dotted line 20 ms^{-1}.

Fig. 13: (top) African Easterly Jet mean latitude in Aug-Sep for each model and reanalysis over 1980-2014; (middle) Mean number of African Easterly Waves over May-Oct for each model, counted at 15°W using the algorithm described in Bain et al. 2014; (bottom) AEW vorticity at 15°W using the algorithm described in Bain et al 2014.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

Figures

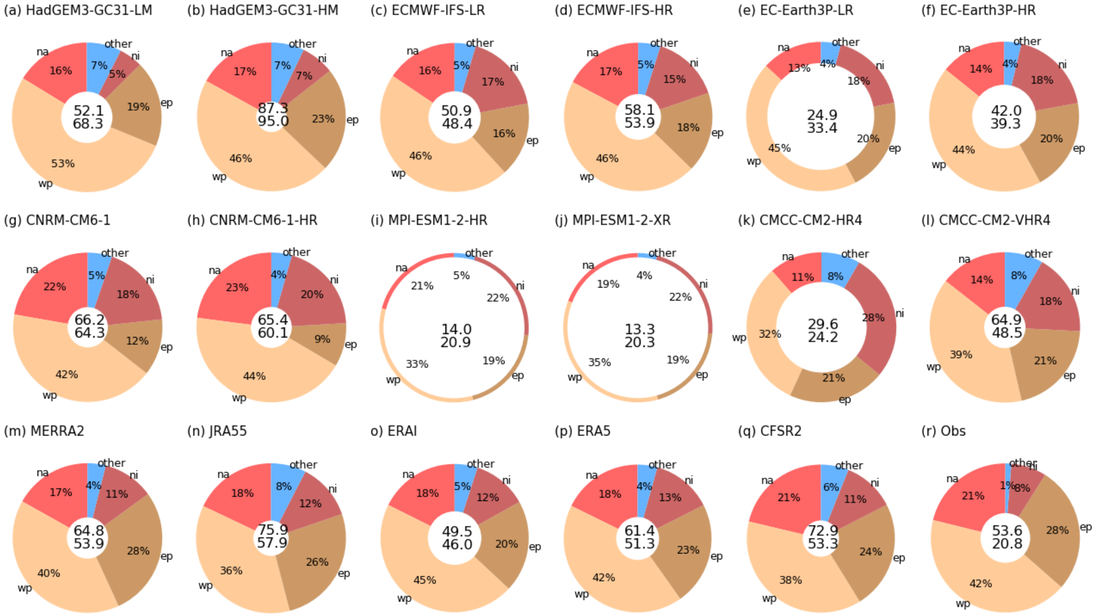


Fig. 1: Northern Hemisphere tropical cyclone frequency (mean storms per year during May-November, 1979-2014) from models, reanalyses and observations, as diagnosed using the TRACK algorithm. The donut chart is divided into NH ocean basins, the totals in the centre are (NH, SH) mean storms per year (Southern Hemisphere uses October-May period). The thickness of the donut is scaled to the total NH TC observed frequency (i.e. donuts thicker than in panel r indicate more NH TCs while thinner indicate fewer NH TCs.).

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

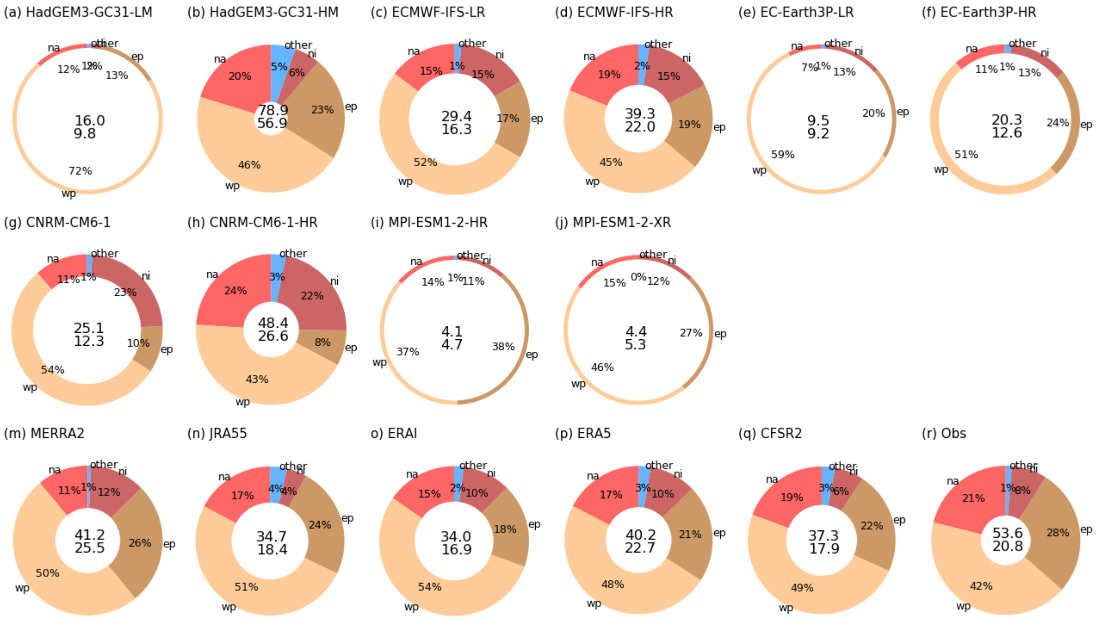


Fig. 2: As Fig. 1 but using the TempestExtremes algorithm. Note that the required diagnostics are not available for the CMCC-CM2-(V)HR models.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

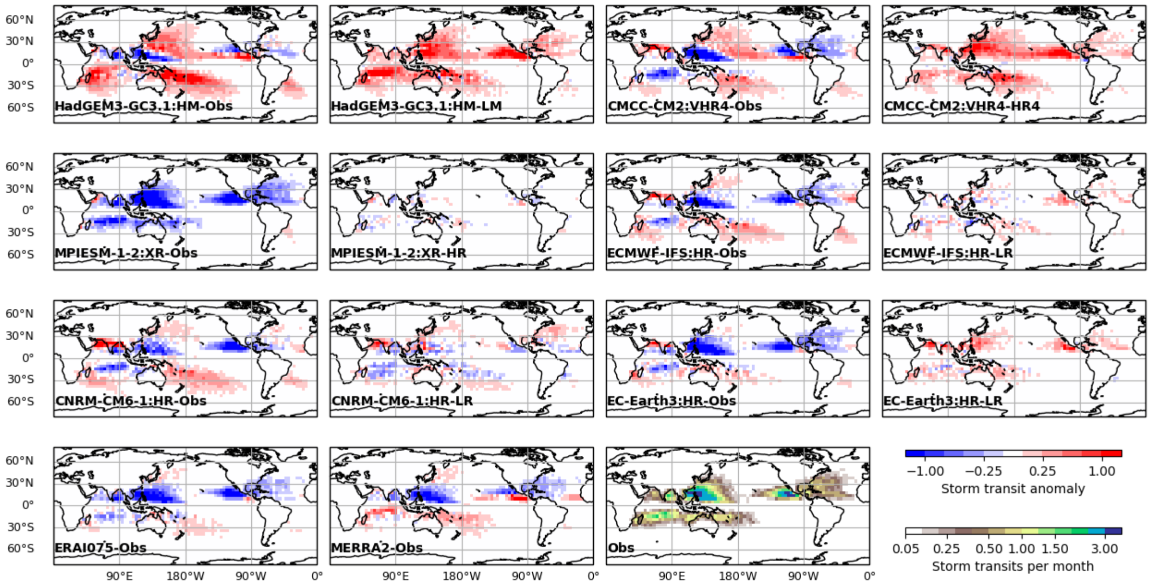


Fig. 3: Model tropical cyclone track density (storm transits per month per 4 degree cap): for each pair of models, the bias for the higher resolution model, and the difference between higher and lower resolution models, are shown respectively, compared to observations (last plot). The period used is 1979-2014. Note the two reanalyses products (ERA-Interim, MERRA2).

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

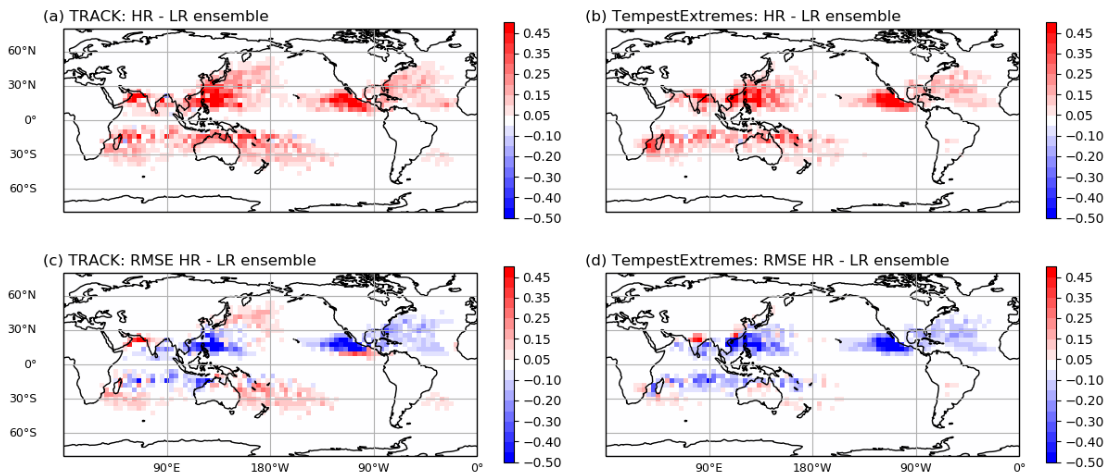


Fig. 4: (a), (b) Ensemble mean of the track density difference between pairs of high and low resolution models using TRACK and TempestExtremes respectively; (c), (d) Ensemble mean of the track density RMSE difference between pairs of high and low resolution models using TRACK and TempestExtremes respectively.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

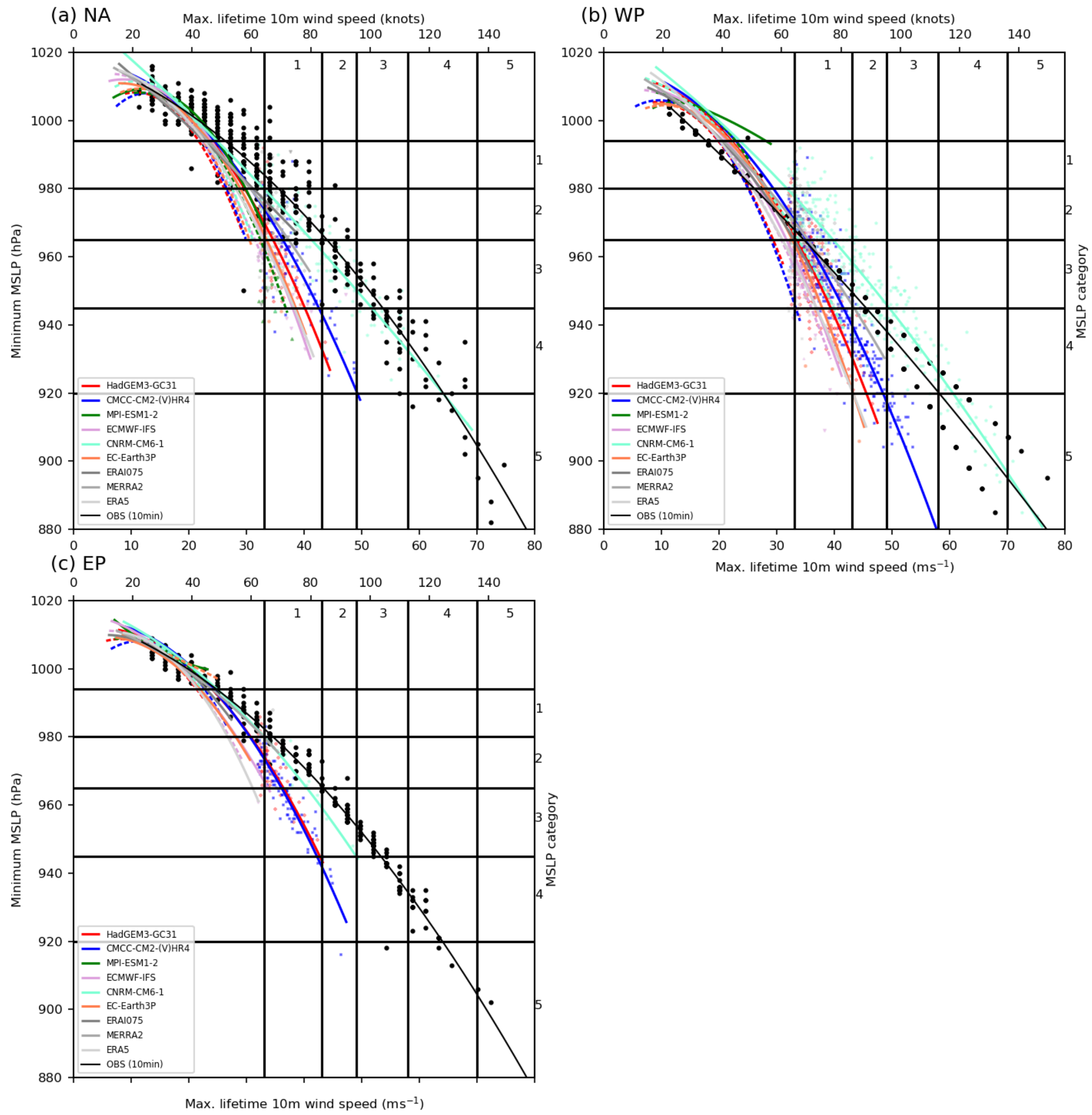


Fig. 5: Scatter plot of the 10 m wind speed vs minimum MSLP of (a) North Atlantic, (b) North Western Pacific and (c) Eastern Pacific tropical cyclones at the peak of 925 hPa wind speed. Each model is indicated (in pairs of lower and higher resolution, dashed and solid lines respectively), together with best-fit curves to all storms (indicated by symbols). Reanalyses from ERA-Interim, MERRA2 and ERA5 (in gray), and observations, are also included. For clarify the model scatter points have not been shown at the lower wind speeds.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

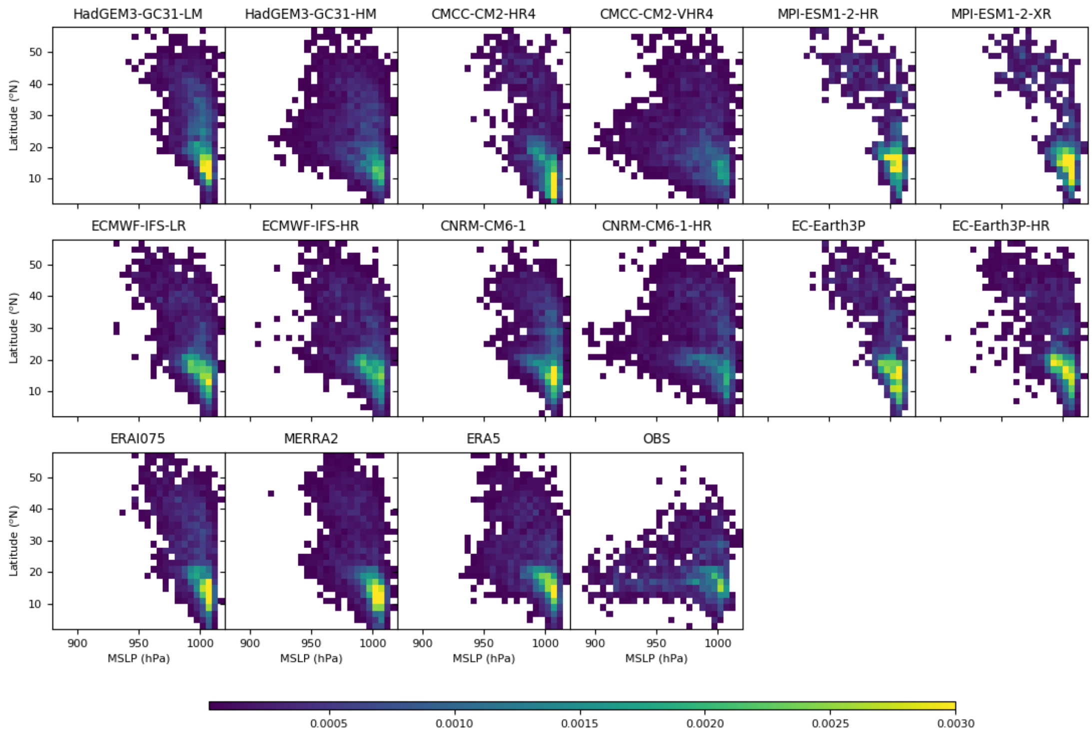


Fig. 6: Joint pdf of the normalised frequency of the MSLP and latitude at peak storm intensity from models, reanalyses and observations for all Northern Hemisphere tropical cyclones over 1979-2014.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

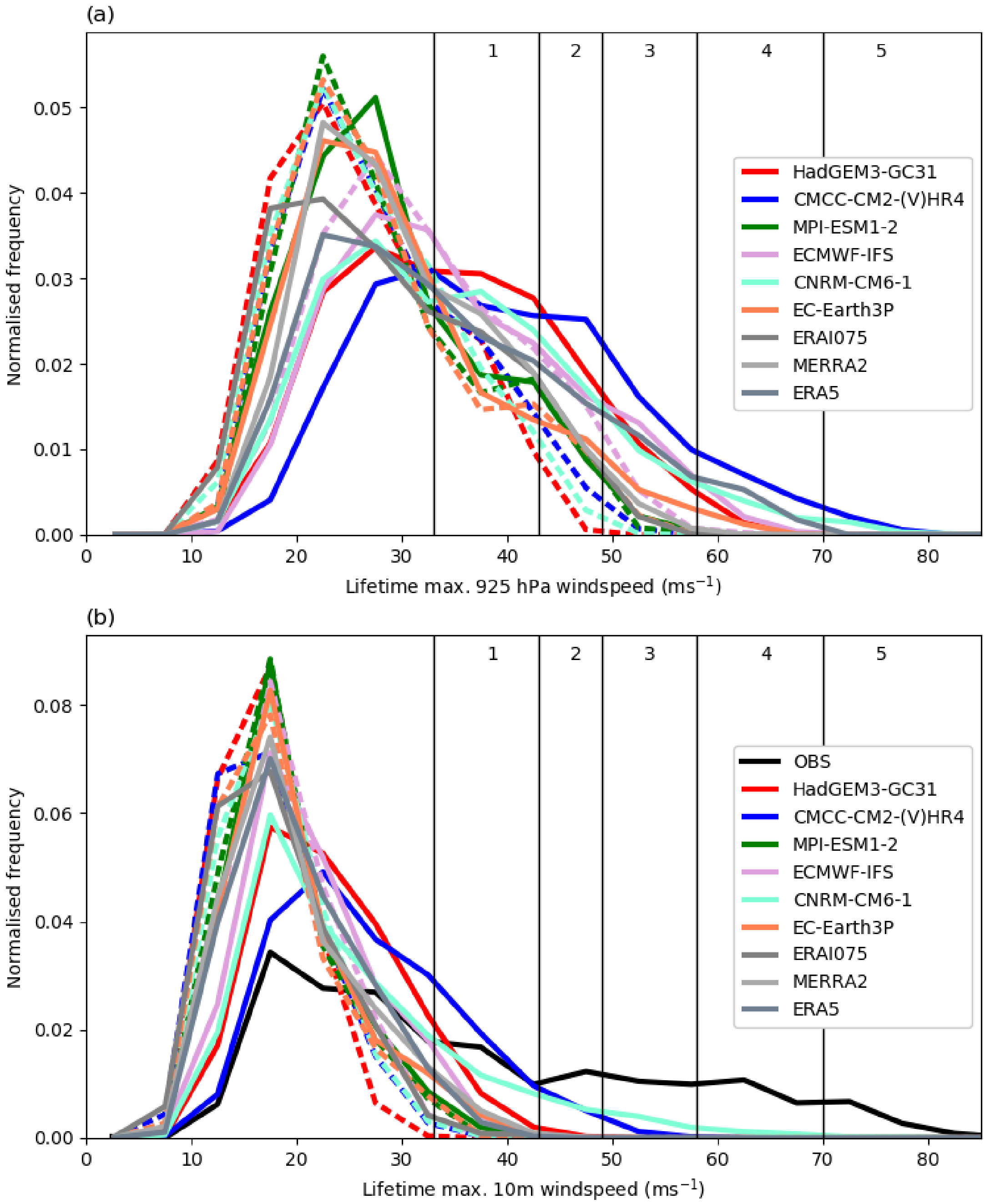
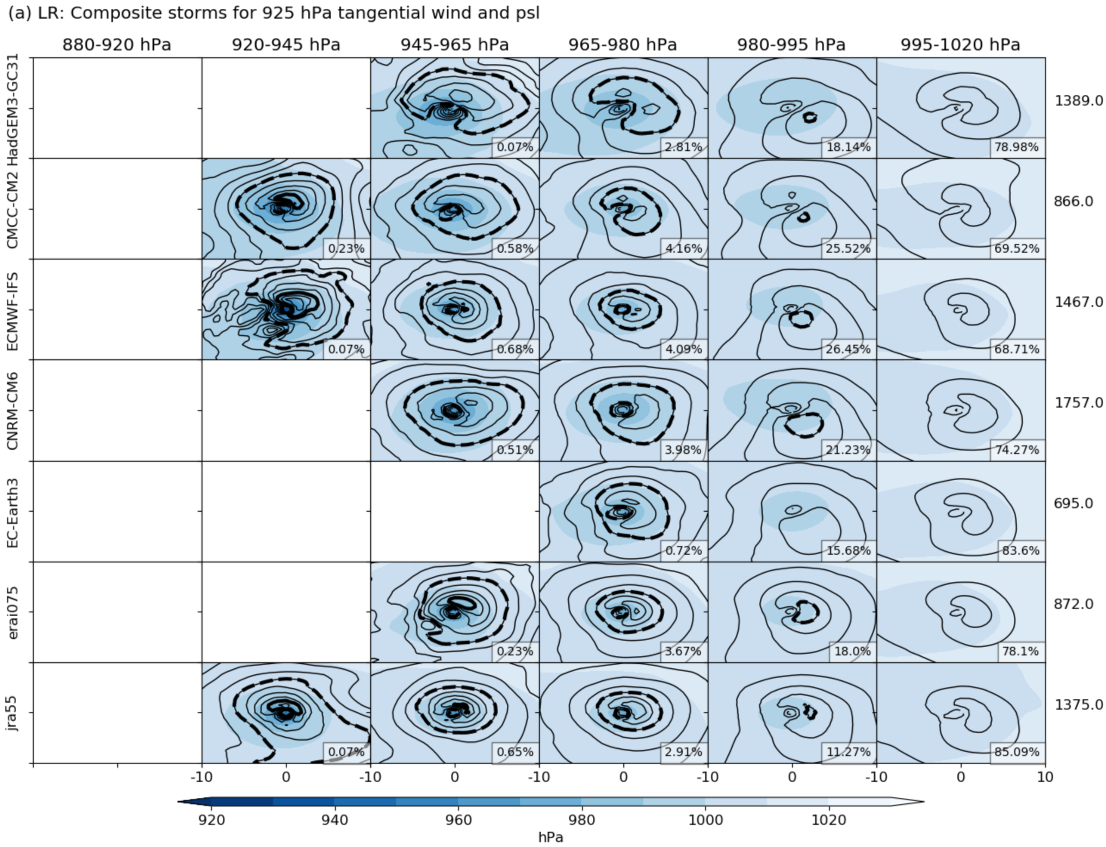


Fig. 7: Normalised probability density function of wind speeds at (a) 925 hPa (vmax) and (b) 10 m, taken at the lifetime peak of the tropical cyclone intensity, for models, reanalyses and observations for Northern Hemisphere storms. Dashed lines show the low resolution models and solid lines are high resolution.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.



Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

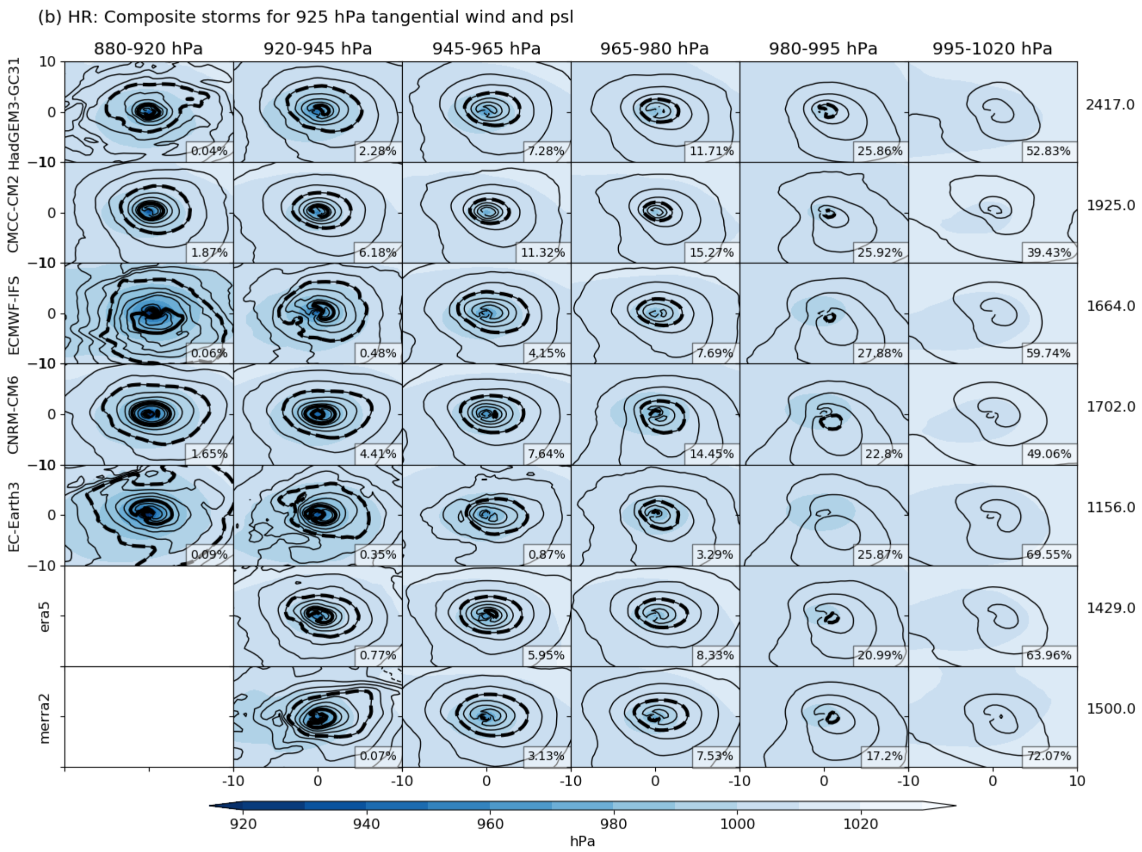


Fig. 8: Composite storm structures from (a) lower and (b) higher resolution models, together with ERA-I, JRA55, CFSR and MERRA2 reanalyses, stratified by minimum surface pressure at peak storm intensity. Colour indicates the surface pressure, and contours the tangential velocity at 925 hPa. The dashed contour is 20 ms^{-1} and the solid contours are at 40, 60 ms^{-1}. The numbers on the right are the total number of tropical cyclones over the period, of which the percentage inset indicates how many occur for each category.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

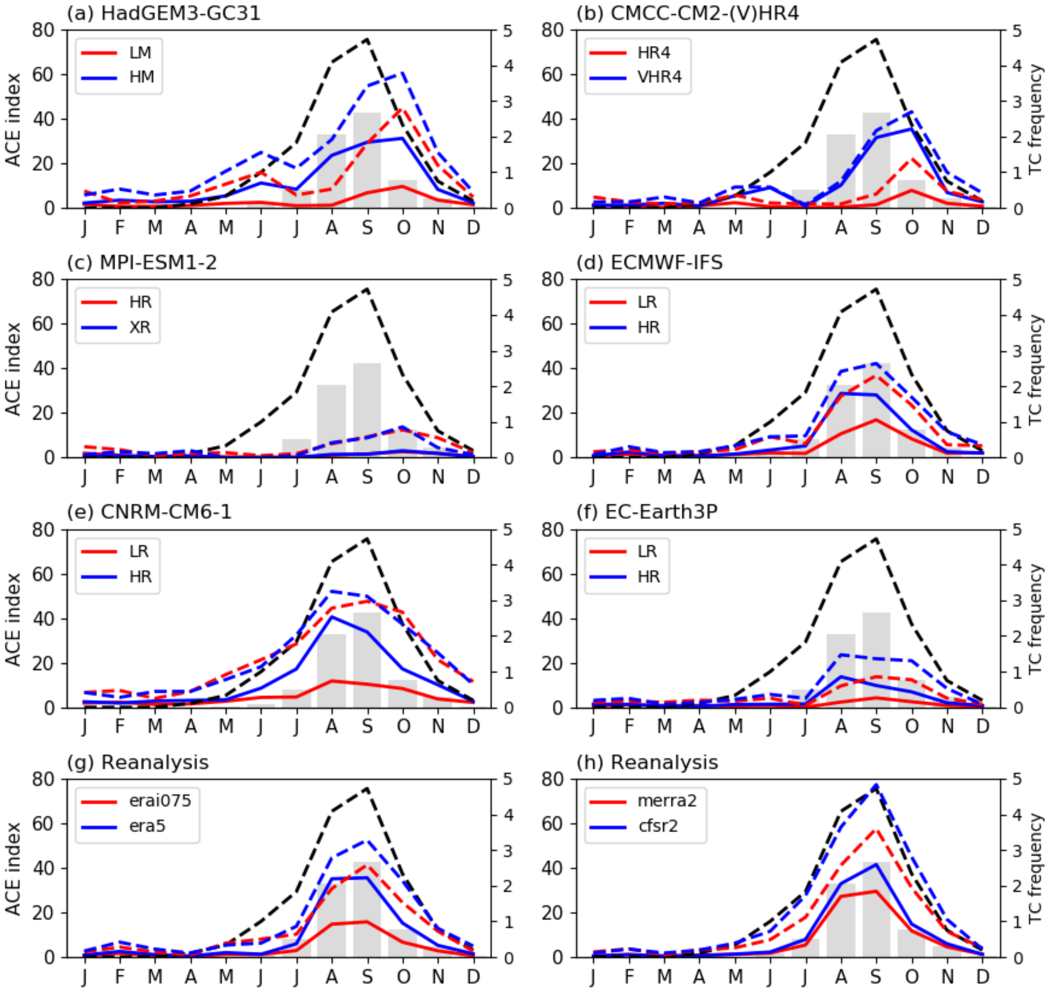


Fig. 9: Mean seasonal cycle of tropical cyclone ACE and frequency in the North Atlantic for models and reanalyses (using TRACK) and observations. In each subplot, the gray bars represent the observed monthly mean ACE over the 1979-2014 period, with the solid lines representing the modelled ACE_{925}. The dashed lines show the TC frequency for observations (black) and models. The red line is the lower resolution and the blue line is the higher resolution for each model or reanalysis.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

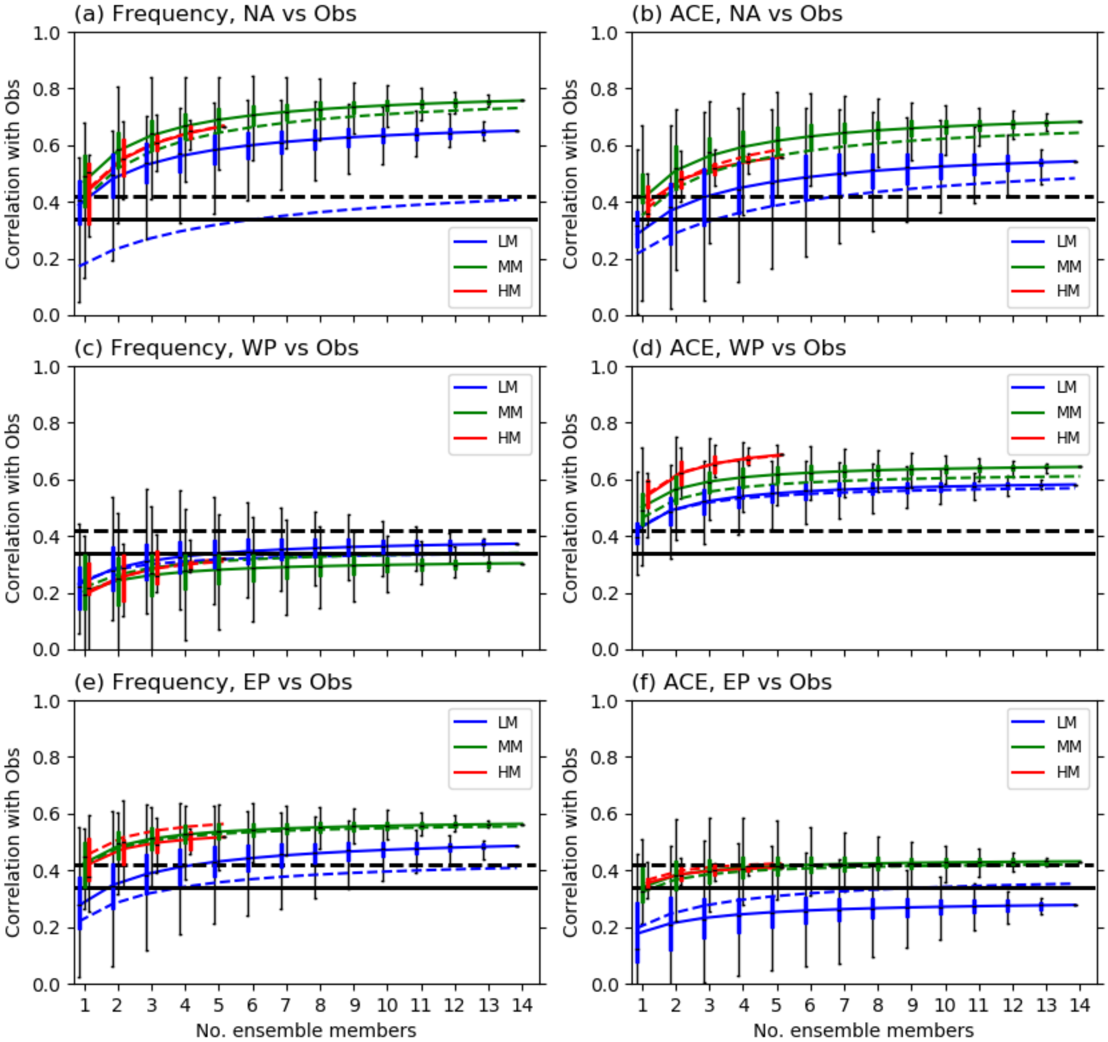


Fig. 10: Correlation of model tropical cyclone frequency (left column) and ACE_{925} (right column) for the North Atlantic (NA), NW Pacific (WP) and NE Pacific (EP) over 1979-2014 against observations for ensembles of HadGEM3-GC31 simulations (a total of 14 members at both MM (100 km) resolution and LM (250 km), and 5 members at HM (50 km) resolution). For each combination of n ensemble members (x axis), a box and whiskers are plotted (the box showing the lower to upper quartile

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

range, with a line at the median, while the whiskers show the range of the data). The mean correlations for each n ensemble member correlation are joined up by the line. The solid lines are for TRACK and the dashed lines for TempestExtremes. The solid and dashed black lines are approximations of the 95% and 99% confidence levels (assuming each of the 36 years are independent samples).

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.

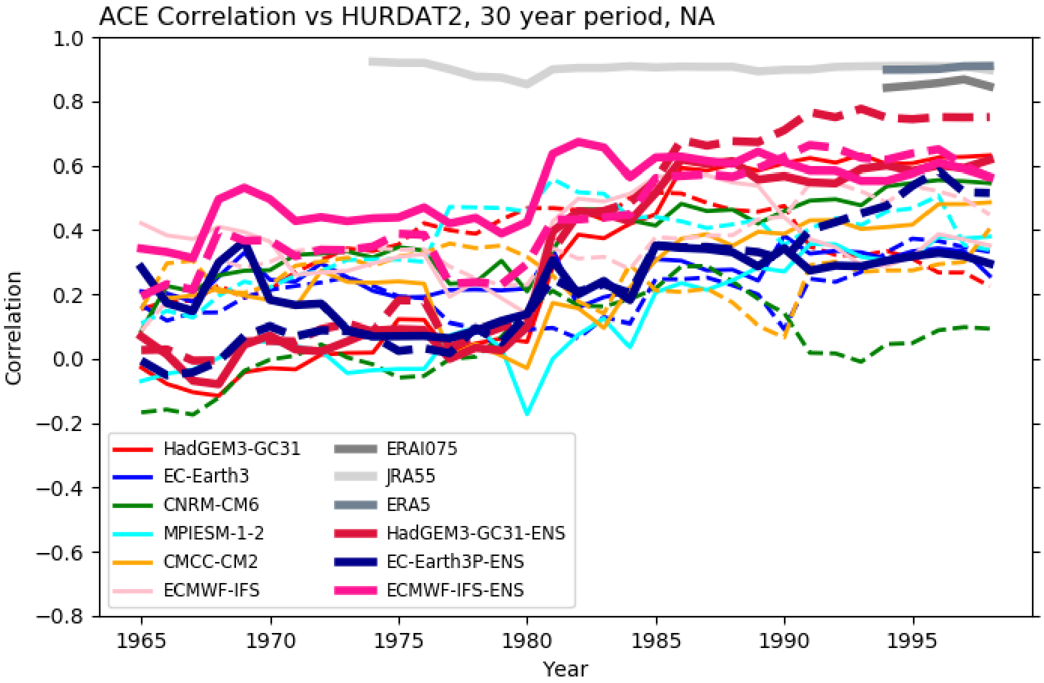


Fig. 11: Correlation of TRACK ACE_{925} from models and reanalyses for North Atlantic tropical cyclone variability against Observed ACE as a function of time, using a moving 30 year period centred on the year shown. The dashed lines are for lower resolution, and solid lines for higher resolution models and reanalyses. The -ENS lines are for up to 3 member ensemble means from the available models.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

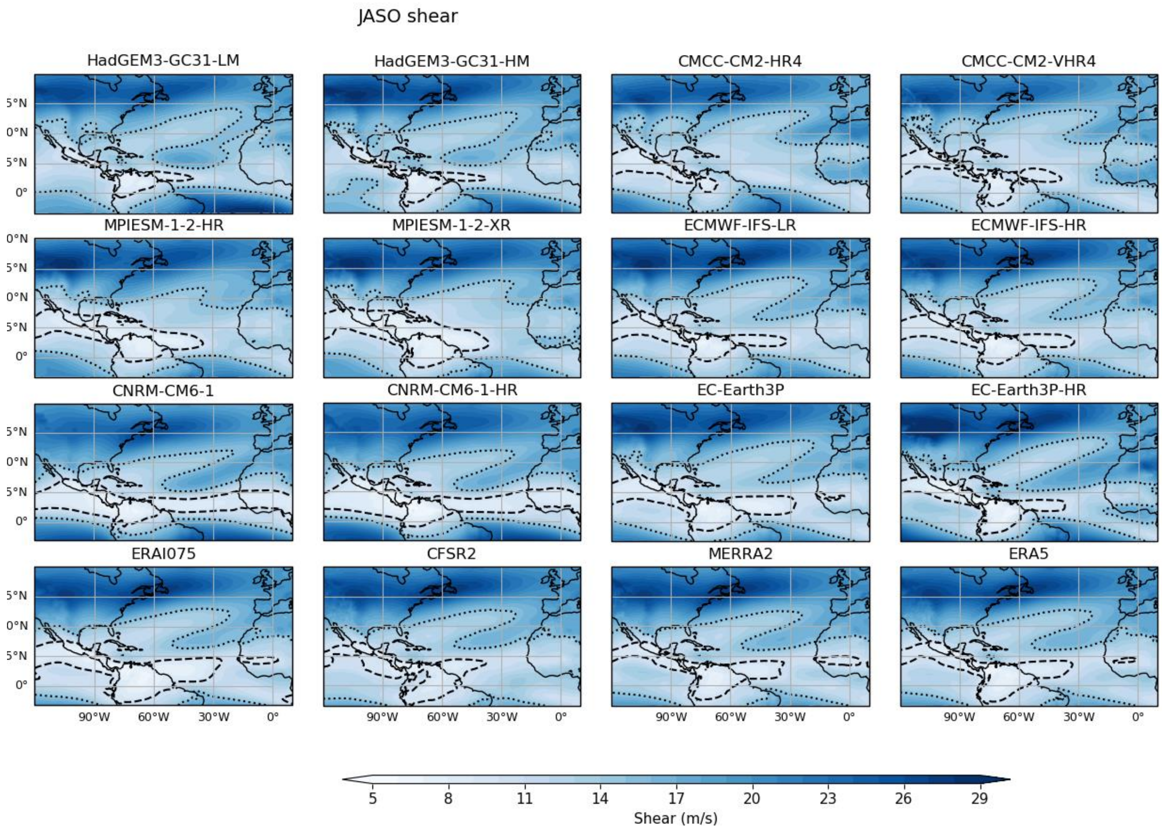
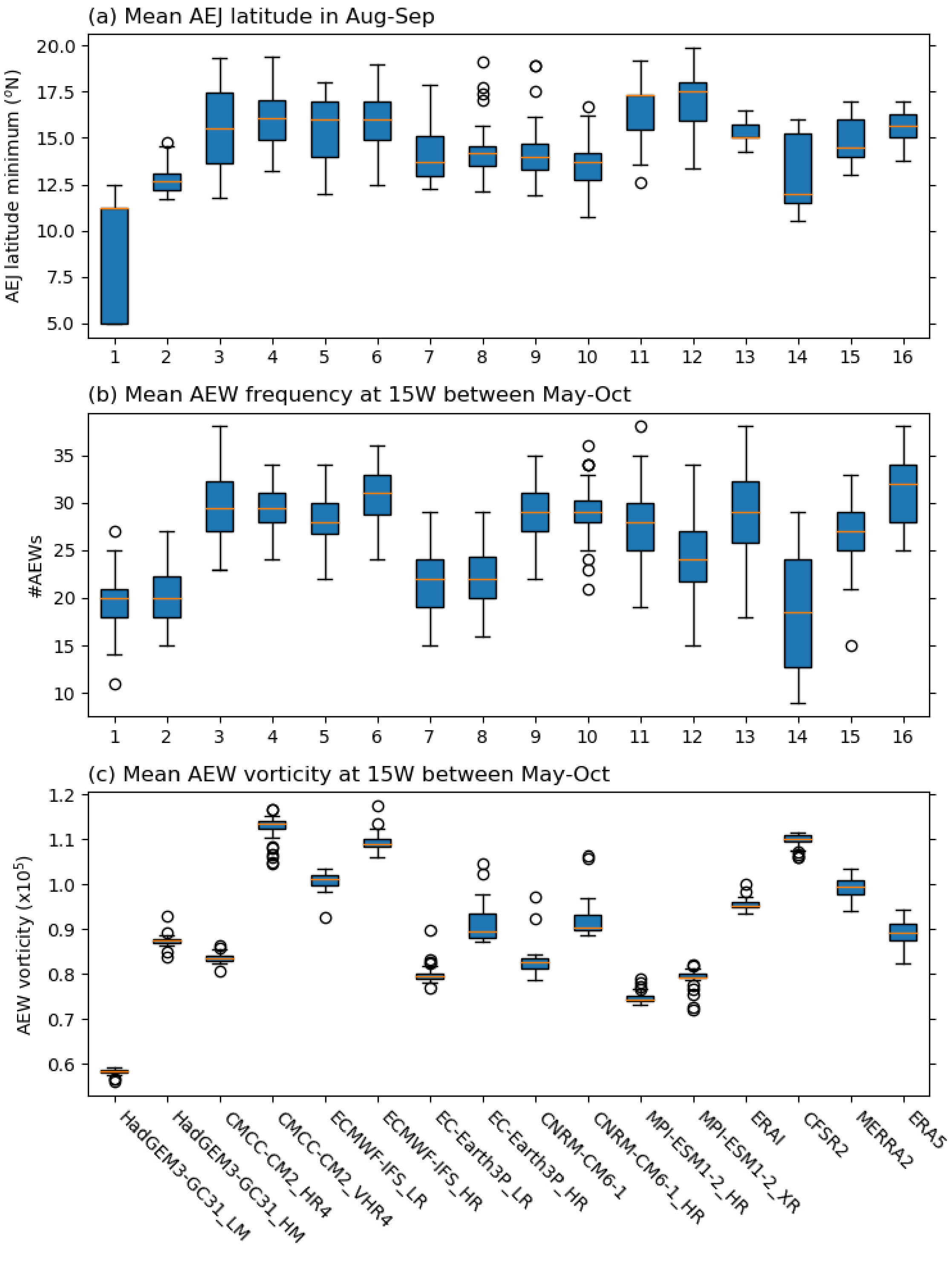


Fig. 12: Wind shear between 850 and 250 hPa for models and reanalyses. Mean over July-October 1980-2013. The dashed line shows 10 ms^{-1}, and the dotted line 20 ms^{-1}.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.



1401

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-19-0639.1.

Fig. 13: (a) African Easterly Jet mean latitude in Aug-Sep for each model and reanalysis over 1980-2014; (b) Mean number of African Easterly Waves over May-Oct for each model, counted at 15°W using the algorithm described in Bain et al. 2014; (c) AEW vorticity at 15°W using the algorithm described in Bain et al 2014.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-0639.1.