"Future changes in the Intensity and Duration of Marine Heat and Cold Waves: Insights from Coupled Model Initial-Condition Large Ensembles"

Clara Deser[1,*], Adam S. Phillips[1], Michael. A. Alexander[2], Dillon J. Amaya[2], Antonietta Capotondi[2,3], Michael G. Jacox[2,4,5] and James D. Scott[2,3]

1 National Center for Atmospheric Research, Boulder CO

2 NOAA Physical Sciences Laboratory, Boulder CO

3 CIRES, University of Colorado, Boulder, CO

4 NOAA Southwest Fisheries Science Center, Monterey, CA

5 University of California Santa Cruz, Santa Cruz, CA

Submitted to *Journal of Climate*

May 9, 2023

Revised October 8, 2023

* Corresponding author: Dr. Clara Deser, Climate and Global Dynamics Laboratory, NCAR, 1850 Table Mesa Drive, Boulder CO 80305, cdeser@ucar.edu

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-23-0278.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Abstract

The future evolution of sea surface temperature (SST) extremes is of great concern, not only for the health of marine ecosystems and sustainability of commercial fisheries, but also for precipitation extremes fueled by moisture evaporated from the ocean. This study examines the projected influence of anthropogenic climate change on the intensity and duration of monthly SST extremes, hereafter termed marine heatwaves (MHWs) and marine coldwaves (MCWs), based on Initial-condition Large Ensembles with seven Earth System Models. The large number of simulations (30-100) with each model allows for robust quantification of future changes in both the mean state and variability in each model.

In general, models indicate that future changes in variability will cause MHW and MCW events to intensify in the northern extra-tropics and weaken in the tropics and Southern Ocean, and to shorten in duration in many areas. These changes are generally symmetric between MHWs and MCWs, except for the longitude of duration change in the tropical Pacific and sign of duration change in the Arctic. Projected changes in ENSO account for a large fraction of the variability-induced changes in MHW and MCW characteristics in each model and are responsible for much of the inter-model spread as a result of differences in future ENSO behavior. The variability-related changes in MHW and MCW characteristics noted above are superimposed upon large mean-state changes. Indeed, their contribution to the total change in SST during MHW and MCW events is generally < 10% except in polar regions where they contribute upwards of 50%.

1. Introduction

Prolonged sea surface temperature (SST) extremes, henceforth referred to as "marine heatwaves" (MHWs) and "marine coldwaves" (MCWs), severely impact the health of aquatic ecosystems and have widespread economic repercussions for commercial fisheries (see the recent review by Oliver et al. 2021 and references therein; Smith et al. 2023). Such events last from weeks to months, and often recur in consecutive years over a given region, compounding their physical and biogeochemical effects. These SST extremes may also be accompanied by temperature anomalies at depth (e.g., Scannell et al. 2020; Amaya et al. 2023a; Schaeffer et al. 2023). As climate change accelerates due to human emissions of carbon dioxide and other greenhouse gases, the oceans will continue to warm, thereby placing additional stress on temperature-dependent

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

physiological processes in marine organisms (e.g., Alexander et al. 2018; Smale et al. 2019; Cheung and Frolicher, 2020).

Observational studies have documented the geographical extent, intensity, duration and frequency of MHWs based on daily SST data from satellite records (1982-present), augmented with monthly SSTs from ship-based archives going back to the early 20th century (Hobday 2016; Holbrook et al. 2019; Oliver et al. 2021 and references therein). These studies find that historical MHW thresholds have been increasingly exceeded in recent years due to an underlying warming trend attributable to anthropogenic influences (Oliver et al. 2018, 2019, and 2021; Frolicher et al. 2018; Xu et al. 2022; Frolicher and Laufkotter, 2018; Laufkotter et al. 2020; Thoral et al. 2022). Modeling studies have also highlighted the role of long-term anthropogenic warming on recent and projected trends in MHW historical threshold exceedance (Frolicher et al. 2018; Alexander et al. 2018; Yao et al. 2022; Guo et al. 2022).

However, anthropogenic effects are not limited to mean state changes alone. As global warming progresses, the upper ocean is expected to become increasingly stratified and the mixed layer is anticipated to shoal (e.g., Capotondi et al. 2012; Kwiatkowski et al. 2020). A shallower mixed layer will promote increased variability and decreased persistence of SST anomalies due to its lower thermal inertia, all other factors being equal. According to this heuristic paradigm, anthropogenic effects on long-term trends in SST extremes (e.g., MHWs and MCWs) can result from both shifts in the mean (e.g., "a rising tide lifts all ships") and from changes in the shape of the full SST distribution as measured by the variance, skewness and kurtosis. In addition, atmospheric circulation, air-sea exchange of heat and momentum, oceanic thermal advection and turbulent mixing, and remote impacts from changing El Niño – Southern Oscillation (ENSO) teleconnections, are all likely to be altered under anthropogenic climate change, further complicating the drivers of SST extremes including MHWs and MCWs (Alexander et al. 2018; Oliver et al. 2018; Sen Gupta et al. 2020; Amaya et al. 2021; Li and Thompson 2021; Oliver et al., 2021; Xu et al. 2022; Shi et al. 2022; Vogt et al. 2022; O'Brien and Deser, 2023). Thus, the physical mechanisms underlying observed and simulated changes in MHWs and MCWs are challenging to quantify due to their complexity, varied geographical and seasonal expression, and lack of adequate data. Only very recently have attempts been made to quantitatively assess the various factors contributing to changes in historical and future MHW characteristics (e.g., Oliver et al. 2021; Amaya et al. 2021; Vogt et al. 2022).

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

Distinguishing between changes in the mean state and changes in variability on the evolving characteristics of marine temperature extremes is important not only from the standpoint of physical understanding, but also for prediction and resource management purposes (Oliver et al. 2019; Amaya et al. 2023b). However, attributing changes in variability to anthropogenic effects requires large amounts of data to refute the null hypothesis that they are due to random sampling fluctuations. An innovative study by Xu et al. (2022) employed an empirically-based dynamical linear inverse model to generate thousands of synthetic realizations of SST variability consistent with the spatio-temporal statistics of the actual observed record over the past 60 years. With so many realizations, they were able to circumvent sampling issues and successfully isolate forced changes in variability from forced changes in mean state on the historical evolution of cold and warm SST extremes. They found that forced changes in the mean state dominated those in variability; a similar conclusion was reached by Oliver (2019) based on an auto-regressive model fit to observations.

Studies of future changes in MHW characteristics have generally relied on the collection of CMIP5 and CMIP6 models, using one simulation per model to construct a multi-model average (e.g., Frolicher et al. 2018; Plecha and Soares, 2020; Qiu et al. 2021; Oliver et al. 2021; Yao et al. 2022). While this approach provides an overall assessment, it confounds model structural uncertainty (e.g., model-dependent forced responses in both the mean state and variability) and sampling uncertainty. In contrast, "Initial-condition Large Ensemble" simulations (hereafter referred to as "Large Ensembles" or LEs) allow for a robust determination of the evolving forced responses in any single model at each location and time (e.g., Deser et al. 2020). This is made possible by conducting a large number of simulations (generally 30 or more) with a given model under a given radiative forcing protocol, where each simulation starts from a different initial condition. As the memory of the initial condition fades, the simulations diverge from each other due to unpredictable internally-generated variability. Thus, each ensemble member has its own unique sequence of internal fluctuations superimposed upon an evolving forced response that is common to all of the ensemble members. With a sufficient ensemble size, the evolving forced response can be isolated by averaging across the ensemble members at each time step and at each location. Importantly, the forced response contains not only evolving changes in the mean background climate, but also any evolving changes in the statistical characteristics of internal variability, for example amplitude, autocorrelation and spatial structure. The power of LEs has yet

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

to be fully exploited for the study of marine temperature extremes and their projected changes. Frolicher et al. (2018) documented future changes in MHW frequency, duration, spatial extent and intensity based on a 10-member LE, while Alexander et al. (2018) used a 30-member LE to study changes in SST extremes for Large Marine Ecosystem regions within the North Pacific and Atlantic. More recently, Burger et al. (2022) quantified compound MHW and ocean acidity extremes under global warming using a 30-member LE.

Here, we make use of seven different model LEs, each of which contains between 30-100 simulations spanning the period 1950-2100, to isolate future changes in MHW and MCW intensity and duration due to forced changes in variability *vs*. mean state in each model separately. We further examine the evolving influence of forced changes in ENSO variability on forced changes in MHW and MCW characteristics by sub-sampling the LEs during ENSO-neutral states. While this sub-sampling procedure necessarily reduces the number of MHW and MCW events for analysis, the large number of realizations in any given model LE ensures that we maintain an adequate number of events for robust results. While previous studies have shown a connection between ENSO events and MHW events (e.g., Oliver et al. 2018; Holbrook et al. 2019; Xu et al. 2021; Capotondi et al. 2022), none has explicitly assessed the influence of ENSO in this way using LEs.

Finally, we analyze cold and warm extremes separately so as not to build-in any assumptions regarding linearity with respect to sign. Previous studies have generally focused exclusively on MHWs (e.g., Oliver et al. 2018; Frolicher et al. 2018; Frolicher and Laufkotter, 2018; Laufkotter et al. 2020; Guo et al. 2022) or on the symmetric component of SST variability in general (Xu et al. 2022; Li and Thompson 2021; Shi et al. 2022). Schlegel et al. (2021) provide a global synthesis of observed MCW events (termed "marine cold spells" in their study) based on daily SSTs over the satellite era (since 1982), highlighting their physical characteristics, ecological impacts and recent trends. Metrics of observed MCW and MHW events based on daily satellite SSTs are compared in Wang et al. (2022) and Yao et al. (2022). The latter study also compared future changes in MCW and MHW properties using daily SSTs from the CMIP6 archive; however, their multi-model mean assessment is based on averaging one simulation per model and thus does not separate the effects of variability-induced changes from mean-state changes.

Our study is based on monthly SST data rather than 5-day running means as in the original MHW definition of Hobday et al. (2016). The use of monthly data can be justified based on the

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

relatively slow decorrelation timescale of SST anomalies and the outsized ecological impacts from long-lasting marine temperature extremes (Jacox et al. 2020; Amaya et al. 2023b). However, shorter-lived MHW and MCW events will be missed in our analyses. We encourage future studies to make use of daily SST (and subsurface temperature) output from model LEs for added insight (see Le Grix et al., 2022 for a recent example).

The rest of this study is organized as follows. The model LEs and our methods for defining MHW and MCW events, separating forced changes due to mean state *vs.* variability, and classifying ENSO-neutral states are described in Section 2. Results are presented in Section 3, beginning with an evaluation of model biases in simulating present-day MHW and MCW intensity and duration (Section 3.1), followed by an analysis of models' future changes in MHW and MCW characteristics (Section 3.2) and the role of ENSO (Section 3.3), and ending with an assessment of the relative contributions from mean-state changes *vs.* changes in variability (Section 3.4). A summary and discussion are provided in Section 4.

2. Data and methods

2.1. Data from model LEs and observations

We analyze seven global coupled model LEs with ensemble sizes ranging from 30 to 100 spanning the years 1950-2100 under historical and future radiative forcing (see Table 1 for model and forcing details). We chose all available models that had at least 30 ensemble members for the entirety of the analysis period 1950-2100 (a full listing of model LEs is provided at https://www.cesm.ucar.edu/community-projects/mmlea). Three of the models (CESM1, CanESM2 and MPI-ESM-LR) are CMIP5 generation and the others (CESM2, CanESM5, GFDL-SPEAR and MIROC6) are CMIP6. Thus, our multi-model LE collection encompasses a range of forcing scenarios and associated global warming responses. In addition, it includes two versions of CESM and two of CanESM, allowing for comparison across CMIP generations of the same model. In total, our multi-model LE collection contains 420 individual simulations. All models employed in this study utilize relatively coarse resolution (around 1.0° - 1.5°) in their ocean component and thus do not explicitly resolve mesoscale processes that have been shown to be important for SST variability along western boundary currents, the Antarctic circumpolar current and other eddy-rich regions (Pilo et al. 2019; Hayashida et al. 2020; Guo et al. 2022).

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

Table 1. Model Large Ensembles used in this study and their salient characteristics. Global SST change refers to the difference between 2070-2100 and 1970-2000. MHW and MCW results for CESM2 are not sensitive to differences in biomass burning aerosol emission protocols used for the first and second 50 members.

Model Name **CMIP Generation**	**Duration** **Future radiative forcing scenario** **Global SST change**	**Ensemble size**	**Resolution** Atmosphere Ocean	**MHW/MCW Sample Size** (per grid box per 31 years) **All years** **ENSO-neutral years**	**Reference Paper**
CESM2 (CMIP6)	1850-2100 SSP3-7.0 2.5°C	100	1.3°x0.9° ~1.0°	1553 / 1493 391 / 377	Rodgers et al. (2021)
CanESM5 (CMIP6)	1850-2100 SSP5-8.5 4.0°C	50	2.8°x2.8° 1.4x0.9°	816 / 793 183 / 175	Swart et al. (2019)
GFDL-SPEAR (CMIP6)	1921-2100 SSP5-8.5 2.7°C	30	50km ~1.0°	553/ 546 121/ 113	Delworth et al. (2020)
MIROC6 (CMIP6)	1850-2100 SSP5-8.5 2.0°C	50	1.4°x1.4° ~1.0°	699 / 710 218/ 210	Tatebe et al. (2019)
CESM1 (CMIP5)	1920-2100 RCP8.5 2.6°C	40	1.3°x0.9° ~1.0°	619/ 608 171/ 164	Kay et al. (2015)
CanESM2 (CMIP5)	1950-2100 RCP8.5 3.1°C	50	2.8°x2.8° 1.4x0.9°	847/ 824 189/ 180	Kirchmeier - Young et al. (2017)
MPI-ESM-LR (CMIP5)	1850-2099 RCP8.5 1.9°C	100	1.9°x1.9° ~1.5°	1619/ 1618 412/ 404	Maher et al. (2019)

Our analyses are based on monthly mean SST on the native grid of each model (see Table 1 for details on model resolution). For the models, we use the temperature of the top ocean model level ("tos), which is the true SST, as opposed to surface temperature ("TS") to avoid contamination from sea ice cover. Specifically, for a grid cell that includes fractional sea ice, "tos" is the water temperature directly beneath the sea ice, whereas "TS" is the area-weighted average of the surface temperature of the ice and the surface temperature of the water within the grid cell. To benchmark the models against observations, we make use of SSTs from ERSSTv5 on a 2° latitude/longitude grid (Huang et al. 2017); similar results (not shown) are obtained with HadISSTv1 on a 1° latitude/longitude grid (Rayner et al. 2003). When comparing models against observations, and forming multi-model averages, we have regridded the data to the CESM1 grid.

2.2. Methods

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

a. Model Large Ensembles

As mentioned above, we separate future changes in MHWs and MCWs into two components: 1) changes in variability; and 2) changes in the mean state. To isolate changes in variability, we define the internal component of SST (iSST) in each ensemble member of a given model LE as iSST(x,t,e) = SST(x,t,e) – SST(x,t,em) where x is grid box location, t is time, e is ensemble member, and em is the ensemble-mean. Note that this procedure removes the monthly ensemble-mean climatology from each ensemble member. We then compute the 10th and 90th percentiles of iSST(x) separately for each calendar month based on pooling together all ensemble members of a given model LE. We compute these percentiles for the 71-year historical period 1950-2020 in order to compare with observations, and also for three 31-year periods (1970-2000, 2020-2050 and 2070-2100) in order to evaluate how climate change influences MHWs and MCWs. For example, for the 100-member CESM2 LE, we determine these thresholds for the month of January for the period 1970-2000 from the set of 31x100 iSST samples at each grid box, and repeat this procedure for each calendar month. Similarly, we determine the thresholds as a function of calendar month for the period 2020-2050 and for the period 2070-2100. In this way, we obtain seasonally-varying thresholds for each period separately. Then, for each period, we define a MHW event if iSST(x,t,e) reaches or exceeds the 90th percentile; similarly, we define a MCW event if iSST(x,t,e) is equal to or less than the 10th percentile. We define the intensity of each MHW and MCW by its corresponding monthly iSST value (°C), and the duration of each MHW (MCW) event by determining the number of consecutive months with MHW (MCW) conditions at a given location. We then form MHW and MCW composites of intensity (°C) and duration (months) based on averaging all MHW and MCW events for each period separately. To give an idea of the sample sizes for these composites, the 100-member CESM2 LE yields 3720 (12 months x 31 years x 100 members x 0.1) monthly MHW samples (and MCW samples) in any given 31-year period at each grid box. The number of discrete MHW and MCW events (e.g., taking into account the lifetime of each event) depends on the duration characteristics, which vary spatially. For the CESM2 LE, each grid box contains, on average, 1553 discrete MHW events and 1493 discrete MCW events in any given 31-year period. Sample size information for all seven model LEs is provided in Table 1 and spatial maps of the number of discrete MHW and MCW events are shown in Fig. S1. The large sample sizes afforded by each model LE allow for a robust determination of MHW and MCW characteristics and any future changes thereof.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

We have repeated our analyses using the 5th and 95th percentile thresholds to define MCW and MHW events, respectively, and find qualitatively similar results to those based on the 10th and 90th percentile thresholds. Thus, we adopt the 10th and 90th percentile thresholds in order to provide a robust comparison with observations for which the 5th and 95th percentile thresholds would be too restrictive (i.e., result in too few samples; see also Jacox et al. 2020). In addition, the 10th and 90th percentile thresholds allow us to robustly quantify future changes in MHW and MCW events during the subset of years that are in an ENSO-neutral state (see below).

To assess the influence of ENSO on future changes in MHWs and MCWs, we have repeated our analyses for ENSO-neutral states as follows. First, we compute the leading Empirical Orthogonal Function (EOF1) of 3-month running means of iSST(x,t) over the tropical Pacific domain (10ºN-10ºS, 120ºE-85ºW) for each time period separately (1970-2000, 2020-2050 and 2070-2100), using the pooled set of ensemble members for a given model LE. [Note that the use of EOF1 as opposed to a fixed regional index such as Nino3.4 SST ensures that the pattern of ENSO-related iSST specific to each model is accommodated.] We then identify the 30th and 70th percentile values of the associated principal component (PC1) for each month separately. If the PC1 value in a given ensemble member e at time t lies within the 30th- 70th percentile range, and if the PC1 values in each of the preceding 5 months also fall within this range, then we define the corresponding iSST(x,t,e) as being in an "ENSO-neutral" state [we choose 5 months as a conservative estimate of the cumulative lagged response of extra-tropical SST anomalies to ENSO: see Alexander et al. 2002; however, a choice of 3 months yields very similar results (not shown)]. We repeat this procedure for all time steps and ensemble members. We then determine the 10th and 90th percentile thresholds of iSST(x) for each month and time period from this set of "ENSO-neutral" samples, and then use these thresholds to construct MHW and MCW composites from the ENSO-neutral samples. Sample size information for the ENSO-neutral composites in each model LE is provided in Table 1 and Fig. S2. We have checked that the results do not change appreciably if we consider both PC1 and PC2 of Tropical Pacific iSSTs in our ENSO-neutral selection procedure; however, the use of two PCs greatly reduces the number of ENSO-neutral samples available for compositing, resulting in less stable statistics and noisier patterns (not shown). While more (or less) sophisticated methods could be used to define "ENSO-neutral" states, we view our approach as a valuable step in assessing the influence of future changes in Tropical Pacific SST variability on MHW and MCW characteristics.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

b. Observations

We first compute the climatological SST for each month at each grid box based on the period 1950-2020. We then define monthly SST anomalies by subtracting the climatological monthly SST from each corresponding month. We then quadratically detrend the data to produce the estimated observed iSST(x,t), hereafter iSST_Obs(x,t). We then compute the 10th and 90th percentiles of iSST_Obs(x,t) as a function of calendar month, and define MHW and MCW events following the procedure outlined for the model LEs above. We used quadratic detrending in lieu of linear detrending to accommodate the accelerating pace of global warming since 1950 (NOAA National Centers for Environmental Information, 2023).

We have tested whether quadratic detrending effectively removes the forced component by applying this procedure to each member of a given model LE individually and comparing the results against the "true" forced component estimated by the ensemble-mean of the LE. We find that for the period 1950-2020, quadratic detrending produces nearly identical MHW and MCW composites as the method outlined in Section 2.2a for each model LE (not shown), lending confidence to the observational results.

2.2.3. Statistical significance testing

We apply the False Discovery Rate (FDR; Wilks 2016) to a 2-sided t-test at the 95% confidence level to assess whether differences in observed and simulated MHW and MCW composites are statistically significant, and whether simulated MHW and MCW composites in the future (2020-2050 or 2070-2100) differ significantly from those in the historical reference period (1970-2000). We use the number of discrete events in each composite for the t-test sample sizes. Note that the t-statistic (critical t-value) is relatively insensitive to the precise number of samples for samples sizes > 75 (Student, 1917), and that our model composites always contain at least 75 discrete samples at every grid box.

3. Results

3.1. Model validation

We begin by comparing the simulated and observed characteristics of MHW and MCW composites over the 71-year period 1950-2020. For conciseness, we present results from the 100-

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

member CESM2 LE in the main paper, and the remaining six model LEs in the Supporting Information. Figure 1a shows MHW composite intensity (ºC) averaged across all 100 members of the CESM2 LE. The model simulates a maximum in intensity along the equatorial Pacific (values ~ 3.5 °C) and secondary maxima along the western boundary currents and their eastward extensions in the North Pacific and North Atlantic (values ~ 2 °C). Similar features are evident in the observations, but with reduced amplitudes, especially in the equatorial Pacific (Fig. 1b). To assess whether the model is significantly biased in its representation of MHW composite intensity, or whether the differences between the simulated and observed MHW composites arise from sampling uncertainty (the observational record samples only one of many possible realizations), we evaluate whether the observed composite values lie outside the 5^{th}-95^{th} percentile range of the distribution of 100 composite values based on each member of the CESM2 LE. According to this measure, CESM2 shows a significant warm bias in composite MHW intensity over much of the Tropical Indo-Pacific and along Antarctica, and a significant cold bias in the Arctic, temperate portions of the Southern Ocean, and along the west coasts of North America and Africa (Fig. 1c). The amplitudes of these model biases are generally < 0.2 °C except in the western equatorial Pacific, off-equatorial eastern Pacific and along the Siberian coast where they reach 0.4 - 1.0 °C. Similar model biases are found for MCW composite intensity, with an even larger overestimation across the full width of the equatorial Pacific basin than is the case for MHW intensity [Fig. 1i; note that negative (positive) values indicate that the model overestimates (underestimates) MCW intensity]. The biases in MHW and MCW intensity within the tropical Indo-Pacific are likely related to the model's overestimation of ENSO amplitude (Capotondi et al. 2020), while those in the Arctic and Antarctic may be associated with deficiencies in the representation of sea ice cover (Danabasoglu et al. 2020).

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

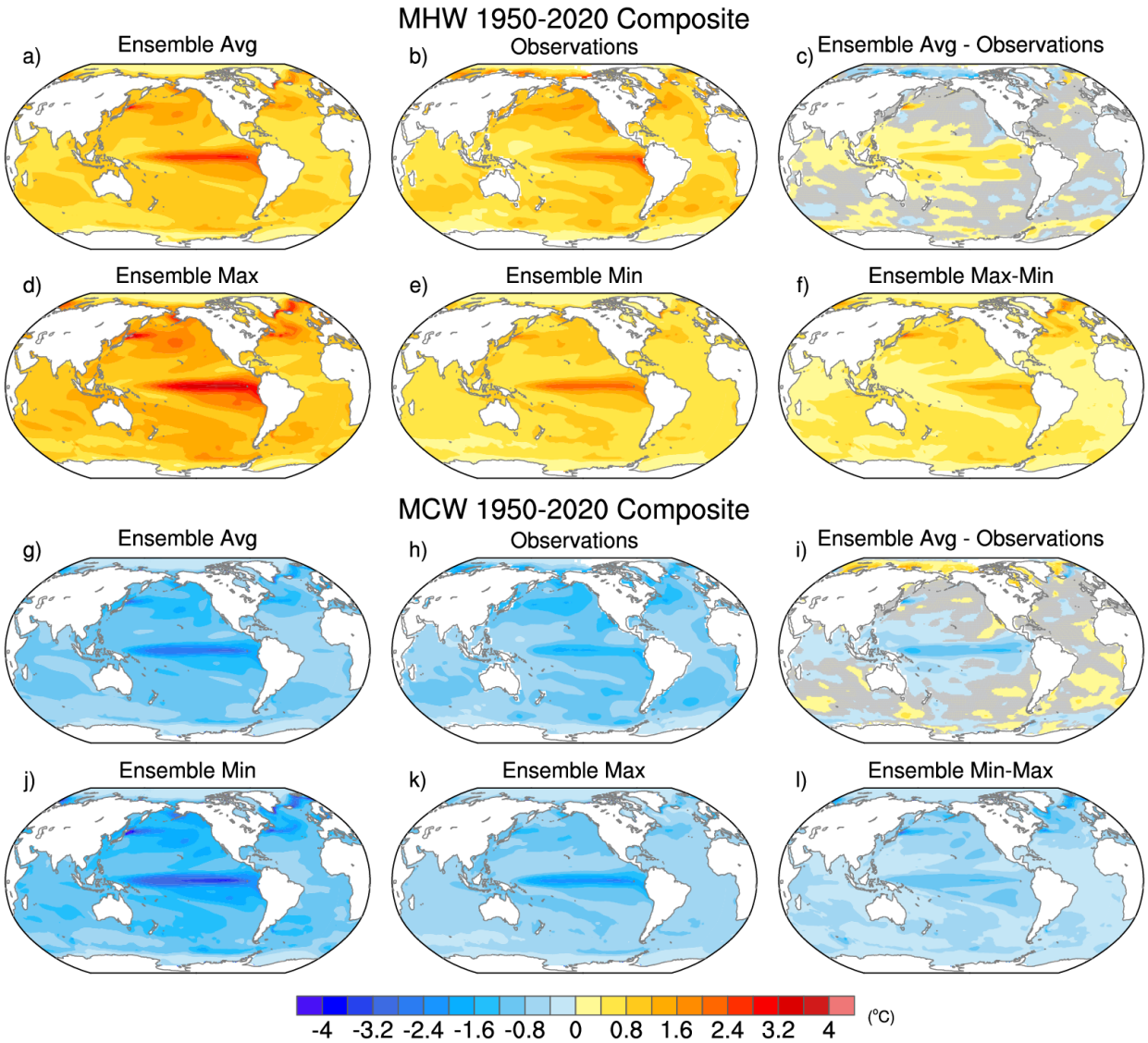

Figure 1. Composite MHW and MCW intensity (ºC) during 1950-2020 from the 100-member CESM2 Large Ensemble and Observations. (a,g) Ensemble average; (b,h) Observations; (c,i) Ensemble average minus Observations; (d,j) Ensemble maximum; (e,k) Ensemble minimum; (f,l) Ensemble maximum minus minimum. Gray shading in (c,i) indicates that observations lie within the 5th-95th percentile range of the CESM2 Large Ensemble.

We also examine composite MHW and MCW intensity in each ensemble member for a direct assessment of the sampling uncertainty due to the finite record length (note that the thresholds are still based on those obtained from all members). Figs. 1d and e show the maximum and minimum values of MHW composite intensity, respectively, across the 100 members of the CESM2 LE. To construct these figures, we have selected the ensemble member with the maximum composite value and the one with the minimum composite value at each grid box separately (i.e.,

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

the ensemble member with the maximum composite value at one grid box may differ from the ensemble member with the maximum composite value at another grid box). The two renditions of MHW composite intensity show striking differences in amplitude throughout the global oceans, with a typical range of 0.4-0.8 ºC (Fig. 1f). Larger differences (up to 1.6 ºC) are found in the equatorial eastern Pacific, the Kuroshio and Gulf Stream Extensions, and along the sea ice edge between Iceland and Svalbard. Similar results are found for MCW composite intensity (Figs. 1j-l), although the ensemble spread in the equatorial Pacific maximizes farther west compared to the MHW results (Fig. 1*l*). Such sampling variations in both MHW and MCW composite intensities over a 71-year period are noteworthy and imply that the observed composites may also be subject to similar levels of sampling uncertainty. [Note that the patterns shown in Figs. 1d,e and Figs. 1j,k may not necessarily obtain in any individual simulation.]

Next, we compare the simulated and observed MHW and MCW composites of duration. The CESM2 ensemble-mean MHW composite duration exhibits largest values (up to 6-7 months) in the tropical eastern Pacific, and secondary maxima (up to 3-4 months) in the western tropical Atlantic and the Pacific sector of the Southern Ocean (Fig. 2a). These maxima are associated with El Niño events (see Section 3.3). Elsewhere, the simulated composite durations are typically between 2-3 months. Observations show a similar, albeit noisier, pattern of MHW composite duration as the model ensemble-mean, but with generally smaller magnitudes (1-2 months) except in the tropical northeast Pacific (Fig. 2b). In particular, the observed values lie outside the model's 5^{th}-95^{th} percentile ensemble spread (indicative of a significant model bias) in the southeast and northeast tropical Pacific, south tropical Atlantic, central north Pacific, Arctic and portions of the Southern Ocean (Fig. 2c).

The ensemble spread in MHW composite duration in CESM2 is striking (Figs. 2d,e). In general, MHW composite duration varies by approximately a factor of two between the minimum and maximum of the 100 ensemble members throughout the global oceans. The absolute range is largest in the eastern tropical Pacific and the Greenland Sea (up to 7 months), with secondary maxima in the Amundsen-Bellingshausen Sea, and parts of the tropical south Atlantic and subtropical northeast Pacific (Fig. 2f). In the eastern tropical Pacific, MHW composite duration can exceed 10 months or be less than 4 months, depending on the realization (Figs. 2d,e). This sampling variability is associated with low-frequency modulation of ENSO (Capotondi et al. 2020).

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

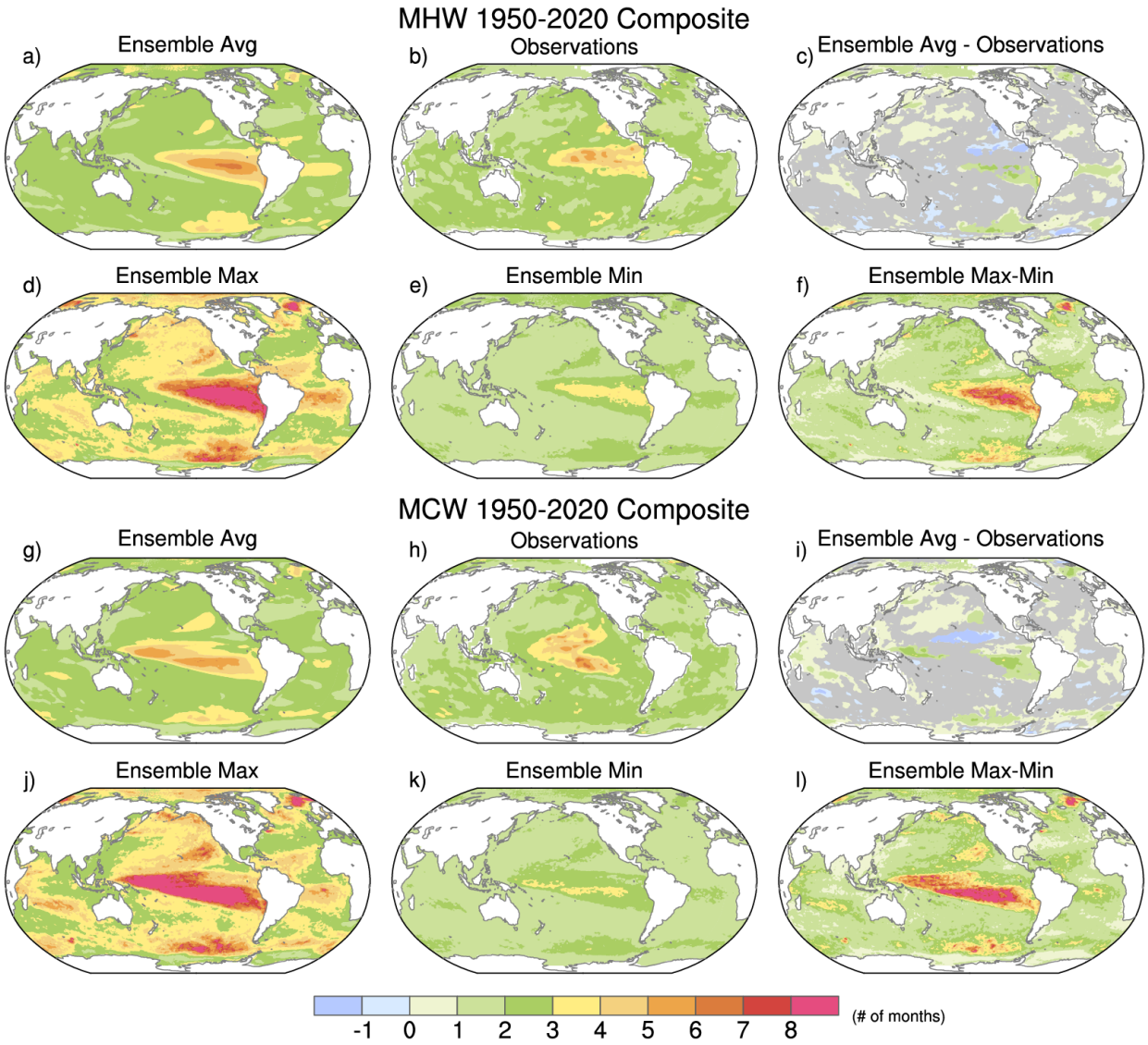

Figure 2. As in Fig. 1 but for composite MHW and MCW duration (months).

Unlike intensity, duration exhibits noticeable asymmetries between MHW and MCW composites (Fig. 2). In particular, the observed MCW composite duration shows a boomerang-shaped pattern in the central tropical Pacific, with off-equatorial maxima in both hemispheres (Fig. 2h), whereas the observed MHW composite duration exhibits peak values along the equator in the eastern Pacific (Fig. 2b). In CESM2, the duration asymmetry takes the form of a westward-extended maximum in the ensemble-mean MCW composite (Fig. 2g) compared to its MHW counterpart (Fig. 2a). Model biases in MCW composite duration are broadly similar to those in

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

MHW composite duration, with additional areas of significant bias in the far western equatorial Pacific and south of Hawai'i, where the values are overestimated and underestimated by 2-3 months, respectively (Fig. 2i). The ensemble spread in composite duration is even more pronounced for MCW than MHW, especially in the tropical Pacific where the range between the maximum and minimum of the individual members can exceed 8 months at many locations (Figs. 2j-l).

Analogous comparisons between simulated and observed MHW and MCW composites for the other six model LEs are shown in Figs. S3-S8 for intensity and Figs. S9-S14 for duration. Figure 3a-g compares the model biases in MHW composite intensity for all 7 LEs, depicted as the difference between each model's ensemble-mean composite and the observed composite. While the regional details of the patterns of significant bias vary across models, there are some common large-scale features. For example, all models show a significant overestimation of MHW composite intensity in the tropical Indian Ocean and the western half of the tropical Pacific. This positive bias is smallest for the two CanESM LEs, and is largest in CESM2 and MIROC6 where it extends across most of the Pacific basin. Another common feature across models is the significant underestimation of MHW composite intensity in the Arctic, and a significant overestimation in the Antarctic. Many of the models show a significant underestimation of MHW composite intensity along the west coast of North America, and all models except the two CESM LEs show a significant positive bias in the North Atlantic. The fractional area of significant model bias in MHW composite intensity regardless of sign is similar across models, ranging from 51% in CanESM5 to 68% in MIROC6 (numbers in the upper right of each panel in Fig. 3; see also Fig. 4). Positive biases make up most of this total areal coverage, except in CESM1 and CanESM2 where positive and negative biases contribute in roughly equal proportion (Fig. 4a). To summarize model behavior, we map the locations where at least 5 out of 7 models (> two-thirds) show a significant model bias and agree on the sign of the bias (Fig. 3h). Areas of majority agreement include the tropical northern Indian and western Pacific and high-latitude Southern Ocean where MHW intensities are significantly overestimated, and the Arctic Ocean where they are significantly underestimated; these areas comprise 34% of the world oceans.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

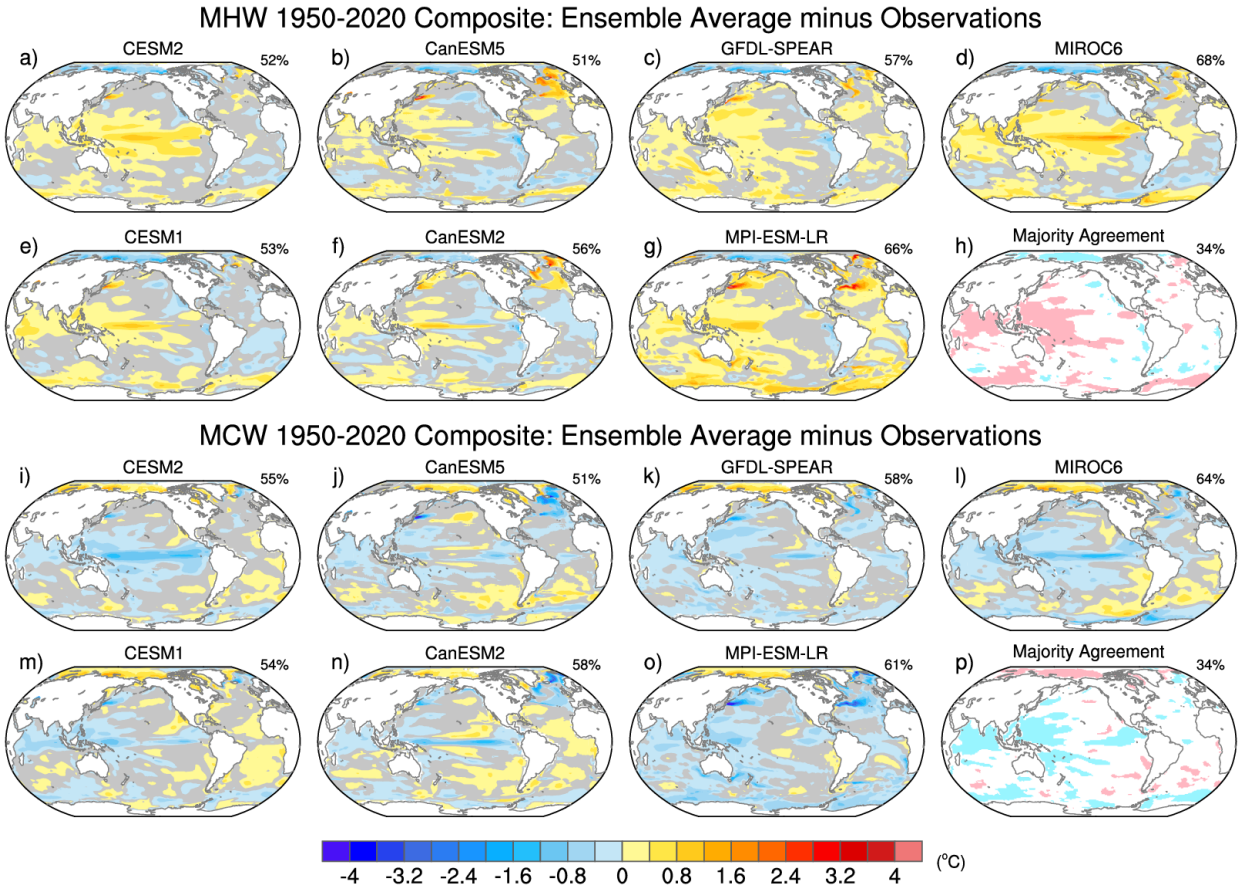

Figure 3. Composite MHW and MCW intensity (ºC) during 1950-2020 for the ensemble-mean of each model Large Ensemble minus observations: (a,i) CESM2; (b,h) CanESM5; (c,i) GFDL-SPEAR; (d,j) MIROC6; (e,k) CESM1; (f,l) CanESM2; (g,o) MPI-ESM-LR. Gray shading indicates that observations lie within the 5th-95th percentile range of the model Large Ensemble. The number in the upper right of each panel denotes the fractional area (%) of significant model bias (e.g., non-gray areas). Shading in panels (h,p) show locations where at least two-thirds of the models show a significant bias (pink shading for positive bias and blue shading for negative bias); the number in the upper right denotes the fractional area (%) of the pink and blue shading.

Model biases in MCW composite intensity are generally analogous to their MHW counterparts, but with slightly larger amplitudes (Figs. 3i-o). [Note that the opposite sign between model biases in MCW and MHW composite intensity indicates that the biases are the same in a relative sense: i.e., a negative bias in MCW intensity and a positive bias in MHW intensity both denote overestimation.] There are also some asymmetries in regional expression of the biases. For example, the significant overestimation of composite intensity in the equatorial Pacific extends farther east for MCW compared to MHW in most models. The areal coverage of significant model bias regardless of sign is also similar for MCW as MHW composite intensity, with a comparable

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

proportion of over- and underestimation within each model (Fig. 4b). Areas of majority agreement in the sign of significant model bias in MCW composite intensity are similar to those for MHW (Fig. 3p).

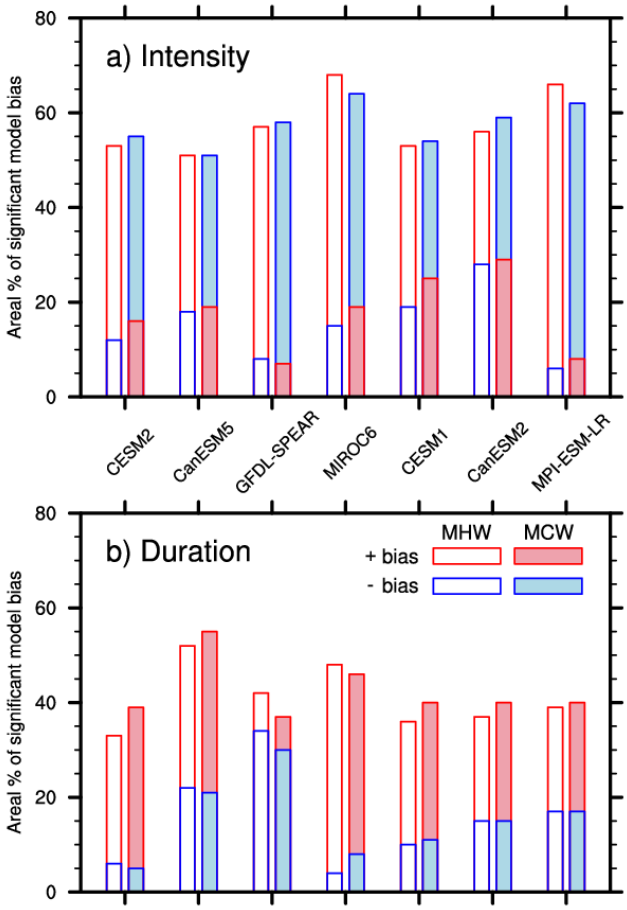

Figure 4. Fractional area (%) of significant positive (red bars) and negative (blue bars) model bias in composite MHW (open bars) and MCW (filled bars) (a) intensity and (b) duration during 1950-2020. Note that a positive (negative) bias in MHW (MCW) intensity indicates model overestimation.

Unlike intensity, there are considerable differences amongst models in the pattern and amplitude of significant bias in MHW (Figs. 5a-g) and MCW (Figs. 5i-o) composite duration. For example, in the tropics, four of the models (CanESM5, GFDL-SPEAR, CanESM2 and MPI-ESM-LR) show significant underestimation MHW and MCW duration (by 1-2 months), while the remaining three models (CESM2, MIROC6 and CESM1) show insignificant or opposite-signed biases. The former group of models also exhibits consistent asymmetries in their patterns of

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

duration bias between MHW and MCW composites, with an equatorial maximum for MHW and off-equatorial ("boomerang" shaped) maxima for MCW. There is greater agreement amongst the 7 LEs in the sign of their extra-tropical biases, with significant overestimation (by 1-2 months) of composite duration for both MHW and MCW within the Arctic, North Pacific and Atlantic, and Southern Ocean, although the details within each region vary considerably. There is less consensus amongst models on the significance and sign of bias in duration compared to intensity for both MHW and MCW (Figs. 5 h,p). Only 17% and 21% of the area of the world oceans show model consensus on MHW and MCW duration bias, respectively. Further, the regions of model consensus are less spatially coherent for duration biases compared to intensity biases, with the North Pacific showing the most widespread areas of model agreement.

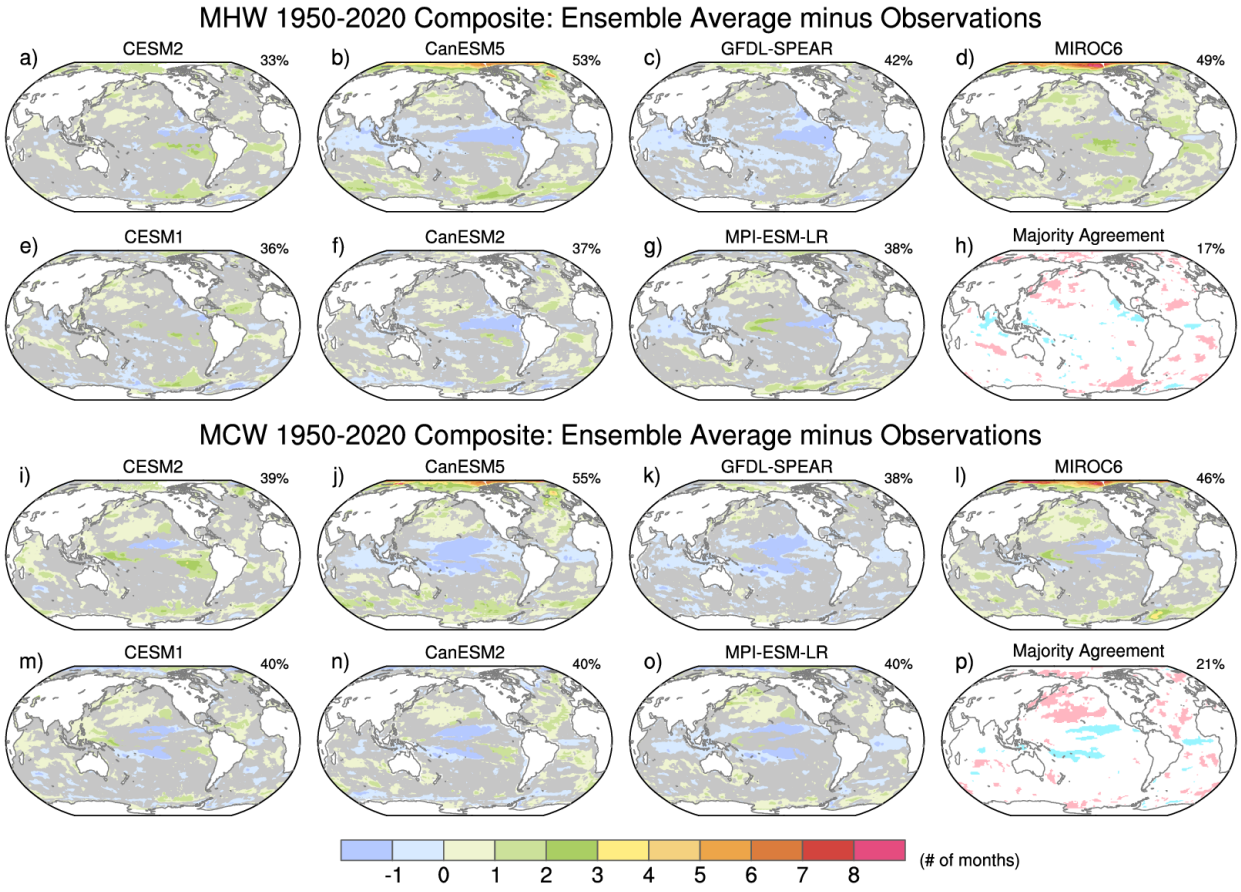

Figure 5. As in Fig. 3 but for composite MHW and MCW duration (months).

The areal coverage of significant model bias regardless of sign is generally smaller for duration than intensity, ranging from 33% for MHW and 39% for MCW in CESM2 to 53% for

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

MHW and 55% for MCW in CanESM5 (Fig. 4b). The proportion of positive *vs*. negative duration bias contributing to the total areal coverage for both MHW and MCW composites varies by model, with CESM2, GFDL-SPEAR, MIROC6 and CESM1 showing a substantial imbalance and the remaining models showing more equivalent values (Fig. 4b).

In summary, all model LEs examined in this study show widespread statistically significant biases in their historical simulation of composite MHW and MCW intensity and duration defined on the basis of monthly SSTs and taking into account sampling uncertainty due to the limited length of the observational record. These biases cover ~50-70% of the area of the global oceans for intensity and ~35-55% for duration in every model. The majority of model LEs significantly overestimate MHW and MCW intensity in the tropical western Pacific and Indian Oceans and across portions of the Southern Ocean, and significantly underestimate it in the Arctic. The models show less consistency in their duration biases both in terms of sign and location, although the North Pacific is a region of general overestimation.

3.2. Future changes

Next, we examine future changes in MHW and MCW composite intensity and duration due to changes in variability (mean state changes are considered in Section 3.4). As before, we present results from the 100-member CESM2 LE in the main paper and the remaining 6 model LEs in the Supporting Information. Note that all model LEs use relatively high-emissions scenarios for their projections (Table 1). Figure 6 shows composite intensity maps for three periods (1970-2000, 2020-2050 and 2070-2100) and the differences between each future period and present-day (e.g., 2020-2050 minus 1970-2000, and 2070-2100 minus 1970-2000) for the CESM2 LE. All three periods show similar spatial patterns of MHW composite intensity, but their amplitudes evolve over time (Figs. 6a-c). Relative to present-day, MHW composite intensity at mid-century is projected to increase significantly in the Arctic and along the Antarctic coastline, and within portions of the North Pacific, North Atlantic and the tropics, and decrease significantly within the Southern Ocean and parts of the tropical Indian, western Pacific, and eastern Atlantic basins, with amplitudes generally < 0.2 ºC (Fig. 6d). The increase in MHW intensity within the Arctic and along the Antarctic coastline is likely related to the melting of sea ice and subsequent exposure of the ocean to the atmosphere, allowing for variability in heat exchange across the air-sea interface. We speculate that the patterns of MHW intensity change within the North Pacific and North

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

Atlantic are potentially related to altered patterns of atmospheric circulation variability (O'Brien and Deser, 2023), while the decreases in MHW intensity in the Southern Ocean may be due in part to a poleward shift of the storm track and associated reduction in air-sea heat and momentum flux variability (Wu et al. 2012; Guo et al. 2022). A similar pattern of future change is projected for late-century as mid-century, with an approximately two-fold increase in amplitude, except in the tropical Pacific where MHW composite intensity is projected to decrease significantly (up to 0.6 ºC), opposite to the increase seen at mid-century (Fig. 6e). The diverging mid- and late-century projections in the tropical Pacific are related to changes in the behavior of ENSO as will be shown in Section 3.3. In CESM2, ENSO is projected to intensify until about 2050 and weaken thereafter (Maher et al. 2023), consistent with the changes in MHW intensity in the tropical Pacific. Future changes in MCW composite intensity are largely similar to those in MHW intensity, both in pattern and amplitude (Fig. 6i,j). Analogous comparisons between present-day and future MHW and MCW composite intensity for each of the six other model LEs are shown in Figs. S15-20.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

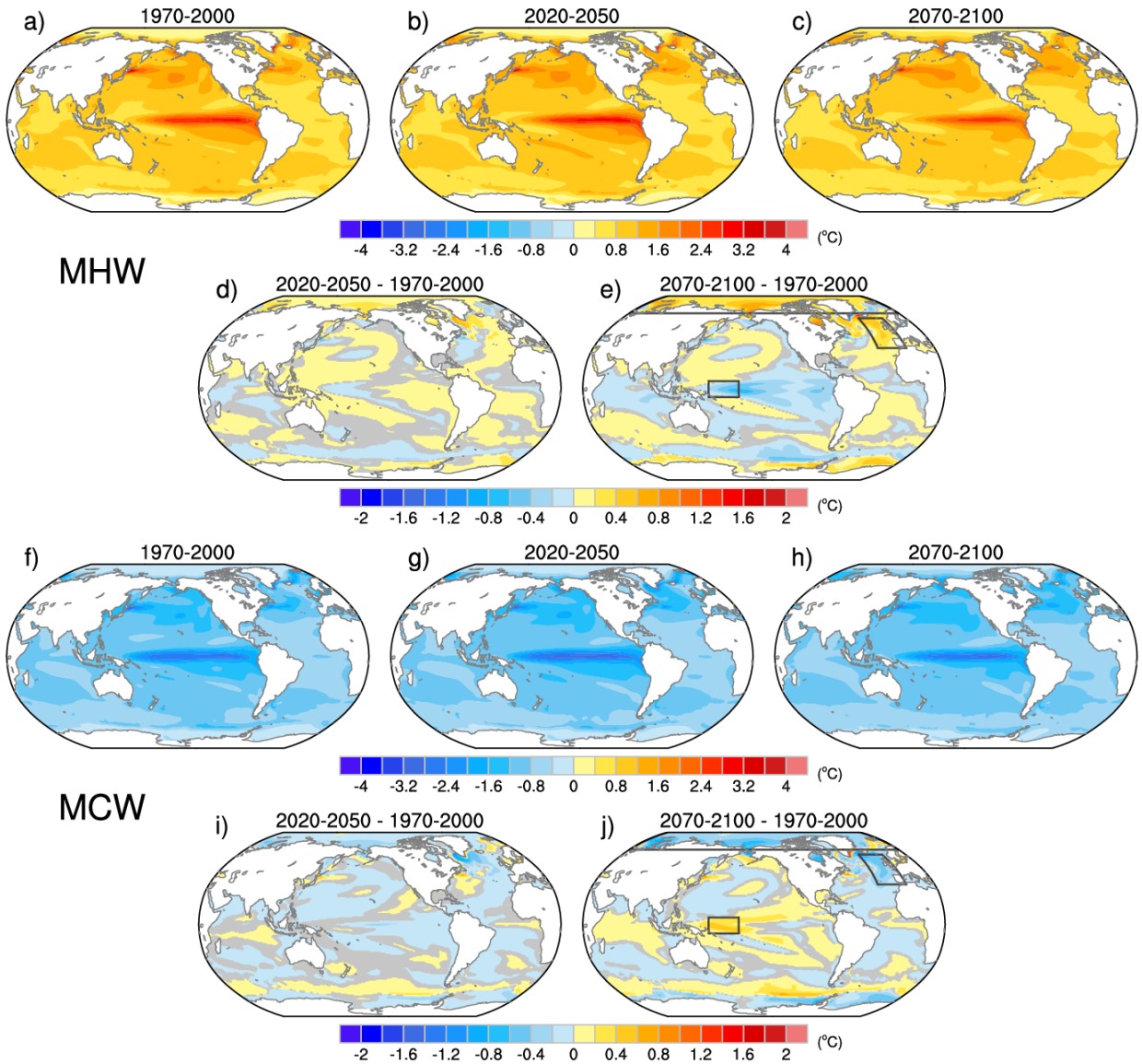

Figure 6. Composite MHW and MCW intensity (ºC) from the 100-member CESM2 Large Ensemble during (a,f) 1970-2000, (b,g) 2020-2050, (c,h) 2070-2100, and differences (d,i) 2020-2050 minus 1970-2000, and (e,j) 2070-2100 minus 1970-2000. Note that the color bar range is twice as large in a-c, f-h compared to d-e, i-j. Gray shading in d-e, i-j indicates that the differences are not statistically significant according to the False Discovery Rate applied to a 2-sided t-test at the 95% confidence level. Boxes in panels e,j outline regions used for Figs. 17 and 18. Note that positive (negative) values in d,e and negative (positive) values in i,j indicate an increase (decrease) in intensity.

Future changes in MHW and MCW composite duration in CESM2 are shown in Fig. 7. Statistically significant changes begin to emerge mid-century and become widespread by late

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

century for both MHW and MCW (Figs. 7d,e and i,j). The late-century changes in composite duration, which are largely similar between MHW and MCW, include reductions over the tropical Indo-Pacific of up to 2-3 months and smaller decreases over the Southern Ocean and eastern North Pacific, accompanied by increases over the North Atlantic and near Antarctica (Figs. 7e and j). Notable asymmetries occur over the tropical Pacific, where late-century duration changes maximize in the east for MHW and in the west for MCW, likely related to differences in their present-day climatological mean states (Figs. 7a *vs.* f). Another asymmetry is apparent in the western tropical Atlantic, where significant late-century increases in duration occur for MCW but not for MHW. We speculate that this difference may be related to the significant reduction in MHW duration over the eastern tropical Pacific, which may act to offset (and mask) the increase that would otherwise occur over the western tropical Atlantic via Pacific-to-Atlantic inter-basin teleconnection mechanisms (similar to those which occur for ENSO events; see for example Alexander et al. 2002). Analogous maps of MHW and MCW composite duration in present-day and future periods for each of the six other model LEs are shown in Figs. S21-S26.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

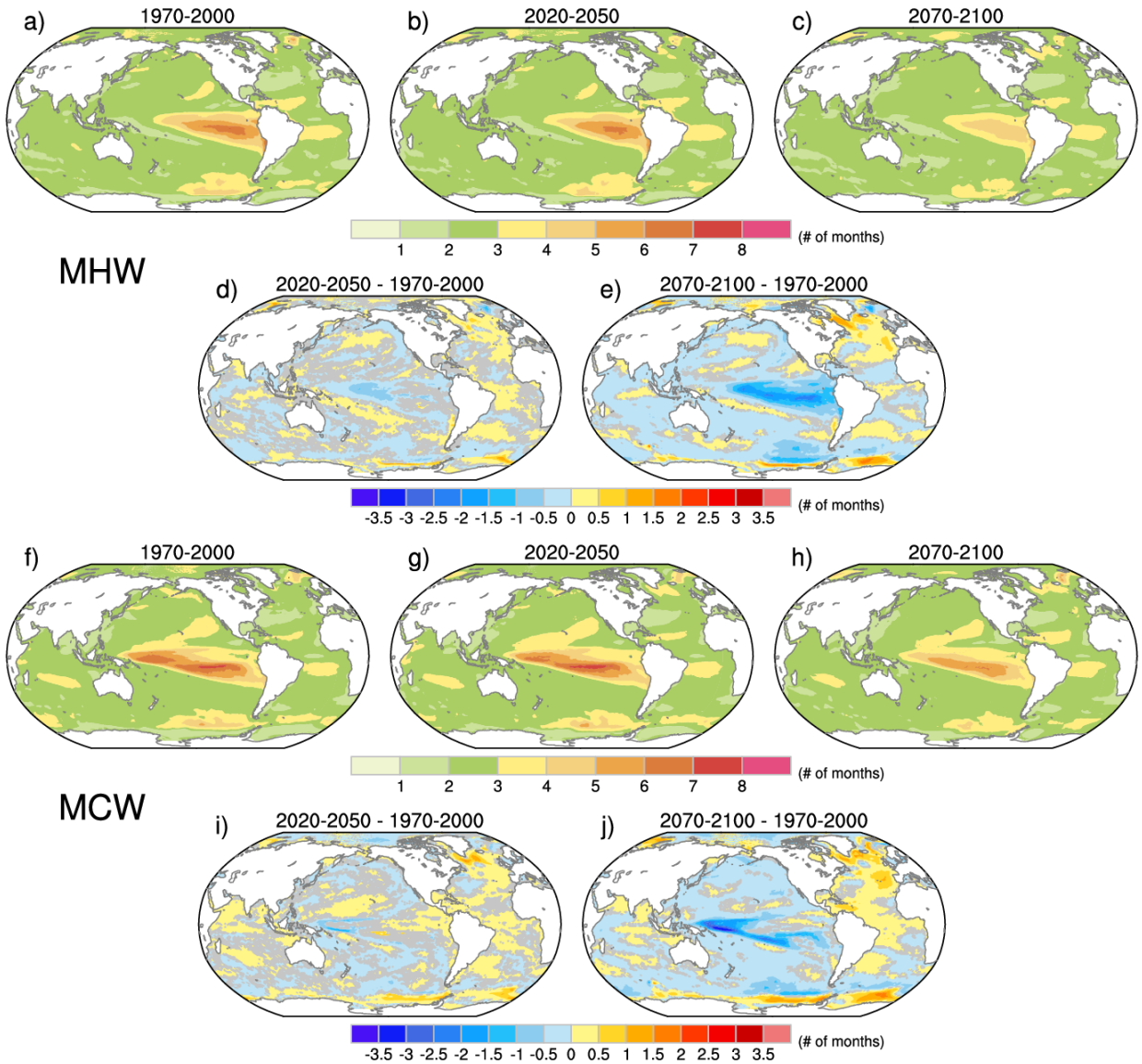

Figure 7. As in Fig. 6 but for composite MHW and MCW duration (months).

Next, we compare the projected late-century changes in composite intensity across the seven model LEs for both MHW and MCW (Fig. 8); mid-century changes are shown in Fig. S27. The models show considerable diversity in their patterns of projected change, especially at low latitudes. Over the tropical Pacific, for example, CanESM2 and CESM2 and MPI-ESM-LR show relatively pronounced reductions in intensity for both MHW and MCW, while the other models show more muted changes of mixed sign (Fig. 8). There is some correspondence between the patterns of future change in MHW/MCW intensity over the tropical Pacific and the projected behavior of ENSO across models, but it is not exact. For example, CanESM2 and CESM2 show

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

pronounced late-century reductions in ENSO variance (Maher et al. 2023), consistent with decreasing MHW/MCW intensity, but CanESM5 and MIROC6 show increases in ENSO variance and MPI-ESM-LR shows little change in ENSO variance (Maher et al. 2023), yet these models also simulate decreasing MHW/MCW intensity (albeit less than in CanESM2 and CESM2). Models agree that MHW and MCW composite intensity will increase in the Arctic and along Antarctica, and diminish in the Indo-Pacific sector of the Southern Ocean by late century (Figs. 8h,p). There is also model consensus that composite intensity will increase over portions of the subtropics and that it will decrease in the tropics on either side of the equator, especially for MHWs (Figs. 8h,p). The fractional area of the world ocean that is projected to experience a significant late-century change in MHW and MCW composite intensity (regardless of sign) varies from a low of 64-65% in GFDL-SPEAR to a high of 84-85% in CESM2 (Fig. 8). Of this fractional area, the proportion of significant increase outweighs that of significant decrease in four of the models, and vice versa in two of the models (Figs. 9a,b). Model consensus on the significance and sign of projected changes in MHW and MCW intensity is found over 36-38% of the area of the world ocean (Figs. 8h,p).

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

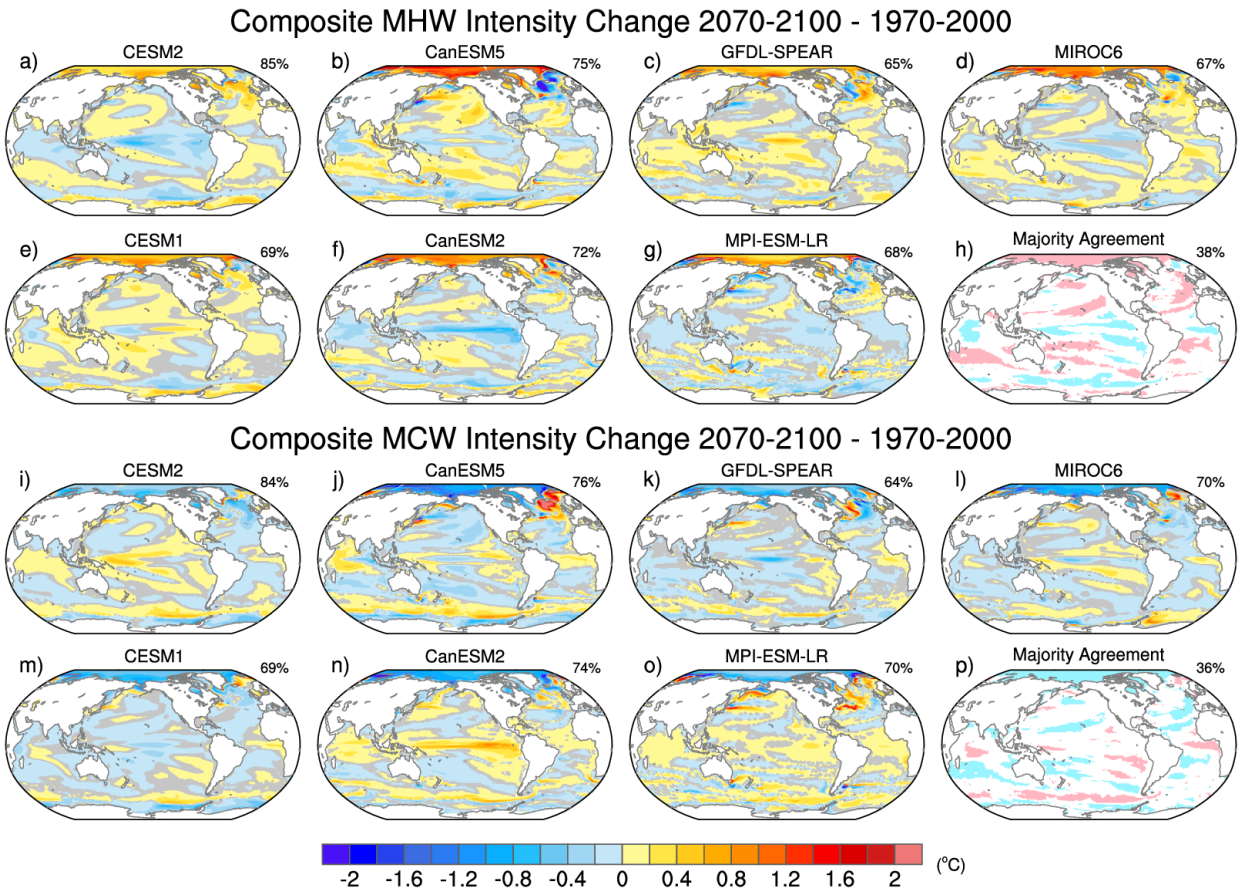

Figure 8. Composite MHW and MCW intensity (ºC) differences between 2070-2100 and 1970-2000 for the ensemble-mean of each model Large Ensemble: (a,h) CESM2; (b,i) CanESM5; (c,j) GFDL-SPEAR; (d,k) MIROC6; (e,l) CESM1; (f,m) CanESM2; (g,n) MPI-ESM-LR. Gray shading indicates that the differences are not statistically significant according to the False Discovery Rate applied to a 2-sided t-test at the 95% confidence level. The number in the upper right of each panel denotes the fractional area (%) of significant differences (e.g., non-gray areas). Note that positive (negative) values in d,e and negative (positive) values in i,j indicate an increase (decrease) in intensity. Shading in panels (h,p) show locations where at least two-thirds of the models show statistically significant values (pink for positive and blue for negative); the number in the upper right denotes the fractional area (%) of the pink and blue shading.

Like intensity, models show a variety of patterns and amplitudes of projected late-century change in MHW and MCW composite duration (Fig. 10; mid-century changes are shown in Fig. S28). For example, CESM2 features an inter-basin contrast between the North Atlantic and the tropical Pacific, while CanESM5 shows out of phase changes between the tropics and extra-tropics (excluding the Antarctic coastline). Models disagree on the sign of duration change over the Arctic, where significant increases are simulated by CESM1, CanESM2 and MPI-ESM-LR, and significant decreases by CanESM5, MIROC6, and GFDL-SPEAR. The reasons for this are

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

unclear. There is little agreement among models on the spatial pattern of duration changes within the tropical Pacific, although modest decreases in MHW duration are found in the majority of models in the far west (Figs. 10h,p). Likewise, there is model consensus that MHW and MCW duration will decrease over much of the North Pacific and at scattered locations within the Southern Ocean except along Antarctica. Five of the seven models indicate that the fraction of the world oceans projected to experience a significant decrease in MHW and MCW duration will outweigh the fraction with a significant increase (Figs. 9c,d). Finally, model consensus on the sign of significant projected changes in MHW and MCW duration is found for 27-28% of the area of the world ocean, somewhat less than for intensity (Figs. 10h,p).

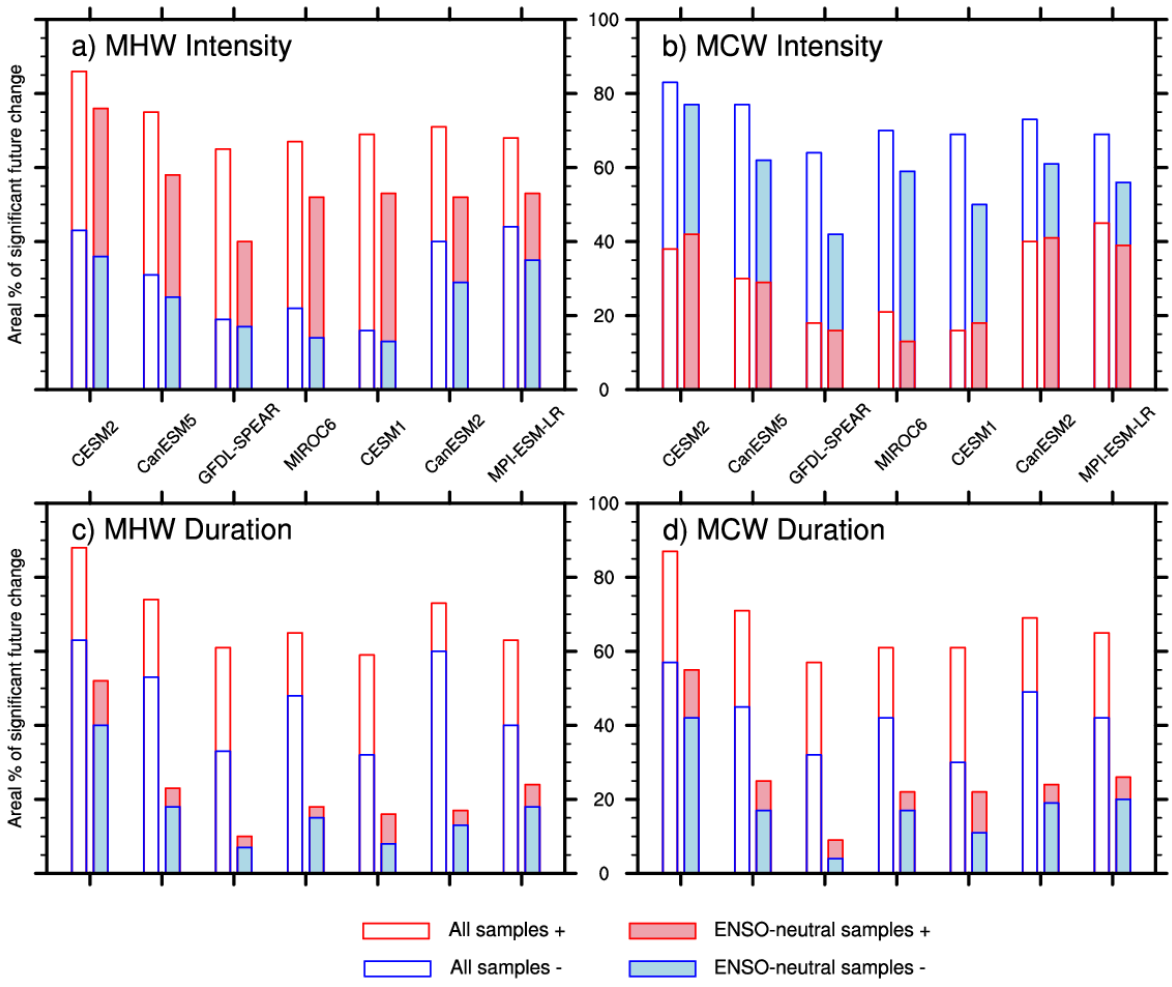

Figure 9. Fractional area (%) of significant positive (red bars) and negative (blue bars) differences between 2070-2100 and 1970-2000 for each model Large Ensemble for composite (a) MHW intensity, (b) MCW intensity, (c) MHW duration and (d) MCW duration. Open (filled) bars are for all (ENSO-neutral) MHW and MCW samples.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

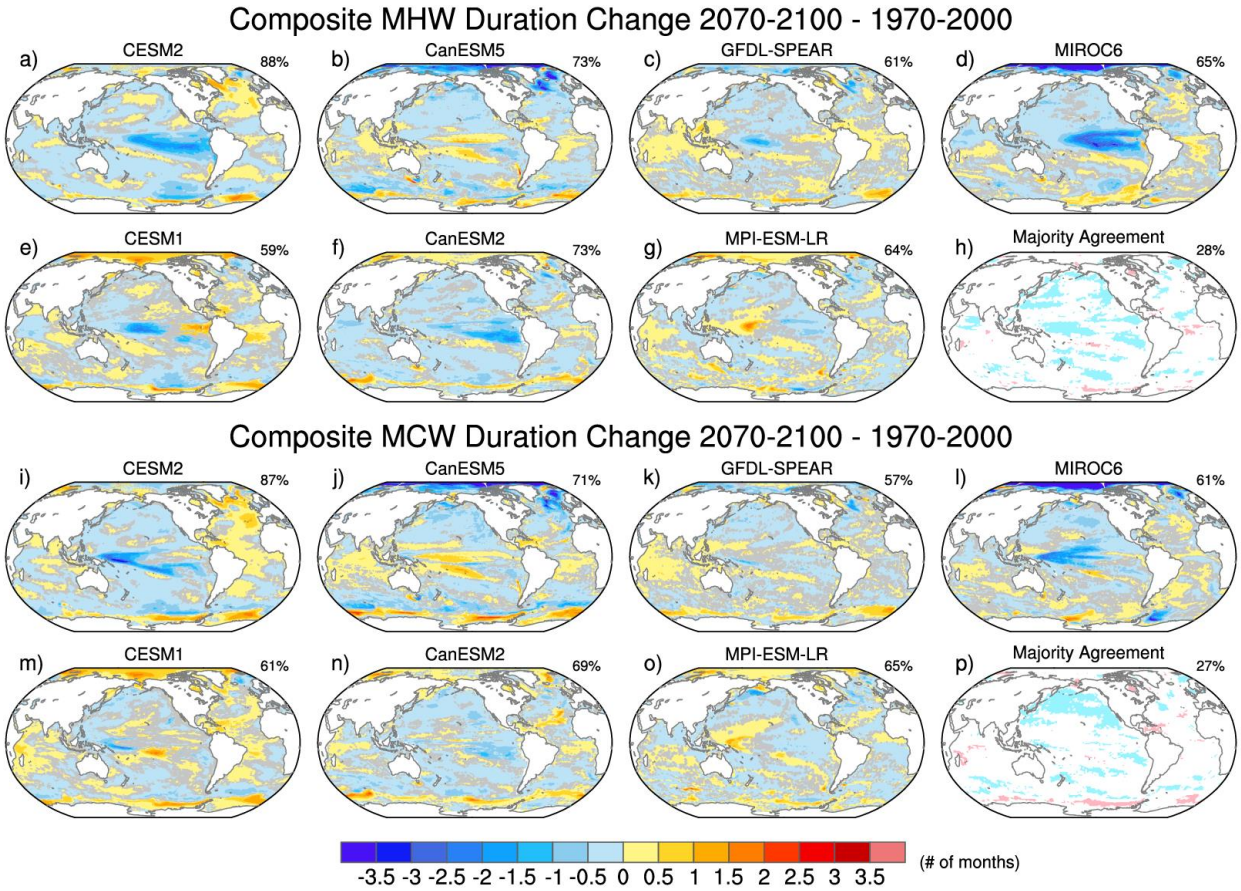

Figure 10. As in Fig. 8 but for composite MHW and MCW duration (months).

3.3. Role of ENSO

It is clear from the preceding results that the tropical Pacific is a region of high-amplitude MHW and MCW intensity and duration, and that the projected changes in this region are subject to large inter-model spread. Model LEs also differ substantially in their ENSO projections, including the sign, spatial pattern and time dependence of future change, for reasons that are not well understood (Maher et al. 2023). Given that SST variability in the tropical Pacific associated with ENSO is known to impact remote regions via atmospheric and oceanic teleconnections, this raises the question, how much do projected changes in ENSO influence projected changes in MHW and MCW characteristics worldwide? To address this question, we recompute the MHW and MCW composites for ENSO-neutral conditions following the procedures outlined in Section 2.2a. As before, we begin by showing results for the CESM2 LE, followed by an intercomparison

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

of late-century changes in ENSO-neutral MHW and MCW composite intensity and duration across all seven model LEs.

As expected, the tropical Pacific maxima in MHW and MCW composite intensity amplitudes in CESM2 are greatly reduced in the ENSO-neutral samples compared to all samples in all three time periods (Figs. 11a-c,f-h compared with Figs. 6a-c,f-h). Other regions, most notably the North Pacific, also show a slight decrease in amplitude in each time period. It is clear that changes in ENSO variability impact future changes in MHW and MCW composite intensity. For example, at mid-century, the areal coverage of significant change is reduced from 70% in the all-sample composites (Figs. 6d,i) to 53% in the ENSO-neutral composites (Figs. 11d,i), along with some alterations in the sign and pattern of change in the Pacific sector. By late century, the differences between the all-sample (Figs. 6e,j) and ENSO-neutral (Figs. 11e,j) composite changes over the North Pacific are particularly striking. The ENSO-neutral composites show a pattern reminiscent of the positive phase of the Pacific Decadal Oscillation (PDO; Newman et al. 2016), with negative values over the western and central North Pacific and positive values in the eastern North Pacific, for both MHW and MCW and opposite in sign to the all-sample composites. The fact that the ENSO-neutral late-century changes in intensity are of the same sign for MHW and MCW indicates that their responses are strongly asymmetric (e.g., opposite directions of change) in this region. The similarity of future changes in intensity over the Arctic, North Atlantic and Southern Ocean between the ENSO-neutral and all-sample composites indicates that changes in ENSO variability play only a minor role in these areas. It is interesting to note that the late-century decrease in ENSO-neutral composite intensity over the tropical Pacific and Atlantic is larger for MCW than MHW (compare yellow shading in Fig. 11j with blue shading in Fig. 11e), a result which is insensitive to whether one or two Tropical Pacific PCs are used in the definition of ENSO-neutral samples (not shown).

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

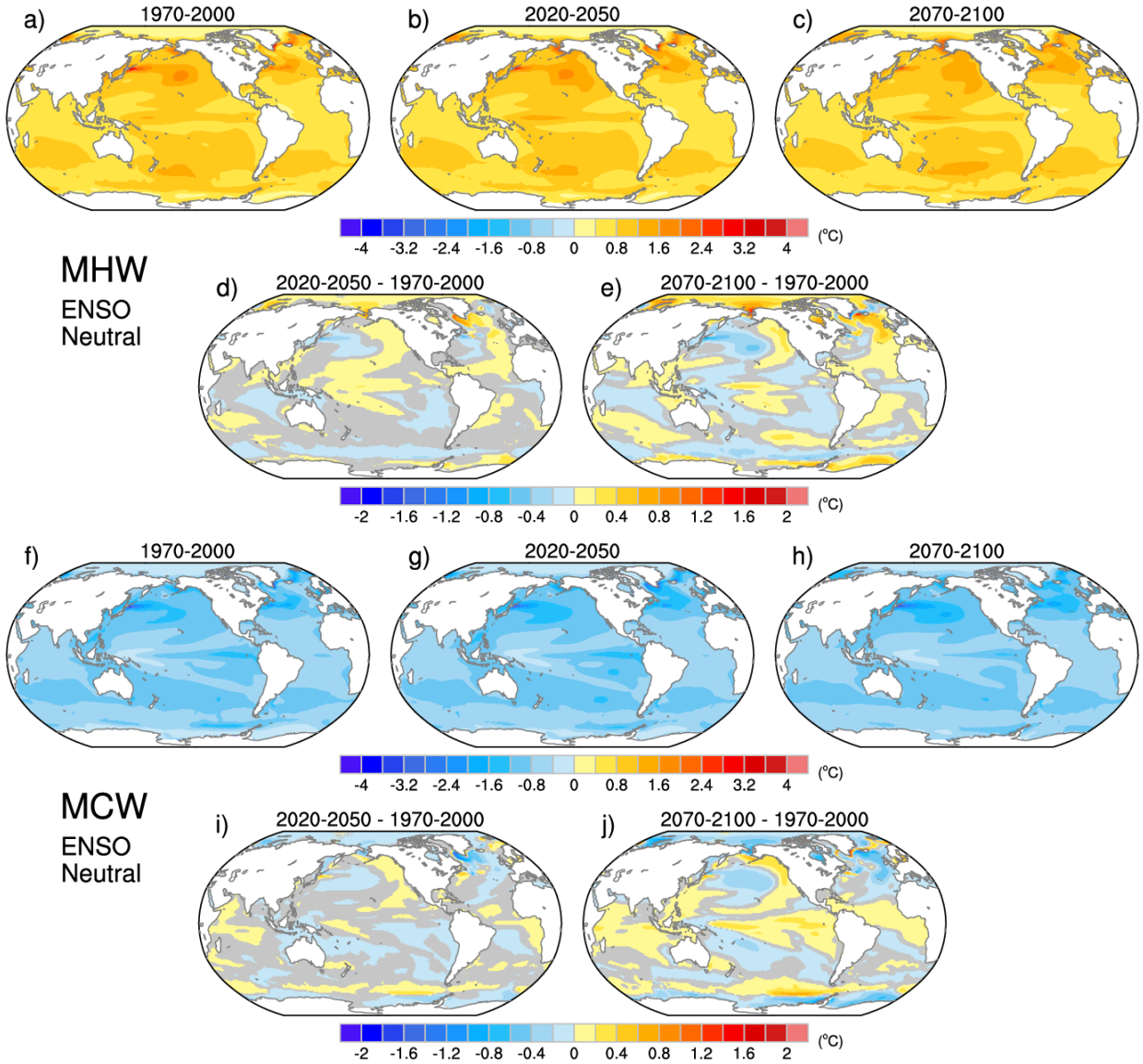

Figure 11. As in Fig. 6 but for ENSO-neutral samples.

MHW and MCW durations in CESM2 are considerably reduced in the ENSO-neutral composites compared to their "all-sample" counterparts in each time period (1970-2000, 2020-2050 and 2070-2100), with values ranging from 1-3 months (Figs. 12a-c, e-f; compare with Figs. 7a-c, f-h). As expected, the largest changes occur in the tropical Pacific, where the prominent duration maximum virtually disappears in the ENSO-neutral composites. Secondary duration maxima in regions with known teleconnections to the tropical Pacific such as the tropical Atlantic,

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

northeast Pacific and Amundsen-Bellinghausen Sea (e.g., west of the Antarctic Peninsula), also show pronounced decreases in the ENSO-neutral composites compared to the all-sample composites. The resulting patterns of ENSO-neutral composite duration are very similar between MHW and MCW for each time period (compare Figs. 12a-c and e-f), more so than their all-sample counterparts (compare Figs. 7a-c and e-f). While these ENSO-neutral patterns do not change appreciably over time, their amplitudes diminish slightly (note, for example, the reduction in areal extent of values between 2-3 months shown in the darker green shading in Fig. 12).

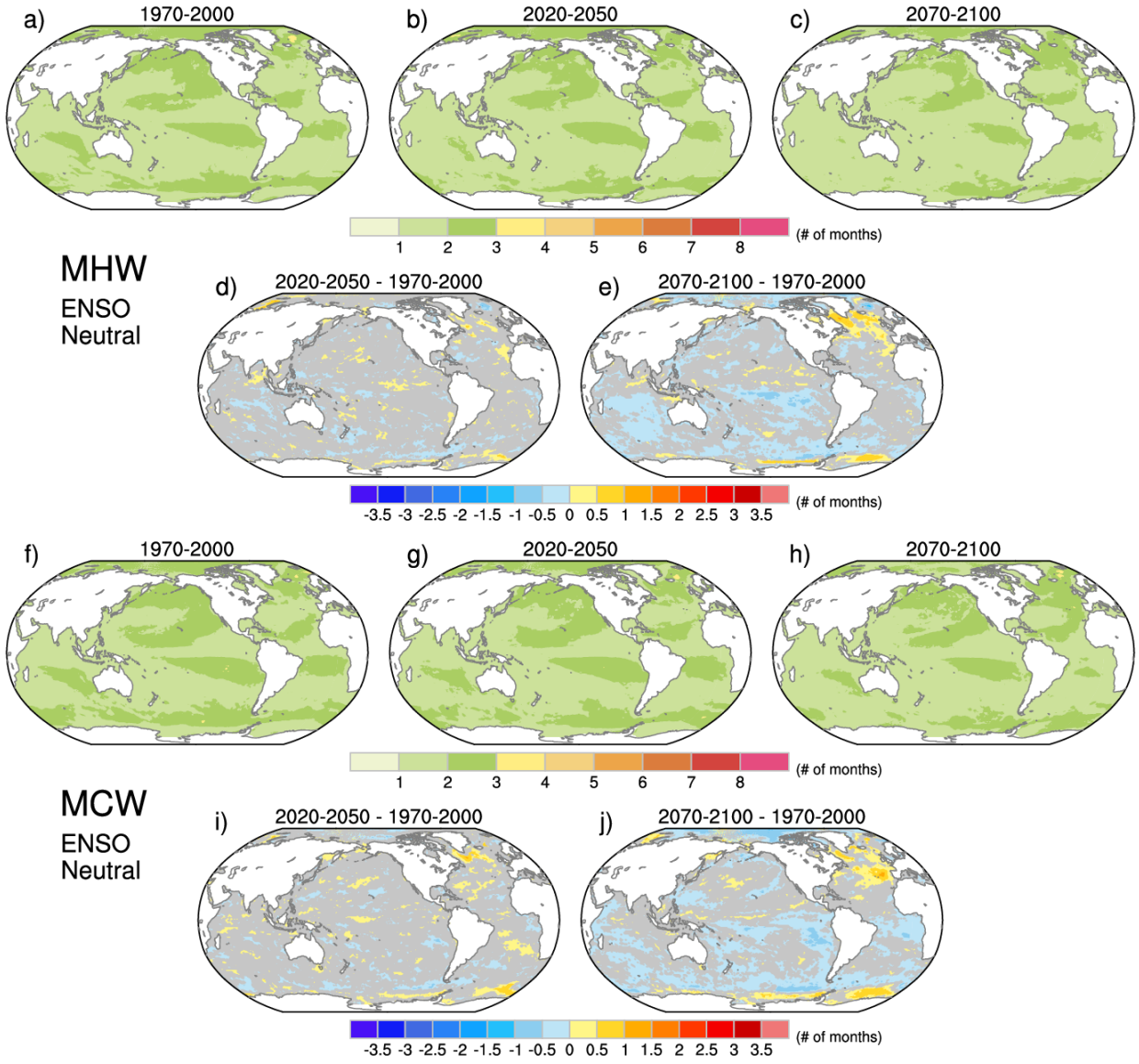

Figure 12. As in Fig. 7 but for ENSO-neutral samples.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

Significant future changes in MHW and MCW composite duration during ENSO-neutral states (Figs. 12d,e and i,j) are considerably weaker in magnitude and less widespread compared to their all-sample counterparts (Figs. 7d,e and i,j). Indeed, significant late-century changes in ENSO-neutral MHW and MCW duration are mainly confined to the Southern Hemisphere, Arctic Ocean and portions of the North Atlantic. It is notable that the pattern of late-century duration change is much more symmetric between MHW and MCW events in the ENSO-neutral composites compared to the all-sample composites, consistent with our earlier conjecture that changes in ENSO variability are responsible for asymmetric changes in duration over the tropical Atlantic. The ENSO-neutral MHW and MCW composites for the six other model LEs are shown in Figs. S29-S34 for intensity and Figs. S35-S40 for duration.

Next, we summarize the late-century changes in ENSO-neutral MHW and MCW composite intensity and duration across all seven model LEs. Like CESM2, the other model LEs all show reduced areal coverage of significant future changes in MHW and MCW composite intensity and duration during ENSO-neutral states (Figs. 13 and 14) compared to their all-sample counterparts (Figs. 8 and 10). For intensity, the areal fraction of significant change is reduced by 10-40% depending on the model (compare open and shaded bars in Figs. 9a,b). Duration shows even more pronounced decreases in areal fraction of significant change, with reductions of 42-84% depending on the model (compare open and shaded bars in Figs. 9c,d). Notably, the areal fraction of significant future change in duration for ENSO-neutral composites falls below 25% for all models except CESM2 (shaded bars in Figs. 9c,d). The relative proportion of positive *vs.* negative contributions to the areal fraction of significant future change remains similar for the ENSO-neutral composites compared to the all-sample composites for a given model (Fig. 9). Finally, the areal fraction of the world oceans with at least two-thirds model consensus on the sign of significant future change in MHW and MCW intensity (duration) decreases from 36-38% (27-28%) in the all-sample composites to 20-21% (1%) in the ENSO-neutral composites (Figs. 13h,p and 14h,p). In terms of amplitude, modest reductions are found between the ENSO-neutral and all-sample MHW and MCW composites for each model, both for intensity (compare Figs. 13 and 8) and duration outside of the Tropical Pacific (compare Figs. 14 and 10). In summary, future changes in ENSO-related variability account for a considerable fraction of the projected late-

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

century MHW and MCW composite changes in intensity, and nearly all of the projected change in duration. Similar conclusions are found for mid-century changes (Figs. S41-S42).

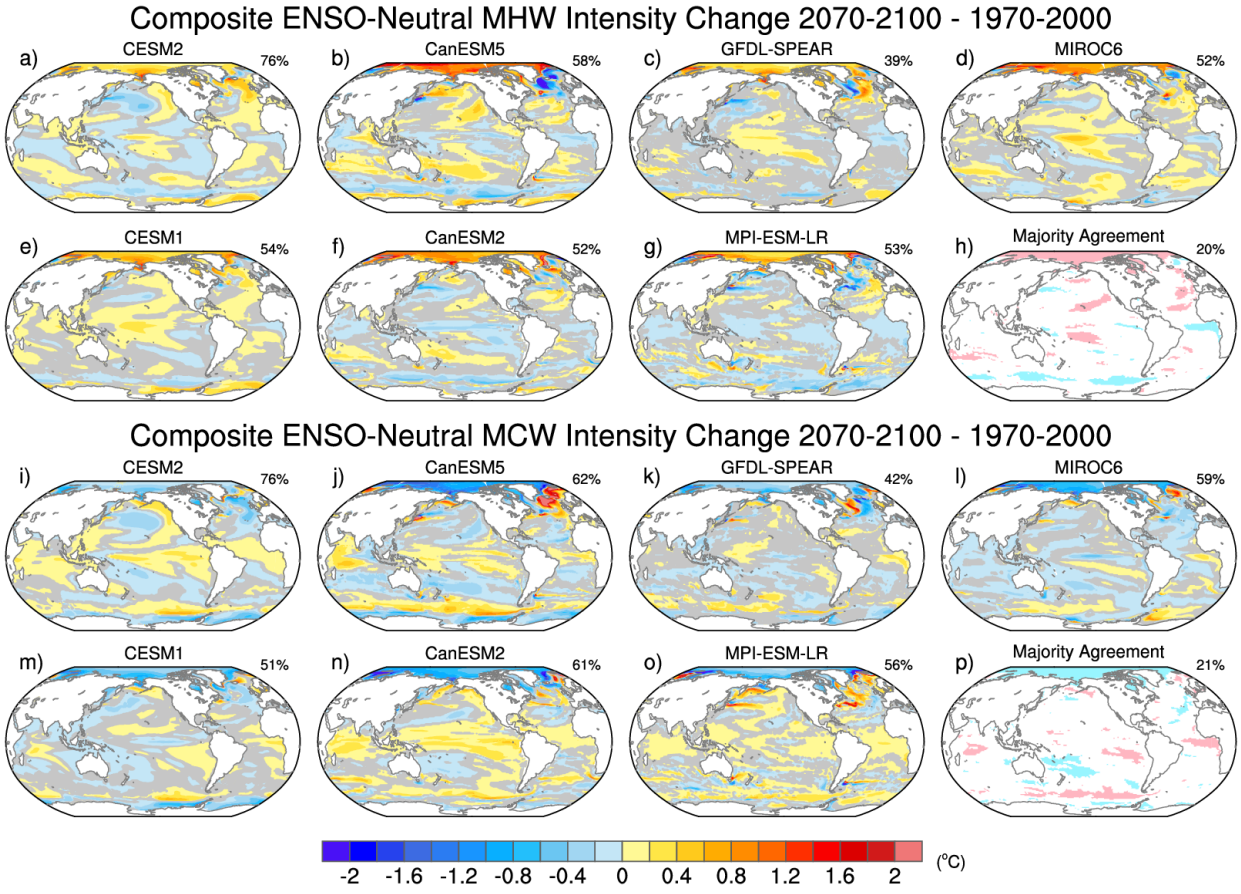

Figure 13. As in Fig. 8 but for ENSO-neutral samples.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

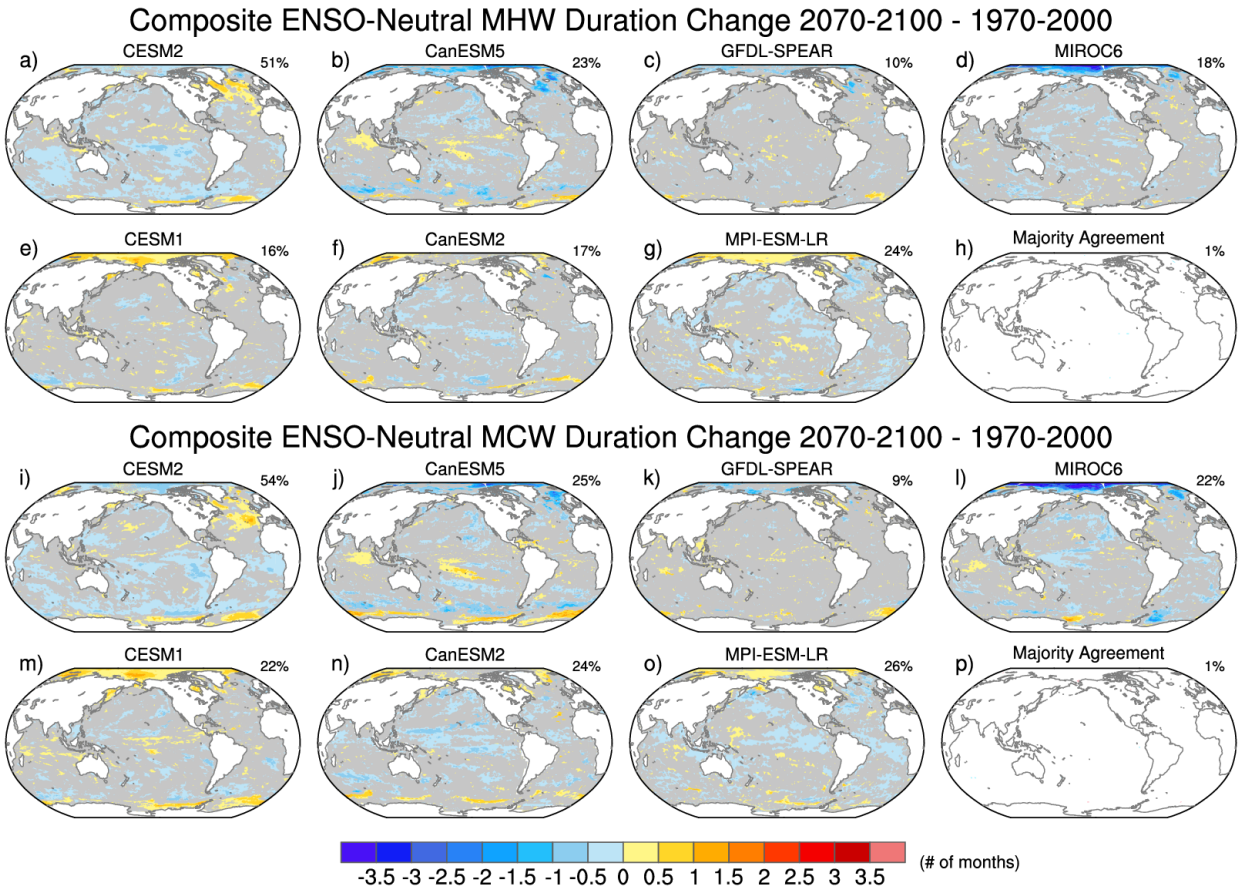

Figure 14. As in Fig. 10 but for ENSO-neutral samples.

3.4. Mean state vs. Variability

Up to now, we have focused on the future evolution of MHW and MCW characteristics due to changes in variability. However, this evolution will be superimposed upon mean state shifts, e.g. "a rising tide lifts all ships". Figure 15 shows the late-century mean state SST change in each model LE computed as the ensemble-mean difference between 2070-2100 minus 1970-2000. The large-scale patterns of mean state change are generally similar across the models [see Manabe et al. (1991), Xie et al. (2010) and Armour et al. (2016) for a discussion of pattern formation mechanisms], but their amplitudes vary considerably due to differences in climate sensitivity and/or radiative forcing scenario (recall Table 1). Of the three CMIP6 models with the same (SSP5-8.5) radiative forcing scenario, CanESM5 shows the largest change in global-mean SST (4.04 ºC) and MIROC6 shows the smallest (2.03 ºC). Similarly, of the three CMIP5 models with the same

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

(RCP8.5) radiative forcing scenario, CanESM2 shows the largest change in global-mean SST and MPI the smallest (3.07 ºC and 1.94 ºC, respectively; Table 1 and Fig. 15). All models simulate intensified warming in the equatorial Pacific and greater SST increases in the northern hemisphere compared to the southern hemisphere. However, they differ in their representation of the North Atlantic "warming hole" response and sign of the SST response in the Weddell and Amundsen-Bellinghausen Seas (Fig. 15). The amplitude of global-mean SST warming does not appear to be directly correlated with the magnitude or spatial extent of significant future changes in MHW and MCW intensity and duration across models (compare Fig. 15 with Figs. 8 and 10).

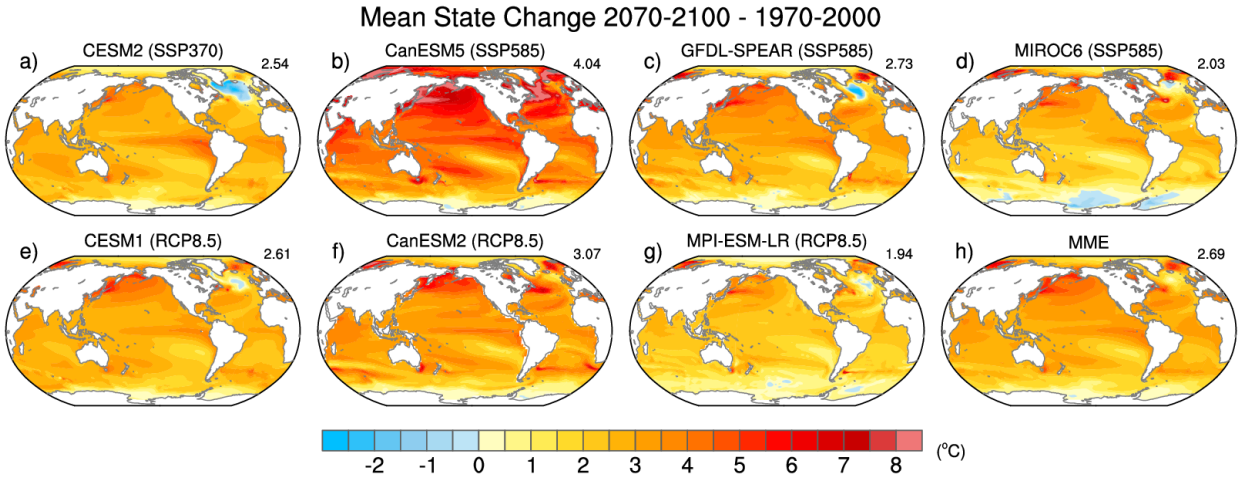

Figure 15. Ensemble-mean SST differences (ºC) differences between 2070-2100 and 1970-2000 for each model Large Ensemble: (a) CESM2; (b) CanESM5; (c) GFDL-SPEAR; (d) MIROC6; (e) CESM1; (f) CanESM2; (g) MPI-ESM-LR; (h) multi-model ensemble (MME) average. The text in parentheses after each model name indicates the radiative forcing scenario applied, and the number in the upper right of each panel denotes the global-mean SST difference.

How large are the relative contributions of changes in variability *vs.* changes in mean state to projected changes in MHW and MCW characteristics? Figure 16 shows maps of the ratio of late-century changes in MHW and MCW composite intensity due to changes in variability divided by those due to changes in variability-plus-mean state for each model LE (e.g., Fig. 8 divided by the sum of Figs. 8 and 15). The number at the upper right of each map indicates the fractional area of the world ocean containing ratios within ±0.1. As can be seen, the models range from 82-91% in this metric for MHW and from 83-90% for MCW. Thus, mean state changes dominate future changes in MHW and MCW composite intensity over most of the world ocean. The Arctic and

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

Antarctic are exceptions, where ratios reach up to ±0.5 for MHW and even higher for MCW in many models, indicating that contributions from changes in variability are comparable to those from changes in the mean state for projected late-century changes in MHW and MCW composite intensity. This is likely due to melting of sea ice and subsequent exposure of the ocean to atmosphere, which then allows for variability in heat exchange across the air-sea interface. Other regions with relatively large contributions from changes in variability include the subpolar North Atlantic, an area influenced by sea ice-related changes in density and deep ocean convection. More modest contributions from changes in variability are found over portions of the tropical Pacific and along the Kuroshio-Oyashio Front in some models, with values up to about ±0.25. The relative contribution of changes in variability tends to be larger for mid-century changes in MHW and MCW composite intensity (Fig. S43) than late-century ones, likely a consequence of the high-emissions scenarios used to force the models. In some models, this contribution can reach up to ±0.4 in parts of the equatorial Pacific and up to ±0.8 in the Arctic (Fig. S43). The fractional area of the world ocean containing ratios within ±0.1 is also correspondingly lower for mid-century changes, ranging from 72-86% for MHW and 69-86% for MCW across models.

To further illustrate the relative contributions from changes in the mean state *vs.* variability, we show histograms of area-averaged iSST from the CESM2 LE for three regions of interest: Arctic (poleward of 67°N), Western Tropical Pacific (8°S-6°N, 155°E-175°W) and Northeast Atlantic (35-62°N, 30-0°W) in Figs. 17a,c,e, respectively. For each regional average, we consider only those grid boxes whose late-century MHW composite intensity changes are statistically significant and of the same sign (positive in the case of the Arctic and Northeast Atlantic, and negative in the case of the Western Tropical Pacific; regional boundaries are outlined in Figs. 6e,j). The distributions shown in Figs. 17a,c,e are based on all monthly iSST samples (not just MHW and MCW samples) from all 100 ensemble members during 1970-2000 (gray bars) and 2070-2100 (blue bars); the purple bars are based on the same set of samples as the blue bars, but with the ensemble-average mean state change (2070-2100 minus 1970-2000) added back in. The 10th and 90th percentiles of each distribution are shown as vertical solid lines, and the 50th percentile as a vertical dashed line.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

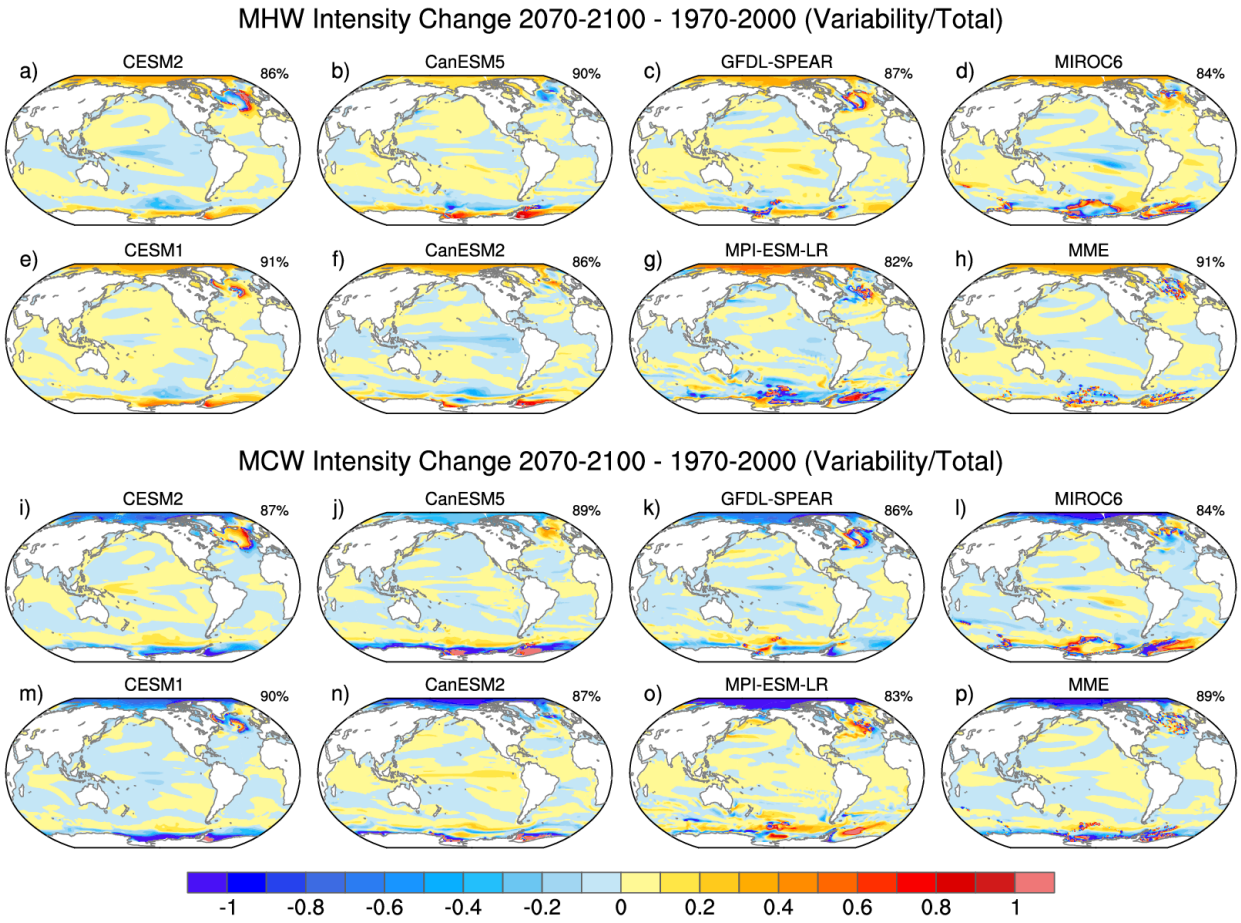

Figure 16. Ratio of late-century changes (2070-2100 minus 1970-2000) in MHW and MCW composite intensity due to changes in variability divided by that due to changes in variability-plus-mean state for: (a,h) CESM2; (b,i) CanESM5; (c,j) GFDL-SPEAR; (d,k) MIROC6; (e,l) CESM1; (f,m) CanESM2; (g,n) MPI-ESM-LR. Panels h,p show the multi-model ensemble (MME) average. The number in the upper right of each panel denotes the fractional area (%) of values within the range -0.1 to +0.1.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

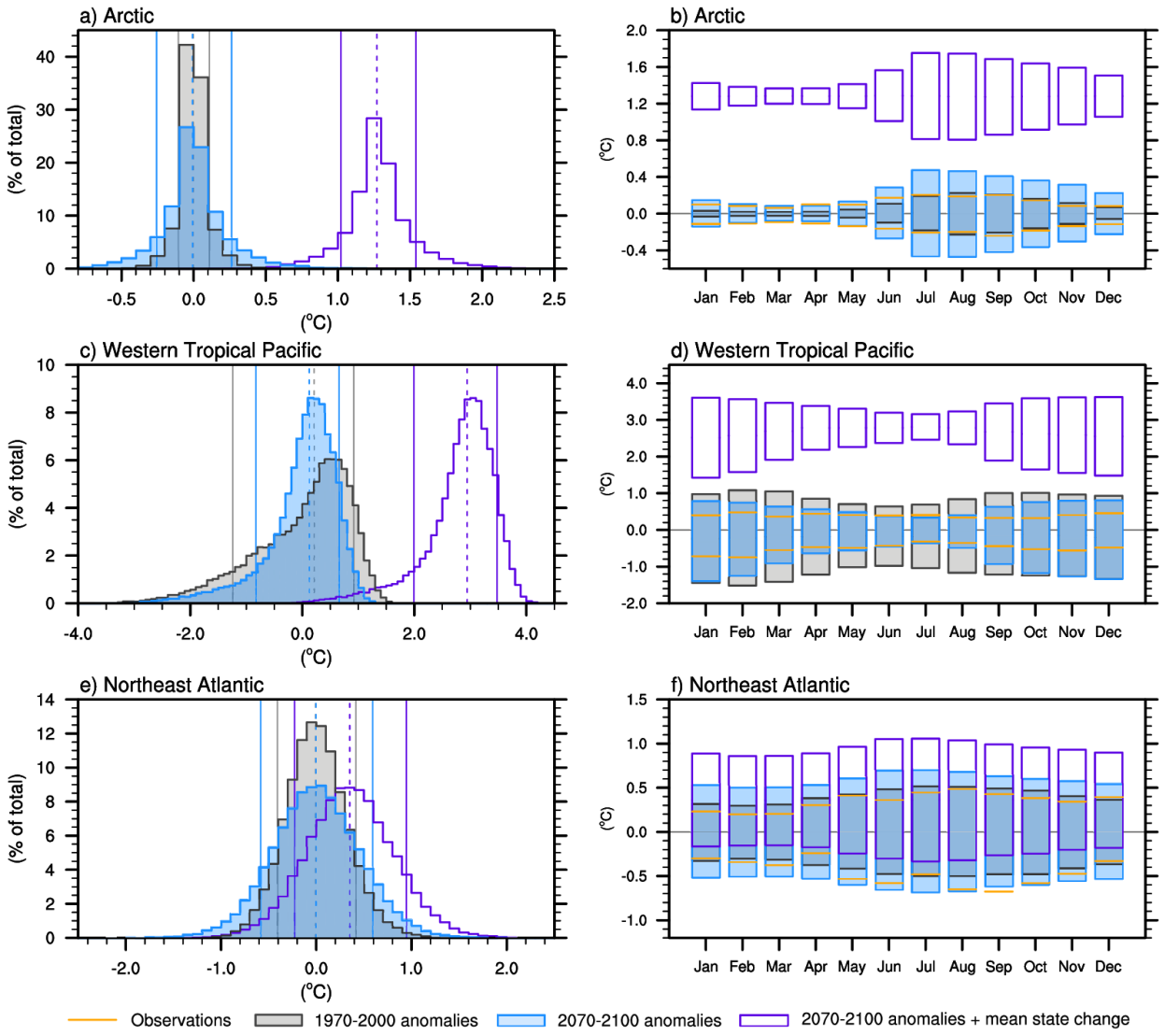

Figure 17. Histograms of area-averaged SST (ºC) from the CESM2 Large Ensemble for (a) Arctic (poleward of 67°N), (c) Western Tropical Pacific (8°S-6°N, 155°E-175°W) and (e) Northeast Atlantic (35-62°N, 30-0°W) based on all months from all 100 ensemble members during 1970-2000 (gray) and 2070-2100 (blue) after removing the ensemble-mean climatological seasonal cycle for each period (see text for details). Purple histograms are the same as the blue histograms but with the mean state change (2070-2100 minus 1970-2000) added back in. The 10th and 90th percentiles of each distribution are shown as vertical solid lines, and the 50th percentile is shown as a vertical dashed line. (b,d,f) Seasonal variation of the 10th and 90th percentiles of the distributions shown in (a,c,e). Observed values for the period 1970-2000 are denoted in orange.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

In the Arctic, there is a notable widening of the iSST distribution in late-century compared to present-day (blue *vs.* gray bars), and this widening is approximately symmetric with respect to sign, as measured by the difference between the 10th percentiles of the future and present-day distributions (0.15 ºC) and between the 90th percentiles of the two distributions (0.16 ºC), see Fig. 17a. Adding in the mean state change shifts the blue distribution by 1.27 ºC, and results in a complete separation of the present-day (gray bars) and late-century (purple bars) histograms (Fig. 17a). Unlike the Arctic, the present-day Western Tropical Pacific iSST distribution is distinctly skewed, with a longer tail of negative values compared to positive ones (gray bars in Fig. 17c). We speculate that this negative skewness is a reflection of the fact that La Niña events extend farther west than El Niño events, both in nature and the model (Capotondi et al. 2020). This distribution narrows in late-century, with a larger change at the lower end than the upper end (the 10th percentile increases by 0.41 ºC while the 90th percentile decreases by 0.25 ºC): however, the negative skewness is maintained (blue bars in Fig. 17c). Adding in the mean state change shifts the blue distribution by 2.2 ºC (purple bars in Fig. 17c). Unlike the Arctic, the Western Tropical Pacific distributions show some overlap between present-day (gray bars) and late-century (purple bars) due to the long negative tail in the future histogram. Like the Arctic, the Northeast Atlantic iSST distributions widen in the future, and this widening is approximately symmetric with respect to sign (0.18 ºC for the 10th percentiles and 0.19 ºC for the 90th percentiles; Fig. 17e). However, due to the pronounced "warming hole" response in this region (recall Fig. 15a), the mean state change shifts the distribution only slightly (0.18 ºC), resulting in a large overlap between present-day and late-century (Fig. 17e, gray *vs.* purple bars). In summary, the three regions show distinctive signatures of future change. In the Arctic, projected changes in variability significantly widen the SST distribution, while those in the mean shift the distribution to an entirely new climatic state (e.g., values that are warmer than any during present-day). In the Western Tropical Pacific, projected changes in variability narrow the negatively-skewed distribution (and slightly reduce the skewness), while changes in the mean state shift the distribution to warmer but not completely unprecedented values. In the Northeast Atlantic, the projected widening of the distribution is equal to the magnitude of the mean state shift.

As stated in Section 2.2, our thresholds for defining MHW and MCW events (based on the 10th and 90th percentiles of the iSST distribution in a given time period, respectively) are computed for each month separately. Here, we briefly show the seasonal variation in these thresholds for the

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

three regions selected above (Figs. 17b,d,f). In the Arctic, the 10th and 90th percentiles of the iSST distribution during 1970-2000 exhibit pronounced seasonal cycles, with minimum amplitudes (± 0.02 ºC) during January-April and largest magnitudes during June-November (maximum of ± 0.22 ºC in August); Fig. 17b (gray bars). The model's seasonal variation is similar to that in observations, although the minimum values are underestimated (Fig. 17b, orange bars). This seasonal variation is likely associated with the amount of open water available for exchanging heat with the atmosphere (i.e., during the months of January-April when sea ice cover is typically around 100%, the water directly beneath the ice is buffered from air-sea interaction and hence is able to maintain a relatively constant temperature). A similar seasonal variation in the 10th and 90th percentiles is found during 2070-2100, but with considerably larger amplitudes compared to present-day (Fig. 17b, blue bars). For example, winter values increase to around ± 0.1 ºC while late summer values reach to nearly ± 0.5 ºC. There is no overlap in any month between the present-day percentiles and the late-century percentiles that include mean-state changes (Fig. 17b, gray and purple bars).

The present-day MHW and MCW thresholds in the Western Tropical Pacific also vary seasonally: the 90th percentile reaches maximum values (1.0-1.1 ºC) from September-March and minimum values from May-July (0.6-0.7 ºC); a similar timing is seen for the 10th percentile but the amplitudes are larger by about 0.4 ºC, reflecting the negative skewness in the full iSST distribution noted earlier (Fig. 17d, gray bars). While negative skewness is also found in observations, the model overestimates the 10th and 90th percentiles in all months, consistent with the model biases shown in Figs. 1c,i. Unlike the Arctic, the seasonal cycle of the projected decrease in threshold magnitudes is inversely related to the present-day seasonal cycle, with maximum decreases from June-August (0.7 ºC and 0.45 ºC for the 10th and 90th percentiles, respectively; Fig. 17d, gray and blue bars). Like the Arctic, the present-day percentiles and the late-century percentiles that include mean-state changes are completely separate in every month (Fig. 17d, gray and purple bars). Compared to the other regions, the Northeast Atlantic shows a muted seasonal cycle of future change in MHW and MCW threshold amplitudes, ranging from 0.2-0.3 ºC (compare gray and blue bars in Fig. 17f), with considerable overlap between the present-day and future distributions in every month (compare gray and purple bars in Fig. 17f). The model realistically simulates present-day thresholds in almost every month (compare orange and gray bars in Fig. 17f).

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

Our MHW and MCW composites convey the average intensity and duration across all samples in a given time period. In Fig. 18, we show the range of intensity values across all individual MHW and MCW events that make up these composite values for each of the three selected regions. The 5^{th}-to-95^{th} percentile range of intensity values for individual MHW events in the Arctic increases from 0.02-0.35 ºC during 1970-2000 to 0.10-0.79 ºC during 2070-2100 (Fig. 18a, gray and blue bars, respectively). The positive skewness in the future intensity distribution of individual MHW events is particularly pronounced. Individual MCW events are largely the mirror image of MHW events, with notable negative skewness in the future compared to present-day (Fig. 18b, gray and blue bars, respectively). The present-day and future distributions of individual MHW (Fig. 18a) and MCW (Fig. 18b) event intensity remain clearly separated even when mean state changes are included (compare gray and purple bars).

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

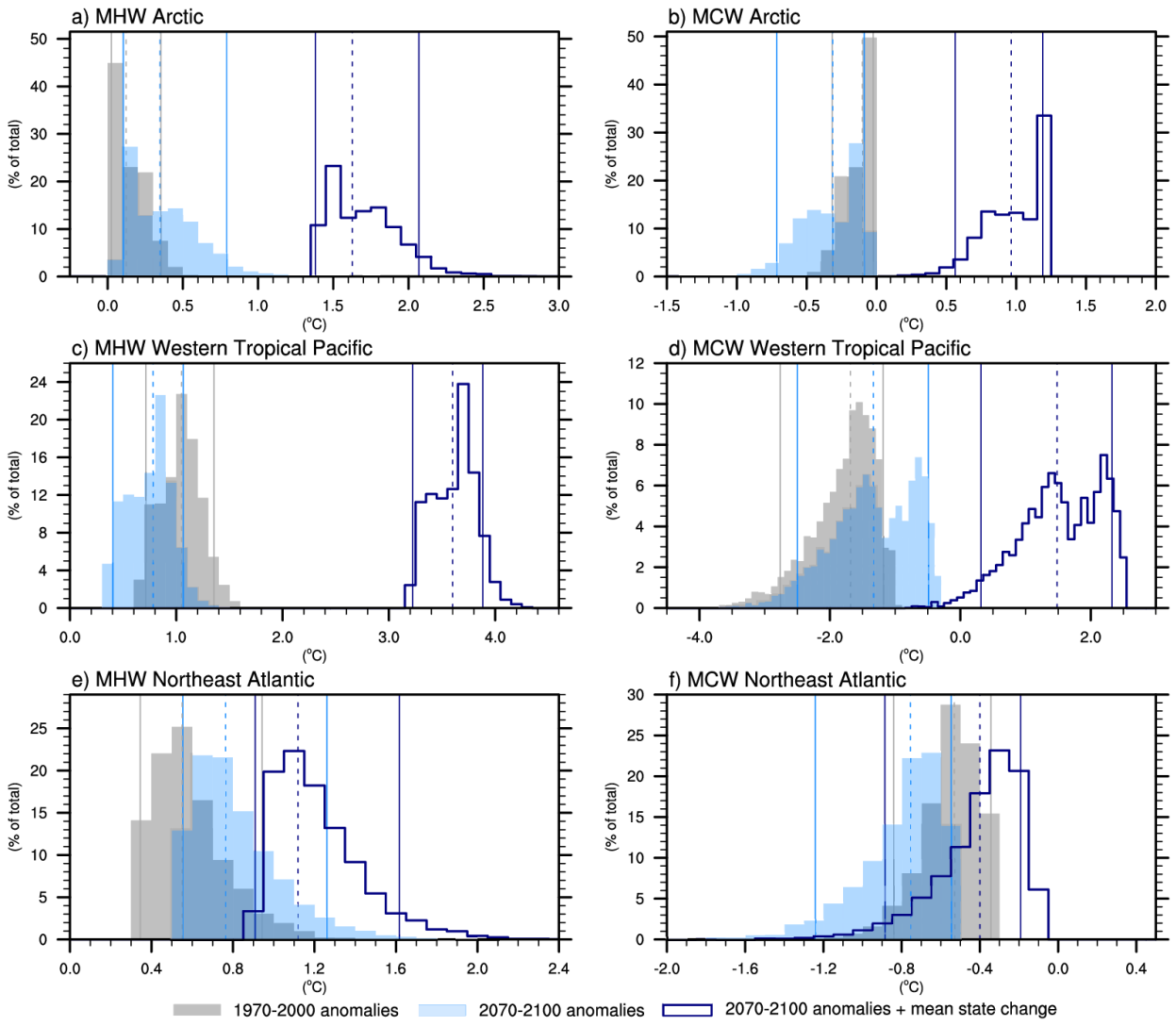

Figure 18. As in Fig. 17 but for (a,c,e) MHW and (b,d,f) MCW intensity (°C) based on all monthly samples. The 5th and 95th percentiles of each distribution are shown as vertical solid lines, and the 50th percentile is shown as a vertical dashed line.

In the Western Tropical Pacific, the distribution of individual MHW event intensity shifts toward lower values in the future compared to present-day, even without mean state changes included (Fig. 18c, gray and blue bars). While this result is somewhat counter-intuitive, it is consistent with the behavior of the upper half of the full iSST distribution shown in Fig. 17b. A similar future shift (toward less negative values) is found for the corresponding distributions of individual MCW events, with a hint of a bimodal shape in the future distribution (Fig. 18d, gray and blue bars). The present-day and future distributions of individual event intensity remain distinct when mean state changes are included (gray and purple bars in Figs. 18c,d).

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

In the Northeast Atlantic, the individual MHW event intensity distributions have a similar degree of positive skewness in present-day and late-century (Fig. 18e, gray *vs.* blue bars). While the MHW distribution is shifted toward warmer values in the future, there remains some overlap with the present-day distribution (Fig. 18e, gray *vs.* purple bars). Analogous behavior is seen for the individual MCW event distributions, with an even greater overlap between future and present-day and negative skewness (Fig. 18f). In summary, the Arctic, Western Tropical Pacific and Northeast Atlantic show diverse behavior in their MHW and MCW intensity distributions, including shape (degree of normality), seasonal dependence and future change as a result of changes in variability in addition to changes in the mean state.

4. Summary and discussion

We have investigated future changes in the intensity and duration of composite MHWs and MCWs based on monthly SSTs in seven coupled model initial-condition Large Ensembles (LEs) over the period 1970-2100 under a range of radiative forcing scenarios. The large number of simulations (30-100) available for each LE provides voluminous samples of MHW and MCW events, allowing for robust quantification of their projected evolution as a result of changes in variability and comparison with changes in the mean state in each model. We find that mean state changes are much larger than variability-related changes in MHW and MCW characteristics, consistent with previous work based on daily SSTs from the CMIP archives (e.g., Frolicher et al. 2018; Oliver et al. 2019; Yao et al. 2022) and a high-resolution coupled model simulation (Guo et al. 2022). However, we also identify widespread statistically significant contributions from projected changes in variability in every model. These changes in variability generally account for $< 10\%$ of the total change in intensity by late-century, except over the Arctic and Antarctic where they contribute up to 50-75%. By mid-century in some models, changes in variability contribute up to 40% in parts of the equatorial Pacific and up to 80% in the Arctic. For ecosystems that have some capacity to adapt to the slowly-evolving SST mean state, variability-induced changes in MHW/MCW properties may pose additional challenges in the coming decades (Guo et al. 2022; Oliver et al. 2019; Oliver et al. 2021).

The patterns and magnitudes of projected forced changes in MHW and MCW intensity and duration due to variability are complex and model dependent. While it is difficult to generalize, the majority of LEs show that future changes in variability will cause MHW and MCW events to

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

intensify in the extra-tropics (excluding the Southern Ocean) and weaken in the tropics and Southern Ocean, although the regional details and spatial extent of these changes differ across models. The majority of LEs also show that future changes in variability will generally cause MHW and MCW events to significantly shorten in duration, although these changes are less geographically widespread than those in intensity. However, there is considerable model disagreement in the sign of significant duration change over the Arctic, the North and South Atlantic and the South Indian Ocean. Future changes in intensity and duration of composite MHWs and MCWs are generally symmetric in all models, except for the longitude of maximum duration change in the tropical Pacific. This asymmetry is likely related to the different spatial and temporal characteristics of El Niño and La Niña events (e.g., Okumura and Deser, 2010; Capotondi et al. 2020). There does not appear to be a direct correspondence between models' amplitude of global-mean SST warming and their magnitude and spatial extent of forced changes in MHW and MCW characteristics due to variability.

Projected changes in ENSO variability have a large impact on future changes in MHWs and MCWs worldwide in every model. All LEs show reduced areal coverage of significant future changes in MHW and MCW intensity and duration during ENSO-neutral states compared to all states (late-century reductions of 10-40% for intensity and 42-84% for duration, depending on the model). Relatedly, the areal fraction of the world oceans with at least two-thirds model consensus on the sign of significant late-century change in MHW and MCW intensity (duration) decreases from 37% (28%) in the all-sample composites to 20% (1%) in the ENSO-neutral composites. Thus, future changes in ENSO exert a major control on the future evolution of MHW and MCW characteristics, underscoring the importance of reducing model spread in ENSO projections, which is considerable and poorly understood (Maher et al. 2023). We note that the collection of models used here encompasses a range of projected ENSO behavior, from little change in the MPI-ESM-LR model to decreasing ENSO variance in CanESM2 to non-monotonically increasing ENSO variance in the remaining models (Maher et al. 2023). The physical mechanisms underlying the diversity of projected ENSO behavior across models remain to be elucidated, although we note that ENSO results from a sensitive balance of competing effects and its response to anthropogenic emissions may likewise depend sensitively on the relative magnitudes and rates of change of different factors including the Bjerknes feedback, atmospheric damping, upper ocean stratification and thermocline depth (see Cai et al. 2021 for a recent review). Without a sound understanding of

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

how and why ENSO may change under anthropogenic emissions, this uncertainty is unlikely to be narrowed.

A potential caveat of our findings is the fact that every model LE shows statistically significant biases in composite MHW and MCW intensity and duration defined on the basis of monthly SSTs over the historical period 1950-2020. These biases are widespread in every model, covering ~50-70% of the area of the global oceans for intensity and ~35-55% for duration. The majority of model LEs significantly overestimate intensity in the tropical western Pacific and Indian Oceans and across portions of the Southern Ocean, and significantly underestimate it in the Arctic. The models show less consistency in their duration biases both in terms of sign and location, although the North Pacific is a region of general overestimation. The physical mechanisms underlying these model biases remain to be uncovered, but deficiencies in the representation of tropical and extra-tropical modes of variability including ENSO and its teleconnections (Fasullo et al. 2020; Maher et al. 2023) and inadequacies in the simulation of mean mixed layer depth (Guo et al. 2022) and sea ice properties (SIMIP community, 2020) are likely to play a role. An important aspect of our assessment of model bias is the consideration of sampling fluctuations inherent in the short (71-year) historical record. All model LEs exhibit substantial ensemble spread in historical MHW/MCW composites, underscoring the importance of having LEs for robust statistics and cautioning against over-interpretation of the observational record, which may be subject to similar levels of uncertainty due to limited sampling.

Another potential caveat of our study is the relatively coarse spatial resolution (approximately 1°-2°) of the model LEs, which limits their ability to represent mesoscale ocean processes and small-scale air-sea interactions (Pilo et al. 2019; Hayashida et al. 2020; Guo et al. 2022). In particular, the dynamically-downscaled ocean model experiments of Pilo et al. (2019) and Hayashida et al. (2020), and the fully coupled high-resolution modeling experiment of Guo et al. (2022), demonstrate that explicit representation of ocean mesoscale processes improves the simulation of historical MHW characteristics based on daily SSTs along the western boundary currents, the Antarctic circumpolar current and other eddy-rich regions of the world oceans. Some of the improvements are also attributable to the better simulation of mixed layer depth (Guo et al. 2022). The degree to which spatial resolution affects the simulation of MHW/MCW characteristics based on monthly SSTs remains to be assessed. The enhanced fidelity of the high-resolution models in turn lends confidence to their future projections, although it does not guarantee that

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

modes of variability such as ENSO and their evolution under global warming are also better simulated.

Our study leaves many important questions for future investigation as follows. What is the seasonal dependence of future changes in MHW and MCW intensity and duration across model LEs? Are future changes larger and more robust in summer than winter due to the shallower depth of the upper ocean mixed layer? What role does ENSO play in setting the seasonality of future changes in MHW and MCW characteristics, and are there any asymmetric influences between El Nino and La Nina events? Do more sophisticated approaches to removing ENSO influences such as multivariate linear inverse modeling (Zhao et al. 2021) reveal additional insights into the role of future changes in ENSO on MHWs and MCWs? Are there additional insights to be gained on the future behavior of MHW and MCW events by using daily data? What is the subsurface structure of the projected changes in intensity and duration, and does it differ between MHWs and MCWs? What are the physical mechanisms governing projected changes in MHW and MCW variability in different locations, and do the dominant processes depend on whether ENSO is playing an active role? Do future changes in large-scale modes of atmospheric circulation variability such as those documented in O'Brien and Deser (2023) affect the future evolution of MHW and MCW characteristics? How do projected changes in MHW and MCW characteristics found here relate to future changes in variance and persistence of daily and monthly SST anomalies documented in Li and Thompson (2021) and Shi et al. (2022)? We hope to address many of these outstanding questions in future studies by leveraging the unique set of information contained in Earth System Model initial-condition Large Ensemble archives.

Acknowledgements

We thank the three anonymous reviewers for their constructive comments and suggestions. We acknowledge the efforts of all those who contributed to producing the model simulations and observational datasets used in this study. A.C. and A.P were supported by the NOAA Climate Program Office Modeling, Analysis, Predictions and Projections (MAPP) program. NCAR is sponsored by the National Science Foundation under Cooperative Agreement 1852977.

Data Availability Statement

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

The ERSSTv5 dataset is publicly available from: www.ncei.noaa.gov/products/extended-reconstructed-sst. All model simulations used in this study are publicly available from: https://www.cesm.ucar.edu/community-projects/mmlea and https://esgf-node.llnl.gov/search/cmip6/. All analyses were performed using NCL; analysis codes are available from Adam S. Phillips upon request.

References

Alexander, M. A., I. Blade, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, 2002: The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air-Sea Interaction Over the Global Oceans. *J. Climate* , **15**, 2205-2231.

Alexander MA, JD Scott, KD Friedland, KE Mills, JA Nye, AJ Pershing, AC Thomas, 2018: Projected sea surface temperatures over the 21st century: Changes in the mean, variability and extremes for large marine ecosystem regions of Northern Oceans. *Elementa: Science of the Anthropocene*, **6(1):9**, DOI: http://doi.org/10.1525/elementa.191

Amaya DJ, MA Alexander, A Capotondi, C Deser, KB Karnauskas, AJ Miller, and NJ Mantua (2021). Are Long-Term Changes in Mixed Layer Depth Influencing North Pacific Marine Heatwaves? [in "Explaining Extremes of 2019 from a Climate Perspective"]. *Bull. Amer. Meteor. Soc.*, 102 (1), S59–S66, https://doi.org/10.1175/ BAMS-D-20-0144.1.

Amaya DJ, MA Alexander, MG Jacox, JD Scott, C Deser, A Capotondi and A Phillips, 2023a: Bottom marine heatwaves along the continental shelves of North America. *Nat. Comm.* 14, 1038. https://doi.org/10.1038/s41467-023-36567-0.

Amaya DJ, MG Jacox, MR Fewings, VS Saba, MF Stuecker, RR Rykaczewski, et al., 2023b: Marine heatwaves need clear definitions so coastal communities can adapt. *Nature*, *616*(7955), 29–32. https://doi.org/10.1038/d41586-023-00924-2.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

Armour, K., Marshall, J., Scott, J., A. Donohoe, and E. Newsom, 2016: Southern Ocean warming delayed by circumpolar upwelling and equatorward transport. *Nature Geosci.* **9**, 549–554fArmou. https://doi.org/10.1038/ngeo2731

Burger, F.A., J. Terhaar, J. and R.L. Frölicher, 2022: Compound marine heatwaves and ocean acidity extremes. *Na.t Commun.* 13, 4722 (2022). https://doi.org/10.1038/s41467-022-32120-7.

Cai, W., Santoso, A., Collins, M. et al., 2021: Changing El Niño–Southern Oscillation in a warming climate. *Nat Rev Earth Environ* **2**, 628–644. https://doi.org/10.1038/s43017-021-00199-z

Capotondi, A., M. A. Alexander, N. A. Bond, E. N. Curchitser, and J. D. Scott, 2012: Enhanced upper ocean stratification with climate change in the CMIP3 models. *J. Geophys. Res*., 117, C04031, doi:10.1029/2011JC007409.

Capotondi, A., C. Deser, A. S. Phillips, Y. Okumura and S. M. Larson, 2020: ENSO and Pacific Decadal Variability in the Community Earth System Model Version 2. *J. Adv. Model. Earth Sy.*, **12**, e2019MS002022, doi:10.1029/2019MS002022.

Cheung, W. W. L., and Frölicher, T. L. (2020). Marine heatwaves exacerbate climate change impacts for fisheries in the northeast Pacific. *Sci. Rep*. 10, 6678. doi: 10.1038/s41598-020-63650-z.

Danabasoglu, G., J. -F. Lamarque, J. Bacmeister, D. A. Bailey, A. K. DuVivier, J. Edwards, L. K. Emmons, J. Fasullo, R. Garcia, A. Gettelman, C. Hannay, M. M. Holland, W. G. Large, D. M. Lawrence, J. T. M. Lenaerts, K. Lindsay, W. H. Lipscomb, M. J. Mills, R. Neale, K. W. Oleson, B. Otto-Bliesner, A. S. Phillips, W. Sacks, S. Tilmes, L. van Kampenhout, M. Vertenstein, A. Bertini, J. Dennis, C. Deser, C. Fischer, B. Fox-Kemper, J. E. Kay, D. Kinnison, P. J. Kushner, M. C. Long, S. Mickelson, J. K. Moore, E. Nienhouse, L. Polvani, P. J. Rasch and W. G. Strand, 2020: The Community Earth System Model version 2 (CESM2). *J. Adv. Model. Earth Sy.*, **12**, e2019MS001916, doi:10.1029/2019MS001916.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., et al. (2020). SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001895. https://doi.org/10.1029/2019MS001895.

Deser, C., F. Lehner, K. B. Rodgers, T. Ault, T. L. Delworth, P. N. DiNezio, A. Fiore, C. Frankignoul, J. C. Fyfe, D. E. Horton, J. E. Kay, R. Knutti, N. S. Lovenduski, J. Marotzke, K. A. McKinnon, S. Minobe, J. Randerson, J. A. Screen, I. R. Simpson and M. Ting, 2020: Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change*, **10**, 277-286, doi:10.1038/s41558-020-0731-2.

Deser, C. and A. S. Phillips, 2023: Spurious Indo-Pacific connections to internal Atlantic Multidecadal Variability introduced by the global temperature residual method. *Geophys. Res. Lett.*, **50**, e2022GL100574, doi: 10.1029/2022GL100574.

Fasullo, J., A. S. Phillips and C. Deser, 2020: Evaluation of Leading Modes of Climate Variability in the CMIP Archives. *J. Climate*, **33**, 5527–5545, doi:10.1175/JCLI-D-19-1024.1.

Frölicher, T. L., Fischer, E. M., and Gruber, N. (2018). Marine heatwaves under global warming. *Nature* 560, 360–364. doi: 10.1038/s41586-018-0383-9

Frölicher, T. L. & Laufkötter, C. Emerging risks from marine heat waves. *Nat Commun* **9**, 650 (2018).

Hayashida, H., Matear, R.J., Strutton, P.G. and X. Zhang, 2020: Insights into projected changes in marine heatwaves from a high-resolution ocean circulation model. *Nat Commun* **11**, 4352 (2020). https://doi.org/10.1038/s41467-020-18241-x

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., et al. (2016). A hierarchical approach to defining marine heatwaves. *Prog. Oceanogr*. 141, 227–238. doi: 10.1016/j.pocean.2015.12.014

Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., et al. (2019). A global assessment of marine heatwaves and their drivers. *Nat. Commun*. 10, 2624. doi: 10.1038/s41467-019-10206-z

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., et al. (2020). Improvements of the daily optimum interpolation sea surface temperature (DOISST) Version 2.1. *J. Clim*. 34, 1–47. doi: 10.1175/JCLI-D-20-0166.1

Jacox, M. G. (2019), Marine heatwaves in a changing climate, *Nature*, 571, 485-487, doi:10.1038/d41586-019-02196-1.

Jacox, M. G., Alexander, M. A., Bograd, S. J. & Scott, J. D. Thermal displacement by marine 497 heatwaves. *Nature* 584, 82–86 (2020).

Kay, J. E., C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. Arblaster, S. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner, J. -F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Met. Soc*., 96, 1333–1349, doi: 10.1175/BAMS-D-13-00255.1.

Kirchmeier-Young, M. C., Zwiers, F. W. and Gillett, N. P. , 2017: Attribution of extreme events in Arctic Sea ice extent. *J. Climate* 30, 553–571.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton, A., Li, H., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., Tagliabue, A., Takano, Y., Tjiputra,

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., and Ziehn, T.: Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, *Biogeosciences*, 17, 3439–3470, https://doi.org/10.5194/bg-17-3439-2020, 2020.

Laufkötter, C., Zscheischler, J., and Frölicher, T. L. (2020). High-impact marine heatwaves attributable to human-induced global warming. *Science* 369, 1621–1625. doi: 10.1126/science.aba0690

Le Grix, N., Zscheischler, J., Rodgers, K. B., Yamaguchi, R., and Frölicher, T. L., 2022: Hotspots and drivers of compound marine heatwaves and low net primary production extremes, *Biogeosciences*, 19, 5807–5835, https://doi.org/10.5194/bg-19-5807-2022.

Li, J., D. W. J. Thompson, 2021: Widespread changes in surface temperature persistence under climate change. *Nature* 599, 425-430. https://doi.org/10.1038/s41586-021-03943-z.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., et al. (2019). The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *J. Adv. Mod. Earth Sys*., 11, 2050– 2069. https://doi.org/10.1029/2019MS001639

Maher, N., Wills, R. C. J., DiNezio, P., Klavans, J., Milinski, S., Sanchez, S. C., Stevenson, S., Stuecker, M. F., and Wu, X.: The future of the El Niño–Southern Oscillation: using large ensembles to illuminate time-varying responses and inter-model differences, *Earth Syst. Dynam*., 14, 413–431, https://doi.org/10.5194/esd-14-413-2023, 2023.

Manabe, S., R. J. Stouffer, M. J. Spelman, and K. Bryan, 1991: Transient Responses of a Coupled Ocean–Atmosphere Model to Gradual Changes of Atmospheric CO2. Part I. Annual Mean Response. *J. Climate*, **4**, 785–818, https://doi-org.cuucar.idm.oclc.org/10.1175/1520-0442(1991)004<0785:TROACO>2.0.CO;2.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

NOAA National Centers for Environmental Information (2023): State of the Climate: Global Climate Report for 2022. Accessed December 21, 2023, from https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202213.

Newman, M., M. A. Alexander, T. R. Ault, K. M. Cobb, C. Deser, E. Di Lorenzo, N. J. Mantua, A. J. Miller, S. Minobe, H. Nakamura, N. Schneider, D. J. Vimont, A. S. Phillips, J. D. Scott, and C. A. Smith, 2016: The Pacific decadal oscillation, revisted. *J. Climate*, **29**, 4399-4427, doi:10.1175/JCLI-D-15-0508.1.

Okumura, Y. M., C. Deser, 2010: Asymmetry in the Duration of El Nino and La Nina. *J. Climate*, **23**, 5826-5843, 10.1175/2010JCLI3592.1.

Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., et al. (2018). Longer and more frequent marine heatwaves over the past century. *Nat. Commun*. 9, 1324. doi: 10.1038/s41467-018-03732-9

Oliver, E. C. J. (2019). Mean warming not variability drives marine heatwave trends. *Clim. Dyn*. 53, 1653–1659. doi: 10.1007/s00382-019-04707-2

Oliver, E. C. J., Benthuysen, J. A., Darmaraki, S., Donat, M. G., Hobday, A. J., Holbrook, N. J., et al. (2021). Marine heatwaves. *Annu. Rev. Mar. Sci*. 13, 313–342. doi: 10.1146/annurev-marine-032720-095144

Pilo, G. S., Holbrook, N. J., Kiss, A. E., and Hogg, A. M., 2019: Sensitivity of marine heatwave metrics to ocean model resolution. *Geophysical Research Letters*, 46, 14604–14612. https://doi.org/10.1029/2019GL084928

Plecha, S. M. and P. M. M. Soares, 2020: Global marine heatwave events using the new CMIP6 multi-model ensemble: from shortcomings in present climate to future projections. *Environ. Res. Lett.* **15** 124058**DOI** 10.1088/1748-9326/abc847

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

Rayner, N., D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice and night marine air temperatures ince the late nineteenth century. *J. Geophys. Res.,* 108, https://doi.org/10.1029/2002JD002670..

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., Edwards, J., Kim, J.-E., Simpson, I., Stein, K., Stuecker, M. F., Yamaguchi, R., Bodai, T., Chung, E.-S., Huang, L., Kim, W., Lamarque, J.-F., Lombardozzi, D., Wieder, W. R., and Yeager, S. G., 2021: Ubiquity of human-induced changes in climate variability, *Earth Syst. Dynam*, https://doi.org/10.5194/esd-2021-50, 2021.

Scannell, H. A., Johnson, G. C., Thompson, L., Lyman, J. M. & Riser, S. C., 2020: Subsurface evolution and persistence of marine heatwaves in the northeast pacific. *Geophys. Res. Lett.* **47**, e2020GL090548.

SIMIP community, 2020: Arctic sea ice in CMIP6. *Geophysical Research Letters*, *47*(10), e2019GL086749.

Sen Gupta A, Thomsen M, Benthuysen JA, Hobday AJ, Oliver E, Alexander LV, Burrows MT, Donat MG, Feng M, Holbrook NJ, Perkins-Kirkpatrick S, Moore PJ, Rodrigues RR, Scannell HA, Taschetto AS, Ummenhofer CC, Wernberg T, Smale DA., 2020: Drivers and impacts of the most extreme marine heatwaves events. *Sci Rep*. doi: 10.1038/s41598-020-75445-3.

Smale, D. A. *et al.* Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nature Climate Change* (2019) doi:10.1038/s41558-019-0412-1.

Smith, K., M. Burrows, A. Hobday, N. King, P. Moore, A. Sen Gupta, M. Thomsen, T. Wernberg and D. Smale, 2023: Biological Impacts of Marine Heatwaves, *Annual Review of Marine Science*, **15**, 1, (119-145), (2023).https://doi.org/10.1146/annurev-marine-032122-121437

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0278.1.

Schaeffer, A., Sen Gupta, A. and Roughan, M., 2023: Seasonal stratification and complex local dynamics control the sub-surface structure of marine heatwaves in Eastern Australian coastal waters. *Commun Earth Environ* **4**, 304. https://doi.org/10.1038/s43247-023-00966-4

Schlegel, R. W., Darmaraki, S., Benthuysen, J. A., Filbee-Dexter, K., and Oliver, E. C., 2021: Marine cold-spells. *Prog. in Oceanog*., 102684. doi:10.1016/j.pocean.2021.102684

Shi H, J Fei-Fei, RCJ Wills, MG Jacox, DJ Amaya, BA Black, RA Rykaczewski, SJ Bograd, M Garcia-Reyes, and WJ Sydeman. Global decline in ocean memory over the 21st century. *Sci. Adv.* 8 (18), eabm3468. https://doi.org/10.1126/sciadv.abm3468.

Student (1917). Tables for estimating the probability that the mean of a unique sample of observations lies between -∞ and any given distance of the mean of the population from which the sample is drawn. *Biometrika* XV, 414-417.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), *Geosci. Model Dev*., 12, 4823– 137 4873, https://doi.org/10.5194/gmd-12-4823-2019, 2019.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'Ishi, R., ... Kimoto, M. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, 12(7), 2727-2765. https://doi.org/10.5194/gmd-12-2727-2019.

Thoral, F., Montie, S., Thomsen, M.S., 2022: Unravelling seasonal trends in coastal marine heatwave metrics across global biogeographical realms. *Sci Rep* **12**, 7740.https://doi.org/10.1038/s41598-022-11908-z

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.

Vogt, L., F. Burger, S. Griffies and T. Frölicher, 2022, Local Drivers of Marine Heatwaves: A Global Analysis With an Earth System Model", *Frontiers in Climate*, https://www.frontiersin.org/articles/10.3389/fclim.2022.847995

Wang, Y., Kajtar, J. B., Alexander, L. V., Pilo, G. S., & Holbrook, N. J., 2022: Understanding the changing nature of marine cold-spells. *Geophys. Res. Lett.* , 49, e2021GL097002. https://doi.org/10.1029/2021GL097002

Wilks, D. S., 2016: "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc*., 97, 2263-2273.

Xie, S. -P., C. Deser, G. A. Vecchi, J. Ma, H. Teng, A. T. Wittenberg, 2010: Global warming pattern formation: sea surface temperature and rainfall. *J. Climate*, **23**, 966-986, doi:101175/2009JCLI3329.1.

Xu, T., M. Newman, A. Capotondi, S. Stevenson, E. DiLorenzo, and M. Alexander, 2022: An increase in marine heatwaves without significant changes in surface ocean temperature variability. *Nat. Commun.*, 13, 7396, doi: 10.1038/s41467-022-34934-x.

Yao, Y., Wang, C., & Fu, Y. (2022). Global marine heatwaves and cold-spells in present climate to future projections. *Earth's Future*, 10, e2022EF002787. https://doi.org/10.1029/2022EF002787

Zhao, Y., M. Newman, A. Capotondi, and E. DiLorenzo, 2021: Removing the Effects of Tropical Dynamics from North Pacific Climate Variability. *J. Climate*, **34**, 9249–9265, doi: 10.1175/JCLI-D-21-0344.1.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0278.1.