Quantifying Spread in Spatiotemporal Changes of Upper-Ocean Heat Content Estimates: An Internationally Coordinated Comparison

Abhishek Savita,[a,b,c,d] Catia M. Domingues,[a,c,e] Tim Boyer,[f] Viktor Gouretski,[g,h] Masayoshi Ishii,[i] Gregory C. Johnson,[j] John M. Lyman,[j,k] Josh K. Willis,[l] Simon J. Marsland,[a,b,c] William Hobbs,[c,m] John A. Church,[n] Didier P. Monselesan,[b] Peter Dobrohotoff,[a,b] Rebecca Cowley,[b] and Susan E. Wijffels[b,o]

[a] *Institute for Marine and Antarctic Studies (IMAS), University of Tasmania, Tasmania, Australia*
[b] *Commonwealth Scientific and Industrial Research Organisation, Oceans and Atmosphere, Victoria/Tasmania, Australia*
[c] *Centre of Excellence for Climate Extremes (CLEX), Australian Research Council, Tasmania, Australia*
[d] *GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany*
[e] *National Oceanography Centre, Southampton, United Kingdom*
[f] *NOAA/National Centers for Environmental Information, Silver Spring, Maryland*
[g] *Center for Earth System Research and Sustainability, CliSAP, Integrated Climate Data Center, University of Hamburg, Hamburg, Germany*
[h] *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*
[i] *Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan*
[j] *NOAA/Pacific Marine Environmental Laboratory, Seattle, Washington*
[k] *Joint Institute for Marine and Atmospheric Research, University of Hawaiʻi at Mānoa, Honolulu, Hawaii*
[l] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California*
[m] *Australian Antarctic Program Partnership, Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania*
[n] *Climate Change Research Centre, University of New South Wales, Sydney, Australia*
[o] *Woods Hole Oceanographic Institution, Woods Hole, Massachusetts*

ABSTRACT: The Earth system is accumulating energy due to human-induced activities. More than 90% of this energy has been stored in the ocean as heat since 1970, with ~60% of that in the upper 700 m. Differences in upper-ocean heat content anomaly (OHCA) estimates, however, exist. Here, we use a dataset protocol for 1970–2008—with six instrumental bias adjustments applied to expendable bathythermograph (XBT) data, and mapped by six research groups—to evaluate the spatiotemporal spread in upper OHCA estimates arising from two choices: 1) those arising from instrumental bias adjustments and 2) those arising from mathematical (i.e., mapping) techniques to interpolate and extrapolate data in space and time. We also examined the effect of a common ocean mask, which reveals that exclusion of shallow seas can reduce global OHCA estimates up to 13%. Spread due to mapping method is largest in the Indian Ocean and in the eddy-rich and frontal regions of all basins. Spread due to XBT bias adjustment is largest in the Pacific Ocean within 30°N–30°S. In both mapping and XBT cases, spread is higher for 1990–2004. Statistically different trends among mapping methods are found not only in the poorly observed Southern Ocean but also in the well-observed northwest Atlantic. Our results cannot determine the best mapping or bias adjustment schemes, but they identify where important sensitivities exist, and thus where further understanding will help to refine OHCA estimates. These results highlight the need for further coordinated OHCA studies to evaluate the performance of existing mapping methods along with comprehensive assessment of uncertainty estimates.

KEYWORDS: Bias; Interpolation schemes; In situ oceanic observations; Uncertainty; Oceanic variability; Trends

---

1. Introduction

The Earth system is accumulating energy due to sustained increase in concentrations of atmospheric greenhouse gases associated with human-induced activities (IPCC 2021; https://www.ipcc.ch; Le Quéré et al. 2018). Since 1971, more than 90% of this energy has been stored in the ocean as heat, with about 60% of that in the upper 700 m (Fox-Kemper et al. 2021;

Corresponding author: Abhishek Savita, asavita@geomar.de

Gulev et al. 2021; Rhein et al. 2013). Upper-ocean heat uptake delays surface warming (Armour et al. 2013; Otto et al. 2013; Raper et al. 2002) but causes a wide range of adverse impacts (IPCC 2019; Stocker 2015), from degradation of marine ecosystems (Bindoff et al. 2019; Hughes et al. 2018; Olsen et al. 2018) to mean sea level rise and associated extreme events (Church et al. 2013; Oppenheimer et al. 2019; Seneviratne et al. 2012; WCRP Global Sea Level Budget Group 2018).

Multidecadal estimates of ocean heat content anomaly (OHCA)—which provide evidence of physical changes in the ocean, mean sea level, and climate—are required along with Coupled Model Intercomparison Project (CMIP) simulations (Eyring et al. 2016) to attribute the detected changes to natural and anthropogenic radiative forcing (Bilbao et al. 2019; Gleckler et al. 2012; Marcos et al. 2017; Slangen et al. 2014; Tokarska et al. 2019) and to constrain uncertainties in CMIP projections used in policy-making and risk management (Carson et al. 2019; Lyu et al. 2021; IPCC 2019; van de Wal et al. 2019).

The first observational estimate of multidecadal increase in global OHCA was compiled by Levitus et al. (2000). Since then, a number of regularly gridded OHCA estimates are produced by different groups (e.g., Boyer et al. 2016; Johnson et al. 2019; Meyssignac et al. 2019; von Schuckmann et al. 2020). Following the findings of Gouretski and Koltermann (2007), these estimates include a diversity of instrumental bias adjustments applied to a large fraction of the historical ocean temperature profiles, collected by expendable bathythermographs (XBTs) (Abraham et al. 2013). With a significant reduction of systematic depth errors in XBT data (Wijffels et al. 2008), Domingues et al. (2008) showed that the rate of multidecadal increase in global upper OHCA and associated thermal expansion was faster than previously reported in the Intergovernmental Panel for Climate Change (IPCC) Fourth Assessment Report (Bindoff et al. 2007). They also showed that CMIP simulations, including both natural (e.g., solar and volcanic) and anthropogenic (e.g., aerosols and greenhouse gases) forcing were in good agreement with their improved observational estimate. Using this improved OHCA (thermal expansion) estimate, Church et al. (2011) and Fox-Kemper et al. (2021) were able to satisfactorily account for the processes causing global mean sea level rise since 1971.

All global OHCA (thermal expansion) estimates show a sustained and statistically significant ocean warming (thermosteric sea level rise) in the upper 700 m, since 1971 (or 1993) (Gulev et al. 2021; Johnson et al. 2019; Oppenheimer et al. 2019; WCRP Global Sea Level Budget Group 2018), despite differences in instrumental bias adjustment, baseline climatology, and methods used to map the uneven spatiotemporal coverage of ocean temperature profiles onto regular fields. These differences in choices of estimation, however, introduce uncertainty in linear rates and spatiotemporal evolution of upper OHCA (Abraham et al. 2013; Boyer et al. 2016; Cheng et al. 2016; Good 2017; Lyman et al. 2010; Meyssignac et al. 2019; Palmer et al. 2010).

Quality control and XBT bias adjustments were thought to be the largest source of spread in global OHCA estimates for 1993–2008 in the upper 700 m (Lyman et al. 2010). Then, the first coordinated study by Boyer et al. (2016), in which research groups used the same dataset protocol, showed that, on average, mapping methods are the largest source of spread for both 1993–2008 and the longer 1970–2008 period, followed by spread due to XBT bias adjustment. This coordinated study also noted small differences in global upper OHCA due to varying definitions in ocean coverage, although based only on two of the eight mapped estimates, and on one of the six XBT bias adjustments (their Fig. 2). Differences in spatiotemporal patterns were seen but the contribution of the influencing factors were not investigated for the globally complete fields nor at the observed locations.

Here, we extend the Boyer et al. (2016) analyses by using the same gridded datasets for the upper 700 m, produced for their coordinated intercomparison, to investigate the sensitivity of 1) global OHCA estimates to the definition of a common ocean mask (i.e., overlapping spatial coverage) and 2) spatiotemporal changes in OHCA to (i) XBT bias adjustment and (ii) mapping method, including a subsampling of mapped estimates where only profiles from surface to 700 m were collected. Unlike Boyer et al. we do not investigate spread due to baseline climatology because the required combinations of gridded datasets were not available from all research groups. Thus, our results are relative to a unique baseline climatology. Two OHCA estimates used in Boyer et al. (2016) were not included in our results because one of them is just a global integral (i.e., no regional maps available) and the other was lost due to hard disk failure. We largely focus on the period from 1970 (or 1993) to 2004, when XBTs were the major instrument type (Abraham et al. 2013), as spread is significantly smaller after 2004 (Boyer et al. 2016).

Section 2 contains an overview of the datasets and approaches, including the intercomparison protocol (section 2a), the definition of the common mask (section 2b), the subsampling based on a specific subset of the global temperature database (section 2c), and the statistical calculations (section 2d). Section 3 describes the results from our sensitivity analyses. A summary and discussion are found in section 4, and conclusions and recommendations in section 5.

2. Dataset protocol

a. Temperature data and intercomparison protocol

We analyzed the same mapped OHCA dataset versions, depth-integrated for the upper 700 m (Table 1), as used in Boyer et al. (2016, their Table 1) with two exceptions: 1) the representative mean from the Pacific Marine Environmental Laboratory (PMEL-R; Lyman and Johnson 2008), which is only a global integral (i.e., no regional fields), and 2) the Gouretski (2012) estimate (GOU), lost after hard disk failure. Despite these two differences from Boyer et al. (2016) when our global OHCA analyses (not shown) were compared to their corresponding analyses (their Figs. 1 and 6), results were nearly equivalent and did not alter any of their conclusions.

The mapped OHCA dataset versions (Table 1) comprise in situ ocean temperature profiles from bottles, conductivity–temperature–depth measurements (CTDs), and XBTs in the EN3v2a global database for 1970–2004 (Ingleby and Huddleston 2007)

TABLE 1. Sensitivity experiments used to create the 42 OHCA dataset versions accounting for variations in XBT bias adjustments (six corrections plus an uncorrected version) and six mapping methods. All experiments are relative to the monthly mean baseline climatology from Alory et al. (2007), which does not include XBT data. Mapping methods are from DOM (Domingues et al. 2008), LEV (Levitus et al. 2012), ISH (Ishii and Kimoto 2009), EN (Ingleby and Huddleston 2007), PMEL (Lyman and Johnson 2008), and WIL (Willis et al. 2004). Periods available are for 1970–2008 or 1993–2008, depending on the mapping method; only the starting year is listed.

XBT bias adjustments		Mapping methods					
Acronym/reference	Adjustment	DOM	LEV	ISH	EN	PMEL	WIL
No_corr	No correction	1970	1970	1970	1970	1970	1993
W08 (Wijffels et al. 2008)	Depth	1970	1970	1970	1970	1970	1993
I09 (Ishii and Kimoto 2009)	Depth	1970	1970	1970	1970	1970	1993
L09 (Levitus et al. 2009)	Temperature	1970	1970	1970	1970	1970	1993
GD11 (Good 2011)	Depth (bathymetry approach)	1970	1970	1970	1970	1970	1993
GK12 (Gouretski 2012)	Depth + temperature	1970	1970	1970	1970	1970	1993
CW13 (Cowley et al. 2013)	Depth + temperature	1970	1970	1970	1970	1970	1993

combined with Argo float profiles for 2000–08 (Argo 2000; Barker et al. 2011). The profiles were mapped by six research groups, including seven permutations of the XBT data, to account for six bias adjustments plus an uncorrected case, totaling 42 members. This combination was specifically produced for the coordinated intercomparison in Boyer et al. (2016). Although it does not necessarily follow the routine approaches of the research groups, the coordinated protocol is a good approximation and one representation that allows us to probe the relative influence of some estimation parameters in causing spread among OHCA estimates. Intercomparison of routine OHCA calculations (e.g., Wang et al. 2018; Liang et al. 2021), not based on coordinated protocols, can only document where robust and inconsistent features exist and so is unable to differentiate sources of spread, as in this study. Beyond differences in mapping methods, XBT bias adjustments, and baseline climatologies, routine OHCA estimations may differ, for example, in horizontal and vertical spacing resolutions, depth integrations, the variety of data from instrument types, and regular updates as global databases are dynamic (i.e., continuously evolving in number of profiles, metadata recovery, quality control, exact and near-duplicates removal, etc.). For instance, see the IQuOD (www.iquod.org) and the Argo floats (https://argo.ucsd.edu) projects.

As described in Boyer et al. (2016), once the in situ temperature profiles from the combined global profile datasets were assembled, they were converted into potential temperature anomalies relative to a monthly mean climatology. Anomaly profiles were depth-integrated for two levels, 0–300 and 300–700 m, and then distributed and separately gridded by each research group using their respective mapping methods, and subsequently added together to obtain gridded anomalies for 0–700 m. To convert potential temperature anomalies into OHCA, the gridded data were multiplied by the seawater density (1020 kg m^{-3}) and heat capacity (4187 J kg^{-1} °C^{-1}) constants.

The 42 OHCA members for 1970–2008 (Table 1) were mapped by DOM (Domingues et al. 2008), LEV (Levitus et al. 2012), ISH (Ishii and Kimoto 2009), EN (Ingleby and Huddleston 2007), PMEL (Lyman and Johnson 2008), and WIL (Willis et al. 2004). WIL, however, starts in 1993, as their

mapping relies on regressions with sea level from satellite altimeter (WCRP Global Sea Level Budget Group 2018). A summary of the six mapping methods is found in Boyer et al. (2016).

All gridded OHCA estimates (Table 1) are relative to the same monthly mean baseline climatology from Alory et al. (2007), corresponding to the "C1_H (or H)" case in Boyer et al. (2016). This climatology comprises bottle and CTD profiles from the EN3v1d (the immediate previous version to EN3v2a) database for 1970–2004 (Ingleby and Huddleston 2007) merged with Argo profile floats for 2000–08 (Barker et al. 2011), and deliberately excludes XBT profiles. Although Boyer et al. (2016) tested the effect of three different climatologies, this was only done for a subset of eight estimates with a single XBT bias adjustment due to limitations in resources. Other insights on the influence of climatology choices can be found in Cheng and Zhu (2015), Good (2017), and Lyman and Johnson (2014).

The 6 (out of 10 or more) proposed XBT bias adjustments used in this study (Table 1) may not correct for all recommended temperature and depth factors and may not apply for all types of XBTs manufactured over the years, as explained in Cheng et al. (2016). The large number of proposed XBT bias adjustments partly reflects the difficulties imposed by missing metadata (Abraham et al. 2013). To minimize this problem, Palmer et al. (2018) developed a deterministic approach to intelligently assign a set of plausible metadata information to XBT profiles, as part of the IQuOD Project (www.iquod.org). Probabilistic efforts underpinned by machine learning are also being explored in the IQuOD project (Leahy et al. 2018).

More recently, Cheng et al. (2018) developed a set of metrics to evaluate XBT bias adjustments and reported that out of the 10 adjustments they examined, four of them—CH14 (Cheng and Zhu 2014), GK12 (Gouretski 2012), GR10 (Gouretski and Reseghetti 2010), and L09 (Levitus et al. 2009)—appear to be the best performing schemes. In this study, both the L09 and GK12 adjustments were applied to temperature profiles (Table 1), along with W08 (Wijffels et al. 2008), I09 (Ishii and Kimoto 2009), GD11 (Good 2011), and CW13 (Cowley et al. 2013). The
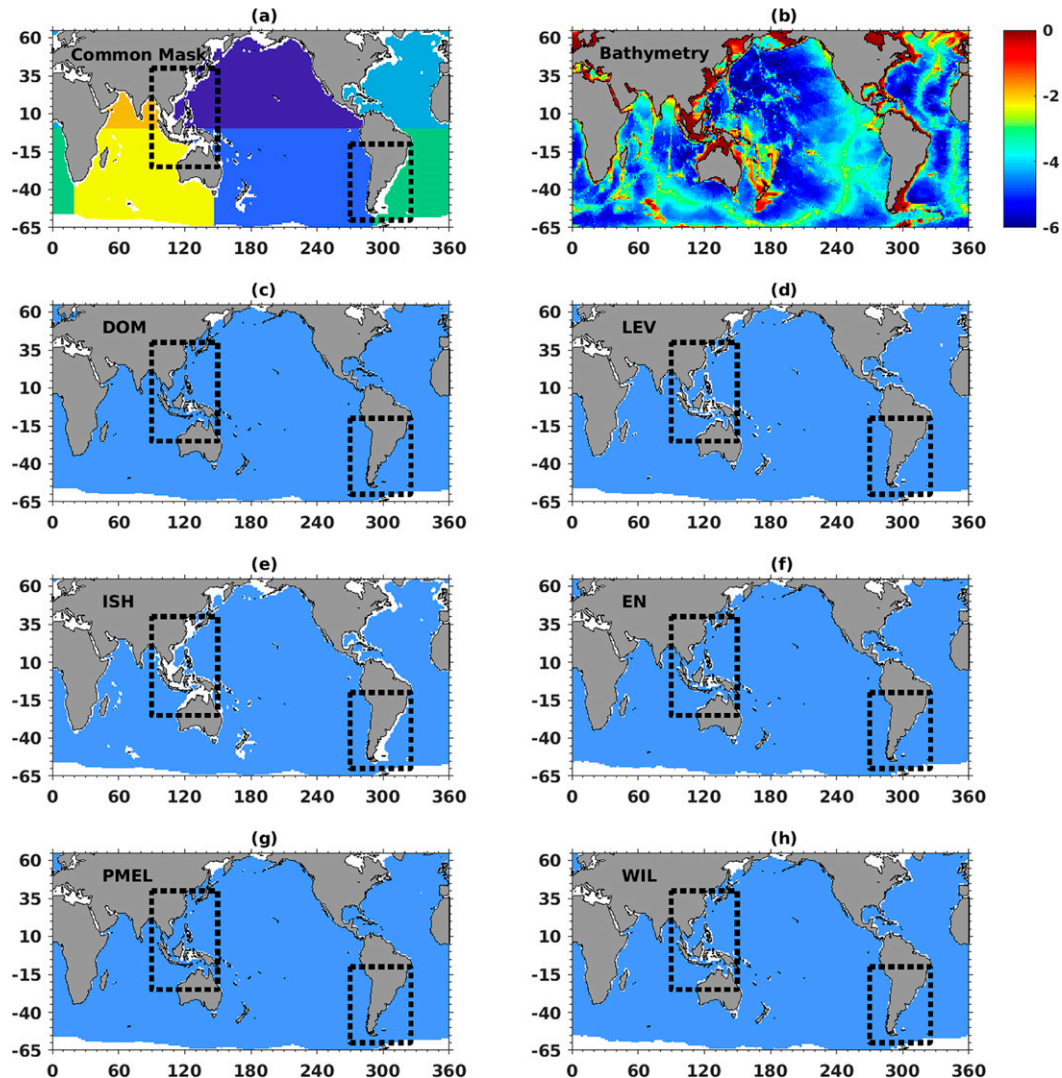
FIG. 1. Ocean mask definitions and bathymetry. (a) Common mask for global and basins (color). (b) Bathymetry (km) from ETOPO5 (https://doi.org/10.7289/V5D798BF). (c)–(h) Original masks from DOM, LEV, ISH, EN, PMEL, and WIL, respectively, where blue denotes where there is data coverage and white where there is not. The Southern Ocean basin is poleward of 35°S (not shown). Boxes (dotted line) illustrate major differences in the original masks among research groups. The total area of the common mask (2.86×10^{14} m²) and individual masks are DOM (3.22×10^{14} m²), LEV (3.04×10^{14} m²), ISH (2.91×10^{14} m²), EN (3.11×10^{14} m²), PMEL (3.18×10^{14} m²), and WIL (3.11×10^{14} m²) (see Table 1 for explanations of abbreviations).

CH14 bias adjustment (Cheng and Zhu 2014), which is recommended by the XBT science community (Cheng et al. 2018; Goni et al. 2019), was not available at the time our coordinated OHCA estimates were produced. The XBT profiles from the recent release of the World Ocean Database 2018 (WOD18; Boyer et al. 2018) as well as from IQuOD's first interim release (v01; Cowley et al. 2021) use the CH14 adjustment.

b. Postprocessing and common ocean mask

For the intercomparisons, we postprocessed the 0–700-m OHCA datasets mapped by the research groups to the same spatiotemporal resolution: annual means, interpolated onto a 1° × 1° spherical grid, area-weighted and relative to the same ocean mask (65°N–65°S) in common to all groups (Fig. 1a). The original masks from each group are shown in Figs. 1c–h, along with bathymetry (Fig. 1b), and their respective surface areas listed in the caption. Most coverage differences are in marginal seas (especially the Indonesian Throughflow region) and shelf areas along the west boundary margins of the North Pacific and Atlantic (particularly off South America), shown as boxes in Fig. 1. Global estimates were derived by integrating the yearly, area-weighted OHCA values for all grid points within the common ocean mask, from 1970 (or 1993) to 2008. Basin integrals follow the color-coded areas in Fig. 1a. Our Southern Ocean definition is poleward of 35°S (not shown).

c. 0–700-m subsampling

Subsampling was performed to quantify the OHCA spread due to XBT bias adjustment without the mapping influence at the observational locations where a subset of temperature profiles was collected from surface to 700 m at least (referred to herein as "non-gridded subsampled profiles"). In addition, to compare the influence of the mapping methods on the OHCA estimates based on these non-gridded subsampled profiles with different XBT bias adjustments, we extracted the corresponding profiles (called gridded subsampled profiles) from the globally complete gridded fields mapped by the research groups, with our definition of ocean mask applied (section 2b).

There are three aspects to highlight for the subsampling exercise. First, since the 0–300-m profiles in this study can be due to a combination of shallow and deep profiles, they were not used in the subsampling. In other words, although the mapped estimates contain grid boxes where observed profiles were available: (i) for the upper 300 m (0–300 m), (ii) for the upper 700 m (0–700 m), or (iii) none at all ("infilled"), only grid boxes that matched (ii) were selected for the subsampling. Second, although the OHCA estimates based on the two subsets of subsampled profiles, gridded and non-gridded, were area-weighted prior to integrating them over the observed locations available for each year (section 3d; Figs. 7–9), these subsampled estimates only contain a small subset of profiles from the global database, and with no analogous analyses in Boyer et al. (2016). Third, when more than one 0–700-m observed profile was available within a 1° × 1° grid box for a certain month and year, they were averaged using a median to create one superobservation ("superobs") for the location, prior to the mapping. Although not all observed profiles averaged together in the superobs were necessarily from XBTs (i.e., could also be from bottles, CTDs, and Argo floats), their spread can only originate from differences in the XBT bias adjustments, so a useful measure of the impact of the XBT bias adjustments on OHCA estimation given that all routine OHCA estimates include data other than just from XBTs.

d. Statistical calculations

We largely focused on two periods, 1970–2004 and 1993–2004. Spread in OHCA estimates after 2004 is much reduced due to Argo data only (Boyer et al. 2016; Liang et al. 2021; Ishii et al. 2017). Spread due to the XBT bias adjustments (section 3b) was calculated on an annual basis by taking the standard deviation (STD) of the datasets with the six XBT bias adjustments (excluding the uncorrected version) for each of the six mapping methods (Table 1). We also estimated the average STD for 1970–2004 and 1993–2004, and in some cases, we calculated the STD for the presatellite altimeter era (1970–93). The satellite altimeter era (1993–2004) coincides with an increase in number of deeper XBT profiles (≥700 m) and wider sampling of the Southern Hemisphere during the World Ocean Circulation Experiment (WOCE; Gould et al. 2013; Wijffels et al. 2008). Finally, we averaged the STD due to XBT bias adjustments obtained for each of the six mapping methods together to estimate an ensemble spread (EnSTD). The ensemble spread due to mapping method

(section 3c) followed the above, except that STD calculations were computed among mapping methods for each of the six XBT adjustments.

The calculation of correlation coefficients for the Taylor diagrams in section 3d followed Taylor (2001). Linear trends were calculated using ordinary least squares regression for 1970–2004 and 1993–2004 (section 3e), with their uncertainty defined as twice the standard error (95% confidence). The standard errors (SE) considered autocorrelation (i.e., reduced degrees of freedom) and were computed by the variance of the residuals about the fit, as in Santer et al. (2008) [see their Eqs. (4)–(6)]. Because our trend periods end in 2004 and the use of the common ocean mask (Fig. 1a), small differences (not shown) are expected between our global OHCA trend values and those in Boyer et al. (2016). For the trend maps, ensemble mean and spread along with the spread-to-mean ratio were also calculated. The ensemble spread is a measure of the disagreement between the corresponding mean values [i.e., can be considered as a proxy for uncertainty, under the assumptions in von Schuckmann et al. (2020)] while the ratio displays their relative weight.

3. Results

a. Effect of common spatial coverage on global estimates

The common mask in Fig. 1a represents the global ocean domain intersected by the original masks from the six research groups (Figs. 1c–h; surface areas listed in the caption). This common domain is mainly determined by the most conservative mask from ISH (Fig. 1e) which excludes the greatest amount of combined area, within marginal seas and shelf zones. Thus, differences in global OHCA between the original and common mask is smallest for ISH, regardless of XBT bias adjustment (Fig. 2i). DOM has the largest OHCA difference (Fig. 2g), about 6 times larger than the other four original masks (LEV, EN, PMEL, and WIL) with similar ocean area (Fig. 1), primarily due to the influence of their mapping method. On the ensemble mean, differences are about 2% for ISH (Fig. 2c) and 13% for DOM (Fig. 2a), comparable to the OHCA contribution below 2000 m to full depth [e.g., ~10% in Rhein et al. (2013); 5%–10% in von Schuckmann et al. (2020); and ~2% in Boyer et al. (2016) and Meyssignac et al. (2019)].

Global OHCA differences due to spatial coverage (individual minus common mask) can vary with XBT bias adjustment but overall tend to increase with time for all mappings, particularly after 1990 (Fig. 2, right panels). Compared to the work of Boyer et al. (2016), where global OHCA estimates were based on individual masks, we see similar short-term variability but smaller multidecadal increases (Fig. 2, left panels). The effect of the common spatial coverage on trends is presented in section 3f. All analyses in the rest of this paper are based on the definition of the common ocean mask (Fig. 1a).

b. Spread due to XBT bias adjustment

Global STD maps (Fig. 3, left panels) show the spread in gridded OHCA regional patterns for the upper 700 m due to the six choices of XBT bias adjustments for time series starting in 1970 (DOM, LEV, ISH, EN, PMEL; Table 1). EnSTD maps
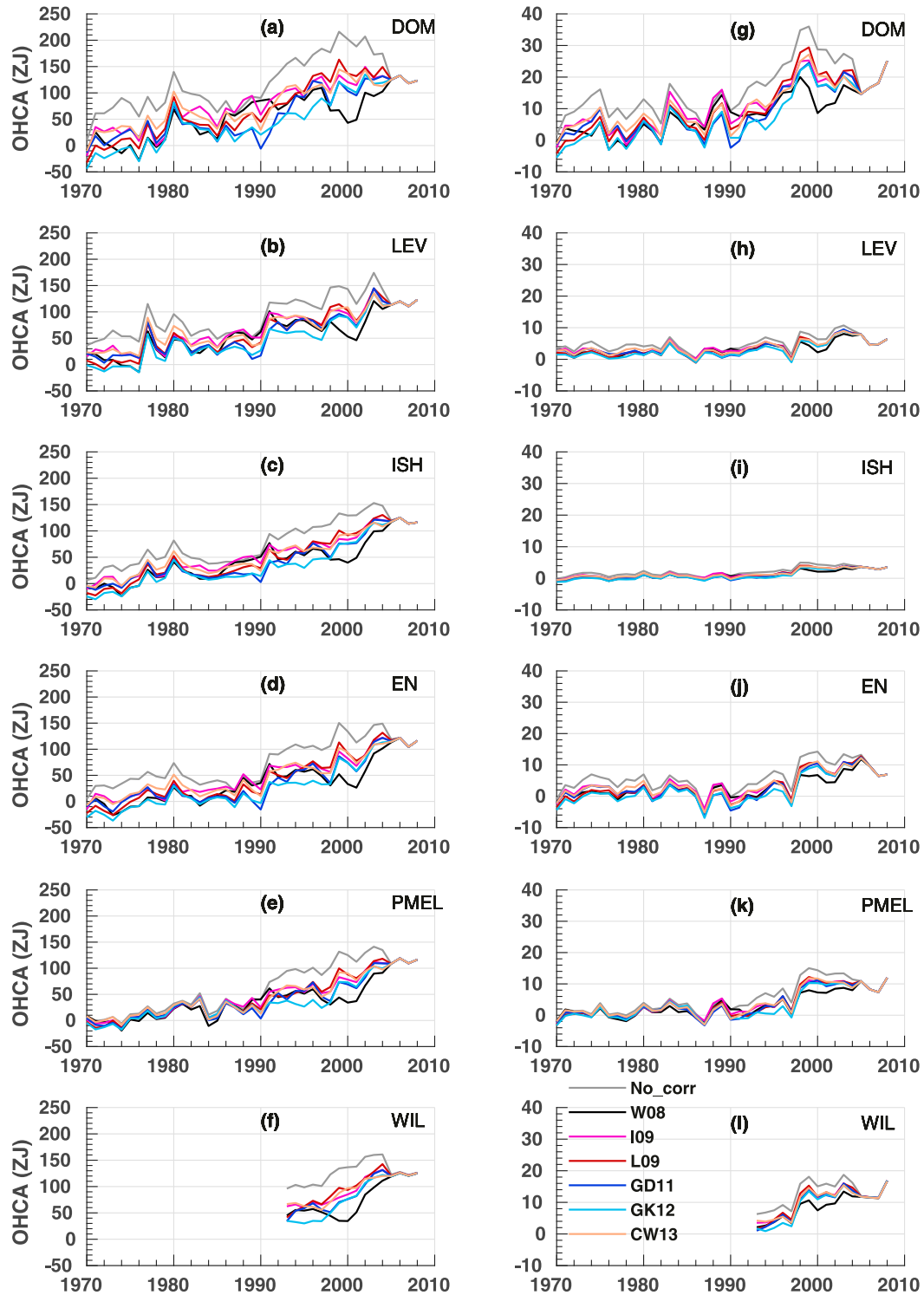
FIG. 2. Global OHCA time series (ZJ) for six XBT bias adjustments as well as uncorrected version (see legend). (left) Time series based on the common mask for each mapping method: (a) DOM, (b) LEV, (c) ISH, (d) EN, and (e) PMEL for 1970–2008, and (f) WIL for 1993–2008. (right) Differences based on the original mask minus common mask.
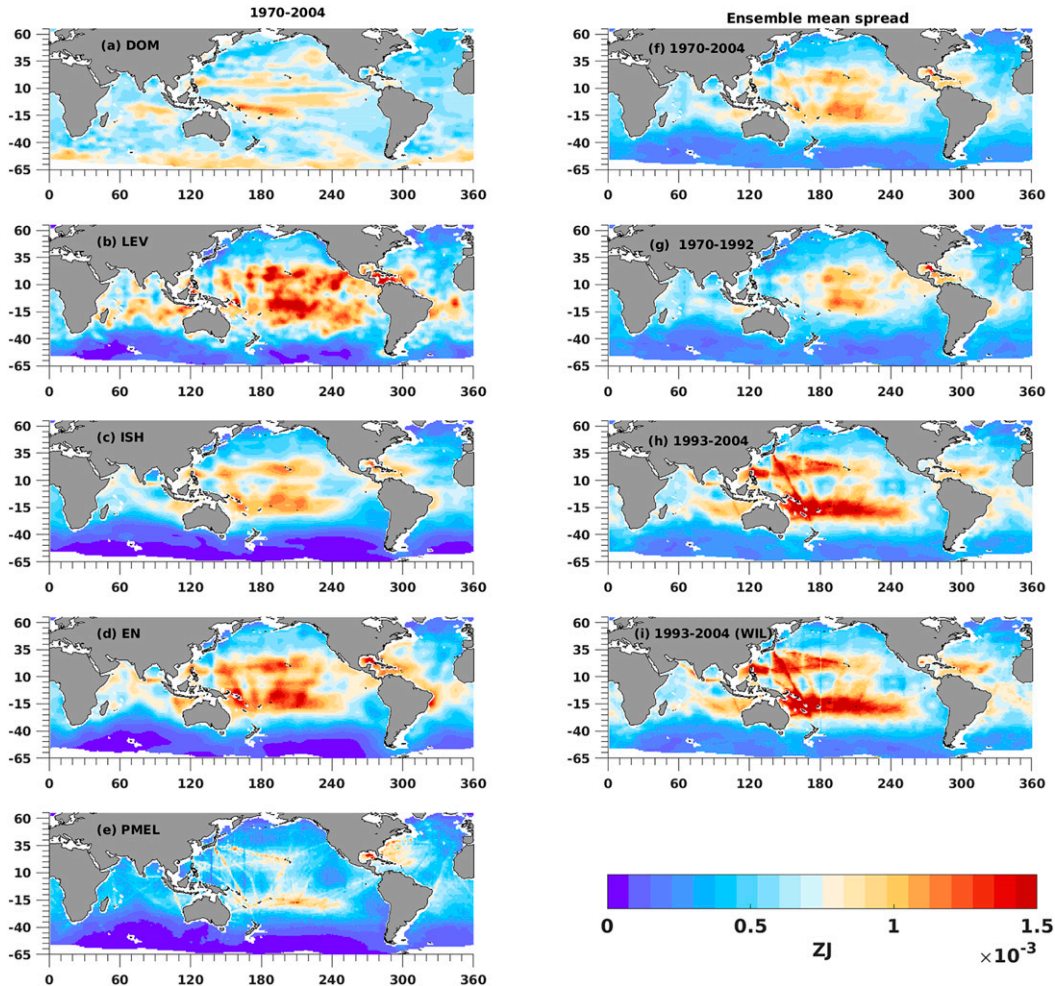
FIG. 3. OHCA spread (ZJ) due to XBT bias adjustment. (left) Results for each mapping method [(a) DOM, (b) LEV, (c) ISH, (d) EN, and (e) PMEL] averaged over 1970–2004. (right) Ensemble mean spread (EnSTD) across mapping methods for different time periods: (f) 1970–2004, (g) 1970–92, (h) 1993–2004 excluding WIL (only available since 1993), and (i) 1993–2004 including WIL.

were estimated by averaging the global STD patterns across the five mapping methods (Fig. 3, right panels) for three periods: 1970–2004 (longest), 1970–92 (prealtimeter era), and 1993–2004 (altimeter era). Inclusion of WIL's estimates in the EnSTD for 1993–2004 (Fig. 3i) did not modify the results in Fig. 3h.

EnSTD is maximum across all basins within 30°N–30°S, with the highest values in the Pacific and for 1993–2004 compared to 1970–92 (Fig. 3, right panels). While the Pacific maximum for 1970–92 is centralized (Fig. 3g), the pattern for 1993–2004 is broken into two zonally extended cells, found farther from the equator (Fig. 3h). Consequently, the longest period, 1970–2004, reflects their combined imprint (Fig. 3f).

Over 1970–2004, the maximum in the EnSTD pattern across 30°N–30°S (Fig. 3f) is mainly influenced by LEV, ISH, and EN (Figs. 3b–d) and their decorrelation radii (shape and size). The imprint of the radius of influence used by these mapping methods becomes obvious after comparing with

PMEL (Fig. 3e). PMEL has a clearer delineation of the repeated XBT lines in their STD maps as their choice of physically based correlation length scales and signal-to-noise ratios in their objective mapping relaxes toward the initial guess of zero anomalies in data-sparse regions (Boyer et al. 2016; Lyman and Johnson 2008). In contrast, maxima in the STD along XBT lines are not evident in DOM (Fig. 3a) but are instead found across the Southern Ocean, where XBT measurements are limited to a small number of meridional repeat lines (Goni et al. 2019), and where seawater is colder and vertical temperature stratification is weaker relative to lower latitudes. DOM is the only method that can project signals to other far-reaching places, not necessarily associated with the XBT sampling, by minimizing a cost function at global and local scales simultaneously, relying on statistics from a reduced set of empirical orthogonal functions (EOFs) from satellite altimeter along with a global constant mode (EOF0) (e.g., Church et al. 2004; Pittman 2016). All other methods
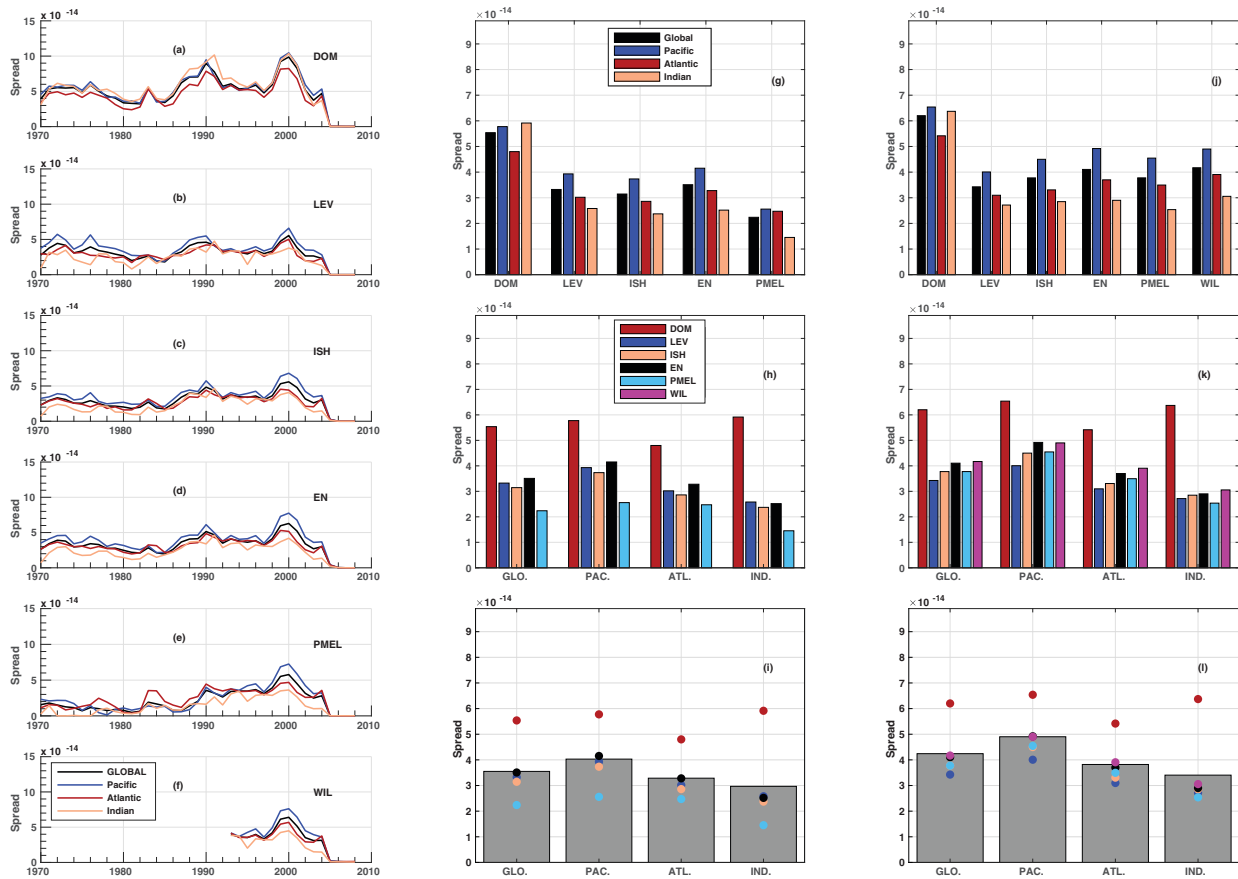
FIG. 4. Global and basin OHCA spread due to XBT bias adjustment per square meter (ZJ m^{-2}). (left) Annual time series from 1970 to 2008 for each mapping method: (a) DOM, (b) LEV, (c) ISH, (d) EN, (e) PMEL, and (f) WIL. (center) Spread averaged over 1970–2004 organized by (g) mapping method, (h) basin, and (i) ensemble mean spread (EnSTD; gray bars) across mapping methods organized by basin. (right) As in the center panels, but for 1993–2004.

only depend on local fitting to infill the sparser in situ ocean temperature observations.

Overall, per square meter, DOM has the highest sensitivity to the differences in XBT bias adjustments, regardless of ocean basin, and well above the EnSTD for both 1970–2004 and 1993–2004 (Fig. 4, right panels). PMEL is the least sensitive for 1970–2004 but not necessarily for 1993–2004. From a basin-average perspective, the Pacific has the highest spread per square meter, followed by the Atlantic and Indian Oceans. DOM is an exception, in which the Pacific and Indian basins have similar sensitivities, around 60 MJ m^{-2} or 6 × 10^{-14} ZJ m^{-2} (1 MJ = 10^6 J; 1 ZJ = 10^{21} J), and higher than the Atlantic. The STD time series for individual basins (Fig. 4, left panels) have similar variability to the global analysis in Boyer et al. (2016). XBT spread per square meter is higher during 1989–2000 compared to previous years, with a maximum peak around 2000 that decays to zero in 2005, when only Argo data were included in the data protocol (section 2a).

c. Spread due to mapping method

Global maps of STD patterns due to the six choices in mapping methods (Table 1) are similar across XBT bias adjustments

(not shown), as XBT bias adjustment differences at grid scale are negligible. So, we only present the EnSTD patterns (Fig. 5, left panels), obtained by averaging the STD patterns across the six XBT bias adjustments. In general, the STD maxima largely coincide with highly energetic eddy regions and frontal systems seen in altimeter sea level (Fu et al. 2010), including the Gulf Stream and Kuroshio–Oyashio boundary current extensions, the Brazil–Malvinas Confluence, the Agulhas and East Australian Current retroflections, and along the Antarctic Circumpolar Current (ACC), particularly in the Indian sector. In contrast with the EnSTD due to XBT bias adjustment (Figs. 3h,i), the EnSTD due to mapping method increases after inclusion of WIL's estimates over all basins for 1993–2004 (Fig. 5d compared to Fig. 5c), also further evident in the zonal integrals (Fig. 5, right panels).

The zonally integrated EnSTD have similar patterns across the latitudinal bands for the three time-average periods and is maximum in the Southern Ocean (35°–60°S) (Fig. 5, right panels). The highest STD contributions for the Southern Ocean peak are from the Indian and Atlantic sectors (40°–50°S), followed by the Pacific sector (50°–60°S). In the Pacific, the largest peak lies around 0°–20°N, followed by
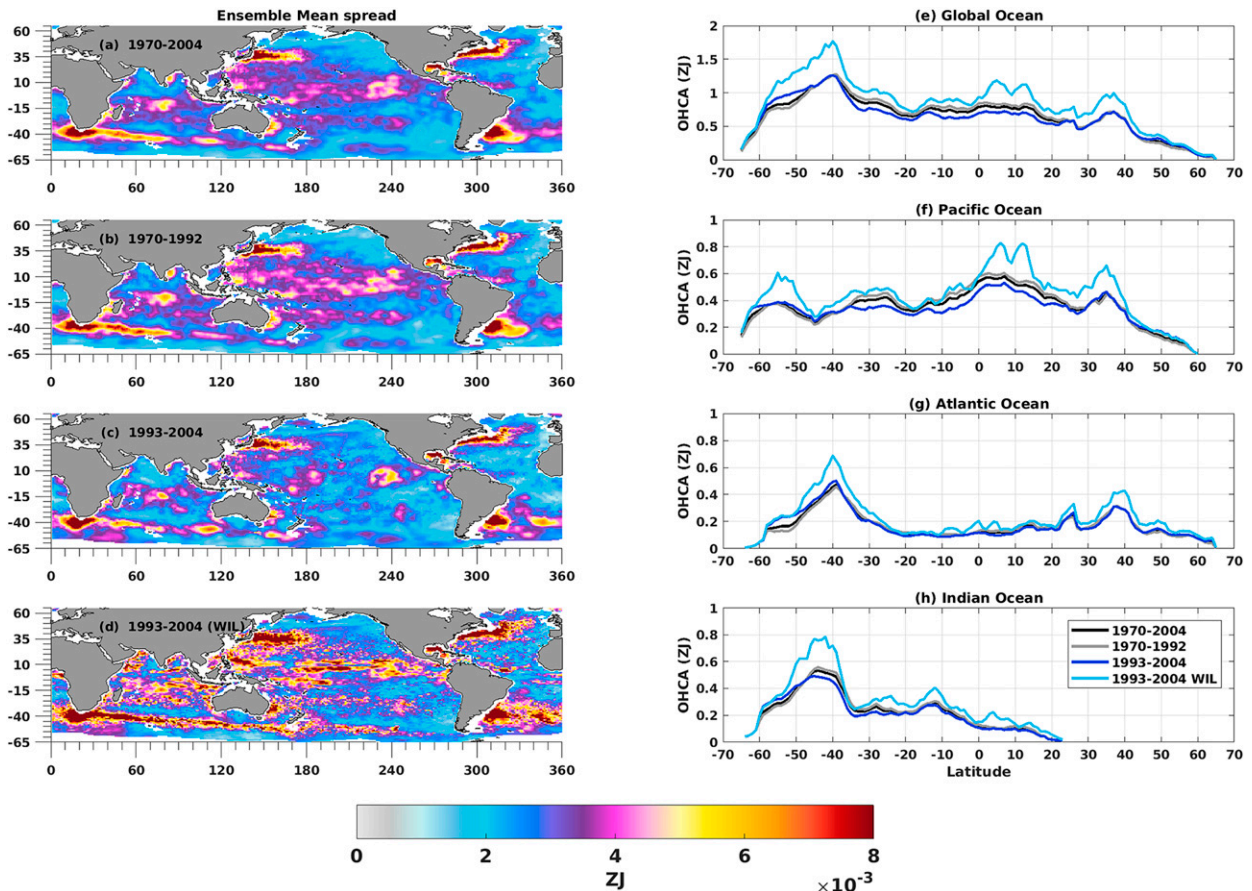
FIG. 5. OHCA ensemble mean spread (EnSTD; ZJ) due to mapping method across XBT bias adjustments for different periods. (left) Global patterns for (a) 1970–2004, (b) 1970–92, (c) 1993–2004 excluding WIL, and (d) 1993–2004 including WIL. (right) Zonal integrals (e) globally and for the (f) Pacific, (g) Atlantic, and (h) Indian Oceans.

three secondary peaks, 30°–40°N, 20°–40°S, and 50°–60°S. The Atlantic has only one secondary peak (30°–40°N) while the Indian has a plateau from 10° to 30°S, with STD values decaying north of 10°S.

Overall, per square meter, the largest spread in OHCA due to mapping method is in the Indian Ocean for all XBT bias adjustments, over 1970–2004 and 1993–2004 (Fig. 6, right panels). In fact, the Indian has the highest spread during most years except in the mid-1980s, when the Pacific has two maxima (Fig. 6, left panels). Over 1993–2004, the EnSTD for the Indian Ocean basin is almost twice as large as for the other basins, mainly due to the two maxima in 1997/98 and 2001/02, seen across all XBT bias adjustments. The EnSTD for the Atlantic has the smallest mapping spread. In terms of individual XBT bias adjustments, L09 and W08 have the highest and lowest spread respectively. Over 1970–2004, the EnSTD difference between the Indian and the other basins is not as large as during 1993–2004, and the lowest spread is for the global ocean. Note that, as the EnSTD of each basin was calculated individually, and so their sum is not necessarily equal to the EnSTD for the global integral. Rapid decrease in spread after 2004 is mainly associated with the increasing number of Argo float array

toward the spatiotemporal sampling designed to resolve ocean climate change signals (Riser et al. 2016; Roemmich et al. 2019).

d. Subset of 0–700-m profiles: The effect of mapping methods

Mapping methods can influence the evolution of global integrals and the associated spatiotemporal variability of OHC estimates in diverse ways, thus causing spread, due to varying covariance length scales, smoothing, etc. Differently from Boyer et al. (2016) and section 3b, here we examine OHCA spread due to XBT bias adjustment without mapping influence, based on a subset of temperature profiles (section 2c), extending from the surface to 700 m at least (Fig. 7a; non-gridded subsampled profiles). We extracted the equivalent data from the gridded fields (Figs. 7b–g; gridded subsampled profiles) to examine the differences (Figs. 7h–m; gridded minus non-gridded). The subsampled profiles underpinning Figs. 7–9 were area-weighted before integrating them over the locations available for each year. Note that, the OHCA integrals based on these non-gridded and gridded subsampled profiles do not represent global integrals, as they only contain a specific fraction of the observed profiles from the global database used in Fig. 2 (section 3b) and in Boyer et al. (2016).
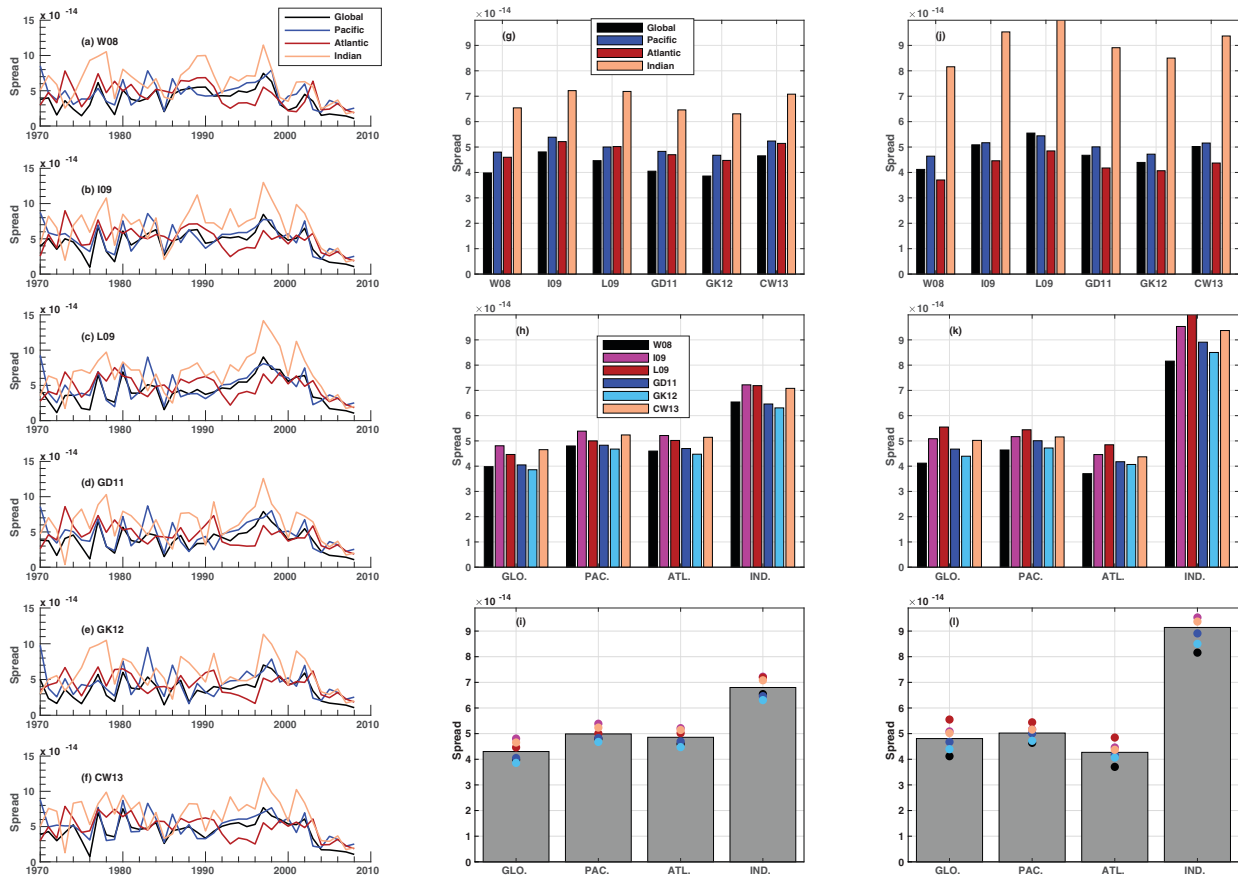
FIG. 6. Global and basin OHCA spread due to mapping method per square meter (ZJ m^{-2}). (left) Annual time series from 1970 to 2008 for each XBT bias adjustment: (a) W08, (b) I09, (c) L09, (d) GD11, (e) GK12, and (f) CW13. (center) Spread averaged over 1970–2004 organized (g) by XBT adjustment and (h) by basin, and (i) the ensemble mean spread (EnSTD; gray bars) across XBT bias adjustments organized by basin. (right) As in the center panels, but for 1993–2004.

Therefore, this subsampling exercise is not fit for the purpose of investigating sampling errors (i.e., out of scope) and was performed to solely examine the influence of mapping methods on OHCA estimates from a subset of 0–700-m profiles.

The differences in OHCA across the XBT adjustments is largest during 1990–2004, in both the non-gridded and gridded subsampled time series (Fig. 7a and Figs. 7b–m, respectively). The larger spread in recent decades could be explained by a number of factors, and should be investigated in a separate study. Potential reasons include the relatively higher ratio of measurements from bottles and CTDs compared to deep XBTs prior to 1990 (Fig. 3 in Pittman 2016), the increased number of deep XBT probe measurements in the 1990s (Fig. 9), with a greater diversity in probe types (Fig. 1 in Cheng and Zhu 2014), and a larger ratio of unknown-to-known probes (Fig. 2 in Abraham et al. 2013). Furthermore, there is additional uncertainty about whether the XBT fall-rate adjustment recommended by Hanawa et al. (1995) was considered or not due to lack of metadata as well as digital archiving of the full-vertical profile resolution (and not only inflexion points) (Abraham et al. 2013). Overall, even though some XBT bias adjustments for deep probes reaching to 700 m

might be larger prior to 1990 (e.g., Cheng and Zhu 2014), Fig. 7 shows that their impact on OHCA spread is smaller compared to the 1990–2004 period. This result suggests that if the community prioritizes understanding and refinements of deep XBT bias adjustments in the 1990s, it is very likely to have the best-cost benefit in reducing OHCA spread.

Interestingly, although L09 and GK12 made the top four XBT bias adjustments (Cheng et al. 2018), their associated OHCA estimates differ from each other over 1990–2004, for the non-gridded (Fig. 7a; without mapping) and corresponding gridded (Figs. 7b–g; with mapping) subsampled time series. The same pattern is observed for the global OHC estimates in Figs. 2a–f, which include all profiles in the coordinated protocol (section 2a). These results suggest that even though XBT bias adjustments might be similar with respect to metrics, they may not equally translate to global OHCA estimates.

The CW13 non-gridded subsampled time series has a noticeable OHCA spike during 1999–2000 (Fig. 7a), even greater than the uncorrected time series (No_corr). This spike is largely reduced after the mapping procedures were applied (Figs. 7b–g). CW13 differs from the other schemes in that they were based on a small fraction of the global dataset with only
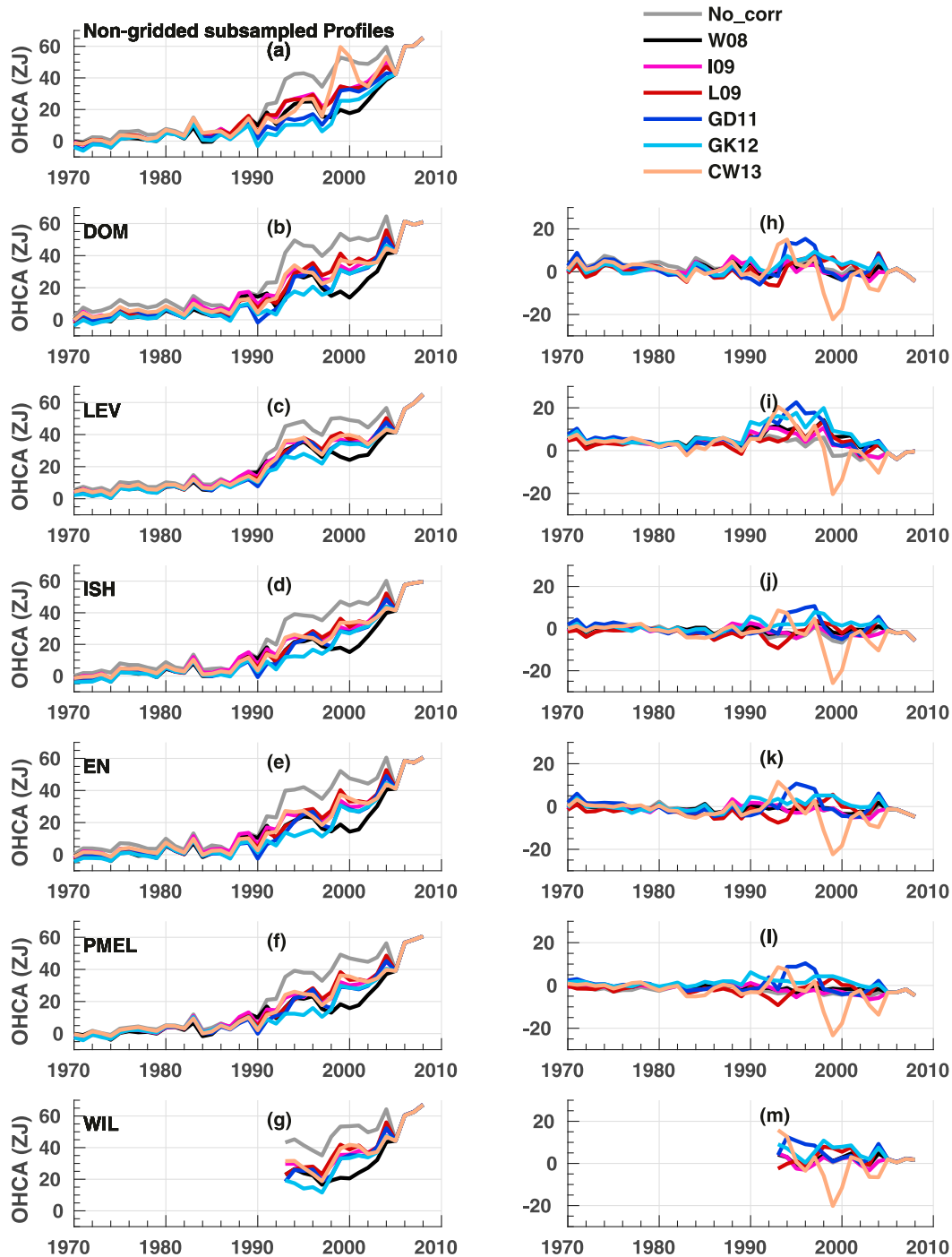
FIG. 7. Global OHCA annual time series (ZJ) for six XBT bias adjustments as well as uncorrected version (see legend) based on subsampled profiles for 0–700 m only. (left) For observed profiles (a) with no mapping applied ("non-gridded profiles"), and subsampled profiles ("gridded subsampled") from each mapping method: (b) DOM, (c) LEV, (d) ISH, (e) EN, and (f) PMEL for 1970–2008, and (g) WIL for 1993–2008. (right) Differences based on the original mask minus common mask.

the highest quality-controlled XBT/CTD profiles. Further investigations (out of the scope of this study) are required to understand this underadjustment, likely associated with a combination of a relatively small number of comparison data in the CW13 study, lack of manual quality-control in the EN3v2a dataset, missing metadata for XBT types, and the impact of the vertical temperature gradient (dependent on latitude) on depth (fall rate) and thermal biases.

Overall, the gridded subsampled time series from LEV and DOM have the largest root-mean-square error (RMSE) and minimum correlation with the corresponding non-gridded subsampled time series for most XBT bias adjustments, PMEL has the smallest RMSE and largest correlation (suggesting that they retain most of the high-frequency small-scale variability features), and the other mapping methods lie in between (Fig. 8).

e. Subset of 0–700-m profiles: Spread due to XBT bias adjustment and mapping method

The OHCA spread for the gridded subsampled time series due to XBT bias adjustment (Fig. 9a) can be higher than spread due to mapping method (Fig. 9b), particularly in the 1990s, in contrast to the globally complete gridded fields (Figs. 4 and 6) for which mapping spread is greater. Although the globally complete OHCA time series include the same 0–700-m profiles as in the subsampled time series in Fig. 9, they also contain extra 0–300-m profiles along with infilled grid areas (section 2c). So, these two extra factors do alter the relative influence of the mapping methods and XBT bias adjustments to OHCA spread. The annual variations in OHCA spread in Fig. 9 are directly correlated with the annual number of 0–700-m profiles prior to 2005 (e.g., higher spread with a higher number of profiles). An exception is when the number of profiles decreases around 1998–2001 while the XBT-related spread increases (Fig. 9a). This STD peak reaches 12 ZJ for the time series based on the 0–700-m non-gridded subsampled profiles and coincides with a peak in OHCA for the CW13 adjustment (Fig. 7a). After CW13 is excluded from the 0–700-m non-gridded subsampled profile estimates, the STD is halved to 6 ZJ (Fig. 9a). After 2005, there are no XBTs, only Argo floats, so XBT-related spread is zero. On average, mapping-related spread for the Argo period (since 2005) is 2 ZJ compared to 4–5 ZJ during the 1990s (Fig. 9b).

The annual maps of 0–700-m non-gridded subsampled profiles (Fig. 10) reveal that the higher XBT-related spread seen in their integrated OHCA time series during the 1990s (Fig. 9a) originates from the XBT lines crossing the ocean basins within 30°S–30°N, and that the maxima during 1998–2001 mainly originate in the Pacific. When CW13 is removed from the STD calculations, the Pacific maxima (orange to red) reduce to the same STD levels (cyan) observed across other years in the 1990s (not shown).

As expected, mapping-related spread decreases with improved spatiotemporal coverage during the Argo era (e.g., WCRP Global Sea Level Budget Group 2018, their Fig. 4). The surprising increase in spread with number of profiles before Argo in Fig. 9b, however, is explained by the spatiotemporal evolution of the sampling in Fig. 11, based on the 0–700-m gridded subsampled OHCA fields for L09. These maps reveal that STD maxima stem from eddy-rich regions and frontal systems of both hemispheres but only the energetic regions of the northwestern Pacific and Atlantic were measured before the 1990s, with a significant data void in the Southern Hemisphere.

f. Linear trends

Global OHCA trend maps were estimated for 1970–2004 (DOM, LEV, ISH, EN, and PMEL) and 1993–2004 (the same plus WIL), including ensemble mean and spread, and shown for the L09 adjustment (Fig. 12). At grid scale, trend maps are insensitive to XBT bias adjustment (not shown).

For 1970–2004, the OHCA trend patterns (and related ensemble mean) are generally consistent across mapping methods, with warming rates everywhere, except in some patches of the tropical–subtropical western Pacific, the southeast Indian Ocean, and the southwestern Atlantic (Fig. 12, left column). Enhanced warming in the North Atlantic and North Pacific western boundary extensions is evident in all mappings (and their ensemble mean), except in DOM, which shows a weaker trend, particularly in the North Atlantic. ISH shows a weaker warming trend along the ACC pathway in the Indian sector of the Southern Ocean compared to the other mappings, although PMEL has incomplete coverage in most of the Southern Hemisphere. In the South Pacific, DOM, LEV, and EN have some cooling patches not evident in ISH. Some differences are found in the Indian and South Atlantic for the cooling patches from LEV, ISH, and EN (not fully sampled by PMEL) whereas DOM displays basin-scale warming. DOM and ISH tend to have smoother and more zonally elongated large-scale features. In LEV and EN, the large-scale features are embedded in smaller noise-like features, associated with the shape and size of their respective decorrelation radii.

For 1993–2004, the global OHCA trend patterns (and their ensemble mean) are also generally robust across mappings (Fig. 12, middle column), and mostly resemble altimeter sea level ENSO-like variability (Hamlington et al. 2020a; Lyu et al. 2017). The most significant inconsistencies are in the Southern Ocean and North Atlantic between DOM and the other methods. DOM has the strongest wavelike cooling pattern (i.e., two cooling patches interposed by warming south of Australia) and the weakest warming rate in the subpolar North Atlantic. Although thermosteric variability is a primary driver of sea level patterns (Hamlington et al. 2020b; Vivier et al. 2005), DOM cannot rule out the influence of other contributions (Fasullo and Gent 2017; Durack et al. 2014)—such as halosteric contributions to density compensation (e.g., North Atlantic) and ocean bottom pressure (e.g., Southern Ocean) leading to barystatic (i.e., mass) changes—because they rely on EOFs from altimeter to interpolate across the sparsely in situ observations.

Maxima in the EnSTD trend patterns are found in the highly energetic eddy and frontal regions during both trend periods (Fig. 12, right column) but are more obvious during 1993–2004 (see Fig. 5). Notably, despite the maximum in spread across the Gulf Stream Extension over both periods, the ratio of the spread-to-mean is smaller than 1. The highest EnSTD regions adjacent to Antarctica in the southern Indian and Pacific are mainly associated with the trend differences in DOM, as explained above, and where we also see large-scale maxima (>2) in the spread-to-mean ratio (Fig. 12, right column).
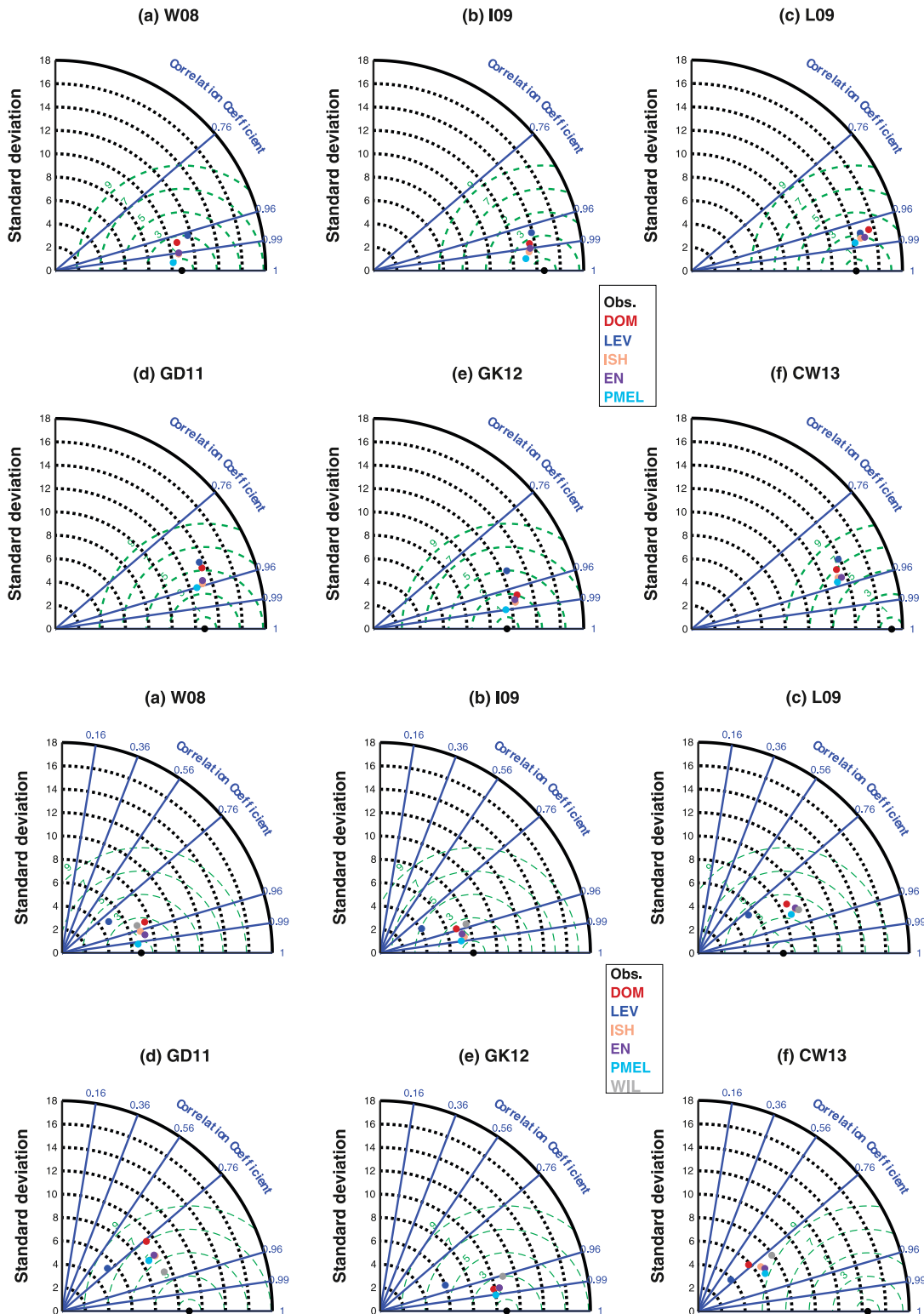
FIG. 8. Taylor diagrams for each XBT bias adjustment comparing "non-gridded profiles" for 0–700 m (only) and respective subsampled profiles mapped by six methods (color legend), for 1970–2004 in the top two rows and 1993–2004 in the bottom two rows. XBT bias adjustments are (a) W08, (b) I09, (c) L09, (d) GD11, (e) GK12, and (f) CW13. STD (black axis; ZJ) and root-mean-square error (RMSE; green axis; ZJ). Correlation coefficients (blue axis) are normalized.
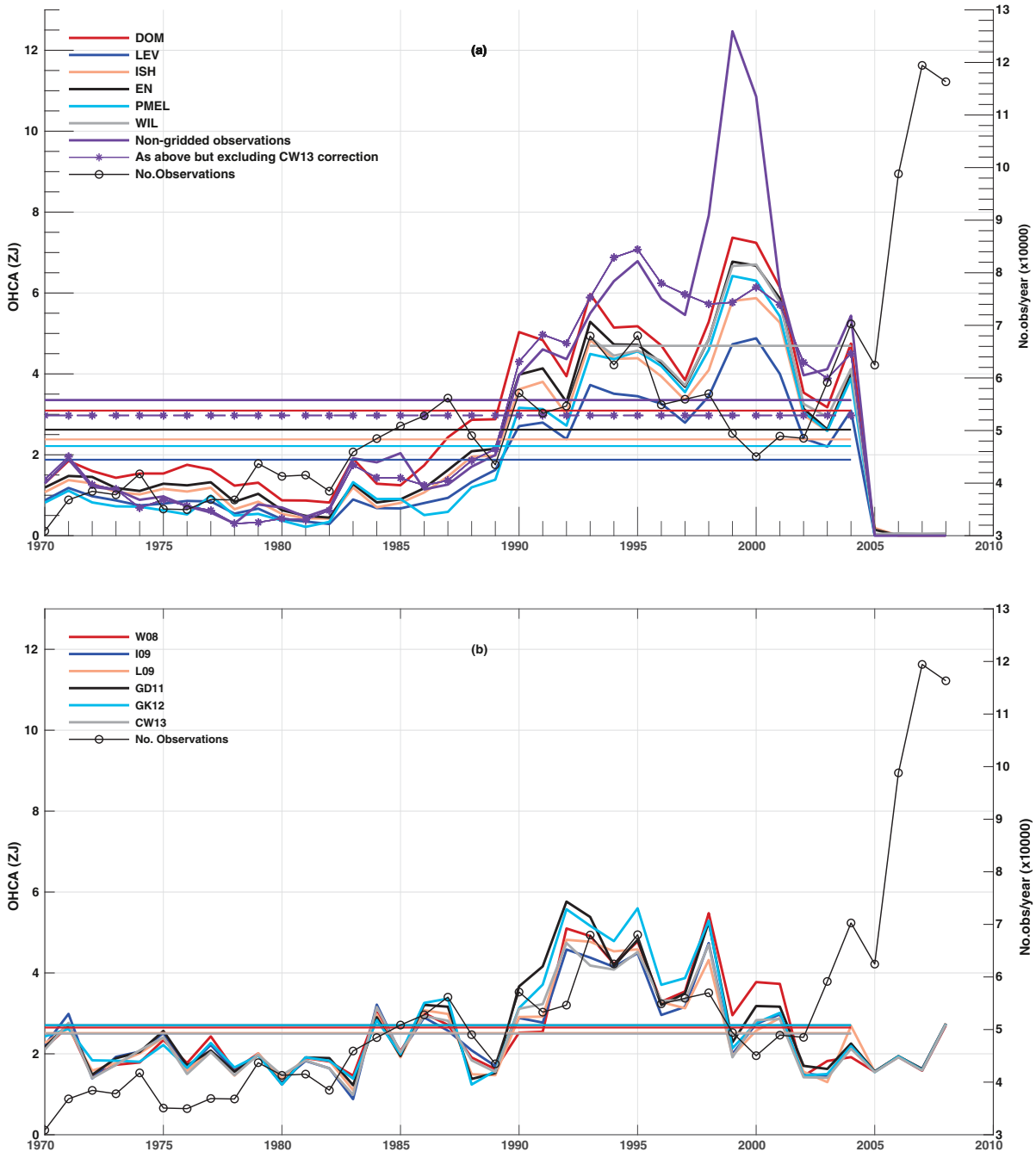
FIG. 9. Global OHCA spread time series (ZJ) based on subsampled profiles for 0–700 m only (left axis), and profile numbers per year (black line with circle, right axis). (a) Spread due to XBT bias adjustment across six mapping methods ("gridded subsampled") and for observed profiles without mapping ("non-gridded"), and with and without the XBT adjustment (correction) from CW13. (b) Spread due to mapping methods across six XBT bias adjustments. Horizontal lines represent the time mean for 1970–2004 (or 1993–2004 for WIL).

The impact of the various XBT bias adjustments only becomes relevant to linear trends when grid scale differences are integrated over large domains (e.g., from zonal to basin to global) for 1970–2004 (Fig. 13) and 1993–2004 (Fig. 14). Because of the common ocean mask (section 3a) and slightly shorter periods (up to 2004), our trend values are not identical

to those in Boyer et al. (2016). For instance, the smallest global trend for 1970–2004 is from PMEL rather than ISH for 1970–2008, as in Boyer et al. (2016).

For 1970–2004 (Fig. 13), the two most striking features are from DOM for the various XBT bias adjustments: (i) the largest warming trends are for the globe, North Pacific, and
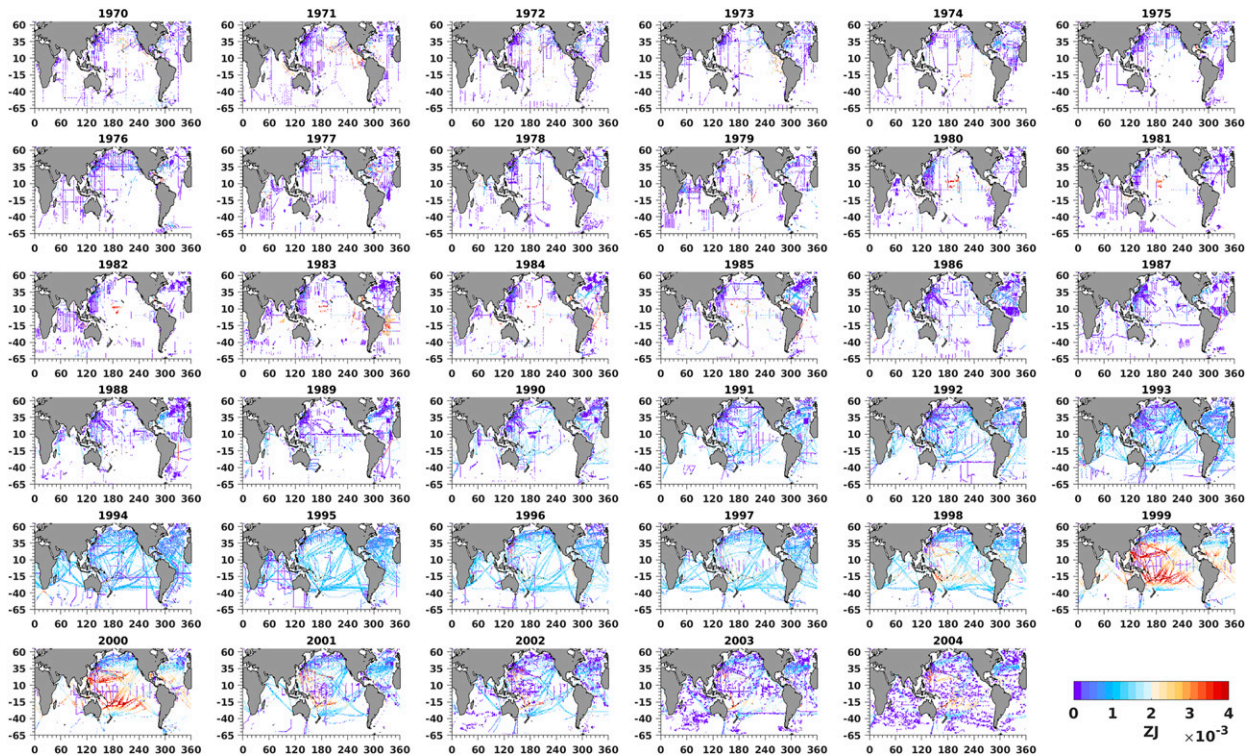
FIG. 10. Annual global maps of OHCA spread (ZJ) due to XBT bias adjustment from non-gridded profiles for 0–700 m only, from 1970 to 2004.

Southern Hemisphere oceans, statistically different in many instances from the trends estimated by the other mappings; and (ii) the smallest warming trends are for the North Atlantic, also statistically different from the other methods. Generally, the smallest OHCA trends are from PMEL because they revert to zero anomalies in data-sparse regions, especially in the Southern Hemisphere (Fig. 12). The fastest warming rates occur in the North Atlantic for all mapping methods (note the different vertical axis scale), where EN has the largest value and DOM the smallest (Fig. 13). Although the North Atlantic is particularly well observed, the trend differences are mostly confined to the highly eddy energetic regions (Fig. 5). Except for DOM's W08, GD11, and CW13 adjustments, warming trends in the poorly observed Southern Ocean for the various mappings are not statistically different from each other.

For 1993–2004 (Fig. 14), WIL generally has the highest warming trends, but in many instances they are not statistically different from other methods. The cooling pattern observed in DOM (Fig. 12) is associated with the cooling trend in the Southern Ocean, which can be statistically different from the other mappings depending on XBT adjustment. The North Indian has a cooling trend for all mappings and for most XBT adjustments. Overall, the higher range in OHCA trends for 1993–2004 (11 years) compared to 1970–2004 (34 years) is explained by the greater influence of interannual to decadal variability over the shorter period (Johnson and Lyman 2020).

4. Summary and discussion

Direct comparison of spatiotemporal changes of upper OHCA estimates is generally complicated because research groups do not necessarily use the same bias-corrected datasets, baseline climatologies, mapping methods, time periods, etc. Boyer et al. (2016) is the first coordinated study on the sensitivity of global upper OHCA estimates to XBT bias adjustments, mapping methods, and baseline climatologies. Here, we focus on extending their global study. We have analyzed the effect of applying a common ocean mask as well as quantifying the spatiotemporal sensitivity (i.e., spread) of upper OHCA estimates and multidecadal trends due to six choices in XBT bias adjustments and six mapping methods, using Boyer et al.'s (2016) internationally coordinated dataset protocol (section 2a), although with the four minor differences described below.

Our OHCA gridded analyses comprise an ensemble of 42 dataset members (including six members with uncorrected XBT data), depth-integrated in the upper 700 m, and mapped from 1970 (or 1993) to 2008 (Table 1). In relation to Boyer et al. (2016), our OHCA ensemble members (i) excluded the estimates from GOU (Gouretski 2012; lost due to disk failure) and PMEL-R (Lyman and Johnson 2008; no regional fields available), (ii) are relative to one of their three climatologies (due to limitations in resources), and (iii) are relative to an ocean mask that is common to the six mapping methods/research groups; also, (iv) our linear trends end in 2004 (rather than 2008) to coincide with the last year that XBT
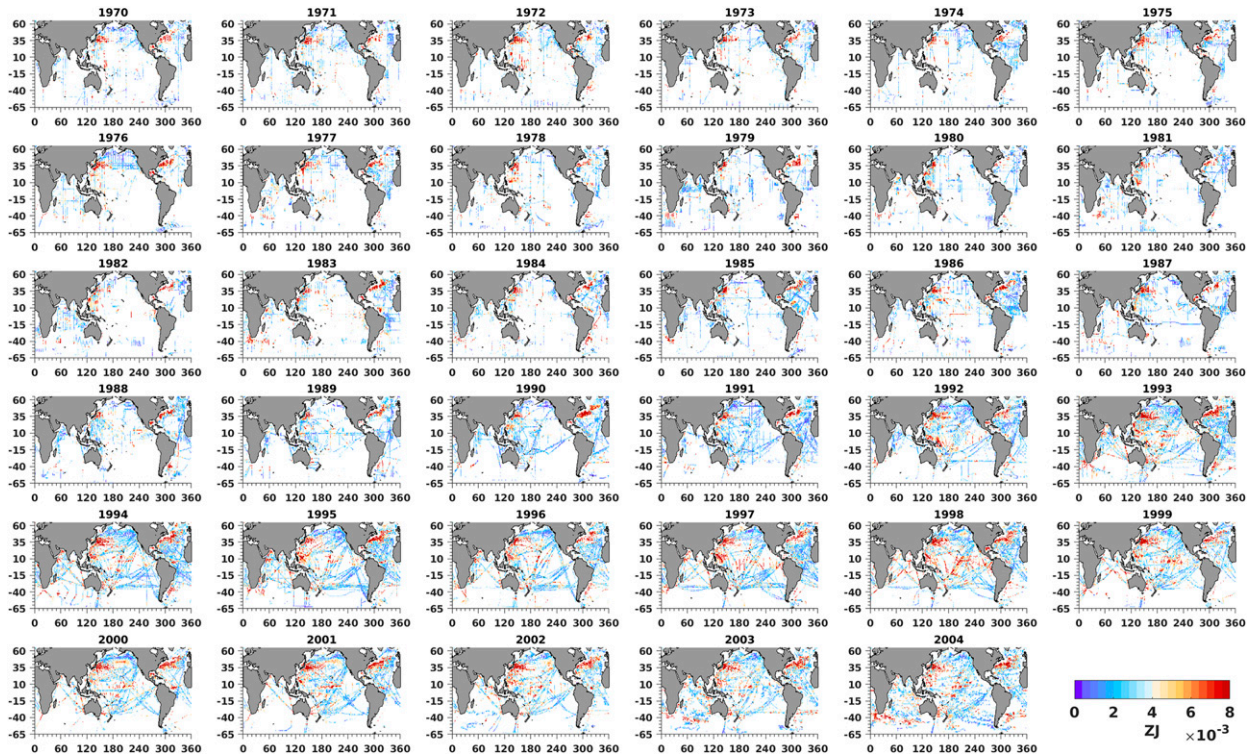
FIG. 11. Annual global maps of OHCA spread (ZJ) due to mapping method using the L09 XBT bias adjustment based on gridded subsampled profiles for 0–700 m only, from 1970 to 2004.

data were included in the combined database (section 2a). Furthermore, we have subsampled the gridded OHCA estimates at locations where temperature profiles were collected from surface to 700 m (i.e., excluding 0–300-m profiles and infilled areas) to probe the influence of the mappings on observed data (section 3d). We also looked at the OHCA spread due to mapping method and XBT bias adjustment at those 0–700-m sampled locations, using both non-gridded (i.e., without mapping) and gridded (i.e., with mapping) estimates (section 3e).

The switch to an ocean mask common to the six research groups reveals that exclusion of shallower oceans (Fig. 1) may decrease global OHCA estimates by 2%–13%. This range is equivalent to the contribution from the global ocean below 2000 m to the full-depth change (Meyssignac et al. 2019), larger than the ~2% implied for shallow areas in some studies (Boyer et al. 2016; Meyssignac et al. 2019) but closer to other studies (Rhein et al. 2013; von Schuckmann et al. 2020). The global OHCA differences among the originators' masks and the common mask increase during the 1990s, and are more dependent on mapping method than XBT bias adjustment, with DOM being the most sensitive method (Fig. 2). The increasing differences since the 1990s are likely related to the observed surface ocean warming of western boundary currents (WBCs)—due to intensification and poleward shift with a warming climate (Wu et al. 2012; Yang et al. 2016)—which cannot be fully captured after excluding a large portion of the

shallower oceans from the common mask (Fig. 1a). The combination of two factors may explain the largest sensitivity in addition to the highest increase in global ocean warming for DOM compared to the other groups with similar spatial coverage. DOM's mapping is conditional to a simultaneous fitting of spatial EOF patterns from satellite altimeter and a global constant mode for interpolation/extrapolation across areas where the coordinated protocol did not include in situ data, such as in depths < 300 or 700 m for 1970–2004 (i.e., the historical era) or depths < 1500 m after 2004 (i.e., the Argo era). The other five mapping methods in this study only use local fitting. Furthermore, at times and in specific regions, DOM's estimates may be subjected to the influence of signals other than upper-ocean thermal patterns, expressed in satellite altimeter sea level, such as from halosteric and barystatic (see Gregory et al. 2019; Landerer et al. 2007; Wu et al. 2019) as well as thermal patterns below 700 m (e.g., Liang et al. 2021).

By examining OHCA estimates (Figs. 2, 6–8, and 12–14), using two (L09 and GK12) of the four XBT bias adjustments highly recommended by Cheng et al. (2018), two (I09 and GD11) not recommended by them, and another two (W08 and CW13) widely used schemes, our results surprisingly show that the top-rated L09 and GK12 yield OHCA estimates similar to those of the not recommended I09 and GD11 schemes, respectively. The best agreement between L09 and I09 or between GK12 and GD11 is found
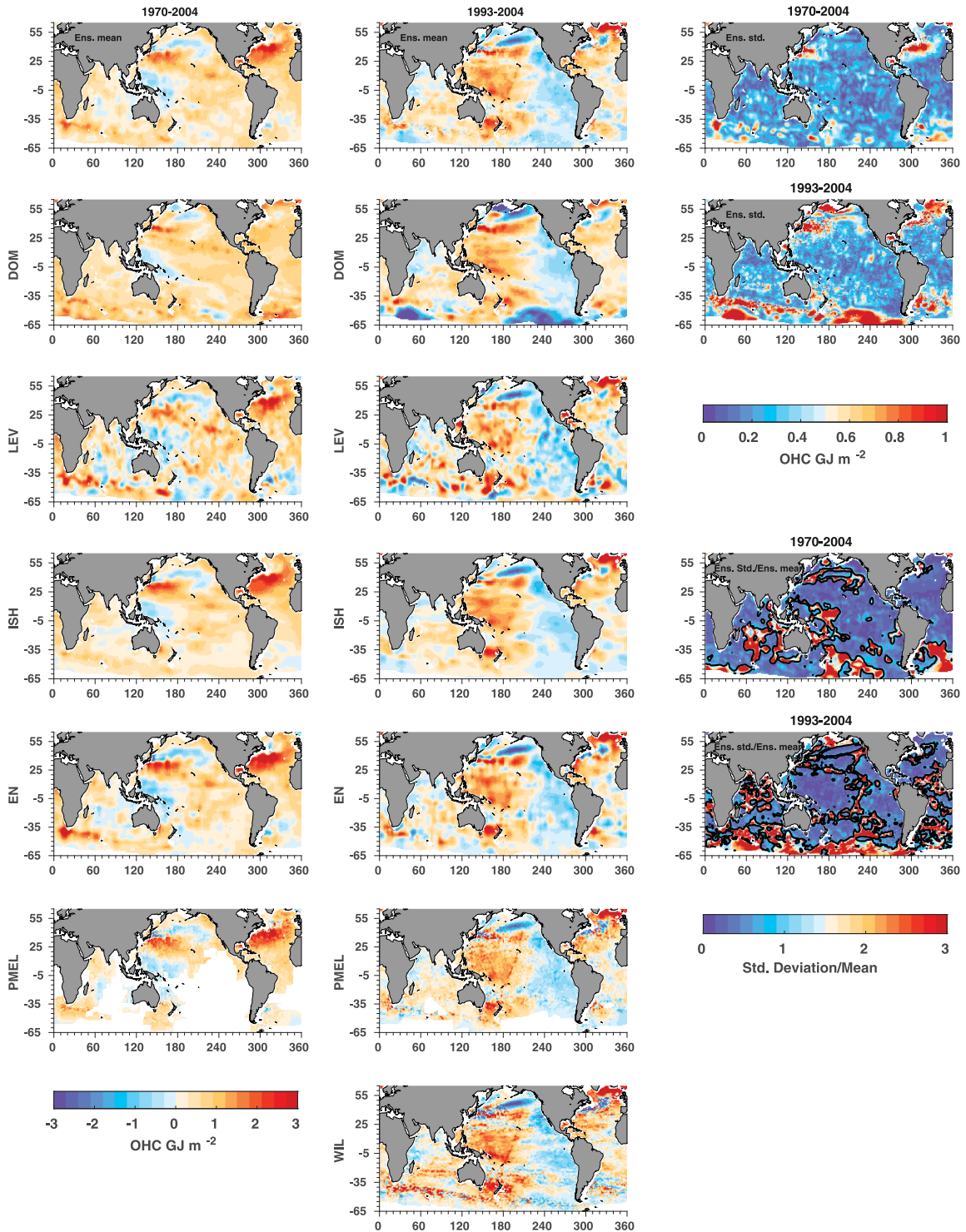
FIG. 12. OHCA linear trend maps (GJ m^{-2}; 1 GJ = 10^{9} J) based on six mapping methods (DOM, LEV, ISH, PMEL, and WIL) using the L09 XBT bias adjustment and their ensemble mean of the mapping methods, for (left) 1970–2004 and (center) 1993–2004. (right) Ensemble mean spread (EnSTD) for 1970–2004 in the top two panels and 1993–2004 in the bottom two panels.
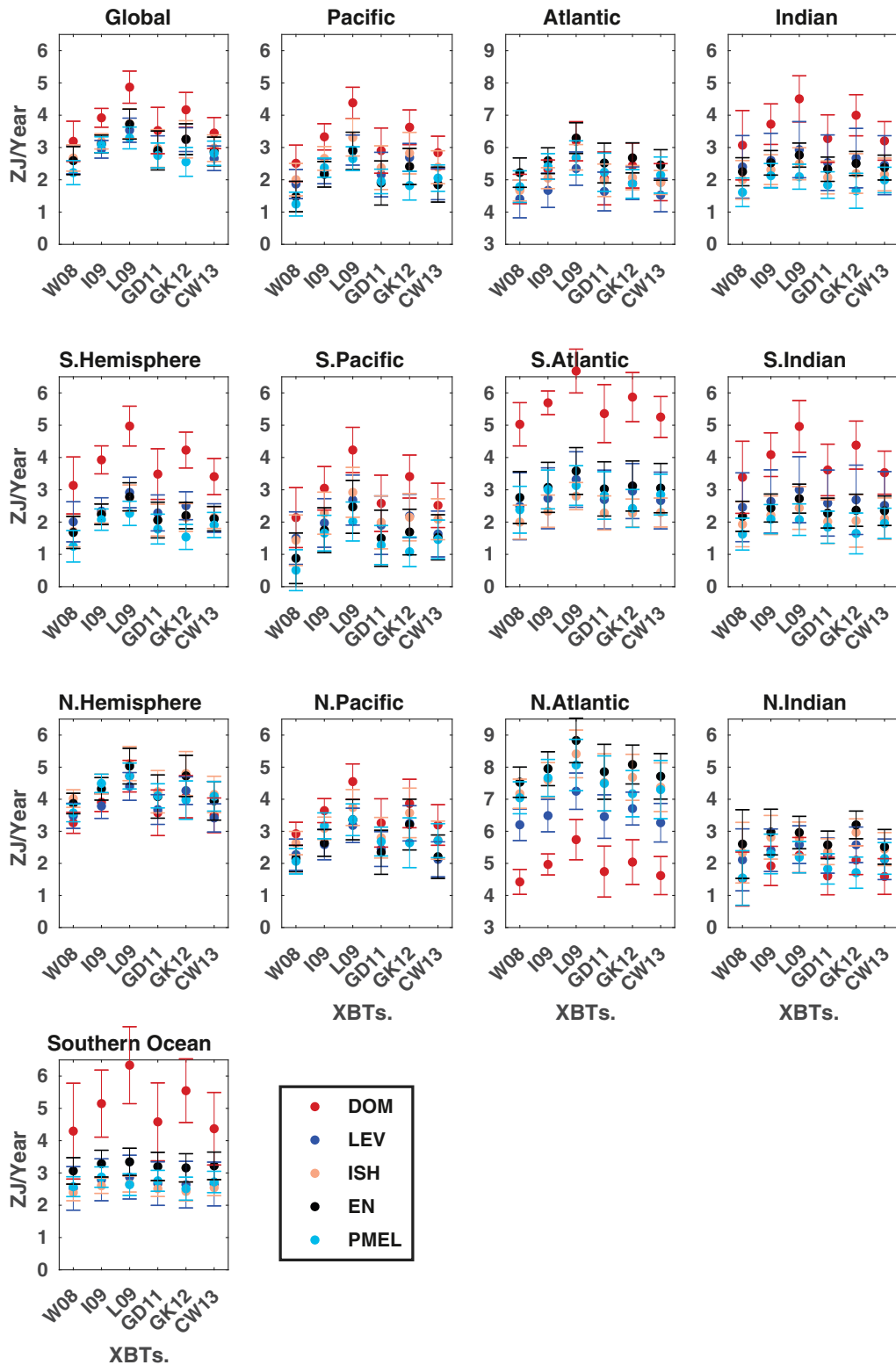
FIG. 13. Linear OHCA trends (ZJ yr^{-1}) for different basins (panels) and for each XBT bias adjustment (*x* axis) and for each mapping method (color legend) for 1970–2004. Error bars take into account the reduction in the degrees of freedom due to the temporal correlation in the residuals (section 2d).
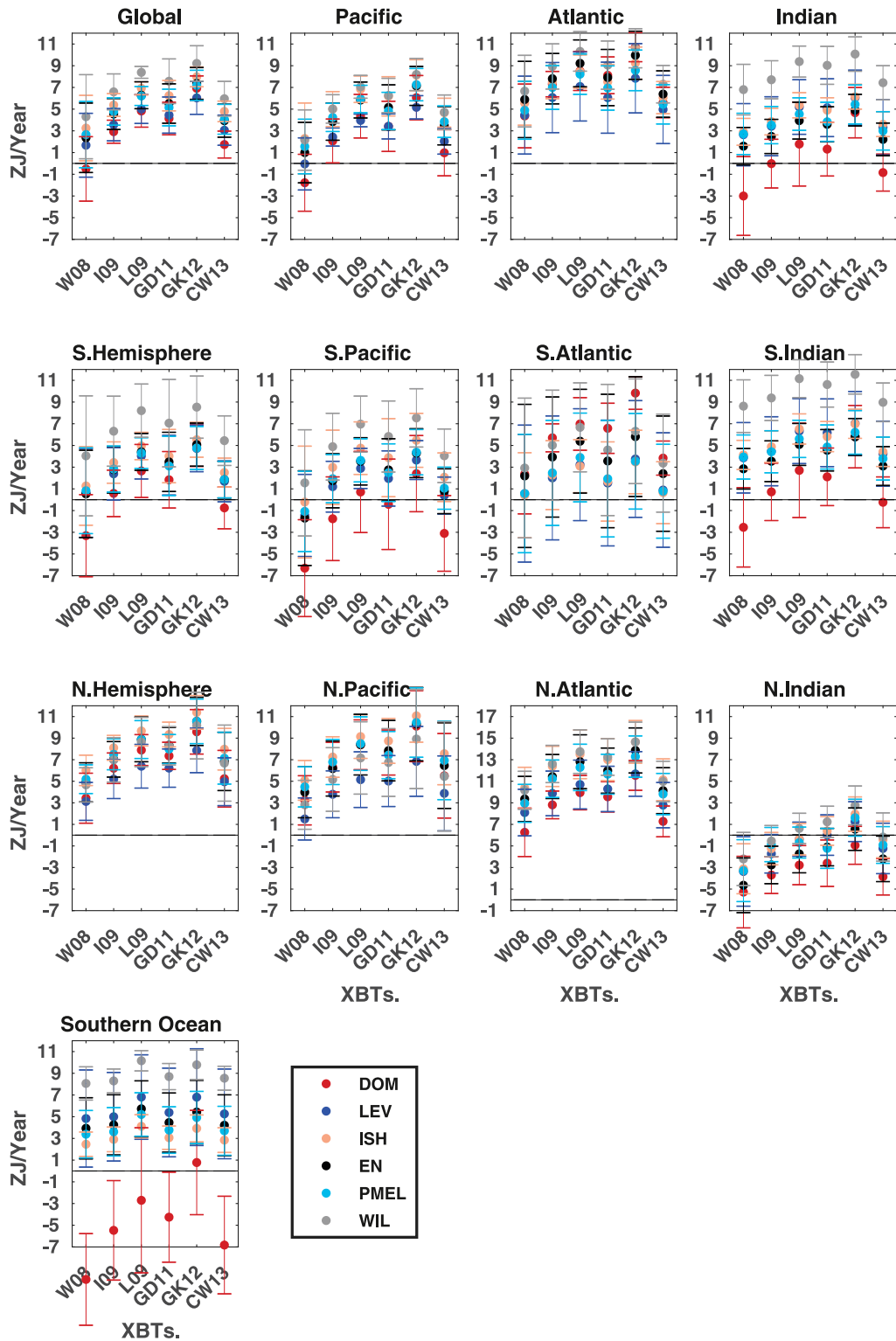
FIG. 14. As in Fig. 13, but for 1993–2004, and including WIL (only available since 1993).

at the locations of the observed 0–700-m profiles (i.e., with no mapping influence; Figs. 7a, 8b, and 8c for L09 and I09 and Figs. 7a, 8d, and 8e for GK12 and GD11), and next at the corresponding OHCA values subsampled from the gridded estimates (more variable due to dependency on mapping method; Figs. 7 and 8), and to a lesser extent if all

gridded values are considered (Fig. 2). More intriguingly, our results show that the OHCA estimates from the top L09 and GK12 schemes differ to a higher degree from each other in comparison to L09/I09 and GK12/GD11 respectively, and especially over the 1990s. The pairings for L09/I09 and GK12/GD11 found in this study are also evident in the global OHCA estimates from Cheng et al. (2016, their Fig. 3), who additionally noted that GK12 and CH14 XBT bias adjustments per se are closer to each other than L09 and CH14.

So, even though the top XBT bias adjustments per se are generally alike with respect to the proposed metrics in Cheng et al. (2018), our results demonstrate that their impact on global OHCA estimates are dissimilar. It might be possible that Cheng et al.'s (2018) metrics may still be imperfect, having in mind the disagreements they reported for the global and side-by-side related adjustments. It might also be possible that adjustments for all known bias factors, although desirable, may not be the single best performance indicator. Other unknown factors that can contribute toward XBT biases are not yet understood or accounted for in global schemes (e.g., Tan et al. 2021). Not all schemes provide adjustments to all types of XBTs, including the unknown types (Cheng et al. 2016), which make up about 50% of the XBT profiles from the global historical database (Abraham et al. 2013). The XBT science community has achieved tremendous progress to better understand and develop bias adjustments, greatly reducing artifacts in OHCA estimates. But until we better understand the reasons why the top recommended adjustments do not necessarily translate to similar OHCA estimates and what the causes of disagreement are in XBT bias adjustments developed using global and side-by-side datasets, and until we have studies corroborating Cheng et al. (2018), it is plausible to assume that spread due to XBT bias adjustment for global integrals is not yet smaller than reported in Boyer et al. (2016). To the best of our knowledge, the six XBT bias adjustments used in this study were correctly implemented in the EN3v2a database (section 2a), but we cannot rule out discrepancies in case there is database dependency (e.g., W08 comparison for EN3v1d and EN3v2a in Boyer et al. 2016; EN4 and WOD13 comparisons in Cheng et al. 2018).

The most significant spread in OHCA due to XBT bias adjustment lies within 30°N–30°S, especially in the Pacific and during the 1990s (Fig. 3), except for DOM and PMEL. DOM is particularly exceptional by being by far the most sensitive mapping method to XBT-related spread (Fig. 4) as well as by its spread patterns not being necessarily related to XBT data. Both results probably arise from the nature of DOM's cost function, which minimizes errors simultaneously at regional and global scales. PMEL's decorrelation radius better highlights the spread along the repeat XBT lines, and thus makes quite evident how the other mappings with local fitting propagate the XBT-related spread to broader areas based on their own radii of influence. As in Boyer et al. (2016), the caveat for the spread due to XBT bias adjustment using the gridded OHCA estimates is that it is subjected to the inherent interpolation/extrapolation of each mapping, and thus it can be affected, to a certain extent, in both magnitude and spatial

distribution. We find that XBT-related spread is smaller in cooler and less stratified waters at high latitudes (e.g., Fig. 9; see also Hutchinson et al. 2013; Ribeiro et al. 2018), although the Southern Ocean is quite undersampled. The OHCA analyses from the subset of non-gridded subsampled profiles for 0–700 m (i.e., no mapping influence) reveal that the maxima in spread in the Pacific around 1999–2000 (Fig. 10) is dominated by CW13. Future OHCA analyses without mapping influence and with all XBT profiles of the global database may help to refine bias adjustments where their impact on OHCA estimates is greatest.

Wang et al. (2018) reported significant OHCA spread in eddy-rich regions, with differences up to 10 times larger along the WBCs and across the ACC frontal regions compared to other ocean regions. Liang et al. (2021) arrived at a similar conclusion for objectively mapped OHCA Argo-related products in the upper 2000 m. In this study, significant OHCA spread (relative to a unique climatology) in eddy-rich regions is due to mapping method (Fig. 5), including the well-observed northwest Atlantic and the poorly observed Southern Ocean. WIL can better account for the energetic mesoscale features than LEV, ISH, EN, and PMEL (e.g., Fig. 12), as resolved by altimeter sea level. Martínez-Moreno et al. (2021) showed that the eddy-rich regions are becoming richer over the altimeter era. Mesoscale eddy variability can affect the slower-varying larger-scale climate signals through an inverse cascade of kinetic energy (Penduff et al. 2019, 2018). In DOM, however, mesoscale variability from altimeter sea level is considered as "high-frequency weather noise" and used as one of the sources of uncertainty in their reconstruction of the larger-scale OHCA patterns (i.e., low-frequency climate signal). The near-global coverage of the Argo floats array is designed to capture low-frequency climate signals and so its spatiotemporal resolution (e.g., 10 days, 3° × 3°) is not enough to resolve mesoscale signals (Fu et al. 2010; Hughes and Williams 2010; Penduff et al. 2011). The cost involved in sustaining such an in situ near-global array to monitor mesoscale variability is currently unattainable (e.g., Palmer et al. 2019). The best approach is to continue exploring synergies with complementary observing systems, such as satellite altimeter (e.g., Roemmich et al. 2019).

The Indian Ocean by far has the largest basin-scale spread due to mapping method, over the 1990s, regardless of XBT bias adjustment (Figs. 5 and 6). Our basin-average OHCA spread estimates are per square meter, separated due to XBT bias adjustment and mapping method, and relative to one baseline climatology. Major spread in OHCA basin estimates from Wang et al. (2018) are found in the Pacific and Southern Oceans; however, their values are not per square meter, and they cannot distinguish sources of spread. During 1993–2004, our OHCA basin averages per square meter for the Pacific and Atlantic, inclusive of their respective Southern Ocean sectors, are similarly influenced by both XBT and mapping spread (e.g., cf. Figs. 4 and 6). Overall, OHCA spread due to XBT adjustments is largest in the Pacific whereas spread due to mapping is largest in the Indian Ocean.

At observed 0–700-m locations, spread in global OHCA in the 1990s is highly correlated with number of observed

profiles (Figs. 9a,b) and explained by more extensive and deeper sampling of the energetic regions in the Southern Hemisphere (Figs. 10 and 11). This counterintuitive correlation, prior to inclusion of Argo data (section 2a), highlights the importance of the spatiotemporal design of the observing system in relation to the ocean dynamics to successfully resolve climate signals of interest.

The ensemble mean spatial patterns from our multidecadal trend maps for 1970–2004 and 1993–2004, based on the coordinated OHCA protocol (Fig. 12), have been previously discussed in numerous studies, based on the routine estimation of the various research groups [e.g., see Johnson and Lyman (2020) and references therein]. This is the first time, however, that we can directly compare individual trend maps, along with ensemble mean and spread, and attribute the inconsistency in patterns solely to differences in mapping methods. Major spread is found in the highly energetic eddy and frontal regions, within WBCs and across the subpolar Southern Ocean, over both periods (Fig. 12, right column), although with stronger and more widespread features during 1993–2004. The ensemble spread is about 2–3 times larger than the ensemble mean in the subpolar Southern Ocean for 1993–2004, and primarily caused by the cooling patterns in DOM. Except for DOM, the warming (cooling) trends in Fig. 14 (Fig. 13) for the poorly observed Southern Ocean basin from the various mappings, and largely regardless of XBT bias adjustment, are not statistically different. The effect of the various XBT bias adjustments only becomes visible when OHCA grid scale differences are integrated over large-scale areas (e.g., from zonal to basin to global). Overall, the global ocean in the upper 700 m is warming over longer/shorter multidecadal periods (e.g., Johnson and Lyman 2020), with the fastest rates in the North Atlantic. However, caution should be exercised in the interpretation of spatial patterns, as differences in individual mapping schemes per se can lead to statistically distinct trends (in sign and amplitude), from regional to basin scales.

5. Conclusions and recommendations

Our coordinated results cannot identify the best mapping or bias adjustment schemes, but they do provide insights into where and when the greatest spatiotemporal sensitivities (i.e., spread) in upper OHCA exist. Hence, our results provide valuable guidance for future developments aiming to have the greatest impact in understanding and reducing uncertainty caused by spread in upper OHCA estimates. By reconciling differences in OHCA estimation, we hope to achieve a more accurate spatiotemporal picture of the physical changes, mechanisms involved, and causes. To promote improved knowledge on the role of the subsurface ocean in the climate system, it is important to understand and quantify the full range of uncertainty and not just the contribution from spread between estimates (e.g., Kent and Kennedy 2021; Palmer et al. 2021; Thorne et al. 2005). In this larger context, our key messages and recommendations are the following:

Users should be aware of larger spread in the Pacific and Atlantic Oceans during the 1990s (due to spread in both XBT bias adjustment and mapping method), in the Indian Ocean (largely due to spread in mapping method), and in highly energetic eddy and frontal regions (e.g., where altimeter sea level has the largest variances). Undersampling is a limitation that should be recognized, when considering times and ocean volumes over which to estimate variability and trends. Continued recovery of actual historical profile data and metadata is valuable to reduce spread in estimates [e.g., Global Oceanographic Data Archeology and Rescue (GODAR), https://www.iode.org]. Now and into the future, having a sustained observing system fit for monitoring ocean climate change is the critical course of action—the Global Climate Observing System (GCOS; WMO 2018).

Further development and refinement of XBT bias adjustments should consider the same database version for both calibration and benchmarking of schemes. Soon, it will be possible to benefit from the first internationally coordinated quality-controlled ocean database (IQuOD), with assigned observational uncertainty (e.g., Cowley et al. 2021) and intelligent metadata (e.g., Palmer et al. 2018).

Future coordinated intercomparisons are necessary to evaluate the performance of mapping methods, including other methods not examined in this study (Barth et al. 2014; Cheng et al. 2017; Kuusela and Stein 2018; Su et al. 2021, 2020), based on synthetic profiles from Argo and/or altimeter observations as well as from de-drifted model simulations that conserve tracer properties (Allison et al. 2019; Cheng and Zhu 2014; Cheng et al. 2017; Garry et al. 2019; Good 2017; Palmer et al. 2019).

Lessons can be learned from the marine surface temperature community, which has been subjected to the same problem of quantifying the imprint of real climate signals from ocean data analyses based on noisy and unevenly distributed measurements, including the quantification of instrumental/methodological biases and the reduction of spurious signals (Kennedy 2014; Kent and Kennedy 2021). Synergistic efforts across multidisciplinary communities should be encouraged to speed up new developments, such as the integration of surface and subsurface ocean data with the full-range of uncertainty stated (e.g., Atkinson et al. 2014).

Finally, the relevance of improving estimation of OHCA and providing full quantification of uncertainty is justified by the uses in a wide range of climate science, such as hindcast reanalyses assimilating ocean data (Storto et al. 2019), ocean and/or climate model analyses investigating the mechanisms of ocean heat uptake (Couldrey et al. 2020; Dias et al. 2020a,b; Gregory 2000; Gregory et al. 2016; Saenko et al. 2021; Savita et al. 2021), detection and attribution of changes to natural and anthropogenic drivers (e.g., Gleckler et al. 2012), and constraining model projections of climate and sea level change relevant for policymakers and decision-makers (Lyu et al. 2021; IPCC 2007; Oppenheimer et al. 2019).

REFERENCES

Abraham, J. P., and Coauthors, 2013: A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Rev. Geophys.*, **51**, 450–483, https://doi.org/10.1002/rog.20022.

Allison, L., and Coauthors, 2019: Towards quantifying uncertainty in ocean heat content changes using synthetic profiles. *Environ. Res. Lett.*, **14**, 084037, https://doi.org/10.1088/1748-9326/ab2b0b.

Alory, G., S. Wijffels, and G. Meyers, 2007: Observed temperature trends in the Indian Ocean over 1960–1999 and associated mechanisms. *Geophys. Res. Lett.*, **34**, L02606, https://doi.org/10.1029/2006GL028044.

Argo, 2000: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE, accessed 2021, http://doi.org/10.17882/42182.

Armour, K. C., C. M. Bitz, and G. H. Roe, 2013: Time-varying climate sensitivity from regional feedbacks. *J. Climate*, **26**, 4518–4534, https://doi.org/10.1175/JCLI-D-12-00544.1.

Atkinson, C. P., N. A. Rayner, J. J. Kennedy, and S. A. Good, 2014: An integrated database of ocean temperature and salinity observations. *J. Geophys. Res. Oceans*, **119**, 7139–7163, https://doi.org/10.1002/2014JC010053.

Barker, P. M., J. R. Dunn, C. M. Domingues, and S. E. Wijffels, 2011: Pressure sensor drifts in Argo and their impacts. *J. Atmos. Oceanic Technol.*, **28**, 1036–1049, https://doi.org/10.1175/2011JTECHO831.1.

Barth, A., J.-M. Beckers, C. Troupin, A. Alvera Azcarate, and L. Vandenbulcke, 2014: divand-1.0: *n*-dimensional variational data analysis for ocean observations. *Geosci. Model Dev.*, **7**, 225–241, https://doi.org/10.5194/gmd-7-225-2014.

Bilbao, R. A. F., J. M. Gregory, N. Bouttes, M. D. Palmer, and P. Stott, 2019: Attribution of ocean temperature change to anthropogenic and natural forcings using the temporal, vertical and geographical structure. *Climate Dyn.*, **53**, 5389–5413, https://doi.org/10.1007/s00382-019-04910-1.

Bindoff, N. L., and Coauthors, 2007: Observations: Oceanic climate change and sea level. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 385–432.

——, and Coauthors, 2019: Changing ocean, Marine ecosystems, and dependent communities. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner et al., Eds., IPCC, 447–588.

Boyer, T., and Coauthors, 2016: Sensitivity of global upper-ocean heat content estimates to mapping methods, XBT bias corrections, and baseline climatologies. *J. Climate*, **29**, 4817–4842, https://doi.org/10.1175/JCLI-D-15-0801.1.

——, and Coauthors, 2018: *World Ocean Database 2018*, NOAA Atlas NESDIS 87, https://www.ncei.noaa.gov/products/world-ocean-database.

Carson, M., K. Lyu, K. Richter, M. Becker, C. M. Domingues, W. Han, and L. Zanna, 2019: Climate model uncertainty and trend detection in regional sea level projections: A review. *Surv. Geophys.*, **40**, 1631–1653, https://doi.org/10.1007/s10712-019-09559-3.

Cheng, L., and J. Zhu, 2014: Artifacts in variations of ocean heat content induced by the observation system changes. *Geophys. Res. Lett.*, **41**, 7276–7283, https://doi.org/10.1002/2014GL061881.

——, and ——, 2015: Influences of the choice of climatology on ocean heat content estimation. *J. Atmos. Oceanic Technol.*, **32**, 388–394, https://doi.org/10.1175/JTECH-D-14-00169.1.

——, and Coauthors, 2016: XBT science: Assessment of instrumental biases and errors. *Bull. Amer. Meteor. Soc.*, **97**, 924–933, https://doi.org/10.1175/BAMS-D-15-00031.1.

——, K. E. Trenberth, J. Fasullo, T. Boyer, J. Abraham, and J. Zhu, 2017: Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.*, **3**, e1601545, https://doi.org/10.1126/sciadv.1601545.

——, and Coauthors, 2018: How well can we correct systematic errors in historical XBT data? *J. Atmos. Oceanic Technol.*, **35**, 1103–1125, https://doi.org/10.1175/JTECH-D-17-0122.1.

Church, J. A., N. J. White, R. Coleman, K. Lambeck, and J. X. Mitrovica, 2004: Estimates of the regional distribution of sea level rise over the 1950–2000 period. *J. Climate*, **17**, 2609–2625, https://doi.org/10.1175/1520-0442(2004)017<2609:EOTRDO>2.0.CO;2.

——, and Coauthors, 2011: Revisiting the Earth's sea-level and energy budgets from 1961 to 2008. *Geophys. Res. Lett.*, **38**, L18601, https://doi.org/10.1029/2011GL048794.

——, and Coauthors, 2013: Sea level change. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1137–1216.

Couldrey, M. P., and Coauthors, 2020: What causes the spread of model projections of ocean dynamic sea-level change in response to greenhouse gas forcing? *Climate Dyn.*, **56**, 155–187, https://doi.org/10.1007/s00382-020-05471-4.

Cowley, R., S. Wijffels, L. Cheng, T. Boyer, and S. Kizu, 2013: Biases in expendable bathythermograph data: A new view based on historical side-by-side comparisons. *J. Atmos. Oceanic Technol.*, **30**, 1195–1225, https://doi.org/10.1175/JTECH-D-12-00127.1.

——, and Coauthors, 2021: International Quality-controlled Ocean Database (IQuOD) v0. 1: The temperature uncertainty specification. *Front. Mar. Sci.*, **8**, 689695, https://doi.org/10.3389/fmars.2021.689695.

Dias, F. B., C. M. Domingues, S. J. Marsland, S. Griffies, S. Rintoul, R. Matear, and R. Fiedler, 2020a: On the superposition of mean advective and eddy-induced transports in global

ocean heat and salt budgets. *J. Climate*, **33**, 1121–1140, https://doi.org/10.1175/JCLI-D-19-0418.1.

——, and Coauthors, 2020b: Ocean heat storage in response to changing ocean circulation processes. *J. Climate*, **33**, 9065–9082, https://doi.org/10.1175/JCLI-D-19-1016.1.

Domingues, C. M., J. A. Church, N. J. White, P. J. Gleckler, S. E. Wijffels, P. M. Barker, and J. R. Dunn, 2008: Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, **453**, 1090–1093, https://doi.org/10.1038/nature07080.

Durack, P. J., S. E. Wijffels, and P. J. Gleckler, 2014: Long-term sea-level change revisited: The role of salinity. *Environ. Res. Lett.*, **9**, 114017, https://doi.org/10.1088/1748-9326/9/11/114017.

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

Fasullo, J. T., and P. R. Gent, 2017: On the relationship between regional ocean heat content and sea surface height. *J. Climate*, **30**, 9195–9211, https://doi.org/10.1175/JCLI-D-16-0920.1.

Fox-Kemper, B., and Coauthors, 2021: Ocean, cryosphere and sea level change. *Climate Change 2021: The Physical Science Basis*, Cambridge University Press, in press.

Fu, L.-L., D. B. Chelton, P.-Y. Le Traon, and R. Morrow, 2010: Eddy dynamics from satellite altimetry. *Oceanography*, **23**, 14–25, https://doi.org/10.5670/oceanog.2010.02.

Garry, F., E. McDonagh, A. Blaker, C. D. Roberts, D. Desbruyères, E. Frajka-Williams, and B. King, 2019: Model-derived uncertainties in deep ocean temperature trends between 1990 and 2010. *J. Geophys. Res. Oceans*, **124**, 1155–1169, https://doi.org/10.1029/2018JC014225.

Gleckler, P. J., and Coauthors, 2012: Human-induced global ocean warming on multidecadal timescales. *Nat. Climate Change*, **2**, 524–529, https://doi.org/10.1038/nclimate1553.

Goni, G. J., and Coauthors, 2019: More than 50 years of successful continuous temperature section measurements by the global expendable bathythermograph network, its integrability, societal benefits, and future. *Front. Mar. Sci.*, **6**, 452, https://doi.org/10.3389/fmars.2019.00452.

Good, S. A., 2011: Depth biases in XBT data diagnosed using bathymetry data. *J. Atmos. Oceanic Technol.*, **28**, 287–300, https://doi.org/10.1175/2010JTECHO773.1.

——, 2017: The impact of observational sampling on time series of global 0–700 m ocean average temperature: A case study. *Int. J. Climatol.*, **37**, 2260–2268, https://doi.org/10.1002/joc.4654.

Gould, J., B. Sloyan, and M. Visbeck, 2013: In-situ ocean observations: A brief history, present status, and future directions. *Ocean Circulation and Climate: A 21st Century Perspective*, G. Siedler et al., Eds., Elsevier, 59–81.

Gouretski, V., 2012: Using GEBCO digital bathymetry to infer depth biases in the XBT data. *Deep-Sea Res. I*, **62**, 40–52, https://doi.org/10.1016/j.dsr.2011.12.012.

——, and K. P. Koltermann, 2007: How much is the ocean really warming? *Geophys. Res. Lett.*, **34**, L01610, https://doi.org/10.1029/2006GL027834.

——, and F. Reseghetti, 2010: On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep-Sea Res. I*, **57**, 812–833, https://doi.org/10.1016/j.dsr.2010.03.011.

Gregory, J. M., 2000: Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dyn.*, **16**, 501–515, https://doi.org/10.1007/s003820000059.

——, and Coauthors, 2016: The Flux-Anomaly-Forced Model Intercomparison Project (FAFMIP) contribution to CMIP6: Investigation of sea-level and ocean climate change in response to CO_2 forcing. *Geosci. Model Dev.*, **9**, 3993–4017, https://doi.org/10.5194/gmd-9-3993-2016.

——, and Coauthors, 2019: Concepts and terminology for sea level: Mean, variability and change, both local and global. *Surv. Geophys.*, **40**, 1251–1289, https://doi.org/10.1007/s10712-019-09525-z.

Gulev, S. K., and Coauthors, 2021: Changing state of the climate system. *Climate Change 2021: The Physical Science Basis*, Cambridge University Press, in press.

Hamlington, B. D., T. Frederikse, R. S. Nerem, J. T. Fasullo, and S. Adhikari, 2020a: Investigating the acceleration of regional sea level rise during the satellite altimeter era. *Geophys. Res. Lett.*, **47**, e2019GL086528, https://doi.org/10.1029/2019GL086528.

——, and Coauthors, 2020b: Origin of interannual variability in global mean sea level. *Proc. Natl. Acad. Sci. USA*, **117**, 13 983–13 990, https://doi.org/10.1073/pnas.1922190117.

Hanawa, K., P. Rual, R. Bailey, A. Sy, and M. Szabados, 1995: A new depth–time equation for Sippican or TSK T-7, T-6 and T-4 expendable bathythermographs (XBT). *Deep-Sea Res. I*, **42**, 1423–1451, https://doi.org/10.1016/0967-0637(95)97154-Z.

Hughes, C. W., and S. D. P. Williams, 2010: The color of sea level: Importance of spatial variations in spectral shape for assessing the significance of trends. *J. Geophys. Res.*, **115**, C10048, https://doi.org/10.1029/2010JC006102.

Hughes, T. P., and Coauthors, 2018: Global warming transforms coral reef assemblages. *Nature*, **556**, 492–496, https://doi.org/10.1038/s41586-018-0041-2.

Hutchinson, K., S. Swart, I. Ansorge, and G. Goni, 2013: Exposing XBT bias in the Atlantic sector of the Southern Ocean. *Deep-Sea Res. I*, **80**, 11–22, https://doi.org/10.1016/j.dsr.2013.06.001.

Ingleby, B., and M. Huddleston, 2007: Quality control of ocean temperature and salinity profiles—Historical and real-time data. *J. Mar. Syst.*, **65**, 158–175, https://doi.org/10.1016/j.jmarsys.2005.11.019.

IPCC, 2007: *Climate Change 2007: Impacts, Adaptation, and Vulnerability*. Cambridge University Press, 976 pp., https://www.ipcc.ch/pdf/assessment-report/ar4/wg2/ar4_wg2_full_report.pdf.

——, 2019: Summary for policymakers. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. H.-O. Pörtner et al., Eds., https://www.ipcc.ch/srocc/chapter/summary-for-policymakers/.

——, 2021: Summary for policymakers. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/.

Ishii, M., and M. Kimoto, 2009: Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *J. Oceanogr.*, **65**, 287–299, https://doi.org/10.1007/s10872-009-0027-7.

——, Y. Fukuda, S. Hirahara, S. Yasui, T. Suzuki, and K. Sato, 2017: Accuracy of global upper ocean heat content estimation expected from present observational data sets. *SOLA*, **13**, 163–167, https://doi.org/10.2151/sola.2017-030.

Johnson, G. C., and J. M. Lyman, 2020: Warming trends increasingly dominate global ocean. *Nat. Climate Change*, **10**, 757–761, https://doi.org/10.1038/s41558-020-0822-0.

——, S. G. Purkey, N. V. Zilberman, and D. Roemmich, 2019: Deep Argo quantifies bottom water warming rates in the

southwest Pacific Basin. *Geophys. Res. Lett.*, **46**, 2662–2669, https://doi.org/10.1029/2018GL081685.

Kennedy, J. J., 2014: A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.*, **52** (1), 1–32, https://doi.org/10.1002/2013RG000434.

Kent, E. C., and J. J. Kennedy, 2021: Historical estimates of surface marine temperatures. *Annu. Rev. Mar. Sci.*, **13**, 283–311, https://doi.org/10.1146/annurev-marine-042120-111807.

Kuusela, M., and M. L. Stein, 2018: Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proc. Roy. Soc. London*, **474A**, 20180400, https://doi.org/10.1098/rspa.2018.0400.

Landerer, F. W., J. H. Jungclaus, and J. Marotzke, 2007: Regional dynamic and steric sea level change in response to the IPCC-A1B scenario. *J. Phys. Oceanogr.*, **37**, 296–312, https://doi.org/10.1175/JPO3013.1.

Leahy, T. P., F. P. Llopis, M. D. Palmer, and N. H. Robinson, 2018: Using neural networks to correct historical climate observations. *J. Atmos. Oceanic Technol.*, **35**, 2053–2059, https://doi.org/10.1175/JTECH-D-18-0012.1.

Le Quéré, C., and Coauthors, 2018: Global carbon budget 2017. *Earth Syst. Sci. Data*, **10**, 405–448, https://doi.org/10.5194/essd-10-405-2018.

Levitus, S., J. I. Antonov, T. P. Boyer, and C. Stephens, 2000: Warming of the world ocean. *Science*, **287**, 2225–2229, https://doi.org/10.1126/science.287.5461.2225.

——, ——, ——, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov, 2009: Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.*, **36**, L07608, https://doi.org/10.1029/2008GL037155.

——, and Coauthors, 2012: World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett.*, **39**, L10603, https://doi.org/10.1029/2012GL051106.

Liang, X., C. Liu, R. M. Ponte, and D. P. Chambers, 2021: A comparison of the variability and changes in global ocean heat content from multiple objective analysis products during the Argo period. *J. Climate*, **34**, 7875–7895, https://doi.org/10.1175/JCLI-D-20-0794.1.

Lyman, J. M., and G. C. Johnson, 2008: Estimating annual global upper-ocean heat content anomalies despite irregular in situ ocean sampling. *J. Climate*, **21**, 5629–5641, https://doi.org/10.1175/2008JCLI2259.1.

——, and ——, 2014: Estimating global ocean heat content changes in the upper 1800 m since 1950 and the influence of climatology choice. *J. Climate*, **27**, 1945–1957, https://doi.org/10.1175/JCLI-D-12-00752.1.

——, and Coauthors, 2010: Robust warming of the global upper ocean. *Nature*, **465**, 334–337, https://doi.org/10.1038/nature09043.

Lyu, K., X. Zhang, J. A. Church, J. Hu, and J.-Y. Yu, 2017: Distinguishing the quasi-decadal and multidecadal sea level and climate variations in the Pacific: Implications for the ENSO-like low-frequency variability. *J. Climate*, **30**, 5097–5117, https://doi.org/10.1175/JCLI-D-17-0004.1.

——, ——, and ——, 2021: Projected ocean warming constrained by the ocean observational record. *Nat. Climate Change*, **11**, 834–839, https://doi.org/10.1038/s41558-021-01151-1.

Marcos, M., B. Marzeion, S. Dangendorf, A. B. Slangen, H. Palanisamy, and L. Fenoglio-Marc, 2017: Internal variability versus anthropogenic forcing on sea level and its components. *Integrative Study of the Mean Sea Level and Its Components*, Springer, 337–356.

Martínez-Moreno, J., A. M. Hogg, M. H. England, N. C. Constantinou, A. E. Kiss, and A. K. Morrison, 2021: Global changes in oceanic mesoscale currents over the satellite altimetry record. *Nat. Climate Change*, **11**, 397–403, https://doi.org/10.1038/s41558-021-01006-9.

Meyssignac, B., and Coauthors, 2019: Measuring global ocean heat content to estimate the Earth energy imbalance. *Front. Mar. Sci.*, **6**, 432, https://doi.org/10.3389/fmars.2019.00432.

Olsen, E., and Coauthors, 2018: Ocean futures under ocean acidification, marine protection, and changing fishing pressures explored using a worldwide suite of ecosystem models. *Front. Mar. Sci.*, **5**, 64, https://doi.org/10.3389/fmars.2018.00064.

Oppenheimer, M., and Coauthors, 2019: Sea level rise and implications for low lying islands, coasts and communities. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner et al., Eds., Cambridge University Press, 321–445.

Otto, A., and Coauthors, 2013: Energy budget constraints on climate response. *Nat. Geosci.*, **6**, 415–416, https://doi.org/10.1038/ngeo1836.

Palmer, M., and Coauthors, 2010: Future observations for monitoring global ocean heat content. *OceanObs' 09: Sustained Ocean Observations and Information for Society*, J. Hall et al., Eds., ESA Publication WPP-306, 1–13, https://doi.org/10.5270/OceanObs09.cwp.68.

——, T. Boyer, R. Cowley, S. Kizu, F. Reseghetti, T. Suzuki, and A. Thresher, 2018: An algorithm for classifying unknown expendable bathythermograph (XBT) instruments based on existing metadata. *J. Atmos. Oceanic Technol.*, **35**, 429–440, https://doi.org/10.1175/JTECH-D-17-0129.1.

——, and Coauthors, 2019: Adequacy of the ocean observation system for quantifying regional heat and freshwater storage and change. *Front. Mar. Sci.*, **6**, 416, https://doi.org/10.3389/fmars.2019.00416.

——, C. Domingues, A. Slangen, and F. B. Dias, 2021: An ensemble approach to quantify global mean sea-level rise over the 20th century from tide gauge reconstructions. *Environ. Res. Lett.*, **16**, 044043, https://doi.org/10.1088/1748-9326/abdaec.

Penduff, T., M. Juza, B. Barnier, J. Zika, W. K. Dewar, A. Treguier, J. Molines, and N. Audiffren, 2011: Sea level expression of intrinsic and forced ocean variabilities at interannual time scales. *J. Climate*, **24**, 5652–5670, https://doi.org/10.1175/JCLI-D-11-00077.1.

——, and Coauthors, 2018: Chaotic variability of ocean heat content: Climate-relevant features and observational implications. *Oceanography*, **31**, 63–71, https://doi.org/10.5670/oceanog.2018.210.

——, W. Llovel, S. Close, I. Garcia-Gomez, and S. Leroux, 2019: Trends of coastal sea level between 1993 and 2015: Imprints of atmospheric forcing and oceanic chaos. *Surv. Geophys.*, **40**, 1543–1562, https://doi.org/10.1007/s10712-019-09571-7.

Pittman, N., 2016: Upper-ocean thermosteric sea level (heat content): Exploring the sensitivity of the CSIRO-ACE CRC-IMAS estimates to 10 XBT bias corrections. Master's thesis, Institute for Marine and Antarctic Studies, University of Tasmania, 97 pp.

Raper, S. C., J. M. Gregory, and R. J. Stouffer, 2002: The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *J. Climate*, **15**, 124–130, https://doi.org/10.1175/1520-0442(2002)015<0124:TROCSA>2.0.CO;2.

Rhein, M., and Coauthors, 2013: Observations: Ocean. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al.,

Eds., Cambridge University Press, 255–316, https://doi.org/10.1017/CBO9781107415324.010.

Ribeiro, N., M. M. Mata, J. L. L. de Azevedo, and M. Cirano, 2018: An assessment of the XBT fall-rate equation in the Southern Ocean. *J. Atmos. Oceanic Technol.*, **35**, 911–926, https://doi.org/10.1175/JTECH-D-17-0086.1.

Riser, S. C., and Coauthors, 2016: Fifteen years of ocean observations with the global Argo array. *Nat. Climate Change*, **6**, 145–153, https://doi.org/10.1038/nclimate2872.

Roemmich, D., and Coauthors, 2019: On the future of Argo: A global, full-depth, multi-disciplinary array. *Front. Mar. Sci.*, **6**, 439, https://doi.org/10.3389/fmars.2019.00439.

Saenko, O. A., J. M. Gregory, S. M. Griffies, M. P. Couldrey, and F. B. Dias, 2021: Contribution of ocean physics and dynamics at different scales to heat uptake in low-resolution AOGCMs. *J. Climate*, **34**, 2017–2035, https://doi.org/10.1175/JCLI-D-20-0652.1.

Santer, B. D., and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703–1722, https://doi.org/10.1002/joc.1756.

Savita, A., and Coauthors, 2021: Super residual circulation: A new perspective on ocean vertical heat transport. *J. Phys. Oceanogr.*, **51**, 2443–2462, https://doi.org/10.1175/JPO-D-21-0008.1.

Seneviratne, S. I., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230.

Slangen, A. B. A., J. A. Church, X. Zhang, and D. Monselesan, 2014: Detection and attribution of global mean thermosteric sea level change. *Geophys. Res. Lett.*, **41**, 5951–5959, https://doi.org/10.1002/2014GL061356.

Stocker, T. F., 2015: The silent services of the world ocean. *Science*, **350**, 764–765, https://doi.org/10.1126/science.aac8720.

Storto, A., and Coauthors, 2019: Ocean reanalyses: Recent advances and unsolved challenges. *Front. Mar. Sci.*, **6**, 418, https://doi.org/10.3389/fmars.2019.00418.

Su, H., H. Zhang, X. Geng, T. Qin, W. Lu, and X.-H. Yan, 2020: OPEN: A new estimation of global ocean heat content for upper 2000 meters from remote sensing data. *Remote Sens.*, **12**, 2294, https://doi.org/10.3390/rs12142294.

——, T. Qin, A. Wang, and W. Lu, 2021: Reconstructing ocean heat content for revisiting global ocean warming from remote sensing perspectives. *Remote Sens.*, **13**, 3799, https://doi.org/10.3390/rs13193799.

Tan, Z., and Coauthors, 2021: Examining the influence of recording system on the pure temperature error in XBT data. *J. Atmos. Oceanic Technol.*, **38**, 759–776, https://doi.org/10.1175/JTECH-D-20-0136.1.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, https://doi.org/10.1029/2000JD900719.

Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears, 2005: Uncertainties in climate trends: Lessons from upper-air temperature records. *Bull. Amer. Meteor. Soc.*, **86**, 1437–1442, https://doi.org/10.1175/BAMS-86-10-1437.

Tokarska, K. B., G. C. Hegerl, A. P. Schurer, A. Ribes, and J. T. Fasullo, 2019: Quantifying human contributions to past and future ocean warming and thermosteric sea level rise. *Environ. Res. Lett.*, **14**, 074020, https://doi.org/10.1088/1748-9326/ab23c1.

van de Wal, R. S. W., and Coauthors, 2019: Uncertainties in long-term twenty-first century process-based coastal sea-level projections. *Surv. Geophys.*, **40**, 1655–1671, https://doi.org/10.1007/s10712-019-09575-3.

Vivier, F., K. A. Kelly, and M. Harismendy, 2005: Causes of large-scale sea level variations in the Southern Ocean: Analyses of sea level and a barotropic model. *J. Geophys. Res.*, **110**, C09014, https://doi.org/10.1029/2004JC002773.

von Schuckmann, K., and Coauthors, 2020: Heat stored in the Earth system: Where does the energy go? *Earth Syst. Sci. Data*, **12**, 2013–2041, https://doi.org/10.5194/essd-12-2013-2020.

Wang, G., L. Cheng, J. Abraham, and C. Li, 2018: Consensuses and discrepancies of basin-scale ocean heat content changes in different ocean analyses. *Climate Dyn.*, **50**, 2471–2487, https://doi.org/10.1007/s00382-017-3751-5.

WCRP Global Sea Level Budget Group, 2018: Global sea-level budget 1993–present. *Earth Syst. Sci. Data*, **10**, 1551–1590, https://doi.org/10.5194/essd-10-1551-2018.

Wijffels, S. E., and Coauthors, 2008: Changing expendable bathythermograph fall rates and their impact on estimates of thermosteric sea level rise. *J. Climate*, **21**, 5657–5672, https://doi.org/10.1175/2008JCLI2290.1.

Willis, J. K., D. Roemmich, and B. Cornuelle, 2004: Interannual variability in upper ocean heat content, temperature, and thermosteric expansion on global scales. *J. Geophys. Res.*, **109**, C12036, https://doi.org/10.1029/2003JC002260.

WMO, 2018: Systematic observations and the Paris Agreement. Report of the Task Team on the Paris Agreement, GCOS-222, 53 pp., https://library.wmo.int/doc_num.php?explnum_id=5417.

Wu, L., and Coauthors, 2012: Enhanced warming over the global subtropical western boundary currents. *Nat. Climate Change*, **2**, 161–166, https://doi.org/10.1038/nclimate1353.

Wu, Q., X. Zhang, J. A. Church, and J. Hu, 2019: ENSO-related global ocean heat content variations. *J. Climate*, **32**, 45–68, https://doi.org/10.1175/JCLI-D-17-0861.1.

Yang, H., G. Lohmann, W. Wei, M. Dima, M. Ionita, and J. Liu, 2016: Intensification and poleward shift of subtropical western boundary currents in a warming climate. *J. Geophys. Res. Oceans*, **121**, 4928–4945, https://doi.org/10.1002/2015JC011513.