

# *Historical ocean heat uptake in two pairs of CMIP6 models: global and regional perspectives*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Kuhlbrodt, T. ORCID: <https://orcid.org/0000-0003-2328-6729>, Voldoire, A., Palmer, M. D., Geoffroy, O. and Killick, R. E. (2023) Historical ocean heat uptake in two pairs of CMIP6 models: global and regional perspectives. *Journal of Climate*, 36 (7). pp. 2183-2203. ISSN 1520-0442 doi: 10.1175/JCLI-D-22-0468.1 Available at <https://centaur.reading.ac.uk/109126/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1175/JCLI-D-22-0468.1>

Publisher: American Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

## Historical Ocean Heat Uptake in Two Pairs of CMIP6 Models: Global and Regional Perspectives

TILL KUHLBRODT<sup>a</sup>, AURORE VOLDOIRE<sup>b</sup>, MATTHEW D. PALMER<sup>c,d</sup>, OLIVIER GEOFFROY<sup>b</sup>, AND RACHEL E. KILLICK<sup>c</sup>

<sup>a</sup> National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom

<sup>b</sup> CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

<sup>c</sup> Met Office Hadley Centre, Exeter, United Kingdom

<sup>d</sup> University of Bristol, Bristol, United Kingdom

(Manuscript received 7 July 2022, in final form 2 November 2022)

**ABSTRACT:** Ocean heat content (OHC) is one of the most relevant metrics tracking the current global heating. Therefore, simulated OHC time series are a cornerstone for assessing the scientific performance of Earth system models and global climate models. Here we present a detailed analysis of OHC change in simulations of the historical climate (1850–2014) performed with two pairs of CMIP6 models: U.K. Earth System Model 1 (UKESM1.0) and HadGEM3-GC3.1-LL, and CNRM-ESM2-1 and CNRM-CM6-1. The small number of models enables us to analyze OHC change globally and for individual ocean basins, making use of a novel ensemble of observational products. For the top 700 m of the global ocean, the two CNRM models reproduce the observed OHC change since the 1960s closely. The two U.K. models (UKESM1.0-LL and HadGEM3-GC3.1-LL) compensate a lack of warming in the 0–700 m layer in the 1970s and 1980s with warming below 2000 m. The observed warming between 700 and 2000 m is substantially underestimated by all models. An increased relevance for ocean heat uptake in the Atlantic after 1991—suggested by observations—is picked up by the U.K. models but less so by the CNRM models, probably related to an AMOC strengthening in the U.K. models. The regional ocean heat uptake characteristics differ even though all four models share the same ocean component (NEMO ORCA1). Differences in the simulated global, full-depth OHC time series can be attributed to differences in the model's total effective radiative forcing.

**KEYWORDS:** Ocean circulation; Climate change; Radiative forcing; Climate models; General circulation models; Anthropogenic effects/forcing

### 1. Introduction

Ocean heat content (OHC) is one of the most relevant metrics for tracking the rate of global climate change; it is arguably more relevant than surface temperature (von Schuckmann et al. 2016). Approximately 89% of the Earth energy imbalance between 1971 and 2018 is stored in the global ocean (von Schuckmann et al. 2020), raising its heat content. Changes and variability in OHC closely track Earth's energy imbalance (Meyssignac et al. 2019; Palmer and McNeill 2014). We define OHC as the integral over a volume of water multiplied by the density and the specific heat capacity (details below), and ocean heat uptake (OHU) as the change of the global OHC with time.

Observations of global OHC from the recent decades show a clear upward trend. The global ocean has been warming since at least the 1960s, with an acceleration in the rate of OHU since the 1990s (Cheng et al. 2017, 2020; Ishii et al.

2017; Levitus et al. 2012). Before the 1960s the underlying observations of ocean temperature are spatially and temporally very sparse, resulting in large uncertainties (Abraham et al. 2013; Palmer 2017). New reconstructions of global OHC that use estimates of ocean circulation to propagate surface temperature anomalies into the ocean interior suggest global ocean heat gain since the beginning of the twentieth century (Gebbie and Huybers 2019; Zanna et al. 2019).

Individual simulations of the historical climate with climate and Earth system models do not consistently trace the OHC observations within their uncertainty range. The models used for phase 5 of the Climate Model Intercomparison Project (CMIP5; Flato et al. 2013) display a wide range of OHC change in the upper 700 m over the latter part of the twentieth century (Cheng et al. 2016; Gleckler et al. 2016), ranging from 30 to 280 ZJ for the period 1971–2005 (Flato et al. 2013, their Fig. 9.17), with three different observations-based time series indicating between 100 and 120 ZJ (1 ZJ =  $10^{21}$  J). One of the possible reasons for this wide spread is the varying vertical stratification in the model oceans, linked to the capacity for downward heat transport (Kuhlbrodt and Gregory 2012). At the same time, the CMIP5 multimodel mean (Gleckler et al.

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

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-22-0468.s1>.

Corresponding author: Till Kuhlbrodt, [t.kuhlbrodt@reading.ac.uk](mailto:t.kuhlbrodt@reading.ac.uk)



This article is licensed under a Creative Commons Attribution 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

DOI: 10.1175/JCLI-D-22-0468.1

© 2023 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

2016) and multimodel median (Cheng et al. 2016) are close to the observed trend in OHC over the late twentieth century, lending confidence that there is no systematic OHU bias in the CMIP5 models. A first analysis of the 1971–2014 heat uptake in the CMIP6 models (Eyring et al. 2021, Fig. 3.26) suggests that for the 0–700 m layer the model ensemble spread (90% probability) is comparable to the spread in an ensemble of observations, with the model ensemble mean close to the observational ensemble mean as in CMIP5. However, for the 700–2000 m layer the CMIP6 models tend to underestimate OHU, and the model spread is much wider than the observational spread.

In this paper we assess the global and the basinwise OHC change of four models used in CMIP6 (Eyring et al. 2016): U.K. Earth System Model 1 (UKESM1.0), HadGEM3 GC3.1-LL, CNRM-ESM2-1, and CNRM-CM6-1. The global perspective serves as a powerful model metric and ensures comparability of the results presented here with other CMIP6 models. The basinwise, or regional, perspective allows for a deeper understanding of the model processes in comparison to the observations.

## 2. Models and data

### a. Employed models

In the following we give a brief description of the models used in this study. HadGEM3 GC3.1 is the state-of-the-art U.K. Met Office global coupled climate model (Williams et al. 2018). The component models of HadGEM3 are the Unified Model (UM) for the atmosphere, Nucleus for European Modelling of the Ocean (NEMO) for the ocean, CICE for the sea ice, and the Joint U.K. Land Environment Simulator (JULES) model for land surface processes. The components interact through the OASIS3-Model Coupling Toolkit (MCT) coupler. The low-resolution configuration used in this study is HadGEM3 GC3.1-LL (Kuhlbrodt et al. 2018). With its nominal resolution of 135 km in the atmosphere and 1° in the ocean it was specifically developed for forming the physical core of the U.K. Earth System Model, UKESM1.0. The lower resolution of HadGEM3 GC3.1-LL gives adequate computational efficiency for an Earth system model while ensuring comprehensive traceability with the higher resolution versions of HadGEM3 GC3.1.

UKESM1.0 (within CMIP6: UKESM1.0-LL) is a new, comprehensive Earth system model (ESM) comprising both physical and biogeochemical aspects of the Earth's atmosphere, ocean, cryosphere and land systems (Sellar et al. 2019). The physical core, HadGEM3 GC3.1-LL, is extended through the addition of ocean and land biogeochemistry and interactive stratospheric-tropospheric trace gas chemistry. A large number of couplings between the Earth system model components mean that UKESM1.0 is currently the most process-comprehensive ESM in CMIP6.

Similarly, CNRM-CM6-1 is the CNRM-CERFACS global climate model (Voldoire et al. 2019) physical core, which is the base for the ESM version, CNRM-ESM2-1. CNRM-CM6-1 is composed of ARPEGE-Climate v6 for the atmosphere;

Surface Externalisée (SURFEX), version 8, for the land surface; NEMO v3.6 for the ocean; and Global Experimental Leads and Ice for Atmosphere and Ocean (GELATO), version 6, for the sea ice. As for HadGEM3 GC3.1, the components are coupled using OASIS3-MCT, the atmospheric resolution is 140 km, and the ocean is 1°. CNRM-ESM2-1 (Séférian et al. 2019) is an extension of CNRM-CM6-1 that includes interactive aerosols representation, a stratospheric chemistry scheme as well as carbon fluxes between all components, through land and ocean biogeochemistry. In the following we will use the term “CNRM models” for CNRM-ESM2-1 and CNRM-CM6-1, and “U.K. models” for UKESM1.0 and HadGEM3 GC3.1-LL.

We use the ensembles of CMIP6 historical simulations from these four models for our analysis, pointing out that the number of ensemble members varies greatly across the models: it is 16 for UKESM1.0, 4 for GC3.1-LL, 10 for CNRM-ESM2-1, and 30 for CNRM-CM6-1. For an individual model, each historical simulation was branched off from the piControl simulation at a specific year, ensuring that each individual historical simulation has different initial conditions. The method of choosing the branching-off years varies across models as detailed in the documentation of these experiments (Andrews et al. 2020; Sellar et al. 2019; Voldoire et al. 2019).

### b. Traceability between the ocean components

All four models studied in the present paper use the same ocean model, NEMO3.6, in the same *shaconemo* eORCA1 configuration (Boucher et al. 2020) with only few differences between the U.K. and the CNRM models regarding parameterizations and parameter settings.

The sea surface temperature (SST) biases in the piControl simulation (Fig. 1) across the models show similarities and differences. In the Atlantic basin, the four bias patterns are fairly similar, with a cold bias in the northwest Atlantic (a known feature of 1° grids; Griffies et al. 2016) and the long-standing warm bias in the eastern South Atlantic (Zuidema et al. 2016). In the Pacific, the equatorial and subtropical cold biases are found in all models. However, the warm bias linked to upwelling in the eastern subtropical basins are more marked in the CNRM models, especially in CNRM-ESM2-1, than in the U.K. models. The eastern subtropical basin SST bias in CNRM-CM6-1 has been shown to be related to a lack of low-cloud (Brient et al. 2019) and the bias increase in the ESM version is due to a reduced aerosols loading and subsequent reduced masking effect (Séférian et al. 2019). The global zonally averaged temperature bias (Fig. 2) shows again some common features and some differences. All models have a cold bias in the top 200 m (coming mainly from the Pacific) and in the Arctic Ocean. The warm bias between 200 and 1000 m is more pronounced in the U.K. models. We will see later that this intermediate depth warm bias in the U.K. models is located in the North Atlantic (section 5), and that the U.K. models tend to have overly strong heat uptake in the Atlantic in the historical simulations (section 4). In the high-latitude Southern Ocean there is a warm bias below 1000 m in the CNRM models (again clearly stronger in CNRM-ESM2-1)

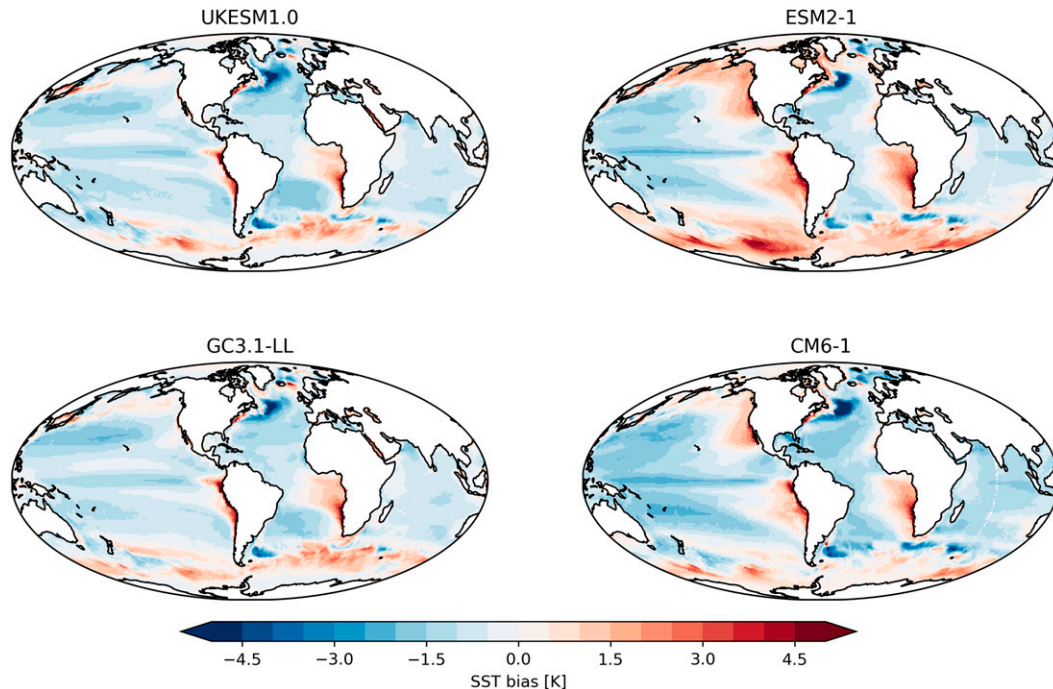


FIG. 1. Annual mean SST biases of the four models against ESA Climate Change Initiative (CCI) observations (Merchant et al. 2014). Shown are average annual means of the first 100 years of the piControl simulations. The bias in this and the next figure is calculated against present-day observations because observations from the mid-nineteenth century with global coverage are not available.

and in HadGEM3-GC3.1-LL, but not in UKESM1.0 which instead displays a cold bias below 3000 m. All four models show a warm bias in the intermediate depth subtropical Southern Ocean, around 30°S and 500 m depth. This is probably a result of too strong wind-driven downwelling, or downwelling of too warm waters, in the regions north of the Antarctic Circumpolar Current (ACC).

Table 1 gives an overview of the differences in the configuration of NEMO ORCA1 between the U.K. models and the CNRM models. The main differences between the U.K. and the CNRM configurations of ORCA1 concern dianeutral mixing parameterizations. The CNRM models use the Fox–Kemper scheme for parameterizing submesoscale mixed layer eddies, and a new parameterization for dianeutral mixing induced by internal waves. The U.K. models employ an asymmetrical (regarding the equator) latitudinal variation of the penetration depth of turbulent kinetic energy beyond the mixed layer. All these differences will affect the dianeutral heat transport, as is known at least from short sensitivity experiments (Kuhlbrodt et al. 2018). Conducting sensitivity experiments to assess their effect on global OHU is a large undertaking because (i) the involved dianeutral processes have a long time scale and hence long experiments are needed, and (ii) any change in parameters and parameterizations may affect the background circulation state and the stratification, and hence new baseline (piControl) simulations would be necessary. It would have been beyond the scope of this study to do such experiments. For this reason, we cannot exclude the possibility of these parameterization

differences on the model's OHU characteristics. However, in this study we present evidence that the radiative forcing and its components have a major effect on OHU in the models, and we indicate links between biases in the model background state and its OHU characteristics.

### c. Ocean heat content and observational datasets

From the model simulations, ocean heat content in a volume of water  $B$  is calculated as

$$\text{OHC}_B(t) = C_p \rho_0 \sum_{x,y,z \in B} \theta[x,y,z(t),t] V[x,y,z(t),t],$$

where  $\theta$  is the potential temperature in an individual grid cell, and  $V$  is its volume. The heat capacity  $C_p = 3991.868 \text{ J kg}^{-1} \text{ K}^{-1}$  and the reference density of seawater  $\rho_0 = 1026 \text{ kg m}^{-3}$  are constants in NEMO as used for these CMIP6 simulations. The summation over the grid cells uses the time-varying cell thickness  $z(t)$  in NEMO ORCA1. The summation volume  $B$  could be the entire World Ocean (WO), an ocean basin, or a specific layer. OHU is simply defined as the change in OHC relative to a given initial value, for instance, the increase in OHC since 1991, diagnosed in 2014.

Time series of global OHC change in the upper 2000 m are estimated based on five different ocean temperature datasets, as shown in Table 2. We note that spatiotemporal sampling is different in observations (near-instantaneous point measurements) and models (full temporal sampling of box volume averages) and therefore observations are subject to

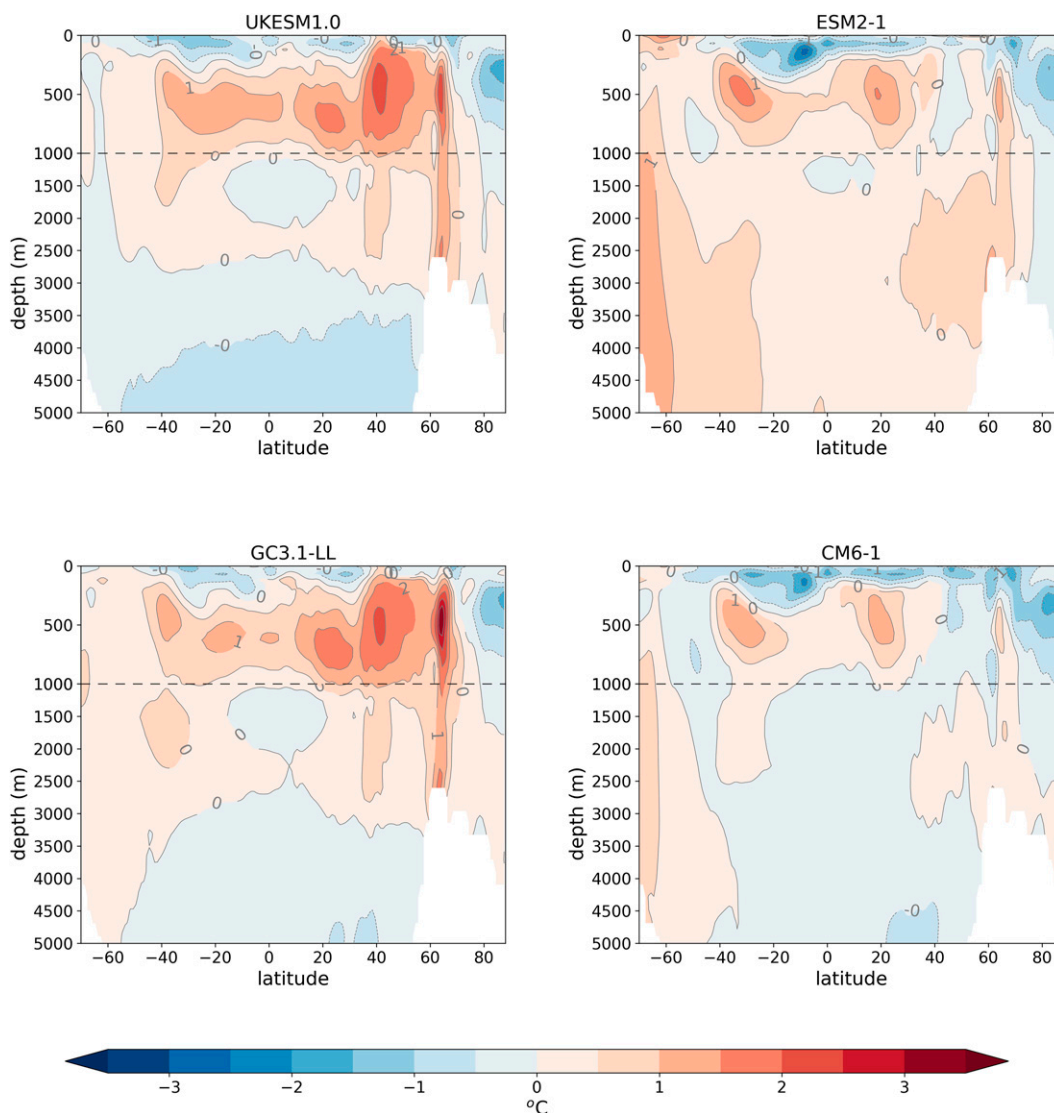


FIG. 2. Zonal mean temperature biases of the four models against WOA18 observations (Locarnini et al. 2018). Shown are average annual means of the first 100 years of the piControl simulations. Simulation data and observations were projected onto the same regular grid prior to the subtraction for calculating the bias.

substantial sampling noise (e.g., Allison et al. 2019). Following Palmer et al. (2021), we construct an ensemble estimate of observed OHC change, relative to a 2005–14 baseline period, for the 0–700 and 700–2000 m layers using all available time series as indicated in Table 2. We apply a 3-yr running mean to all time series for the 0–700 and 700–2000 m layers to reduce sampling noise, and the resultant time series used in our ensemble estimates are identical to those assessed by IPCC AR6 (Gulev et al. 2021). The Palmer et al. (2021) approach combines an estimate of structural uncertainty, based on the standard deviation of the ensemble spread, with the maximum internal/parametric uncertainty (i.e., the uncertainty specified by each dataset originator) across the ensemble to estimate the total uncertainty. Structural and internal uncertainty are assumed to be independent and combined

in quadrature to estimate the total uncertainty. Central estimate time series for each layer are constructed from the ensemble average and have the advantage of reducing sampling noise compared to individual datasets (Gulev et al. 2021). Further information on the 0–700 and 700–2000 m ensemble OHC change estimates is available in the supplementary material (Figs. S1 and S2 in the online supplemental material). For the observational estimates of global OHC below 2000 m we assume a linear increase in OHC of  $1.15 \pm 0.57 \text{ ZJ yr}^{-1}$  from 1992 onward, following Purkey and Johnson (2010), and von Schuckmann et al. (2020).

In much of the following analysis, we use the ocean basins as employed by the NCEI (Boyer et al. 2018; Levitus et al. 2012; Locarnini et al. 2019) and displayed in Fig. 3. NCEI

TABLE 1. Comparison of the NEMO3.6 ORCA1 configurations in the U.K. models and in the CNRM models. A centered entry (e.g.  $1000 \text{ m}^2 \text{ s}^{-1}$  for the isoneutral mixing coefficient) applies to both sets of models identically.

Setting or parameterization	U.K. models	CNRM models
Coupling frequency	3 h	1 h
Time step	45 min	30 min
Horizontal viscosity		$20000 \text{ m}^2 \text{ s}^{-1}$
Vertical variation of horizontal viscosity		Yes
Fox–Kemper scheme for mixed layer eddy parameterization	No	Yes
Parameterization of mesoscale eddies		Held and Larichev (1996)
Isonutral mixing coefficient		$1000 \text{ m}^2 \text{ s}^{-1}$
Vertical mixing scheme	TKE-dependent vertical diffusion	TKE plus internal wave-induced mixing parameterization of de Lavergne, including tidal
Geothermal heat flux climatology		Goutorbe et al. (2011)
Lateral boundary condition for momentum		Free slip
Penetration of TKE below the mixed layer	TKE source below mixed layer (nn_etau=1)	
Latitudinal variation of TKE penetration below the mixed layer	Asymmetrical with regard to the equator (nn_htau = 4)	Symmetrical with regard to the equator (nn_htau = 1)

have partitioned the global ocean in six basins: North Atlantic (NA), north Indian (NI), North Pacific (NP), South Atlantic (SA), south Indian (SI), and South Pacific (SP). The boundary between the northern and southern basins is the equator. Time series of monthly ocean heat content for all NCEI basins were downloaded from the NCEI website.<sup>1</sup> Similar basin time series were provided by L. Cheng (2019, personal communication) for the Cheng et al. (2017) dataset (Table 2). These data products were used to estimate the total uncertainty for basin ocean heat content changes, for the time periods 1971–91 and 1991–2014, following the approach of Palmer et al. (2021). The method for calculating the ensemble mean and the error of the basinwise OHC is the same as for the global time series (see also detail in section 4).

In addition to observational time series of OHC change we also make use of vertical profiles of temperature change based on the NCEI, EN4, and Cheng et al. datasets. While for the NCEI data basinwise averages are provided, spatially complete monthly gridded fields of ocean temperature for EN4 and Cheng were used to generate decadal averages for the periods 1965–74 and 2005–14. These were combined with the NCEI ocean basin masks to provide spatially averaged vertical profiles of ocean temperature for each decade, from which changes could be assessed (see section 5 for more details).

#### d. Drift in the preindustrial control runs

In control runs of climate and Earth system models some variables tend to be nonstationary, a phenomenon referred to as “drift” (Sen Gupta et al. 2013; Hobbs et al. 2016; Irving et al. 2021; Séférian et al. 2016). Specifically, in a preindustrial control simulation (piControl) with near-zero radiative forcing of the Earth system the global OHC should be nearly

constant, but often a trend is found instead. Partly this stems from the finite length of the simulations, and hence ongoing internal adjustments of the oceanic temperature and salinity fields to the surface fluxes and interior mixing processes, but it can also reflect issues with energy conservation in the models (Hobbs et al. 2016; Irving et al. 2021). For these reasons, we briefly discuss the piControl drift in the four models of this study along with the chosen detrending methods, noting that simple forms of drift correction are adequate to remove any issues of nonconservation or ongoing adjustment in climate model simulations (Hobbs et al. 2016).

The drift in globally integrated OHC for the four models is plotted by depth layer in Fig. 4. For the global ocean as a whole (Fig. 4d), UKESM1.0 is the only model that cools, while the other three are warming up at varying degrees. In the abyssal layer (2000 m–bottom; Fig. 4c) the picture is similar to the full global ocean. In the intermediate layer, however (700–2000 m; Fig. 4b), UKESM1.0 does not show a clear trend while in the top 700 m (Fig. 4a) it is cooling again. Conversely, CNRM-ESM2-1 (red lines) warms up in the lower layers, but not in the top 700 m. The marked difference in drift between UKESM1.0 and GC3.1-LL is explained by the different spinup histories of these two models (Menary et al. 2018; Yool et al. 2020). GC3.1-LL was spun up as a coupled model for about 700 years, while UKESM1.0 was spun up in an ocean-only configuration for over 5000 years, followed by 1000 years of a coupled spinup.

In the upper two layers, 0–700 and 700–2000 m, there are clear signs of decadal and centennial variability. The two CNRM models have centennial variability with a phase of about 180 years, more clearly seen in 700–2000 m. UKESM1.0 meanwhile shows a mixed signal of decadal and centennial variability mostly in the 0–700 m layer.

In the CNRM models (amber, red) the drift is positive in the piControl simulations. In the 0–700 and 0–2000 m layers, the

<sup>1</sup> [https://www.ncei.noaa.gov/access/global-ocean-heat-content/basin\\_heat\\_data\\_monthly.html](https://www.ncei.noaa.gov/access/global-ocean-heat-content/basin_heat_data_monthly.html).

TABLE 2. Observations-based datasets used in this OHC study.

Dataset abbreviation	Reference	Notes
EN4	<a href="#">Good et al. (2013)</a> with updates; version 4.2.1 with XBT bias corrections applied ( <a href="#">Gouretski and Reseghetti 2010</a> )	Used in 0–700 m ensemble time series Data available at <a href="https://www.metoffice.gov.uk/hadobs/en4/">https://www.metoffice.gov.uk/hadobs/en4/</a>
NCEI	<a href="#">Levitus et al. (2012)</a> with updates	Used in 0–700 and 700–2000 m ensemble time series Data available at <a href="https://www.ncei.noaa.gov/access/global-ocean-heat-content/">https://www.ncei.noaa.gov/access/global-ocean-heat-content/</a>
ISH	<a href="#">Ishii et al. (2017)</a>	Used in 0–700 and 700–2000 m ensemble time series Data available at <a href="https://climate.mri-jma.go.jp/pub/ocean/ts/v7.3.1/">https://climate.mri-jma.go.jp/pub/ocean/ts/v7.3.1/</a>
CHG	<a href="#">Cheng et al. (2017)</a> with updates	Used in 0–700 and 700–2000 m ensemble time series Data available at <a href="http://www.ocean.iap.ac.cn">http://www.ocean.iap.ac.cn</a>
DOM	<a href="#">Domingues et al. (2008)</a>	Used in 0–700 m ensemble time series

drift is weak during the first 200 years, then it rises abruptly in the CNRM-CM6-1 version and more progressively, in CNRM-ESM2-1. The full-depth OHC drift in the ESM version is  $\sim 600$  ZJ in 500 years, equivalent to  $\sim 0.10 \text{ W m}^{-2}$ , and is larger in the CNRM-CM6-1 version. The drift in the 0–700 m layer is smaller than in UKESM1.0, whereas it is larger than in UKESM1.0 in the 700–2000 m layer. Overall, the differences in global OHC drift between the models indicate significant differences in vertical heat transport.

We note that for all models the trend in global full-depth OHC in the piControl simulations, expressed as the equivalent net TOA radiation imbalance (Table 3), is significantly smaller than the OHC trend in the early twenty-first century (about  $+0.7 \text{ W m}^{-2}$ ; [Johnson et al. 2016](#)).

For the subsequent analysis we subtracted the long-term drift and variability from each historical simulation individually, and for each layer individually, by estimating the trend from the matching section of the piControl. We show in the appendix that varying the degree of the polynomial used for detrending in most cases has a small impact on global OHU

values. For individual basins the effect is not negligible though. We choose quadratic detrending for all four models because of the nonlinear drift, that is, the significant multidecadal and centennial variability in the piControl (Fig. 4). The historical ensemble mean was calculated from these detrended time series.

### 3. Global ocean heat uptake

The temporal evolution of the global OHC in the CMIP6 historical simulations from the four models, together with the ensemble of OHC observations (section 2c), is plotted in Fig. 5. In the 0–700 m layer (Fig. 5b), OHC in the two CNRM ensemble means (yellow and red) is within the observational error while the U.K. models (light blue and dark blue) show virtually no warming between 1970 and 1993 and a too-strong warming afterward. Strikingly, in the 700–2000 m layer all four model ensembles warm substantially too little (CNRM) or basically not at all (U.K.). In the 2000 m–bottom layer the CNRM models display a relatively realistic warming rate while the U.K. models warm too strongly throughout. The result for the total global OHC time series (Fig. 5e) is that the U.K. models and CNRM-CM6-1 are slightly outside the observational uncertainty around 1970 because of a too-small warming rate, whereas CNRM-ESM2-1 displays warming within the observational uncertainty throughout. In the U.K. models we find substantial bias compensation between either the too-small warming (before 1991) or the too-large warming (after 1991) in the 0–700 m layer; the unrealistic cooling in the 700–2000 m layer; and the too-strong warming in the 2000 m–bottom layer. In the CNRM models, the three layers appear much closer to the observational ensemble mean.

For the well-observed period of 2006–16, estimates of full-depth global OHC change show good correspondence with annual variations in the TOA energy budget ([Meyssignac et al. 2019](#)), as expected from climate model simulations ([Palmer and McNeill 2014](#)). The differences between the CNRM models and the U.K. models equally appear in the TOA net radiation time series (Fig. 5a) where the CNRM models have a larger net TOA radiation budget in the 1970s and 1980s while it is smaller than in the U.K. models after 1991.

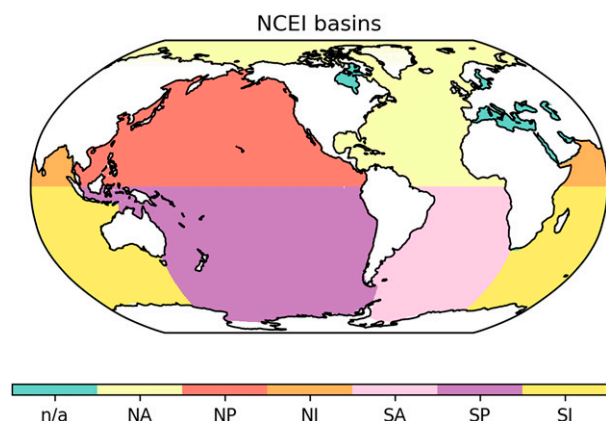


FIG. 3. Partition of the World Ocean into basins following the NCEI dataset. The Arctic is part of the NA basin. Some marginal seas are disregarded (“n/a”), e.g., the Mediterranean and the Hudson Bay.

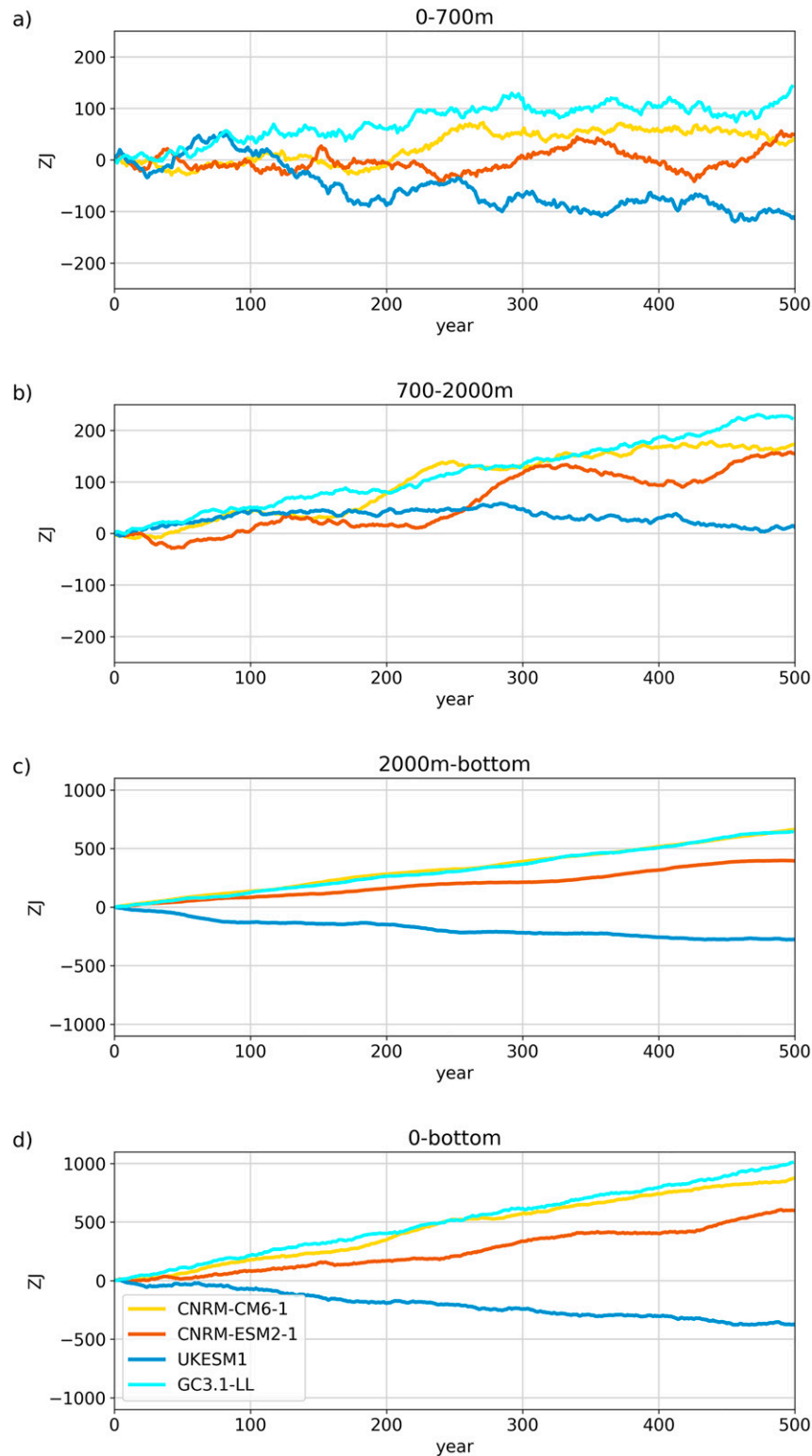


FIG. 4. Time series of global OHC anomaly in the piControl of UKESM1.0, GC3.1, CNRM-CM6-1, and CNRM-ESM2-1 for the layers (a) 0–700 m, (b) 700–2000 m, (c) 2000 m–bottom, and (d) 0 m–bottom.

TABLE 3. Long-term (500 years) global OHC trend (over full depth) in the piControl simulations, expressed as the equivalent net TOA radiation imbalance  $\text{eqR}_{\text{TOA}}$ .

Model	UKESM1.0	HadGEM3-GC3.1	CNRM-ESM2-1	CNRM-CM6-1
$\text{eqR}_{\text{TOA}}$ ( $\text{W m}^{-2}$ )	−0.07	0.17	0.10	0.16

As mentioned above, residual sampling noise can introduce spurious annual-to-decadal variability into observation-based estimates of OHC change (Allison et al. 2019). In addition, processes of internal climate variability will have an imprint on observations that the simulations do not necessarily capture. Therefore, we might expect some mismatch between observations and model simulations on these time scales. However, this cannot account for the discrepancies seen in the simulated multidecadal trends.

In the observational ensemble mean (Figs. 5b,e) and some of the simulations we note a marked loss of OHC from about 1962 to 1969 in the 0–700 m layer and for the global total, before the beginning of the consistent OHC warming trend that lasts throughout the present. It seems likely that the OHC loss in the 1960s is caused by the aerosol radiative forcing becoming more negative in that time (see Fig. 6 further below).

The observed rate of global heat uptake in the 0–700 m layer appears to notably drop off after the year 2004 (Fig. 5b), marking a phase of reduced OHU in the 0–700 m layer in some regions and a shift of heat from the Pacific to the Indian due to internal variability in the climate system (Liu et al. 2016; Gastineau et al. 2019). The simulation ensemble means do not capture this phase of regionally reduced upper-ocean heat uptake since their internal variability is in general not in phase with the real climate system.

We highlight two individual simulations here and will analyze them in more detail in section 5. One simulation from the UKESM1.0 ensemble, labeled r17, shows an OHC time series that more closely matches observations (teal lines in Fig. 5) in the 0–700 m, 2000 m–bottom, and full ocean layers than the other members of the ensemble. It is remarkable that r17 warms more than the UKESM1.0 ensemble mean in the 0–700 m layer (before 1991), but less than the ensemble in 2000 m–bottom. From the CNRM-CM6-1 ensemble, by contrast, r4 (brown lines) stands out for a particular poor performance, with very little OHC before 1991 (like the U.K. models) and a marked cooling in the 700–2000 m layer until the year 2000.

What are the reasons for the varying rate of OHU in the four models in the historical simulations? The different magnitudes of the net TOA radiation (Fig. 5a) indicate a role for the radiative forcing. We investigate the available time series of the total effective radiative forcing (ERF) and its components (Fig. 6) for CNRM-CM6-1, HadGEM-GC3.1-LL and UKESM1.0. For CNRM-CM6-1 and HadGEM-GC3.1-LL, we have used the CMIP6 Radiative Forcing Model Intercomparison Project (RFMIP) experiments piClim-histall, piClim-histghg, piClim-histnat, and piClim-histaer (Pincus et al. 2016). For UKESM1.0-LL and CNRM-ESM2-1, these experiments are not available. However, for UKESM1.0-LL we obtained an approximation of the aerosol ERF from the difference

between the atmosphere-only AerChemMIP experiments histSST-piAer and histSST (Collins et al. 2017).

Figure 6a reveals that the total ERF is smaller in GC3.1 than in CNRM-CM6-1 between 1960 and 1991, but much larger afterward. The next three panels in Fig. 6 explore the processes behind the differences in the total ERF. The greenhouse gas ERF (Fig. 6b) is clearly larger in GC3.1-LL than in CNRM-CM6-1 for most of the time. The negative aerosol ERF (Fig. 6d) is stronger in the U.K. models than in CNRM-CM6-1 between about 1950 and 1990. The natural forcing (Fig. 6c) is more negative in GC3.1-LL than in CNRM-CM6-1 from 1960 to 1993 due to a larger sensitivity to volcanic aerosols. For both models, the natural forcing is adjusted to give a net zero ERF in the time average, taking into account short phases of strongly negative forcing in response to volcano eruptions (Andrews et al. 2019). Because the volcano forcing is more negative in GC3.1-LL, the natural forcing during other times is more positive to compensate (up to about  $0.3 \text{ W m}^{-2}$  before 1960).

To picture the effect of the different total ERF across the models, we applied the total ERF time series to a two-layer energy balance model (EBM; Geoffroy et al. 2013) which we fitted to the full models using an abrupt  $2 \times \text{CO}_2$  experiment for each model respectively. (This experiment was used rather than the abrupt  $4 \times \text{CO}_2$  experiment because it is closest to historical simulations in term of warming magnitude.) The EBM parameters  $\lambda$ ,  $C$ ,  $C_0$ ,  $\varepsilon$ , and  $\gamma$  were determined to be  $0.69 \text{ W m}^{-2} \text{ K}^{-1}$ ,  $9.1 \text{ W yr m}^{-2} \text{ K}^{-1}$ ,  $76 \text{ W yr m}^{-2} \text{ K}^{-1}$ , and 1.31 and  $0.58 \text{ W m}^{-2} \text{ K}^{-1}$  for GC3.1-LL; and  $0.90 \text{ W m}^{-2} \text{ K}^{-1}$ ,  $6.4 \text{ W yr m}^{-2} \text{ K}^{-1}$ ,  $84 \text{ W yr m}^{-2} \text{ K}^{-1}$ , and 1.5 and  $0.81 \text{ W m}^{-2} \text{ K}^{-1}$  for CNRM-CM6-1. We note that EBM parameters are sensitive to the choice of the experiment used for the calibration.

Figure 6e shows the evolution of the deep ocean temperature simulated by the EBM. This deep ocean temperature evolution should picture roughly the ocean heat content evolution. The two solid lines indicate the EBM for CNRM-CM6-1 (yellow) and for GC3.1-LL (cyan) each forced with the total ERF time series from GC3.1-LL. The fact that these two lines are very close throughout the simulated time, and the fact that for the 1960s and 1970s a decrease in deep ocean temperature is simulated, strongly suggests that the total ERF time series from GC3.1-LL leads to an ocean heat loss irrespective of details of the GCM (as represented by the fitted EBM). The dashed lines indicate that, conversely, either EBM fit being forced by the total ERF from CNRM-CM6-1 does not lead to a mid-twentieth-century heat loss. Thus, from Fig. 6e it is clear that the GC3.1-LL ERF leads to a zero or negative OHC trend from 1960 to 1991 and large OHC trend afterward compared to the CNRM-CM6-1 ERF irrespective of the model formulation. In other words, there are strong indications that the absence of warming between 1960 and 1991

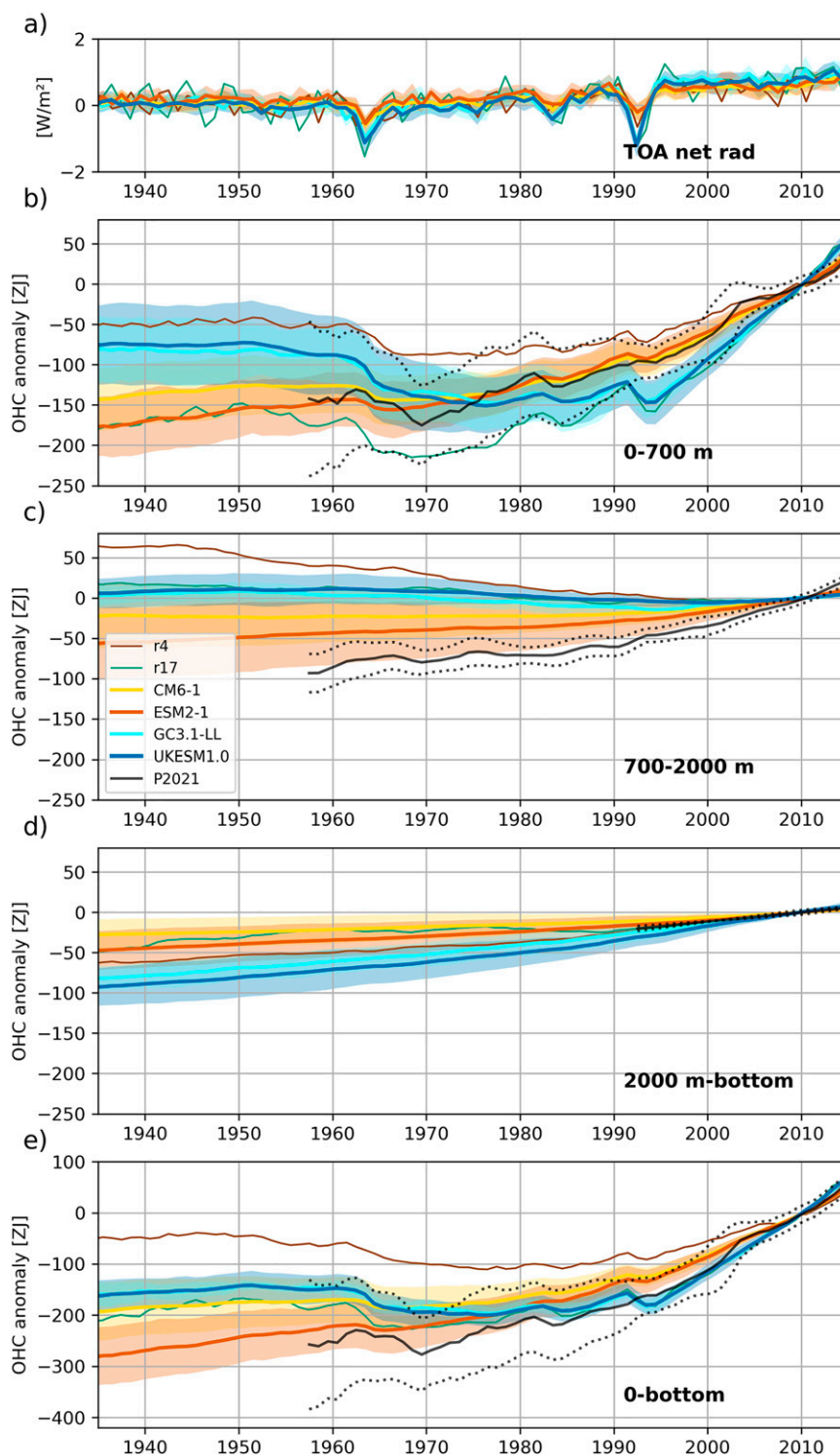


FIG. 5. (a) Net radiation balance and (b)–(e) global OHC for the four models of this study. The historical ensembles (light blue, dark blue, yellow, and red) are compared with an ensemble of the observational datasets available for each layer (black). Thick lines indicate the ensemble means. Shading indicates  $\pm 1$  standard deviation of the individual model ensembles. For the [Palmer et al. \(2021\)](#) observations ensemble, the total error is dotted and the data are available in the supplementary material. Ensemble member r17 from UKESM1.0 is in teal, and ensemble member r4 from CNRM-CM6-1 is in brown. All OHC time series are normalized to the 2005–14 average. The TOA time series are normalized to the 1850–1900 average. Note the different vertical scale in (e).

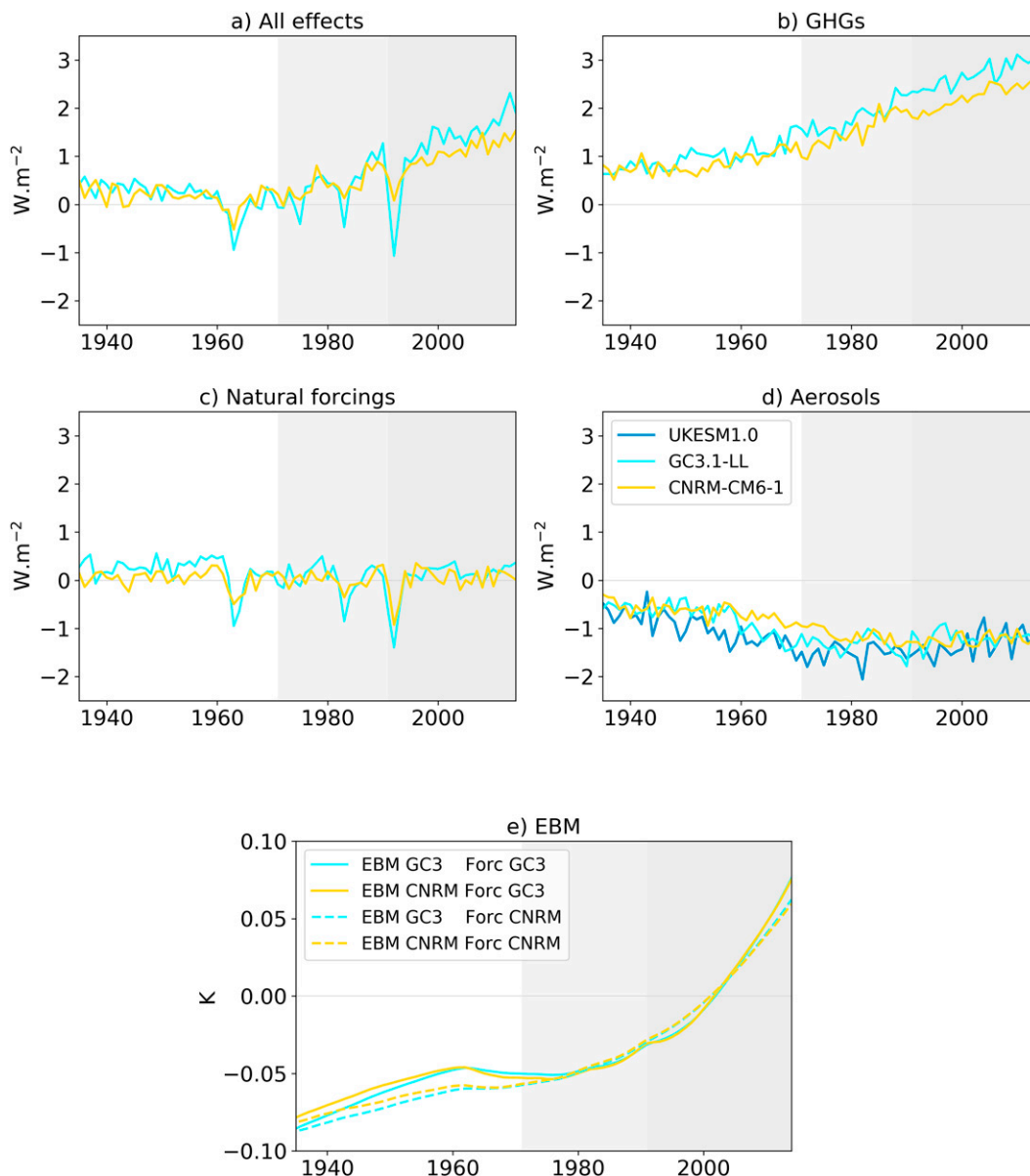


FIG. 6. Time series of (a) the total ERF and its components: (b) GHGs, (c) natural forcings, and (d) aerosols for two of the four models, plus UKESM1.0 in (d). (e) Time series of deep ocean temperature (in K) simulated by a simple two-layer EBM as described in [Geoffroy et al. \(2013\)](#). Temperatures are anomalies to the reference period average 1985–2014. Solid lines indicate the deep ocean temperature simulated by the EBM of each model forced by the GC3.1-LL total ERF, while dashed lines stand for the same EBM forced by the CNRM-CM6-1 total ERF [the total ERF curves are shown in (a)]. The shading indicates the focus periods of the analysis, 1971–91 (light gray) and 1991–2014 (darker gray).

in the U.K. models can be attributed to the strongly negative aerosol ERF, while the rapid warming afterward is a result of the strong positive greenhouse gas (GHG) forcing.

Complementing [Fig. 6](#), relevant components of the aerosol ERF for all four models in the year 2014 are given in [Table 4](#). Between the two CNRM models the aerosol ERF is quite different: strong in CNRM-CM6-1 with  $-1.21 \text{ W m}^{-2}$  (as in [Fig. 6](#)), while weaker (less negative) in CNRM-ESM2-1 with

$-0.82 \text{ W m}^{-2}$ . As seen above, the aerosol ERF is very similar between the two U.K. models. The explanation is that the treatment of aerosols is very similar in UKESM1.0 and GC3.1, using a dynamical aerosol model, while for the CNRM models only CNRM-ESM2-1 has a dynamical aerosol model, with the aerosol treatment being much simpler in CNRM-CM6-1. We also see some variation in the GHG ERF with both ESMs having a somewhat smaller greenhouse gas

TABLE 4. ERF in 2014 for the four models. Values for anthropogenic (ANT), well-mixed greenhouse gas (WMGHG), and aerosol (AER) forcing are from [Andrews et al. \(2019\)](#) for UKESM1.0 and GC3.1 and from [S  f  rian et al. \(2019\)](#) and [Voldoire et al. \(2019\)](#) for CNRM-ESM2 and for CNRM-CM6.

Model	UKESM1.0	HadGEM3-GC3.1	CNRM-ESM2-1	CNRM-CM6-1
$F_{\text{ANT}} (\text{W m}^{-2})$	1.61	1.81	1.59	1.50
$F_{\text{WMGHG}} (\text{W m}^{-2})$	2.89	3.09	2.41	2.64
$F_{\text{AER}} (\text{W m}^{-2})$	−1.13	−1.10	−0.82	−1.21

forcing than their physical climate model counterparts. In CNRM-ESM2-1, both GHG and aerosol ERF are weaker than in its physical core counterpart CNRM-CM6-1, but the aerosols reduction ERF is larger than the GHG reduction, leading to an overall increase in anthropogenic ERF from 1.5 to 1.59  $\text{W m}^{-2}$  in 2014. This increase in ERF in CNRM-ESM2-1 probably explains the larger OHC change simulated globally in this model compared to CNRM-CM6-1.

We note that in all four models, the impact of the major twentieth-century volcano eruptions (Agung in 1963, El Chich  n in 1982, and Pinatubo in 1991) is clearly discernible in the simulated TOA radiation imbalance (Figs. 5a, 6a). In the ensemble of observations-based OHC time series for 0–700 m and the full global ocean (Figs. 5b,e) an impact of the volcano eruptions might be discernible, although it cannot easily be distinguished from interannual and decadal variability. In the simulated 0–700 m OHC time series (Figs. 5b,e) the ensemble means of the U.K. models display a strong response. Especially for the Pinatubo eruption in 1991 the simulated ocean heat loss is too large. This is mirrored in the TOA net radiation time series (Fig. 5a) showing the larger negative radiation balance in the U.K. models compared to the CNRM models. For the year 1991 this difference exceeds 1  $\text{W m}^{-2}$ . Figure 6 suggests that the effect of eruptions on the U.K. ERF forcing is larger than in the CNRM ERF and this has a clear impact on the EBM deep ocean temperature, whichever model is considered. [Gregory et al. \(2020\)](#) discuss the response of AOGCMs to volcanic forcing in greater detail, suggesting that, in comparison to observations, models tend to overestimate the response in the first few years after an eruption but underestimate it afterward.

We have shown above that the different ERFs partly explain the different OHU characteristics in the U.K. models contrasted with the CNRM models, but these differences might also be related to the transient climate response (TCR) and the effective climate sensitivity (EffCS) of these models. Table 5 gives an overview of these quantities across the four models and compares it to the OHU between 1991 and 2014 in the historical simulations. For the actual values of TCR and

EffCS we referred to [Andrews et al. \(2019\)](#) for UKESM1.0 and GC3.1, and [S  f  rian et al. \(2019\)](#) and [Voldoire et al. \(2019\)](#) for CNRM-ESM2 and for CNRM-CM6. Compared to the CNRM models, the U.K. models have a larger 0–700 m ocean heat uptake in 1991–2014, a larger EffCS and a larger TCR. This means that even if forced with the same ERF, the U.K. models would warm more quickly than the CNRM models. This probably reinforces the difference between the two model families (Table 4). We note that [Kuhlbrodt and Gregory \(2012\)](#) found an average TCR of 1.83 for the CMIP5 models, while for the CMIP6 models the average is 2.0 ([Meehl et al. 2020](#)). The U.K. models discussed here have got a larger TCR than these averages.

#### 4. Regional ocean heat content change

An analysis of the OHC change in the 0–700 and 700–2000 m layers by ocean basin and time period is shown in Fig. 7. We chose periods before and after the Pinatubo eruption because of the substantial change in OHU rate in the U.K. models. OHC change over these periods was calculated as the difference in OHC between the last and the first annual mean of the period. For the models, we calculated this difference for each ensemble member individually and then plotted the mean value, using one standard deviation as the structural error estimate (i.e., a measure of ensemble spread).

The model ensembles are compared with an observational ensemble of two datasets, NCEI and CHG (cf. Table 2). Not all OHC observational datasets are available for individual basins, and below 2000 m no basinwise estimates of OHC change exist. As an estimate of the standard error of the individual datasets, we used the “se” variable of the annual means in the case of NCEI and the provided error variable of CHG. The CHG data came in monthly means. For the error of the annual means, we assumed perfectly correlated error across the monthly means and used the annual average. When taking the difference of two annual means for the bar height we used

TABLE 5. TCR, EffCS, and OHU between 1991 and 2014 of the four models covered in this study. EffCS was calculated using all 150 years of an abrupt  $4 \times \text{CO}_2$  experiment using the regression method ([Gregory et al. 2004](#)).

Model	UKESM1.0	HadGEM3-GC3.1	CNRM-ESM2-1	CNRM-CM6-1	Observational ensemble mean
TCR (K)	2.76	2.48	1.92	2.0	—
EffCS (K)	5.36	5.54	4.84	4.95	—
OHU 0–700 (ZJ)	172	171	119	125	128
OHU 700–2000 (ZJ)	7	15	36	26	77

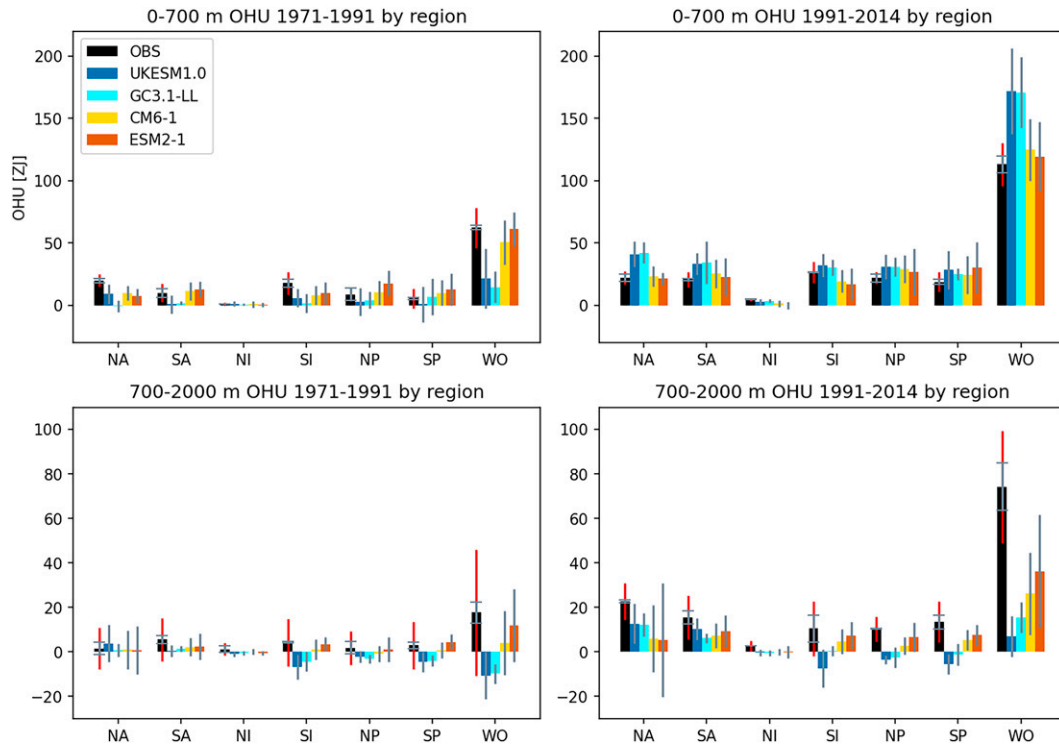


FIG. 7. OHU (left) from 1971 to 1991 and (right) from 1991 to 2014 in the (top) 0–700 and (bottom) 700–2000 m layers by NCEI ocean basin (with WO for global ocean). We use an ensemble of two observational datasets (black): the pentadal NCEI dataset and the CHG dataset. For the observations we plotted the structural error (gray) and the total error (red). Note the different scales on the y axes in the top and bottom rows. The four models are UKESM1.0 in dark blue, GC3.1-LL in light blue, CNRM-CM6-1 in yellow, and CNRM-ESM2-1 in red. The column represents the ensemble mean and the gray bar represents one standard deviation of the ensemble. For WO, the columns represent the entire global ocean as simulated in models, including those marginal seas that are disregarded in the NCEI basin definitions.

the larger of the two individual standard errors as the standard error of this bar.

For the standard error  $\sigma_{\text{std}}$  of the observational ensemble we used again the larger of the two individual standard errors. An estimate of the structural error  $\sigma_{\text{struct}}$  of the ensemble is obtained from the standard deviation of the ensemble. Last, the total error  $\sigma_{\text{tot}}$  is calculated as  $\sigma_{\text{tot}} = \sqrt{\sigma_{\text{std}}^2 + \sigma_{\text{struct}}^2}$ . Figure 7 shows the structural errors in gray (all ensembles) and the total error in red (for the observations).

In the global total of the 0–700 m layer the CNRM models (yellow, red) are within the observational error (black, with red line) for both the pre-1991 and the post-1991 periods. The U.K. models though (light and dark blue) are strongly underestimating the pre-1991 OHC change, and markedly overestimating the post-1991 OHC change. It is clear that the underestimate of pre-1991 OHC change in the U.K. models happens in all six basins (Fig. 7, top left) while the overestimate post-1991 mainly appears to stem from the two Atlantic basins (NA, SA), with smaller contributions from the Pacific basins (NP, SP).

The CNRM models appear to overestimate OHU in the two Pacific basins before 1991 in the 0–700 m layer while OHU in the North Atlantic tends to be too small. After 1991,

this slight overestimation in the Pacific is compensated in the south Indian Ocean. Note that we leave the north Indian Basin out of the discussion here because of the very small contribution to global OHU.

In the 700–2000 m layer the observations suggest some OHU both before and after 1991 (cf. Fig. 5c), although this is statistically significant only in the period after 1991 (Fig. 7, lower row). Before 1991, the U.K. models simulate heat loss, occurring mostly in the Pacific and the south Indian. CNRM-ESM2-1 simulates heat gain in the Southern Hemisphere basins, and CNRM-CM6-1 is not distinguishable from zero. Given the large observational uncertainty however, none of the four models can be labeled as unrealistic. After 1991 there is a clear signal of OHU in the observations, for which the strongest contributions come from the two Atlantic basins (NA, SA). The CNRM models capture about half of the observed total, distributed across all basins, but particularly underestimating the Atlantic basins. The U.K. models capture only very little of the observed OHU in this layer and period; they stand out for continued heat loss in the Pacific and south Indian Basins. Nevertheless, they simulate the dominating role of the Atlantic basins for global OHU in the 700–2000 m layer in the period after 1991.

In summary, on a regional level the CNRM models are markedly different from UKESM1.0. In the period 1971–91 in 0–700 m and in 1991–2014 in the 700–2000 m layer the CNRM models tend to show too much warming in the Pacific basins, and too little warming in the Atlantic basins. The U.K. models, meanwhile, have too-small OHU in the Pacific and the south Indian (except in the 0–700 m layer after 1991), but they capture to some degree the dominant role of the two Atlantic basins for global OHU after 1991.

These differences between the CNRM models and the U.K. models could point to a different role for the Atlantic meridional overturning circulation (AMOC). In the ensemble mean and during the historical simulations, the AMOC in the CNRM models does not increase significantly (Voldoire et al. 2019), while in the historical simulations of the U.K. models it increases by up to 3 Sv ( $1 \text{ Sv} \equiv 10^6 \text{ m}^3 \text{ s}^{-1}$ ) around the year 2000 (Andrews et al. 2020; Yool et al. 2020). At the same time, the AMOC in the piControl is stronger in the CNRM models (between 16 and 20 Sv) than in the U.K. models (around 15 Sv). We will revisit this question in section 6.

### 5. Depth–time structure of ocean temperature trends

The previous sections have shown that the time evolution of OHU significantly differs across layers and basins. We explore the structure of the OHU in depth and time in more detail by calculating the trend of temperature change in each basin and on each level separately.

Observational estimates of these trends are obtained from three datasets—World Ocean Atlas 2018 (WOA18), ISH (Ishii et al. 2017), and CHG—as the difference between two decadal means (1995–2004 and 1965–74). The scarcity of ocean temperature observations below 2000 m and before the Argo period does not allow for a higher time resolution measure like statistical trend estimates from time series of annual means. For the same reason we did not calculate error estimates: these are only available at locations and times where in situ measurements were taken. It is beyond the scope of this paper to extend such error estimates to basinwide averages. Instead, the spread of the three different datasets provides a rough estimate of the structural error. The most recent “decadal” mean available from WOA18 covers the 13 years from 2005 to 2017 inclusive. Because the CMIP6 historical model simulations finish with the year 2014 we could not use this last WOA18 decade for comparison.

Figure 8 shows the temperature change over the last 30 years of the twentieth century in six ocean basins and the global ocean (“WO”) at every depth level for the four models and the three observations-based datasets. The trend ( $^{\circ}\text{C decade}^{-1}$ ) was calculated from the difference between two decadal averages, 1995–2004 minus 1965–74. As in the previous plots we have applied quadratic detrending of the individual simulations, here at each level and each basin individually. The results are robust against the detrending method (subtraction of a linear, quadratic, or fourth-order trend; not shown).

For the global ocean we find that between the surface and 200 m the alignment between the ensemble-mean simulated trends and those from observations looks satisfactory, confirming

the models’ ability to reproduce the observed trends at the sea surface and in the mixed layer for the global mean. However, between about 300 m and about 1300 m depth, the U.K. models cool down while the CNRM models, like the observations, warm up. Note that this is a simulated middepth cooling in the late twentieth century; the models’ drift has been removed by the detrending and is much smaller in magnitude. Below about 1800 m we find the opposite effect: the model oceans warm up while the observations suggest a slight cooling.

The temperature trends in the individual basins reveal more details about the discrepancies between the simulations and the observations. The NA basin features the strongest discrepancies. Below 1500 m the simulated NA trends are positive at all levels, albeit with a large variability within the ensembles. The warming is particularly strong for the U.K. models between 2000 and 3500 m, where the observations indicate a negative temperature trend that is up to a factor of 3 smaller in magnitude than the positive trends in the U.K. models. Water masses in the depth range 2000–3500 m are dominated by the deep southward-flowing branch of the AMOC. Hence the observations indicate a cooling of the southward AMOC branch below 1500 m while it substantially warms up in the models.

Above 1500 m in the NA basin, the picture is quite different. Here the simulations, in the ensemble means, hardly warm at all below 300 m, again with a large ensemble spread, while the observations consistently show a warming trend that is about  $0.02^{\circ}\text{C decade}^{-1}$ . This means that in the simulations the upper, northward-flowing branch of the AMOC is not warming in most ensemble members. This could mean that the additional heat in the simulations in the deeper layers of the NA basin has not been advected by the AMOC from other basins, but rather is the result of reduced surface heat loss within the NA basin. (We remind the reader that this basin comprises the Atlantic north of the equator and the entire Arctic Ocean.) The spread of the ensemble members is large in the NA basin, especially between 700 and 4000 m, which indicates significant internal decadal variability of North Atlantic deep-water formation among the ensemble members.

The SA basin is characterized by a warming of the deep ocean below 2000 m in the simulations, opposing the observed small cooling. Again, the positive trend is larger in the U.K. models than in the CNRM models. Between 300 and 1000 m, most U.K. simulations, along with the ensemble mean, simulate a cooling where the observations suggest a trend close to zero. This simulated subsurface cooling is in line with the global average and the other basins (except for NA), while the observations indicate a subsurface cooling in the SP only. The variance of the UKESM1.0 ensemble is still large in the deep SA, but smaller than in the NA, consistent with an assumed source of the variance in the NA basin. (Note that the ensemble size of GC3.1-LL is much smaller than UKESM1.0, hence the variances of these two ensembles do not compare well.) In the same depth range 300–1000 m, the CNRM models simulate a significant warming trend that indicates strong OHU in the subsurface SA, which is not in line with two of the three observational products. In the top 200 m of the SA,

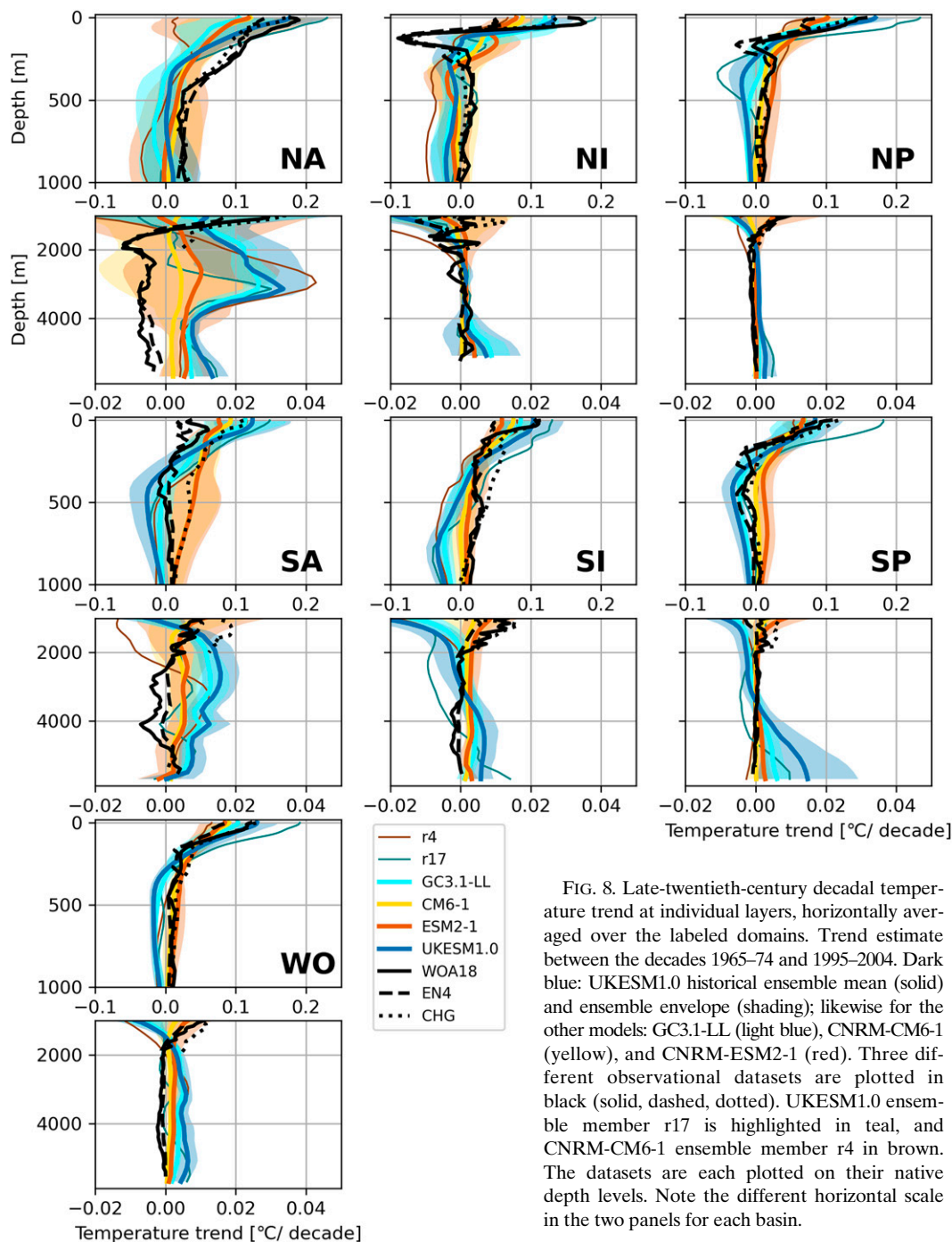


FIG. 8. Late-twentieth-century decadal temperature trend at individual layers, horizontally averaged over the labeled domains. Trend estimate between the decades 1965–74 and 1995–2004. Dark blue: UKESM1.0 historical ensemble mean (solid) and ensemble envelope (shading); likewise for the other models: GC3.1-LL (light blue), CNRM-CM6-1 (yellow), and CNRM-ESM2-1 (red). Three different observational datasets are plotted in black (solid, dashed, dotted). UKESM1.0 ensemble member r17 is highlighted in teal, and CNRM-CM6-1 ensemble member r4 in brown. The datasets are each plotted on their native depth levels. Note the different horizontal scale in the two panels for each basin.

we find too-strong warming in all models, and more so for the U.K. models.

Note that the CHG dataset (dotted) significantly deviates from the other two observations-based datasets in the SA basin (0–2000 m) and the SI basin (100–400 m). The CHG data relax to the CMIP model mean in absence of observational coverage (Cheng et al. 2019). Hence, the often sparse in situ data coverage in the Southern Ocean could explain this discrepancy.

The pattern of subsurface cooling in the U.K. models and subsurface warming in the CNRM models is not only found in the SA but also in the SP. Here the U.K. models appear to be somewhat closer to an observed slight cooling. Furthermore, a similar discrepancy between the two model families is visible in the SI and NP basins, with the CNRM models being closer to observations in these cases. Overall, the CNRM models show a preference to store

heat in the subsurface layers, while the U.K. models store heat below 2000 m.

A further outstanding feature in the UKESM1.0 simulations is the large warming trend and the large variance in Antarctic Bottom Water (AABW). This is particularly strong in the SP (below 4000 m) where the observations suggest a zero trend, as reproduced by the CNRM models, but the UKESM1.0 ensemble mean has trends of up to  $0.015^{\circ}\text{C decade}^{-1}$ , with the largest values in the deepest layers. We suggest that this strong trend and the large variance stem from interdecadal or centennial variability in AABW formation and the associated OHU. We have not related our analysis of the pre-industrial cold bias in the abyssal SP, followed by a historical warming trend, to the results of [Bourgeois et al. \(2022\)](#) who link present-day weak stratification with larger OHU in twenty-first-century simulations. The large heat uptake in the abyss is unrealistic as can be seen from the comparison with observations here in [Fig. 8](#), and indeed as well in [Fig. 5d](#).

Regarding abyssal warming we find a similar picture in the SI basin, with cooling below 3000 m suggested by observations while the simulations show a clear warming signal, and a substantial variance across the ensemble. In the NI basin the warming signal is less strong, but there are still signs of a large variance across the temperature trends in the simulated AABW, below about 4000 m. The fact that the AABW variance is largest in the SP basins suggests that the source of the variance—deep-water formation processes—is in this basin. The CNRM models do not show signs of strong trends or large variance in the abyss.

The r17 simulation of the UKESM1.0 ensemble (marked in teal in [Fig. 8](#)), with its nearly realistic global OHU in the layers 0–700 m, 2000 m–bottom, and 0 m–bottom ([Fig. 4](#)), shows a particularly large warming trend in the top 200 m of the North and South Pacific basins that is too strong compared to the observations. At the same time, the deep warming trends below 1500 m in the NA, SI and SP basins stand out for being clearly smaller than most ensemble members. In other words, in this simulation there is unusually (for this model) strong heat uptake in the top 200 or 300 m, and particularly weak OHU at depths below 2000 m (NA, SI) and in the abyss (SP). Hence, in line with what we said above about the unrealistic downward heat transports in UKESM1.0, this one ensemble member is more realistic because of the much-reduced vertical heat transport during the period analyzed here (1965–2004 overall). The actual vertical temperature profile of r17 in 1965–74 (not shown) indicates colder than (ensemble) average waters in the upper NA and SP, and hence globally, but warmer conditions in the abyssal SP. It appears that subsequently the cold surface layers readily take up heat, while the relatively warm abyssal waters do not take up more heat.

The r4 simulation of the CNRM-CM6-1 ensemble (marked brown in [Fig. 8](#)) is an outlier in displaying particularly small OHU throughout the twentieth century in the 0–700 m layer and the entire World Ocean ([Figs. 5b,e](#)) as well as a marked heat loss in the 700–2000 m layer ([Fig. 5c](#)). In line with this behavior, we see in [Fig. 9](#) that r4 has a very small warming trend in the upper NA and the upper NP, leading to small surface warming globally too. In the intermediate depths (500–1500 m) there is cooling in the NA, SA, and NI basins, again visible

globally too. Contrastingly, in the deep NA (2000–4000 m) there is a very strong warming signal. Given the marked internal decadal AMOC variability in the CNRM models, the (unrealistic) downward redistribution of warming in r4 might well be caused by a particularly strong phase of the AMOC in the NA basins (not shown).

In summary, in [Fig. 8](#) we see evidence for the CNRM models preferentially warming up (storing heat) at intermediate depths (300–1000 m) in the South Atlantic and South Pacific basins and below 2000 m in both Atlantic basins. In all cases, the observations do not suggest a warming trend in these layers and regions. The U.K. models, by contrast, store most of the heat in the North Atlantic below 2000 m, and to a lesser extent in the South Atlantic at the same depth range. In addition, UKESM1.0 warms up in the abyssal South Pacific. There could be various reasons for these differences between the CNRM models and the U.K. models. Even though the ocean components of the models are nearly the same, differences in mixing parameterizations ([Table 1](#)) might contribute to different preferences for storing heat. Different forcing from the atmosphere models (for instance, wind stress over the Southern Ocean or buoyancy fluxes over the North Atlantic) will lead to different circulation states in the models, as illustrated by their different drift behavior in the control runs ([Fig. 4](#)). Indeed, the zonal mean temperature trends in the control runs ([Fig. 2](#)) already indicate a preference for the U.K. models to warm up in the Northern Hemisphere, especially north of  $40^{\circ}\text{N}$ , while the CNRM models show a warming trend in the 300–1000 m subtropical regions along with the high southern latitudes. Finally, differences in radiative forcing ([Fig. 6](#)) might well lead to different responses of the ocean circulation. For instance, it is likely that the AMOC increase during the twentieth century in the U.K. models ([Yool et al. 2021](#)) is a result of Northern Hemisphere cooling from the strong aerosol forcing ([Menary et al. 2020](#)).

We find that, for the global ocean, the depth structure analyzed here for four models, is quite similar to the percentile-based analysis of the depth structure of the historical warming structure ([Sohail et al. 2021](#), their Fig. 3) in the CMIP6 mean. Intriguingly their percentile-based analysis shows a warming trend in the coldest 5% that is not reflected in the depth-based analysis (their Figs. 3e,f). It is tempting to speculate that the abyssal warming we found for UKESM in the South Pacific basin, but not in the global average, is the reason for the coldest 5% warming up in the global average.

Finally, in discussing the vertical profiles of temperature trends we have focused on the three decades from the period of 1965–74 to the period of 1995–2004 to capture the general model behavior in the late twentieth century, during both the lower warming rates before 1990 and the faster warming more recently. An analysis of a different time period would obviously highlight different features, for instance, the near-zero OHU in the U.K. models in the 1970s and 1980s.

## 6. Processes of internal variability

In both CNRM-CM1 and UKESM1.0, there is a rather large spread in global OHU between members. We have even

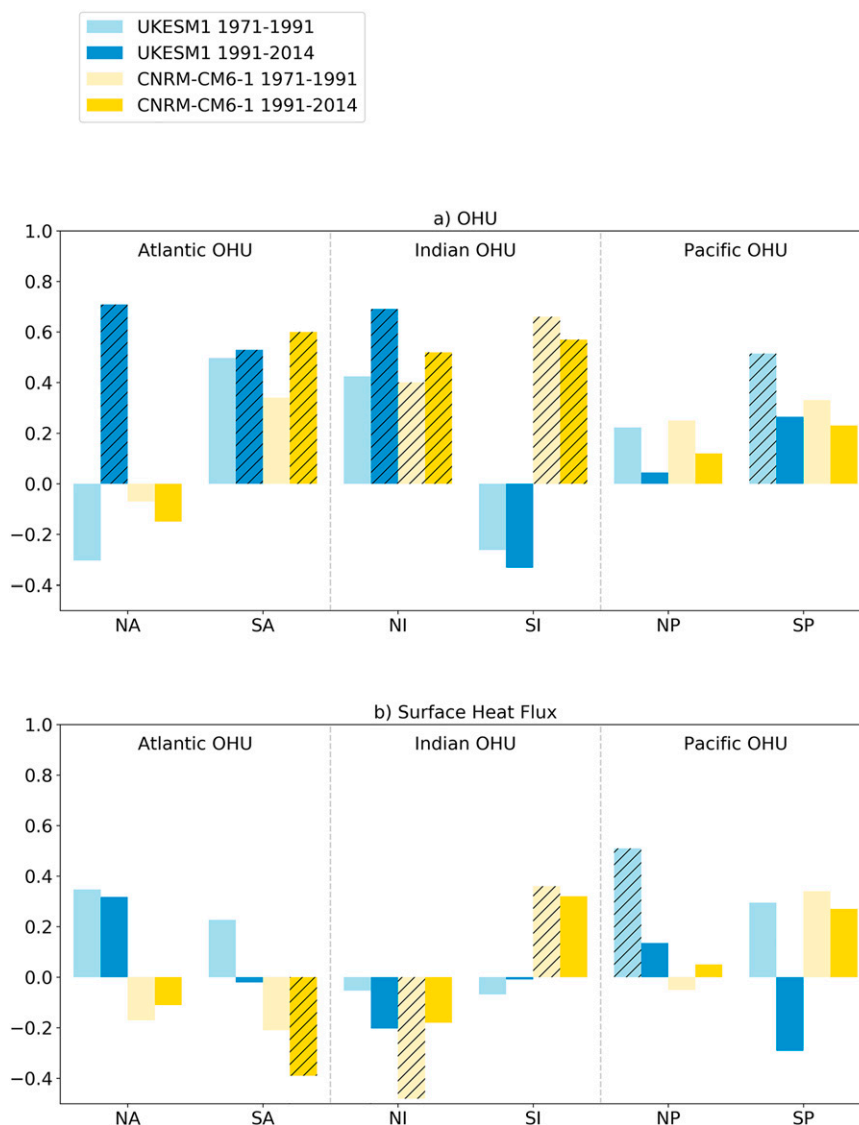


FIG. 9. Ensemble spread correlation between (a) regional OHU (0 m–bottom) and (b) global OHU (0 m–bottom) regionally integrated surface net heat flux (positive when the net flux is downward). Blue bars are for UKESM1.0 and yellow bars are for CNRM-CM6-1; the light colors are for the period 1971–91, and darker colors are for 1991–2014. The hatching indicates significant correlation coefficients ( $p < 0.05$ ).

highlighted that in each model, there are outlier members: in UKESM1.0, one member stands out as more realistic while in CNRM-CM6-1, one member appears as particularly unrealistic relative to the observed trends. To further explore the causes for the differences between members in UKESM1.0 and in CNRM-CM6-1, we analyze in Fig. 9 correlations across the ensemble members between regional OHU spread (0 m–bottom) and 1) global OHU spread (0 m–bottom) and 2) regional net surface heat fluxes spread. This analysis is only done on those two models for which we have a large ensemble size which ensures robustness in the correlation analysis. The objective is to show that processes of internal variability depend on the time period particularly for UKESM1.0, and that the

North Atlantic plays a role in the UKESM1.0 internal variability but not in the CNRM-CM6-1 OHC variability.

We first calculated the correlation coefficient across the UKESM1.0 historical ensemble members and the CNRM-CM6-1 ensemble of global OHU against basinwise ocean heat uptake, over the fixed time periods 1971–91 and 1991–2014. We consider a correlation coefficient significant when the  $p$  value  $< 0.05$  (Fig. 9a). In the following discussion, we have not considered correlations involving the NI basins because the absolute values (in ZJ) are much smaller than in all other basins (cf. Fig. 7).

For UKESM1.0, over the first period, the global OHU spread is mainly correlated with the South Pacific OHU. The

most striking feature is that during the second period the correlation with South Pacific is weaker and nonsignificant, whereas it becomes highly correlated with both South and North Atlantic OHU spread. As for the model mean, this reflects a regime shift in 1991: heat storage in the North and South Atlantic is relevant for global OHU only after 1991, when the AMOC is stronger by about 3 Sv (Yool et al. 2020). For CNRM-CM6-1, there is a significant correlation with the Indian Basin spread over both periods. There is also a strong positive correlation between the global OHU spread and the South Atlantic OHU spread, although significant only after 1991. As for the mean behavior, there is no major regime shift in 1991, no significant role for the North Atlantic and no significant change in the AMOC (Voldoire et al. 2019). This shows that CNRM-CM6-1 internal OHU variability is mainly driven by OHU variability in the Southern Ocean regions. A deeper analysis has revealed that it is mainly related to tropical OHU rather than to high-latitude OHU (not shown).

We have refrained from performing a heat budget analysis given that the net ocean heat budget is the result of a subtle equilibrium between very large quantities and that the resulting budget uncertainty is of the same order of magnitude as the ocean heat budget in itself. In addition, like in most CMIP6 models the energy budgets of the four models analyzed here are not fully closed (Irving et al. 2021). However, we have correlated the regional integrated net surface heat flux term with the regional OHU among members to highlight the regions where the spread in OHU is consistent with the surface heat flux spread (Fig. 9b). Here the integrated net surface heat flux term, converted into ZJ, has been detrended using the same method as for OHU to ensure consistency. A positive correlation means that members with high OHU are also members that get more heat locally at the surface. Thus, the local heat uptake spread could be explained by the local surface heat flux budget. Conversely, a negative correlation means that the OHU spread is probably due to differences in ocean heat transport among members. There are only a few regional correlations in Fig. 9b that are significant. A striking feature is the negative correlation in the South Atlantic for CNRM-CM6-1 after 1991. As the local OHU is well correlated with global OHU, this means that members with the highest OHU are also members with the lowest net surface heat flux in the South Atlantic. In other words, the heat must get transported there from elsewhere. Since the Indian Ocean OHU is positively correlated with the global OHU, we conclude that part of the heat taken up in the Indian Ocean is stored in the South Atlantic.

In UKESM1.0 we find a strong correlation between local OHU in the Atlantic and global OHU, after 1991 (Fig. 9a). This shows us that most of the global OHU after 1991 in UKESM1.0 is stored in the Atlantic basins. At the same time, there is no significant correlation between any regional OHU and regional surface heat flux (Fig. 9b). In other words, there is no preferred ocean basin across the ensemble where the heat enters the ocean. Hence, some of the heat stored in the Atlantic basin must get advected from elsewhere, in contrast to our conjecture about the absence of a warming trend in the upper North Atlantic (previous section, Fig. 8). A more

detailed analysis of the AMOC response and its correlation with the transport and uptake of heat is warranted here. In the period before 1991, we see some positive correlations in the Pacific basin (as said above). Since this is a period of near-zero global OHU (Fig. 5e), we could attribute these correlations to large-scale climate variability in the Pacific region.

In summary, as for the mean behavior, the processes of internal variability differ between the two models. In the UKESM1.0 ensemble, the dominating region for OHU is the Pacific in 1971–91. This switches to both Atlantic basins in 1991–2014, very likely driven by a stronger AMOC. In the CNRM-CM6-1 ensemble, the south Indian Ocean dominates global OHU, and the heat storage extends into the South Atlantic after 1991. There is no significant role for the North Atlantic.

## 7. Conclusions

Global ocean heat uptake (OHU) is an essential metric of Earth's energy imbalance. Basinwise OHU is characterized by the dominant features of global circulation patterns. Both global and basinwise OHU have been reasonably well observed in the late-twentieth and early-twenty-first centuries. In this study we use these observations for a detailed assessment of the simulated historical global and regional OHU in four CMIP6 models, two from the U.K. and two from CNRM. All four models share the same ocean component, yet temporal and spatial features are markedly different across the models.

In simulating the historical global OHU from 1960 to 2014, the CNRM models tend to track the observational record more closely than the U.K. models. In the 0–700 m layer both in 1971–91 and in 1991–2014 the CNRM models are very close to the observations-based data, while the U.K. models simulate nearly zero OHU before 1991 and too strong OHU after 1991. In the 700–2000 m layer the observations show a clear warming trend that none of the models capture adequately. Below 2000 m, the CNRM models capture the warming trend while the U.K. models overestimate it.

Because the differences in performance between the CNRM and the U.K. models in the 0–700 m layer are strongly time dependent (i.e., before and after 1991) we show that the time-dependent historical radiative forcing plays a dominant role. Specifically, compared to the U.K. models, CNRM-CM6-1 has got a weaker (less negative) aerosol forcing before 1980, and a weaker (less positive) greenhouse gas forcing and transient climate response, which matters most after 1991 when the CO<sub>2</sub> concentration increase accelerates. In a simple EBM experiment we show that these properties have a strong imprint on global OHU time series irrespective of whether the EBM is calibrated to HadGEM3-GC3.1-LL or CNRM-CM6-1. We see this as an indication that the aerosol forcing in 1971–91 is too strong (too negative) in the U.K. models.

The full-depth global OHU after 1991 is only slightly overestimated by the U.K. models. In a detailed analysis of the depth structure of the warming we find a compensation between simulated cooling in the depth range 300–1300 m by

warming both above 300 m and, most strongly for UKESM1.0, below 2000 m. This is equivalent to a downward heat transport throughout the water column which we show to be unrealistic. CNRM-ESM2-1 global full-depth OHU is with the observational uncertainty almost at all times. In the 1990s, both CNRM models appear to slightly underestimate the rate of global full-depth OHU.

The analysis of OHU in the individual regional basins reveals that, during the period 1971–91, when in UKESM1.0 the aerosol radiative forcing leads to a near-zero change in global OHC, the variability of global OHU is dominated by the abyssal South Pacific. In the period 1991–2014, where the GHG radiative forcing leads to a rapid warming, it is the two Atlantic basins that dominate the global OHC increase; in the North Atlantic this happens through strong heat accumulation below 2000 m. However, the amount of both the South Pacific and Atlantic heat storage are not in accordance with observations. Conversely, the U.K. models show cooling trends between 200 and 1000 m depth in several basins that are not supported by observations either.

The OHU mechanisms are different in the CNRM ensembles, with only a minor behavior change before and after 1991. In both periods, the global OHU is governed by OHU between 200 and 1000 m in the Southern Ocean, with the South Atlantic gaining relevance after 1991. There is also a warming trend in the deep Atlantic below 2000 m. This is smaller than in the U.K. models, but still not in agreement with observations. In the 700–2000 m layer after 1991 there is a clear warming signal in observations, mostly in the Atlantic basins. The CNRM models simulate that for the layer as a whole, while the U.K. models pick up the regional focus on the Atlantic.

In this paper we have focused on the analysis of late twentieth- and twenty-first-century OHU globally, for individual basins and in detailed resolution of the depth structure. Apart from showing the clear role of the effective radiative forcing for the time-dependent OHU, we have not gone into an in-depth analysis of the reasons for the striking differences between the models. Further work will explore the role of the ocean circulation and the stratification in these differences in more detail, taking into account the different atmospheric forcing, different spinup histories, and large-scale centennial modes of internal variability.

**Acknowledgments.** T. K. was funded by the National Environmental Research Council (NERC) national capability grant for the U.K. Earth System Modelling project, Grant NE/N017951/1. T. K. and A. V. were also supported by the EU Horizon 2020 Research Programme CRESCENDO project, Grant Agreement 641816. A. V. was funded by the EU Horizon 2020 CONSTRAIN project under Grant Agreement 820829. R. K. and M. P. were supported by the Met Office Hadley Centre Climate Programme funded by the BEIS and Defra. We thank Lijing Cheng for providing us with updated

and extended versions of his datasets. We are grateful to Catia Domingues for preprocessing of global OHC time series.

**Data availability statement.** The U.K. simulation data are archived at the Met Office and are available for research purposes through the JASMIN platform ([www.jasmin.ac.uk](http://www.jasmin.ac.uk)) maintained by the Centre for Environmental Data Analysis (CEDA). All the simulations used in this study (CNRM models and U.K. models) are available on the Earth System Grid Federation (ESGF, <https://esgf-node.llnl.gov/search/cmip6/>).

## APPENDIX

### Detrending

To assess the influence of detrending and the degree of the polynomial used for it, we have calculated the ocean heat uptake data in Fig. 7 (top-right panel), with linear, quadratic, and fourth-order detrending, and with no detrending at all, in Tables A1–A4. No detrending means that the historical simulations were used directly, without reference to the piControl.

In the U.K. models, for the ensemble-mean global OHU, the differences from the degree of the polynomial used in detrending are smaller than 3%, and in any case much smaller than the ensemble-mean standard deviation. If no detrending is used, the difference is still smaller than 3% for UKESM1.0, yet can be up to 6% in GC3.1-LL, directly reflecting the much larger drift in this model. For individual basins, the differences from the degree of the polynomial used in detrending can be up to 9% and more, for example, for the North and South Atlantic (NA, SA) basins in Table A2. We note here that for a very small number of individual UKESM1.0 simulations the output of one year of the simulation was unavailable on disk at the time of conducting this analysis. In these cases, we linearly interpolated in time the missing value.

For the CNRM models the impact of detrending is larger because of the strong centennial internal variability in the piControl simulations (see Fig. 4). For the global ensemble-mean OHU, the method of detrending can lead to variations of up to 9% (Table A3), and this variation can be much bigger in individual basins (e.g., NA and SP).

We have detrended each of the seven regions individually, that is, with reference to the trends in the individual basins. This gives more consistent results compared to using the global piControl trend for detrending basins. In the latter case, inconsistencies in the results arise because the global trend can have the opposite sign to the basinwide trend.

In conclusion, whether detrending should be applied or not, and if yes, what type of polynomial should be used, depends on the characteristics of the trend and the internal variability in the control simulation. Here we have chosen to use quadratic detrending, thus subtracting a large part of the internal centennial variability from the analyzed time series. For other studies, a different choice might be more appropriate [e.g., Sohail et al. (2021) and Irving et al. (2021) use cubic].

TABLE A1. UKESM1.0 historical ensemble-mean OHU by basin in the 0–700 m layer, and one standard deviation of the ensemble.

OHU by basin, UKESM1.0, 0–2000 m, 1991–2014	No detrending	Linear	Quadratic	Fourth-order polynomial
NA	41 ± 12	41 ± 11	41 ± 10	41 ± 10
NI	2 ± 2	3 ± 2	3 ± 2	3 ± 2
NP	30 ± 9	31 ± 9	31 ± 10	31 ± 10
SA	31 ± 7	31 ± 7	33 ± 9	33 ± 9
SI	31 ± 9	32 ± 9	32 ± 9	32 ± 9
SP	27 ± 15	28 ± 15	28 ± 15	28 ± 15
Global	168 ± 30	170 ± 30	172 ± 34	171 ± 35

TABLE A2. GC3.1-LL historical ensemble-mean OHU by basin in the 0–700 m layer, and one standard deviation of the ensemble.

OHU by basin, GC3.1-LL, 0–2000 m, 1991–2014	No detrending	Linear	Quadratic	Fourth-order polynomial
NA	45 ± 7	44 ± 7	42 ± 8	42 ± 8
NI	4 ± 1	3 ± 2	3 ± 2	3 ± 2
NP	30 ± 7	29 ± 7	31 ± 7	30 ± 7
SA	37 ± 14	36 ± 15	34 ± 17	34 ± 17
SI	31 ± 5	30 ± 5	30 ± 6	30 ± 6
SP	27 ± 3	25 ± 3	25 ± 5	24 ± 4
Global	179 ± 22	171 ± 22	171 ± 28	169 ± 29

TABLE A3. CNRM-CM6-1 historical ensemble-mean OHU by basin in the 0–700 m layer, and one standard deviation of the ensemble.

OHU by basin, CNRM-CM6-1, 0–700 m, 1991–2014	No detrending	Linear	Quadratic	Fourth-order polynomial
NA	26 ± 7	22 ± 6	23 ± 8	24 ± 7
NI	−1 ± 3	0 ± 3	1 ± 3	−1 ± 2
NP	24 ± 15	29 ± 15	29 ± 11	25 ± 14
SA	21 ± 11	22 ± 10	25 ± 11	21 ± 12
SI	19 ± 16	18 ± 12	19 ± 9	20 ± 17
SP	24 ± 19	29 ± 17	24 ± 15	24 ± 19
Global	115 ± 22	122 ± 22	125 ± 25	116 ± 24

TABLE A4. CNRM-ESM2-1 historical ensemble-mean OHU by basin in the 0–700 m layer, and one standard deviation of the ensemble.

OHU by basin, CNRM-ESM2-1, 0–700 m, 1991–2014	No detrending	Linear	Quadratic	Fourth-order polynomial
NA	26 ± 7	22 ± 6	21 ± 5	24 ± 7
NI	−1 ± 3	0 ± 3	0 ± 3	−1 ± 2
NP	24 ± 15	29 ± 15	27 ± 19	25 ± 14
SA	21 ± 11	22 ± 10	23 ± 15	21 ± 12
SI	19 ± 16	18 ± 12	17 ± 13	20 ± 17
SP	24 ± 19	29 ± 17	30 ± 20	24 ± 19
Global	115 ± 22	122 ± 22	119 ± 28	116 ± 24

## 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. C., 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>.
- Andrews, M. B., and Coauthors, 2020: Historical simulations with HadGEM3-GC3.1 for CMIP6. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001995, <https://doi.org/10.1029/2019MS001995>.
- Andrews, T., and Coauthors, 2019: Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3.1 and UKESM1. *J. Adv. Model. Earth Syst.*, **11**, 4377–4394, <https://doi.org/10.1029/2019MS001866>.
- Boucher, O., and Coauthors, 2020: Presentation and evaluation of the IPSL-CM6A-LR climate model. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002010, <https://doi.org/10.1029/2019MS002010>.
- Bourgeois, T., N. Goris, J. Schwinger, and J. F. Tjiputra, 2022: Stratification constrains future heat and carbon uptake in the Southern Ocean between 30°S and 55°S. *Nat. Commun.*, **13**, 340, <https://doi.org/10.1038/s41467-022-27979-5>.
- Boyer, T. P., and Coauthors, 2018: World Ocean Database 2018. NOAA Atlas NESDIS 87, 207 pp., [https://www.ncei.noaa.gov/sites/default/files/2020-04/wod\\_intro\\_0.pdf](https://www.ncei.noaa.gov/sites/default/files/2020-04/wod_intro_0.pdf).
- Brient, F., R. Roehrig, and A. Voldoire, 2019: Evaluating marine stratocumulus clouds in the CNRM-CM6-1 model using short-term hindcasts. *J. Adv. Model. Earth Syst.*, **11**, 127–148, <https://doi.org/10.1029/2018MS001461>.
- Cheng, L., K. E. Trenberth, M. D. Palmer, J. Zhu, and J. P. Abraham, 2016: Observed and simulated full-depth ocean heat content changes for 1970–2005. *Ocean Sci.*, **12**, 925–935, <https://doi.org/10.5194/os-12-925-2016>.
- , —, 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>.
- , J. Abraham, Z. Hausfather, and K. E. Trenberth, 2019: How fast are the oceans warming? *Science*, **363**, 128–129, <https://doi.org/10.1126/science.aav7619>.
- , and Coauthors, 2020: Record-setting ocean warmth continued in 2019. *Adv. Atmos. Sci.*, **37**, 137–142, <https://doi.org/10.1007/s00376-020-9283-7>.
- Collins, J. W., and Coauthors, 2017: AerChemMIP: Quantifying the effects of chemistry and aerosols in CMIP6. *Geosci. Model Dev.*, **10**, 585–607, <https://doi.org/10.5194/gmd-10-585-2017>.
- 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>.
- 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>.
- , and Coauthors, 2021: Human influence on the climate system. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 423–552, [https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC\\_AR6\\_WGI\\_Chapter03.pdf](https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter03.pdf).
- Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.
- Gastineau, G., A. R. Friedman, M. Khodri, and J. Vialard, 2019: Global ocean heat content redistribution during the 1998–2012 Interdecadal Pacific Oscillation negative phase. *Climate Dyn.*, **53**, 1187–1208, <https://doi.org/10.1007/s00382-018-4387-9>.
- Gebbie, G., and P. Huybers, 2019: The Little Ice Age and 20th-century deep Pacific cooling. *Science*, **363**, 70–74, <https://doi.org/10.1126/science.aar8413>.
- Geoffroy, O., D. Saint-Martin, G. Bellon, A. Voldoire, D. J. L. Olivié, and S. Tytéc, 2013: Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *J. Climate*, **26**, 1859–1876, <https://doi.org/10.1175/JCLI-D-12-00196.1>.
- Gleckler, P. J., P. J. Durack, R. J. Stouffer, G. C. Johnson, and C. E. Forest, 2016: Industrial-era global ocean heat uptake doubles in recent decades. *Nat. Climate Change*, **6**, 394–398, <https://doi.org/10.1038/nclimate2915>.
- Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans*, **118**, 6704–6716, <https://doi.org/10.1002/2013JC009067>.
- Gouretski, V., 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>.
- Goutorbe, B., J. Poort, F. Lucazeau, and S. Raillard, 2011: Global heat flow trends resolved from multiple geological and geophysical proxies. *Geophys. J. Int.*, **187**, 1405–1419, <https://doi.org/10.1111/j.1365-246X.2011.05228.x>.
- Gregory, J. M., and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, <https://doi.org/10.1029/2003GL018747>.
- , T. Andrews, P. Ceppi, T. Mauritsen, and M. J. Webb, 2020: How accurately can the climate sensitivity to CO<sub>2</sub> be estimated from historical climate change? *Climate Dyn.*, **54**, 129–157, <https://doi.org/10.1007/s00382-019-04991-y>.
- Griffies, S. M., and Coauthors, 2016: OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geosci. Model Dev.*, **9**, 3231–3296, <https://doi.org/10.5194/gmd-9-3231-2016>.
- Gulev, S. K., and Coauthors, 2021: Changing state of the climate system. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 287–422, [https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC\\_AR6\\_WGI\\_Chapter02.pdf](https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter02.pdf).
- Held, I. M., and V. D. Larichev, 1996: A scaling theory for horizontally homogenous, baroclinically unstable flow on a beta plane. *J. Atmos. Sci.*, **53**, 946–952, [https://doi.org/10.1175/1520-0469\(1996\)053<0946:ASTFHH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1996)053<0946:ASTFHH>2.0.CO;2).
- Hobbs, W., M. D. Palmer, and D. Monselesan, 2016: An energy conservation analysis of ocean drift in the CMIP5 global coupled models. *J. Climate*, **29**, 1639–1653, <https://doi.org/10.1175/JCLI-D-15-0477.1>.
- Irving, D., W. Hobbs, J. Church, and J. Zika, 2021: A mass and energy conservation analysis of drift in the CMIP6 ensemble. *J. Climate*, **34**, 3157–3170, <https://doi.org/10.1175/JCLI-D-20-0281.1>.

- Ishii, M., 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., J. M. Lyman, and N. G. Loeb, 2016: Improving estimates of Earth's energy imbalance. *Nat. Climate Change*, **6**, 639–640, <https://doi.org/10.1038/nclimate3043>.
- Kuhlbrodt, T., and J. M. Gregory, 2012: Ocean heat uptake and its consequences for the magnitude of sea level rise and climate change. *Geophys. Res. Lett.*, **39**, L18608, <https://doi.org/10.1029/2012GL052952>.
- , and Coauthors, 2018: The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *J. Adv. Model. Earth Syst.*, **10**, 2865–2888, <https://doi.org/10.1029/2018MS001370>.
- Levitus, S., and Coauthors, 2012: World Ocean heat content and thermocline sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett.*, **39**, L10603, <https://doi.org/10.1029/2012GL051106>.
- Liu, W., S.-P. Xie, and J. Lu, 2016: Tracking ocean heat uptake during the surface warming hiatus. *Nat. Commun.*, **7**, 10926, <https://doi.org/10.1038/ncomms10926>.
- Locarnini, R. A., and Coauthors, 2019: *Temperature*. Vol. 1, *World Ocean Atlas 2018*, NOAA Atlas NESDIS 81, 52 pp., [https://www.ncei.noaa.gov/data/oceans/woa/WOA18/DOC/woa18\\_vol1.pdf](https://www.ncei.noaa.gov/data/oceans/woa/WOA18/DOC/woa18_vol1.pdf).
- Meehl, G. A., C. A. Senior, V. Eyring, G. Flato, J.-F. Lamarque, R. J. Stouffer, K. E. Taylor, and M. Schlund, 2020: Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Sci. Adv.*, **6**, eaba1981, <https://doi.org/10.1126/sciadv.aba1981>.
- Menary, M. B., and Coauthors, 2018: Preindustrial control simulations with HadGEM3-GC3.1 for CMIP6. *J. Adv. Model. Earth Syst.*, **10**, 3049–3075, <https://doi.org/10.1029/2018MS001495>.
- , and Coauthors, 2020: Aerosol-forced AMOC changes in CMIP6 historical simulations. *Geophys. Res. Lett.*, **47**, e2020GL088166, <https://doi.org/10.1029/2020GL088166>.
- Merchant, C. J., and Coauthors, 2014: Sea surface temperature datasets for climate applications from phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geosci. Data J.*, **1**, 179–191, <https://doi.org/10.1002/gdj3.20>.
- 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>.
- Palmer, M. D., 2017: Reconciling estimates of ocean heating and Earth's radiation budget. *Curr. Climate Change Rep.*, **3**, 78–86, <https://doi.org/10.1007/s40641-016-0053-7>.
- , and D. J. McNeall, 2014: Internal variability of Earth's energy budget simulated by CMIP5 climate models. *Environ. Res. Lett.*, **9**, 034016, <https://doi.org/10.1088/1748-9326/9/3/034016>.
- , C. M. Domingues, A. B. A. Slangen, and F. Boeira 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/abdaccc>.
- Pincus, R., P. M. Forster, and B. Stevens, 2016: The Radiative Forcing Model Intercomparison Project (RFMIP): Experimental protocol for CMIP6. *Geosci. Model Dev.*, **9**, 3447–3460, <https://doi.org/10.5194/gmd-9-3447-2016>.
- Purkey, S. G., and G. C. Johnson, 2010: Warming of global abyssal and deep Southern Ocean waters between the 1990s and 2000s: Contributions to global heat and sea level rise budgets. *J. Climate*, **23**, 6336–6351, <https://doi.org/10.1175/2010JCLI3682.1>.
- Séférian, R., and Coauthors, 2016: Inconsistent strategies to spin up models in CMIP5: Implications for ocean biogeochemical model performance assessment. *Geosci. Model Dev.*, **9**, 1827–1851, <https://doi.org/10.5194/gmd-9-1827-2016>.
- , and Coauthors, 2019: Evaluation of CNRM Earth system model, CNRM-ESM2-1: Role of Earth system processes in present-day and future climate. *J. Adv. Model. Earth Syst.*, **11**, 4182–4227, <https://doi.org/10.1029/2019MS001791>.
- Sellar, A. A., and Coauthors, 2019: UKESM1: Description and evaluation of the U.K. Earth system model. *J. Adv. Model. Earth Syst.*, **11**, 4513–4558, <https://doi.org/10.1029/2019MS001739>.
- Sen Gupta, A., N. C. Jourdain, J. N. Brown, and D. Monselesan, 2013: Climate drift in the CMIP5 models. *J. Climate*, **26**, 8597–8615, <https://doi.org/10.1175/JCLI-D-12-00521.1>.
- Sohail, T., D. B. Irving, J. D. Zika, R. M. Holmes, and J. A. Church, 2021: Fifty year trends in global ocean heat content traced to surface heat fluxes in the sub-polar ocean. *Geophys. Res. Lett.*, **48**, e2020GL091439, <https://doi.org/10.1029/2020GL091439>.
- Volodire, A., and Coauthors, 2019: Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *J. Adv. Model. Earth Syst.*, **11**, 2177–2213, <https://doi.org/10.1029/2019MS001683>.
- von Schuckmann, K., and Coauthors, 2016: An imperative to monitor Earth's energy imbalance. *Nat. Climate Change*, **6**, 138–144, <https://doi.org/10.1038/nclimate2876>.
- , 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>.
- Williams, K. D., and Coauthors, 2018: The Met Office Global Coupled Model 3.0 and 3.1 (GC3 & GC3.1) configurations. *J. Adv. Model. Earth Syst.*, **10**, 357–380, <https://doi.org/10.1002/2017MS001115>.
- Yool, A., and Coauthors, 2020: Spin-up of UK Earth system model 1 (UKESM1) for CMIP6. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001933, <https://doi.org/10.1029/2019MS001933>.
- , and Coauthors, 2021: Evaluating the physical and biogeochemical state of the global ocean component of UKESM1 in CMIP6 historical simulations. *Geosci. Model Dev.*, **14**, 3437–3472, <https://doi.org/10.5194/gmd-14-3437-2021>.
- Zanna, L., S. Khatiwala, J. M. Gregory, J. Ison, and P. Heimbach, 2019: Global reconstruction of historical ocean heat storage and transport. *Proc. Natl. Acad. Sci. USA*, **116**, 1126–1131, <https://doi.org/10.1073/pnas.1808838115>.
- Zuidema, P., and Coauthors, 2016: Challenges and prospects for reducing coupled climate model SST biases in the eastern tropical Atlantic and Pacific Oceans: The U.S. CLIVAR Eastern Tropical Oceans Synthesis Working Group. *Bull. Amer. Meteor. Soc.*, **97**, 2305–2328, <https://doi.org/10.1175/BAMS-D-15-00274.1>.