Model biases in the atmosphere-ocean partitioning of poleward heat transport are persistent across three CMIP generations

Aaron Donohoe[1], Robert Fajber[2], Tyler Cox[3], Kyle Armour[1], David DBattisti[1], and Gerard Roe[1]

[1]University of Washington
[2]McGill
[3]University of Washington Seattle

October 17, 2023

Abstract

The observed partitioning of poleward heat transport between atmospheric and oceanic heat transports (AHT and OHT) is compared to that in coupled climate models. Poleward OHT in the models is biased low in both hemispheres, with the largest biases in the Southern Hemisphere extratropics. Poleward AHT is biased high in the Northern Hemisphere, especially in the vicinity of the peak AHT near 40$ˆ\circ$N. The significant model biases are persistent across three model generations (CMIP3, CMIP5, CMIP6) and are insensitive to the satellite radiation and atmospheric reanalyses products used to derive observational estimates of AHT and OHT. Model biases in heat transport partitioning are consistent with biases in the spatial structure of energy input to the ocean and atmosphere. Specifically, larger than observed model evaporation in the tropics adds excess energy to the atmosphere that drives enhanced poleward AHT at the expense of weaker OHT

Model biases in the atmosphere-ocean partitioning of poleward heat transport are persistent across three CMIP generations

A. Donohoe[1,3], R. Fajber[2], T. Cox [3], K.C. Armour[3,4], D.S. Battisti [3]and G.H. Roe [4]

[1]Polar Science Center, Applied Physics Laboratory
University of Washington
Seattle, Washington 98195, USA.
[2] Department of Atmospheric and Oceanic Sciences, McGill University
[3] Department of Atmospheric and Oceanic Sciences, University of Washington
[4]School of Oceanography, University of Washington
[5]Department of Earth and Space Sciences, University of Washington

Key Points:

- Climate models simulate too little poleward oceanic heat transport and too much poleward atmospheric heat transport in the extratropics
- Model biases in heat transport partitioning are persistent across model generation and are insensitive to the observational data sets used
- Stronger than observed evaporation in models enhances atmospheric heat transport at the expense of oceanic heat transport

Abstract

The observed partitioning of poleward heat transport between atmospheric and oceanic heat transports (AHT and OHT) is compared to that in coupled climate models. Poleward OHT in the models is biased low in both hemispheres, with the largest biases in the Southern Hemisphere extratropics. Poleward AHT is biased high in the Northern Hemisphere, especially in the vicinity of the peak AHT near 40°N. The significant model biases are persistent across three model generations (CMIP3, CMIP5, CMIP6) and are insensitive to the satellite radiation and atmospheric reanalyses products used to derive observational estimates of AHT and OHT. Model biases in heat transport partitioning are consistent with biases in the spatial structure of energy input to the ocean and atmosphere. Specifically, larger than observed model evaporation in the tropics adds excess energy to the atmosphere that drives enhanced poleward AHT at the expense of weaker OHT.

Plain Language Summary

The equator-to-pole contrast of solar radiation entering the climate system drives the large-scale oceanic and atmospheric circulations that, in turn, move heat from the equator to the poles to moderate latitudinal temperature contrasts. The ocean moves the majority of heat in the tropics whereas the atmosphere moves the vast majority of heat in the mid- and polar-latitudes. We demonstrate that state-of-the-art climate models representing both oceanic and atmospheric circulations systematically simulate too little oceanic heat transport and too much atmospheric heat transport relative to observational estimates. These model biases in the atmosphere-ocean partitioning of poleward heat transport are persistent across three generations of climate model ensembles spanning twenty years of progress in climate modeling and are insensitive to the choice of datasets used to calculate observed heat transports. The model biases are consistent with stronger than observed surface evaporation in the tropics which enhances atmospheric heat transport at the expense of oceanic heat transport.

1 Introduction

The combined meridional heat transport (MHT) by the ocean and atmosphere moderates spatial gradients in temperature on Earth. In the absence of MHT, the equator-to-pole temperature gradient would be approximately three times larger than observed based on radiative considerations alone (Pierrehumbert, 2010), rendering the tropics uninhabitably warm and the high latitudes uninhabitably cold. Observational estimates of the partitioning of MHT between poleward atmospheric heat transport (AHT) and poleward oceanic heat transport (OHT) show that OHT exceeds AHT in the deep tropics (equatorward of 10°) while AHT dominates in the mid- and high-latitudes of both hemispheres (Vonder Haar & Oort, 1973; Oort & Haar, 1976; Trenberth & Caron, 2001; Mayer et al., 2021).

The partitioning of MHT between AHT and OHT impacts climate and its changes. For example, the convergence of OHT in the extratropics is inherently linked to the surface energy budget and thus demands a surface temperature response, whereas the convergence of the same quantity of AHT in the atmosphere can be radiated to space with less impact on surface climate (Cardinale et al., 2020). Indeed, previous work by Enderton & Marshall (2009) has shown that aquaplanets with nearly identical total MHT but different AHT-OHT partitioning can have very different climates (e.g., different surface temperature and sea ice distributions).

Given the dependence of climate on the partitioning between poleward AHT and OHT, we ask here: how well do coupled climate models represent the observed AHT-OHT partitioning? This question was briefly addressed in Chapter 9 of the Intergovernmental Panel on Climate Change 5th assessment report (Flato et al., 2013) which concluded that model OHT was within the wide range of observational OHT estimates. Comparison of observational and model AHT-OHT partitioning is difficult because the standard methodology for partitioning MHT between AHT and OHT differs between observations and models due to the contrasting reliability and availability of the climate fields used to calculate AHT and OHT. Recent work (Donohoe et al., 2020) has demonstrated the near equivalence of the model and observational approaches to AHT-OHT partitioning in a model setting, enabling a comprehensive observational-model comparison. In this study we apply these methods to three generations of coupled model simulations (Phases 3, 5, and 6 of the Coupled Model Intercomparison Project, CMIP) and to several observational radiation and atmospheric reanalysis products. Our aim is to determine whether the models accurately capture the partitioning of AHT and OHT derived from observational datasets.

In Section 2 we provide an overview of the observational and model methodologies for partitioning MHT into AHT and OHT and demonstrate the near equivalence of these two approaches. In section 3, we compare the observational and model MHT partitioning across the three different model generations (CMIP3, CMIP5, and CMIP6) and examine the sensitivity of our findings to the choice of observational data sets used to partition MHT. In Section 4 we consider an alternative method for comparing AHT-OHT partitioning in models and observations from the processes that contribute to spatial gradients in energy input to the atmosphere and ocean. A summary and discussion follows.

2 Methods for partitioning MHT into AHT and OHT in observations and coupled models

The methodology used to partition MHT into AHT and OHT in coupled climate models and observations is described in detail in Donohoe et al. (2020). Here we summarize the conceptual approach.

2.1 Heat transport partitioning in climate models

Near closure of the top of atmosphere (TOA) and surface energy budgets in climate models allows for the energy transport by the atmosphere and ocean across a latitude band to be calculated from the energy input into/out-of the fluid spatially integrated over the polar cap bounded by that latitude:

$$\mathrm{MHT}(\Theta) = 2\pi a^2 \int_{\Theta}^{90} -F^* \cos\theta d\theta, \tag{1}$$

where a is the radius of the Earth, Θ is the latitude (with θ a latitude variable of integration), and F is the net energy input to the atmosphere, ocean, or combined atmosphere-ocean system. The total MHT can be found by taking F to be the radiative flux at the TOA ($\mathrm{RAD_{TOA}}$), OHT by taking F to be the net surface heat flux (SHF = radiative plus turbulent flux into the ocean), and AHT by setting F to be the net energy input to the atmosphere ($\mathrm{RAD_{TOA}}$ - SHF). The * denotes that the global mean of each energy flux term has been removed to ensure heat transport goes to zero at both poles. This adjustment is necessary because climate models do not conserve energy globally (≈ 1 W m^{-2} imbalances) in both the atmosphere and ocean (Lucarini & Ragone, 2011).

2.2 Heat transport partitioning in observations

In contrast to coupled climate models where the surface energy budget is (nearly) closed, the sparsity and uncertainty of observational surface radiative and turbulent energy flux measurements results in an unrealistically large (>10 W m^{-2}) global mean surface energy imbalance (Stephens et al., 2012; Trenberth et al., 2009), which prohibits the evaluation of OHT from the net surface heat flux. Instead, we use a conceptual approach following Vonder Haar & Oort (1973) and Trenberth & Caron (2001): MHT is calculated using Eq. 1 with satellite RAD_{TOA} (Loeb & Coauthors, 2018); AHT is calculated from the time average of the vertically and zonally integrated meridional energy flux in the atmosphere derived from high frequency (6 hourly) atmospheric reanalysis; OHT is then calculated as the residual of satellite derived MHT and reanalysis derived AHT. In the AHT calculation, a vertically integrated moist static energy anomaly is removed before integrating (Donohoe & Battisti, 2013; Cardinale et al., 2020; Donohoe et al., 2020), effectively applying a mass correction needed to make the AHT calculation physically meaningful (Trenberth & Stepaniak, 2003; Liang et al., 2018).

To show that the "observational" and "model" methods are comparable, we partition MHT into AHT and OHT using both methods in a NCAR CESM1 coupled pre-industrial control simulation (see Donohoe et al., 2020, for details). The two approaches give nearly identical partitioning of MHT into AHT and OHT (cf. the dashed and solid red and blue lines in Supporting Information Fig. S1) with a root mean squared difference AHT (and OHT) between the two methods of 0.07 PW. The close correspondence of the two calculations of MHT partitioning suggests that the "observational" and "model" approaches we use here to partition MHT are directly comparable. We use this result to justify the examination of potential model biases in MHT partitioning using these two methodologies.

3 Results: model biases in MHT partitioning

Climate model biases in MHT partitioning are analyzed using pre-industrial control simulations from three different CMIP generations and several different sets of observational products (see Supporting Information for details). The presentation of our results is organized as follows. Section 3.1 presents the observational estimate of MHT partitioning using the most contemporary and high resolution data available, which is compared against the MHT partitioning in the three CMIP ensembles. Section 3.2 analyzes the sensitivity of our results to the observational data used by comparing eight different observational estimates of MHT partitioning against the multi-generation CMIP ensemble mean. The results show that the sign and spatial structure of model biases in MHT partitioning are consistent across model generation and observational data sets used.

3.1 Consistent model biases in AHT-OHT partitioning across three CMIP generations

In this section, we use CERES Energy Balanced and Filled (EBAF) TOA radiation (Loeb et al., 2009) and the ERA5 atmospheric reanalysis (Hersbach et al., 2020) to calculate an observational estimate of MHT and its partitioning over the period 2001-2020. This observational estimate (solid line) is compared against each of the three CMIP ensembles (in each row of Fig. 1; with dashed lines showing individual models and the thick dashed lines showing CMIP ensemble averages).

Poleward MHT peaks near 35° in both hemispheres in both models and observations (Fig. 1), consistent with constraints due to Earth-Sun geometry whereby the meridional distribution of net TOA radiation (RAD^*_{TOA}) is dominated by the second order Legendre polynomial (equator-to-pole scale) as discussed by Stone (1978). However, across all three CMIP generations, the amplitude of poleward MHT in models is biased low in the mid-latitudes of both hemispheres relative to the observational estimate. In the Southern Hemisphere (SH), the observational estimate of maximum poleward MHT is 5.7 PW, which is significantly larger (95% confidence interval of t-test) than the ensemble means of CMIP3 (5.2 PW), CMIP5 (5.3 PW), and CMIP6 (5.4 PW). In the Northern Hemisphere (NH) the observational estimate of maximum poleward MHT is 5.8 PW, exceeding the ensemble mean of CMIP3 (5.6 PW), CMIP5 (5.5 PW), and CMIP6 (5.7 PW), but only for CMIP5 is the difference statistically significant. In the SH, the inter-model spread in peak MHT (2 standard deviations) is as large as 23% of the ensemble mean and has values of 1.2 PW in CMIP3, 0.8 PW in CMIP5, and 0.8 PW in CMIP6. The inter-model spread in peak NH MHT is smaller than its SH counterpart with values of 0.8 PW in CMIP3, 0.6 PW in CMIP5, and 0.6 PW in CMIP6. Donohoe & Battisti (2011) demonstrated that the the inter-model spread and bias in MHT in CMIP3 results from biases and spread in the albedo of clouds which impact the equator-to-pole gradient of absorbed solar radiation. The bias and spread in MHT is only slightly reduced in CMIP5 and CMIP6, and also results primarily from model differences in mean-state shortwave cloud radiative effects (not shown).

We next analyze the partitioning of MHT between OHT and AHT. In the NH, the model ensemble mean is significantly biased toward too little poleward OHT and too much poleward AHT in all three CMIP generations. The observational estimate of peak NH AHT is 4.4 PW as compared to 4.7 $\pm$ 0.2 PW in CMIP3, 4.7 $\pm$ 0.1 PW in CMIP5, and 4.8 $\pm$ 0.1 PW in CMIP6 where the stated uncertainty is two standard deviations of the ensemble mean. The peak in NH OHT is robustly equatorward of the peak AHT, but has significantly larger values for the observational estimate (2.0 PW) than in the model ensemble means (1.7 PW in CMIP3, 1.8 PW in CMIP5, and 1.7 PW in CMIP6). The model bias toward smaller than observed OHT extends poleward to the Arctic where OHT has been demonstrated to have large impacts on sea ice extent (Holland et al., 2006; Seager et al., 2002).

In the SH, poleward OHT in the models is biased low relative to the observational estimate in all three CMIP generations. The largest biases in OHT are found the vicinity of 40°S where the observational OHT is -0.7 PW compared to the ensemble mean OHT at that latitude is -0.3 $\pm$ 0.2 PW in CMIP3, -0.2 $\pm$ 0.1 PW in CMIP5, and -0.1 $\pm$ 0.1 PW in CMIP6. The observational estimate of poleward OHT is only exceeded in three model simulations (two in CMIP3 and one in CMIP5). In contrast, the poleward AHT in the SH is not significantly different between the models and observational estimates.

These results suggest that in the SH, the majority of the model biases in MHT are a result in biases in OHT, whereas in the NH the models generally simulate too much poleward AHT and too little poleward OHT. Alternatively, the fractional contribution of AHT-OHT to total MHT (i.e., normalizing each model by the model specific MHT) is biased toward too much poleward AHT and too little poleward OHT with biases that are nearly hemispherically symmetric between the two hemispheres (not shown). Importantly, the sign and spatial structure of model biases in MHT and AHT-OHT partitioning are remarkably consistent across the three CMIP generations spanning over 20 years of progress in climate modeling.

3.2 Sensitivity of results to observational data sets used

We next consider whether the identified model biases in AHT-OHT partitioning are sensitive to the choice of observational data sets (TOA radiation and atmospheric reanalysis) used to partition MHT. We use the mean of all ensemble members across all three CMIP generations, referred to as the CMIP-mean, as a reference for all analyses in this subsection.

We begin by analyzing the MHT and AHT/OHT partitioning estimated using two additional satellite-derived observational estimates of TOA radiation (see Supporting Information for details): the unadjusted CERES single scanner footprint (SSF) data and the ERBE satellite data which spans the 1984-1990 (left panels of Fig. 2 bordered by the black box). In these three panels, the choice of TOA radiation product alters the calculated observational MHT (solid black line) whereas the AHT is unchanged between panels (ERA5 is used in each). Because the observational OHT is calculated from the difference of MHT and AHT, the observational OHT estimate (solid blue line) also varies between panels. Observational MHT calculated from the three different TOA radiation products is consistently larger than the CMIP-mean in both hemispheres. Model biases in MHT are largest when the CERES SSF product is used (Fig. 2E) and smallest when the ERBE product is used (Fig. 2C). The CMIP-mean OHT is biased low compared to that derived from all three TOA radiation datasets with largest magnitude biases when CERES SSF is used, especially in the SH. Model biases in AHT/OHT partitioning are insensitive to observational TOA radiation data set used which give a consistent estimate of MHT despite their substantial (≈ 5 W m^{-2} = 2.5 PW globally integrated) differences in global mean TOA radiative balance associated with absolute calibration uncertainty (Loeb et al., 2009).

We next analyze the sensitivity of our results to the choice of atmospheric reanalysis used to calculate the AHT (Fig. 2 panels A, B, D, F and H). In these five panels, the MHT is identical (calculated using the CERES EBAF product) whereas the AHT is calculated from the ERA5, ERA-interim, NCEP, MERRA2, and JRA reanalyses. Since OHT is calculated from the residual of MHT and AHT, the OHT difference between the three panels are equal and opposite to the the inter-panel differences in AHT. The CMIP-mean bias toward too much poleward AHT and too little poleward OHT is apparent using all five observational estimates of AHT. Poleward AHT is largest when using ERA5 followed closely by JRA, MERRA2 and then ERA interim, whereas using NCEP produces the smallest poleward AHT with the most notable difference near the peak in the SH at 40°S. Therefore, model biases in the AHT-OHT partitioning are smallest using ERA5 and largest using NCEP. These results suggest that the sign and spatial structure of model biases in MHT partitioning are consistent across atmospheric reanalysis datasets, whereas the magnitude of the bias depends on the reanalyses product used. Differences in AHT calculated between the different reanalyses are not impacted by differences in the spatial resolution (see analysis and Fig. S2 in the Supporting Information) as even the coarsest product (NCEP) resolves the spatial scales responsible for the vast majority of AHT.

Finally, we evaluate whether heat storage due to the transient response to anthropogenic forcing impacts our observational estimates of OHT. The Earth is not in equilibrium but, rather, is accumulating energy at an average rate of 0.7 W m^{-2} globally (Johnson et al., 2016). The vast majority of this energy accumulation is stored in the ocean (Von Schuckmann et al., 2016) and it is possible that the spatial structure of this energy storage projects onto our diagnoses of observational OHT for the following reason: observed 'implied' OHT is calculated from the spatial integral of inferred surface heat fluxes (TOA radiation plus AHT convergence) and the latter is balanced by the sum of OHT divergence and ocean heat storage in a transient system. We diagnose the impact of observed ocean heat storage on the implied OHT ($\mathrm{OHT_{STORAGE}}$) from the trend in ocean heat content, derived from UK Hadley Center EN4 objective ocean analysis (Good et al., 2013) over the CERES period (see Supporting Information for details). $\mathrm{OHT_{STORAGE}}$ is removed from the 'implied' OHT to estimate the 'true' OHT (solid teal line in Fig. 1F) that must be transported laterally in the ocean to close the ocean energy budget. $\mathrm{OHT_{STORAGE}}$ is very small (< 0.1 PW in magnitude) and, thus, the diagnosed 'true' OHT is visually indistinguishable from the observational 'implied' OHT (solid blue line in Fig. 1F). The global mean ocean heat uptake of 0.7 W m^{-2} translates to 0.4 PW of global energy input to the ocean but the implied OHT of ocean heat storage is significantly smaller in magnitudes due to ocean heat uptake being more globally uniform than regionally isolated. The negligible impact of ocean heat storage on 'implied' OHT over the historical period is consistent with the small (< 0.1 PW) differences between OHT in the ensemble mean of historical CMIP5 simulations averaged over the 2000-2018 time period as compared the pre-industrial control simulations using the same models (Supporting Information Fig. S3).

Collectively, these results suggest that the sign of model biases in AHT-OHT partitioning is robust to different observational products (satellite TOA radiation and atmospheric reanalysis) used to partition MHT. Additionally, the spatial pattern of transient heat uptake by the ocean makes a negligible impact on estimated OHT. However, the magnitude of the model bias in AHT-OHT partitioning does vary with observational datasets used. In this regard, the use of CERES EBAF and ERA5 data for our primary analysis (Fig. 1) is a conservative estimate of model biases in AHT-OHT partitioning (a smaller OHT bias is found only when using the combination of ERBE and ERA5 products).

4 Biases in energy input to the atmosphere and ocean and inferred AHT and OHT biases

Here we evaluate potential causes of the persistent model biases in AHT and OHT in terms of model biases in the spatial structure of energy input into the ocean and atmosphere. Starting in the ocean, energy conservation demands that OHT across a latitude band balances the net surface heat flux *out* of the ocean (-SHF by our sign convention) integrated over the polar cap bounded by that latitude, which from Eq. 1 is represented by:

$$\mathrm{OHT}(\Theta) = 2\pi a^2 \int_{\Theta}^{90} \left(-\mathrm{SHF}^*\right) cos(\theta) d\theta. \tag{2}$$

SHF is equal to the net downward surface radiation ($\mathrm{RAD_{SURF}}$) into the ocean minus the upward turbulent energy fluxes of sensible (SENS) and latent heat ($\mathrm{L_vE}$):

$$\mathrm{SHF} = \mathrm{RAD_{SURF}} - \mathrm{SENS} - \mathrm{L_vE}. \tag{3}$$

Substitution of Eq. (3) into Eq. (2) allows the OHT to be decomposed into the implied transports of each term contributing to SHF:

$$\mathrm{OHT} = \mathrm{OHT_{RAD,SURF}} + \mathrm{OHT_{SENS}} + \mathrm{OHT_E}, \tag{4}$$

where, for example, the OHT implied by evaporation ($\mathrm{OHT_E}$) is:

$$\mathrm{OHT}(\Theta)_\mathrm{E} = 2\pi a^2 \int_{\Theta}^{90} \mathrm{L_vE}^* cos(\theta) d\theta, \tag{5}$$

where, as in Eq. 1 and 2, the * indicates that the global (ocean domain) mean has been removed from the term. Because SENS^* is small compared to the other terms (Fig. 3C) and $\mathrm{RAD_{SURF}}$ is dominated by solar input to the surface (Supporting Information Figs. S5E,F), the predominant energy balance in this framework can be summarized as follows: the magnitude of OHT (black line in Fig. 3D) is governed by the imbalance between excess (relative to the global mean) solar radiation entering the tropical ocean (orange line) and excess evaporative loss (green). Perfect local compensation between surface solar input and evaporation implies zero OHT whereas weaker evaporative loss demands a larger fraction of solar input be realized as OHT. We use this framework to understand model biases in OHT in terms of biases in the meridional structure of terms contributing to SHF.

The latitudinal structure of CMIP-mean $\mathrm{L_vE}$, SENS and $\mathrm{SURF_{RAD}}$ over the ocean domain is compared to observational estimates of the same quantities with $\mathrm{L_vE}$ and SENS taken from the WHOI Objectively Analyzed (OA) Air-Sea Flux product (Yu et al., 2004) and $\mathrm{SURF_{RAD}}$ estimates from the CERES EBAF surface product (Kato & Coauthors,

2018) in Fig. 3C. Evaporation is biased high in models (relative to the observational estimate) at all latitudes except the Arctic (Supporting Information Fig. S4). Evaporation biases are largest (> 20 W m^{-2}) in the subtropics of both hemispheres and are much smaller in the high latitudes. These evaporation biases manifest as enhanced subtropical ocean energy loss by E^* in the models (cf. the dashed and solid green lines in Fig. 3C) and an implied model bias toward too little (by approximately 0.4 PW) poleward OHT due to evaporation in each hemisphere (OHT_E, green line in Fig. 3D). Thus, evaporation biases alone explain the majority of the model bias in OHT identified in Section 3 (compare green and dashed black lines in Fig. 3D).

The observational RAD^*_{SURF} has a stronger equator-to-pole gradient than that in climate models (cf. the solid and dashed orange lines in Fig. 3C) especially in the SH. Model biases in RAD^*_{SURF} are associated with larger than observed downwelling solar radiation into the extratropical Southern Ocean (Supporting Information Fig. S5E) due to clouds that are optically thinner than observed (Donohoe & Battisti, 2012). As a result, observed poleward $OHT_{RAD,SURF}$ is larger than that in models with larger magnitude (0.4 PW) biases in the SH. The model biases in $OHT_{RAD,SURF}$ mirror the impact of TOA radiation biases on MHT (left panels of Figure 1) including the partitioning between shortwave and longwave biases within each hemisphere, suggesting that model biases in MHT and OHT in the SH are due to biases in shortwave absorption whereas those in the NH are due to biases in OLR and net surface longwave (Supporting Information Figs. S5B,F).

The sum of model biases in OHT_E, $OHT_{RAD,SURF}$ and OHT_{SENS} (solid black line in Fig. 3D) finds that models would have weaker than observed poleward OHT of 0.6 PW in the NH and 0.8 PW in the SH based on biases in energy input to the ocean. This overall inferred OHT bias is primarily due to a nearly hemispherically mirror-imaged bias in OHT_E which is enhanced by poleward $OHT_{RAD,SURF}$ in the SH. The bias in OHT inferred from surface flux biases matches the spatial structure but exceeds in magnitude the OHT biases calculated in Section 3 from TOA radiation and atmospheric reanalysis (dashed black line in Fig. 3D). These two calculations of model OHT biases do not have to match as they use different conceptual approaches and rely on completely independent observational climate fields. Nonetheless, the consistency of the sign, spatial pattern, and magnitude of the OHT biases calculated using the two different approaches suggest that the model biases in surface energy fluxes are large enough to account for the AHT-OHT partitioning biases inferred from the residual TOA radiation and AHT estimates.

We use a similar calculation of the model biases in implied AHT from the spatial structure of energy input to the atmosphere to compute an alternative estimate of AHT biases to those calculated in Section 3. The AHT analog to Eq. 4 is:

$$\mathrm{AHT} = \mathrm{AHT}_{\mathrm{RAD,ATMOS}} + \mathrm{AHT}_{\mathrm{SENS}} + \mathrm{AHT}_{\mathrm{E}}, \tag{6}$$

where the atmospheric analog to Eq. 5 for the AHT due to evaporation (AHT_E) is:

$$\mathrm{AHT}(\Theta)_E = 2\pi a^2 \int_{\Theta}^{90} -\mathrm{L_v E^*} cos(\theta) d\theta. \tag{7}$$

The spatial integral is over a global (land plus ocean) domain. Here RAD_{ATMOS} is the net radiative heating of the atmospheric column which is equivalent to the net radiation at TOA minus RAD_{SURF}. Fajber et al. (2023) demonstrated that poleward AHT is primarily determined by evaporation ($AHT \approx AHT_E$) because $L_v E^*$ dominates the spatial structure of energy input to the atmosphere. We note that $L_v E^*$ spatially integrated over the ocean domain has opposing impacts on AHT_E versus OHT_E (and likewise for $SENS^*$ and AHT_{SENS} versus OHT_{SENS}). This arises because excess evaporation over the low latitudes ($E^* > 0$) adds energy to the atmosphere to enhance the demand for poleward AHT at the expense of removing energy from the low-latitude ocean to reduce the demand for poleward OHT.

To more clearly see the compensation between biases in AHT-OHT due to model biases in L_vE^* (and SENS*) over the ocean domain, we take the following approach to compare models and observations of AHT via Eqs. 6 and 7. First, $\mathrm{AHT_E}$ and $\mathrm{AHT_{SENS}}$ are calculated from the observational WHOI OA evaporation and sensible heat flux data *over the ocean domain only*, and are compared to analogous model calculations over the ocean domain. Then, the contribution of turbulent energy fluxes over land to the combined $\mathrm{AHT_E}$ and $\mathrm{AHT_{SENS}}$ is estimated from the CERES EBAF net surface radiation spatially integrated over land. This approach assumes that (via surface energy balance) surface radiative gain is balanced by turbulent loss. These calculations are compared to analogous calculations in the models. Finally, $\mathrm{AHT_{RAD,ATMOS}}$ is calculated from the CERES EBAF TOA and surface data *over the global domain* and is compared to the analogous global domain calculation in models (orange lines in Fig. 3A,B). This strategy circumvents the lack of reliable observational estimates of turbulent energy fluxes over land – instead inferring them from a like-with-like observational-to-model comparison of surface radiation over land and assuming that $\mathrm{RAD_{SURF}}$ is balanced by upward turbulent fluxes from the land to the atmosphere (the latter assumption has been validated in models).

Model biases in $\mathrm{AHT_E}$ compose the vast majority of AHT biases diagnosed from Eq. 6 (cf. the green and solid black lines in Fig. 3B) and suggest that the stronger than observed poleward AHT in models is driven by an enhanced equator-to-pole gradient in evaporation. Model $\mathrm{RAD^*_{ATMOS}}$ is more negative in the deep tropics as compared to observations (due to stronger longwave cooling in the models– Supporting Information Fig. S5) which contributes to smaller $\mathrm{AHT_{RAD,ATMOS}}$ export from the tropics in the models that generally opposes the low latitude biases in $\mathrm{AHT_E}$ (orange line in Fig. 3B). Interestingly, shortwave absorption in the atmosphere is biased low in the models, which reduces the demand for poleward AHT by nearly 0.4 PW in both hemispheres (red line in Supporting Information Fig. S5D). However, this model deficit in atmospheric heating of the tropics is nearly compensated for by weaker than observed longwave cooling of the atmosphere such that there is almost no bias in $\mathrm{AHT_{RAD,ATMOS}}$ at the equator-to-pole scale. Turbulent energy fluxes over the land inferred from net surface radiation are nearly identical in models and observations and make a negligible impact on AHT biases (cf. purple dashed and solid lines in Fig. 3C,D).

These calculations demonstrate that the model biases in the partitioning of poleward heat transport between AHT and OHT that were inferred in Section 3 are consistent (in sign, spatial structure, and magnitude) with the model biases in energy input into the atmosphere and ocean by radiative fluxes and turbulent exchange between the atmosphere and ocean. Stronger than observed evaporation in the models contributes to enhanced poleward AHT at the expense of reduced OHT that is nearly hemispherically symmetric whereas radiative biases due to thinner than observed clouds in the extratropical Southern Ocean results in too weak poleward MHT that is primarily manifested in the surface energy budget and implied OHT bias.

5 Summary and discussion

Coupled climate models have too little poleward OHT in both hemispheres and too much AHT in the NH, compared to observational estimates. These model biases are remarkably consistent across three generations of coupled model ensembles (CMIP3, CMIP5, and CMIP6) and across different sets of observational TOA radiation and atmospheric reanalysis data. These conclusions are not impacted by observed transient energy accumulation in the ocean.

The method used here to balance the mass budget of the atmospheric reanalysis differs from that used in the work of Trenberth & Stepaniak (2004) and M. et al. (2017). Specifically, we implicitly assume zero net atmospheric mass flux through a given latitude circle whereas other works adjust the mass flux to balance the polar cap spatial integral of the surface pressure tendency and evaporation minus precipitation. Our choice stems from defining the energy budget with respect to a fixed mass of atmosphere (Donohoe & Battisti, 2013; Liang et al., 2018). The AHT associated with the mass flux due to evaporation minus precipitation is primarily compensated for a return flow of mass and energy in the ocean and requires a consistent treatment of the energy fluxes through the atmosphere, surface and ocean (M. et al., 2017) that depends on the choice of zero point energy (e.g., the units used for temperature). Physically, a poleward (water) mass flux in the atmosphere is balanced by the mass flux of precipitation minus evaporation and an equivalent equatorward mass flux in the ocean. The energy flux of each of these mass fluxes is the product of mass flux and mean energy of the fluid, has a minimal net (AHT+OHT) poleward energy transport but is of order 0.2 PW in magnitude for each the compensating AHT and OHT. The standard definition of SHF in climate models does not include the sensible heat of this net (water) mass flux across the air/sea interface and we believe including this term would create an inconsistency between the model derived and observationally inferred OHT. Our interpretation is supported by the near equivalence of the AHT calculated in CESM via the "observational" and "model" partitioning calculations using our method of calculating AHT from reanalysis data whereas including the net mass flux of water in the AHT creates a substantial mismatch between the two calculations (not shown). We emphasize that all choices made here were aimed at creating a consistent way to compare observational and model MHT and AHT-OHT partitioning despite the different climate fields that go into each calculation.

This work focused on model biases in the vertical zonal and time integral of atmospheric moist static energy fluxes that comprise AHT without regard for biases in the underlying atmospheric circulations and associated temperature and humidity structures of the atmosphere. Donohoe et al. (2020) demonstrated that model biases in poleward AHT primarily result from larger than observed dry (sensible) heat transport by transient eddies in the mid-latitudes of both hemisphere (their Fig. 4D) and in the NH smaller than observed dry heat transport by stationary eddies; the moisture (latent heat) transport has negligible biases. Model biases in evaporation are expected to be manifested as biases in both moist and dry AHT because dry AHT is set by the spatial pattern of condensational heating of the atmosphere which represents the portion of $\mathrm{AHT_E}$ that is not transported poleward as latent heat (Fajber et al., 2023); while spatial patterns of evaporation directly demand poleward moist AHT, the energy input to the atmosphere via evaporation is handed off to dry AHT where precipitation forms and the atmosphere is heated condensationally. Therefore, our finding that model biases toward too much AHT result from stronger than observed evaporation is consistent with the finding that excess poleward AHT in the models is expressed as a bias toward too much dry heat transport.

Remarkably, the model OHT bias inferred from observational estimates from satellite TOA radiation and atmospheric reanalyses is in descent agreement with model biases in the energy exchange between the ocean and atmosphere calculated from independent observational estimates of surface heat fluxes. The latter bias is due primarily to stronger than observed low-latitude evaporation in the models. We note that the community has been reluctant to diagnose OHT from the observed surface energy balance because of uncertainty in the turbulent energy fluxes. Yet, our analysis paints a consistent picture of the model biases in turbulent energy fluxes – whether these are inferred from the residual of TOA radiation and AHT or from bulk formula. We also note that observational estimates of global mean evaporation and its equator-to-pole gradient vary substantially (Stephens et al., 2012) with reanalysis products generally having more evaporation than the bulk formula based estimates such as WHOI OA flux (Yu et al., 2004) and SEAFLUX (Curry et al., 2004). We chose to use WHOI OA flux for the analysis in Section 4 because the bulk formula in this product are optimized to match buoy observations – making it the most observationally constrained estimate of evaporation. Additionally, the global constraint of evaporation balancing precipitation is nearly satisfied from the combination of the WHOI OA FLUX evaporation over the ocean (62.8 W m^{-2} contribution to global mean) plus the ERA5 reanalysis evaporation over land (12.9 W m^{-2} for a global total evaporation of 75.7 W m^{-2}) nearly balancing the best observational estimate of global mean precipitation (77.9 W m^{-2}) from the NOAA GPCP (Adler et al., 2018). The lack of closure of the observed global mean surface energy budget suggests that observational surface radiation and/or turbulent energy fluxes are poorly constrained and one hypothesized solution is that both global mean evaporation and precipitation are substantially underestimated (Stephens et al., 2012). Our analysis circumvents this debate by removing global mean quantities, showing that the equator-to-pole gradient of surface energy fluxes is consistent with that inferred from TOA radiation and AHT divergence. This suggests that the meridional structure of surface energy fluxes constrained by TOA radiation and AHT could be used in conjunction with global mean imbalances to give an additional constraint for reconciling which terms in the observed surface energy budget are most uncertain and/or biased.

6 Open Research

All underlying observational data sets are publicly available. The CMIP data analyzed in this study can be found in the Earth System Grid392 Federation (ESGF) repository at `https://esgf-node.llnl.gov/projects/esgf-llnl`. Observational calculations of AHT from the vertical integral of 6 hourly reanalysis data (MERRA, ERA5, NCEP and JRA) are available at `https://atmos.uw.edu/~aaron/cmip_AHT_partition/`.

Acknowledgments

AD, KCA, TC, and DSB acknowledge support from National Science Foundation Award AGS-2019647. RF was supported by the NOAA Climate and Global Change Postdoctoral Fellowship programs for the Advancement of Earth System Science (CPAESS) under award NA18NWS4620043B.

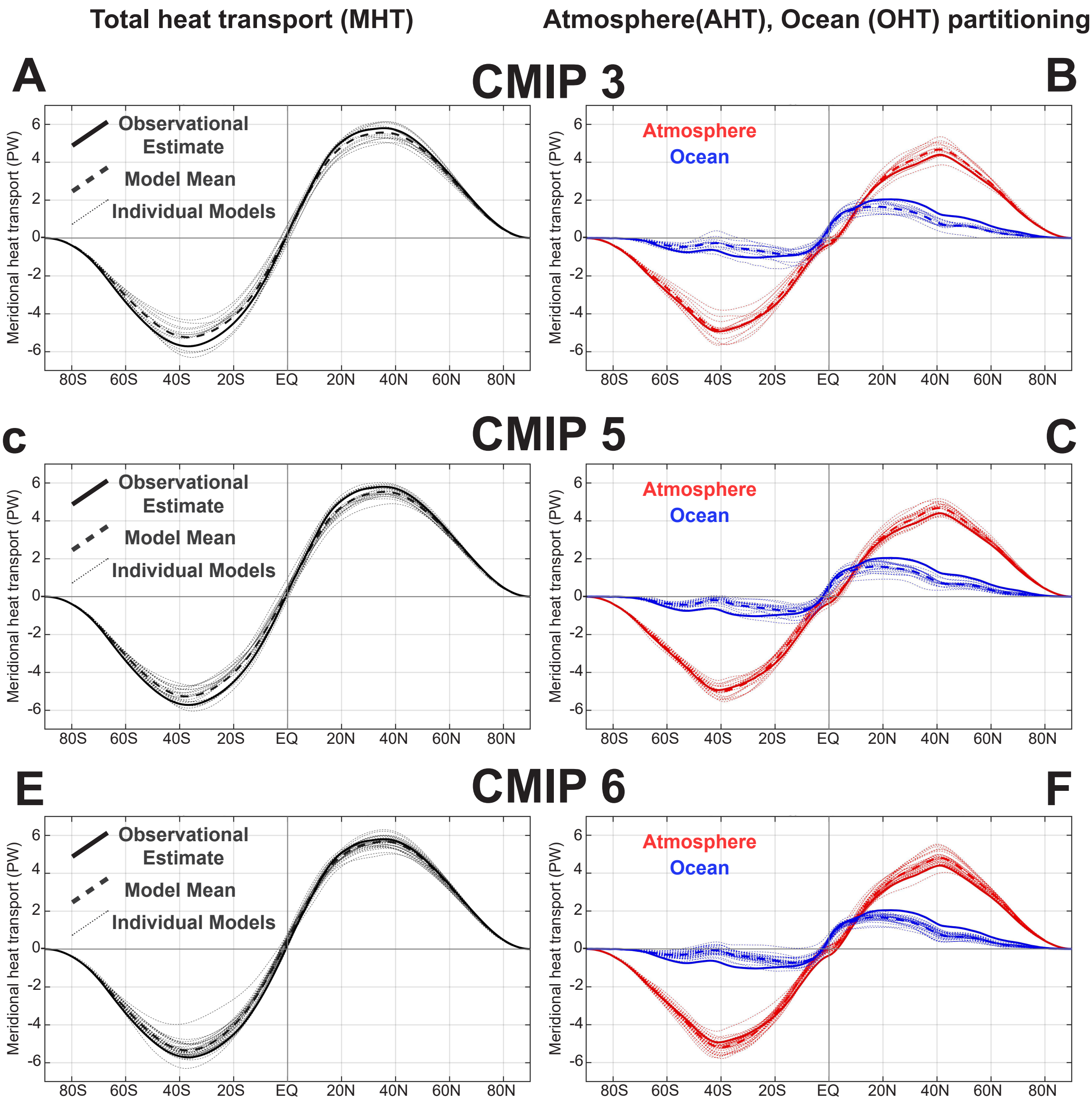

Figure 1. Observational and model (left panels) total meridional heat transport (MHT) and (right panels) its partitioning between the atmosphere (AHT, red) and ocean (OHT, blue). Results from the CMIP3, CMIP5, and CMIP6 models are shown in the top, middle and bottom panels respectively. The observational estimates are shown by the heavy solid line, individual coupled models are shown by the dotted lines and the model ensemble mean is shown by the heavy dashed line.

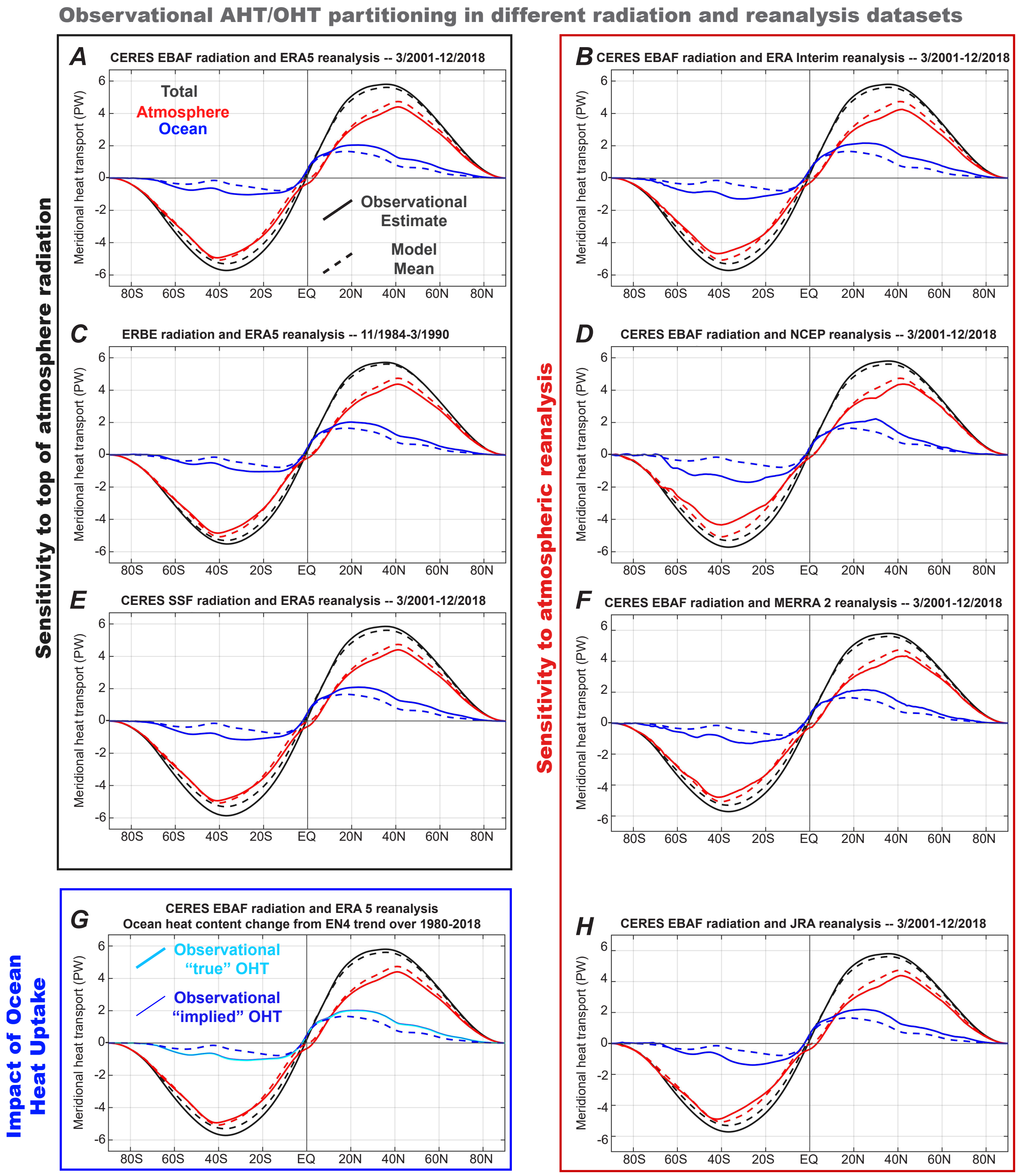

Figure 2. Comparison of MHT, OHT and AHT in models and observations using eight different observational estimates of MHT (black solid), AHT (red), and OHT (blue). The left panels show the sensitivity of the transports to TOA radiation product used with CERES EBAF on the top panel, ERBE in the second panel, and the unadjusted CERES SSF on the bottom and with the ERA5 AHT estimate across all panels. The right panels show the observational transports calculated using CERES EBAF TOA radiation in all panels but using different atmospheric reanalysis products in each panel: (B) ERA Interim; (D) NCEP; (F) MERRA2 and; (H) JRA. Panel (G) shows the impact of observed spatial patterns in ocean heat storage on implied OHT using EN4 ocean heat content changes over 2000-2018. The model mean is the average over all models in CMIP3, CMIP5, and CMIP6 (CMIP-mean).

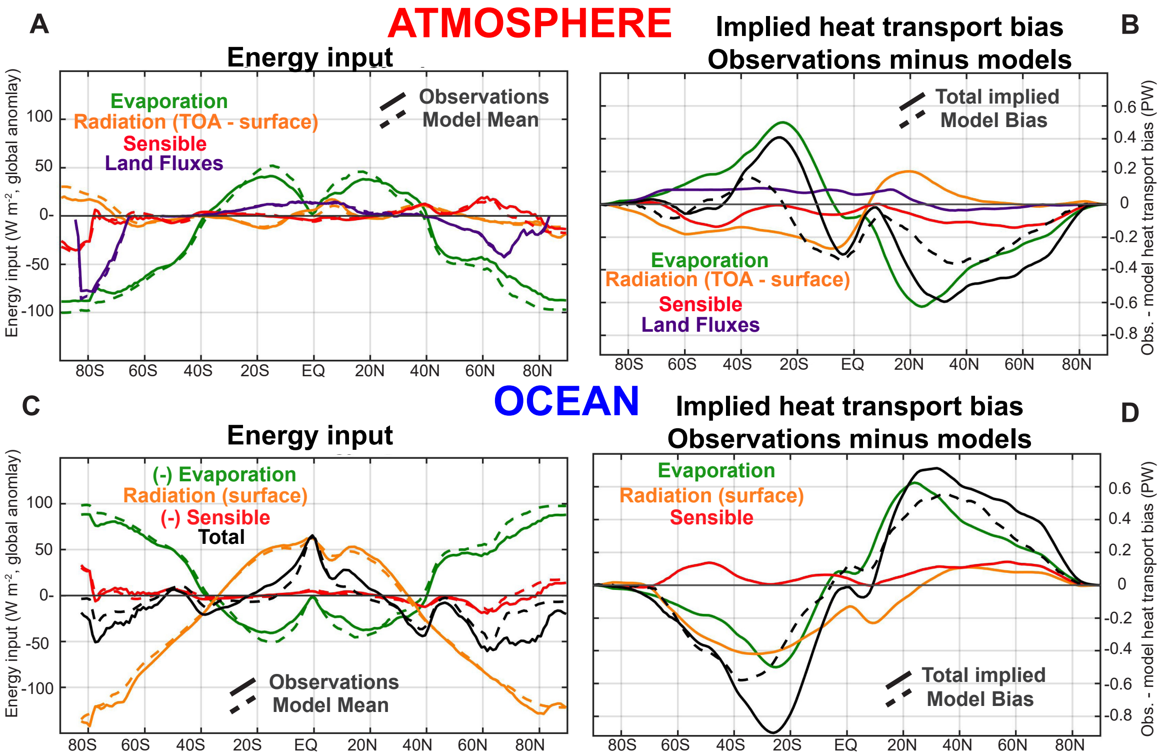

Figure 3. Model and observational estimates of the energy input into the atmosphere and ocean and the implied AHT and OHT biases resulting from each input. (A) Global anomaly energy input into the atmosphere in models (dashed) and observations (solid). See text for definition of terms. (B) Implied AHT bias (observations minus models) due to each energy input. The solid black line shows the sum of all terms. The dashed black line shows the bias in heat transport inferred from CERES and ERA5 data. (C) As in A but for the energy input to the ocean. (D) As in B but for the implied OHT bias.

References

Adler, R., Sapiano, M., Huffman, G., Wang, J., Gu, G., Bolvin, D., … Shin, D. (2018). The global precipitation climatology project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, *9*(4). doi: 10.3390/atmos9040138

Cardinale, C., Rose, B., Lang, A., & Donohoe, A. (2020). Stratospheric and tropospheric flux contributions to the polar cap energy budgets. *J. Climate*, *34*(11), 4261-4278.

Curry, J., Bentamy, A., Bourassa, M., Bourras, D., Bradley, E. F., Brunke, M., … Zeng, X. (2004). Seaflux. *Bull. Amer. Meteor. Soc.*, 490-424. doi: 10.1175/BAMS-85-3-409

Donohoe, A., Armour, K., Roe, G., & Battisti, D. (2020). The partitioning of meridional heat transport from the Last Glacial Maximum to CO_2 quadrupling in coupled climate models. *J. Climate*, *33*(10), 4141-4165.

Donohoe, A., & Battisti, D. (2011). Atmospheric and surface contributions to planetary albedo. *J. Climate*, *24*(16), 4401-4417.

Donohoe, A., & Battisti, D. (2012). What determines meridional heat transport in climate models? *J. Climate*, *25*, 3832-3850.

Donohoe, A., & Battisti, D. (2013). The seasonal cycle of atmospheric heating and temperature. *J. Climate*, *26*(14), 4962-4980.

Enderton, D., & Marshall, J. (2009). Controls on the total dynamical heat transport of the atmosphere and oceans. *J. Atmos. Sci.*, *66*, 1593-1611.

Fajber, R., Donohoe, A., Ragen, S., Armour, K., & Kushner, P. (2023). The hydrologic cycle drive atmospheric heat transport. *Proc. Nat. Acad. Sci.*, *120*(25), e2217202120. doi: 10.1073/pnas.2217202120

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., … Rummukainen, M. (2013). Evaluation of climate models [Book Section]. In T. Stocker et al. (Eds.), *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (p. 659Â–740). Cambridge University Press. doi: 10.1017/CBO9781107415324.018

Good, S. A., Martin, M., & Rayner, N. (2013). En4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res.-Oceans*, *118*, 6704-6716.

Hersbach, H. B. B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz?Sabater, J., Nicolas, J., & Peubey, C. (2020). The era5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*. doi: 10.1002/qj.3803

Holland, M. M., Bitz, C., & Tremblay, B. (2006). Future abrupt reductions in the summer arctic sea ice. *Geophys. Res. Lett.*, *33*(23). doi: 10.1029/2006GL028024

Johnson, G., Lyman, J., & Loeb, N. (2016). Improving estimates of Earth's energy imbalance. *Nat. Clim. Chang.*, *6*(7), 639-640.

Kato, S., & Coauthors. (2018). Surface irradiances of edition 4.0 clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) data product. *J. Climate*, *31*(11), 4501-4527.

Liang, M., Czaja, A., Graversen, R., & Tailleux, R. (2018). Poleward energy transport: is the standard definition physically relevant at all time scales? *Climate Dyn.*, *50*, 1785-1797.

Loeb, N. G., & Coauthors. (2018). Clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition 4.0 data product. *J. Climate*, *31*(2), 895-918.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., … Wong, T. (2009). Towards optimal closure of the Earth's top-of-atmosphere radiation budget. *J. Climate*, *22*, 748-766.

Lucarini, V., & Ragone, F. (2011). Energetics of IPCC4AR4 climate models: energy balance and meridional enthalpy transports. *Rev. Geophys.*, *49*, RG1001.

M., M., Haimberger, L., Edwards, J., & Hyder, P. (2017). Toward consistent diagnostics of the coupled atmosphere and ocean energy budgets. *J. Climate*, *22*(30), 9225-9246.

Mayer, J., Mayer, M., & Haimberger, L. (2021). Consistency and homogeneity of atmopsheric energy, moisture and mass budgets in ERA5. *J. Climate*, *34*(10), 3955-3974.

Oort, A., & Haar, T. V. (1976). On the observed annual cycle in the ocean–atmosphere heat balance over the Northern Hemisphere. *J. Phys. Oceanogr.*, *6*, 781-800.

Pierrehumbert, R. (2010). *Principles of planetary change.* Cambridge University Press.

Seager, R., Battisti, D., Yin, J., Naik, N., Gordon, N., Clement, A., & Cane, M. (2002). Is the gulf stream responsible for europe's mild winters? *Quart. J. Roy. Meteor. Soc.*, *128*, 2563-86.

Stephens, G., Li, J., Wild, M., Clayson, C., Loeb, N., Kato, S., ... Andrews, T. (2012). An update on Earth's energy balance in light of the latest global observations. *Nat. Geo. Sci.*, *5*, 691-696. doi: 10.1038/ngeo15802

Stone, P. (1978). Constraints on dynamical transports of energy on a spherical planet. *Dynam. Atmos. Oceans*, *2*, 123-139.

Trenberth, K. E., & Caron, J. M. (2001). Estimates of meridional atmosphere and ocean heat transports. *J. Climate*, *14*, 3433-3443.

Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bull. Amer. Meteor. Soc.*, *90*(3), 311-324.

Trenberth, K. E., & Stepaniak, D. P. (2003). Co-variability of components of poleward atmospheric energy transports on seasonal and interannual timescales. *J. Climate*, *16*, 3691-3705.

Trenberth, K. E., & Stepaniak, D. P. (2004). The flow of energy through the earth's climate system. *Quart. J. Roy. Meteor. Soc.*, *130*, 2677-2701.

Vonder Haar, T., & Oort, A. (1973). New estimate of annual poleward energy transport by Northern Hemisphere oceans. *J. Phys. Oceanogr.*, *2*, 169-172.

Von Schuckmann, K., Palmer, M., Trenberth, K., Cazenave, A., Chambers, D., Champollion, N., ... Wild, M. (2016). An imperative to monitor earth's energy imbalance. *Nat. Clim. Change.*, *6*, 138-144. doi: 10.1038/nclimate2876

Yu, L., Weller, R., & Sun, B. (2004). Improving latent and sensible heat flux estimates for the Atlantic Ocean (1988-1999) by a synthesis approach. *J. Climate*, *17*, 373-393.

Model biases in the atmosphere-ocean partitioning of poleward heat transport are persistent across three CMIP generations

A. Donohoe[1,3], R. Fajber[2], T. Cox [3], K.C. Armour[3,4], D.S. Battisti [3]and G.H. Roe [4]

[1]Polar Science Center, Applied Physics Laboratory
University of Washington
Seattle, Washington 98195, USA.
[2] Department of Atmospheric and Oceanic Sciences, McGill University
[3] Department of Atmospheric and Oceanic Sciences, University of Washington
[4]School of Oceanography, University of Washington
[5]Department of Earth and Space Sciences, University of Washington

Key Points:

- Climate models simulate too little poleward oceanic heat transport and too much poleward atmospheric heat transport in the extratropics
- Model biases in heat transport partitioning are persistent across model generation and are insensitive to the observational data sets used
- Stronger than observed evaporation in models enhances atmospheric heat transport at the expense of oceanic heat transport

Abstract

The observed partitioning of poleward heat transport between atmospheric and oceanic heat transports (AHT and OHT) is compared to that in coupled climate models. Poleward OHT in the models is biased low in both hemispheres, with the largest biases in the Southern Hemisphere extratropics. Poleward AHT is biased high in the Northern Hemisphere, especially in the vicinity of the peak AHT near 40°N. The significant model biases are persistent across three model generations (CMIP3, CMIP5, CMIP6) and are insensitive to the satellite radiation and atmospheric reanalyses products used to derive observational estimates of AHT and OHT. Model biases in heat transport partitioning are consistent with biases in the spatial structure of energy input to the ocean and atmosphere. Specifically, larger than observed model evaporation in the tropics adds excess energy to the atmosphere that drives enhanced poleward AHT at the expense of weaker OHT.

Plain Language Summary

The equator-to-pole contrast of solar radiation entering the climate system drives the large-scale oceanic and atmospheric circulations that, in turn, move heat from the equator to the poles to moderate latitudinal temperature contrasts. The ocean moves the majority of heat in the tropics whereas the atmosphere moves the vast majority of heat in the mid- and polar-latitudes. We demonstrate that state-of-the-art climate models representing both oceanic and atmospheric circulations systematically simulate too little oceanic heat transport and too much atmospheric heat transport relative to observational estimates. These model biases in the atmosphere-ocean partitioning of poleward heat transport are persistent across three generations of climate model ensembles spanning twenty years of progress in climate modeling and are insensitive to the choice of datasets used to calculate observed heat transports. The model biases are consistent with stronger than observed surface evaporation in the tropics which enhances atmospheric heat transport at the expense of oceanic heat transport.

1 Introduction

The combined meridional heat transport (MHT) by the ocean and atmosphere moderates spatial gradients in temperature on Earth. In the absence of MHT, the equator-to-pole temperature gradient would be approximately three times larger than observed based on radiative considerations alone (Pierrehumbert, 2010), rendering the tropics uninhabitably warm and the high latitudes uninhabitably cold. Observational estimates of the partitioning of MHT between poleward atmospheric heat transport (AHT) and poleward oceanic heat transport (OHT) show that OHT exceeds AHT in the deep tropics (equatorward of 10°) while AHT dominates in the mid- and high-latitudes of both hemispheres (Vonder Haar & Oort, 1973; Oort & Haar, 1976; Trenberth & Caron, 2001; Mayer et al., 2021).

The partitioning of MHT between AHT and OHT impacts climate and its changes. For example, the convergence of OHT in the extratropics is inherently linked to the surface energy budget and thus demands a surface temperature response, whereas the convergence of the same quantity of AHT in the atmosphere can be radiated to space with less impact on surface climate (Cardinale et al., 2020). Indeed, previous work by Enderton & Marshall (2009) has shown that aquaplanets with nearly identical total MHT but different AHT-OHT partitioning can have very different climates (e.g., different surface temperature and sea ice distributions).

Given the dependence of climate on the partitioning between poleward AHT and OHT, we ask here: how well do coupled climate models represent the observed AHT-OHT partitioning? This question was briefly addressed in Chapter 9 of the Intergovernmental Panel on Climate Change 5th assessment report (Flato et al., 2013) which concluded that model OHT was within the wide range of observational OHT estimates. Comparison of observational and model AHT-OHT partitioning is difficult because the standard methodology for partitioning MHT between AHT and OHT differs between observations and models due to the contrasting reliability and availability of the climate fields used to calculate AHT and OHT. Recent work (Donohoe et al., 2020) has demonstrated the near equivalence of the model and observational approaches to AHT-OHT partitioning in a model setting, enabling a comprehensive observational-model comparison. In this study we apply these methods to three generations of coupled model simulations (Phases 3, 5, and 6 of the Coupled Model Intercomparison Project, CMIP) and to several observational radiation and atmospheric reanalysis products. Our aim is to determine whether the models accurately capture the partitioning of AHT and OHT derived from observational datasets.

In Section 2 we provide an overview of the observational and model methodologies for partitioning MHT into AHT and OHT and demonstrate the near equivalence of these two approaches. In section 3, we compare the observational and model MHT partitioning across the three different model generations (CMIP3, CMIP5, and CMIP6) and examine the sensitivity of our findings to the choice of observational data sets used to partition MHT. In Section 4 we consider an alternative method for comparing AHT-OHT partitioning in models and observations from the processes that contribute to spatial gradients in energy input to the atmosphere and ocean. A summary and discussion follows.

2 Methods for partitioning MHT into AHT and OHT in observations and coupled models

The methodology used to partition MHT into AHT and OHT in coupled climate models and observations is described in detail in Donohoe et al. (2020). Here we summarize the conceptual approach.

2.1 Heat transport partitioning in climate models

Near closure of the top of atmosphere (TOA) and surface energy budgets in climate models allows for the energy transport by the atmosphere and ocean across a latitude band to be calculated from the energy input into/out-of the fluid spatially integrated over the polar cap bounded by that latitude:

$$\mathrm{MHT}(\Theta) = 2\pi a^2 \int_{\Theta}^{90} -F^* \cos\theta d\theta, \tag{1}$$

where a is the radius of the Earth, Θ is the latitude (with θ a latitude variable of integration), and F is the net energy input to the atmosphere, ocean, or combined atmosphere-ocean system. The total MHT can be found by taking F to be the radiative flux at the TOA ($\mathrm{RAD_{TOA}}$), OHT by taking F to be the net surface heat flux (SHF = radiative plus turbulent flux into the ocean), and AHT by setting F to be the net energy input to the atmosphere ($\mathrm{RAD_{TOA}}$ - SHF). The * denotes that the global mean of each energy flux term has been removed to ensure heat transport goes to zero at both poles. This adjustment is necessary because climate models do not conserve energy globally (≈ 1 W m^{-2} imbalances) in both the atmosphere and ocean (Lucarini & Ragone, 2011).

2.2 Heat transport partitioning in observations

In contrast to coupled climate models where the surface energy budget is (nearly) closed, the sparsity and uncertainty of observational surface radiative and turbulent energy flux measurements results in an unrealistically large (>10 W m^{-2}) global mean surface energy imbalance (Stephens et al., 2012; Trenberth et al., 2009), which prohibits the evaluation of OHT from the net surface heat flux. Instead, we use a conceptual approach following Vonder Haar & Oort (1973) and Trenberth & Caron (2001): MHT is calculated using Eq. 1 with satellite $\mathrm{RAD_{TOA}}$ (Loeb & Coauthors, 2018); AHT is calculated from the time average of the vertically and zonally integrated meridional energy flux in the atmosphere derived from high frequency (6 hourly) atmospheric reanalysis; OHT is then calculated as the residual of satellite derived MHT and reanalysis derived AHT. In the AHT calculation, a vertically integrated moist static energy anomaly is removed before integrating (Donohoe & Battisti, 2013; Cardinale et al., 2020; Donohoe et al., 2020), effectively applying a mass correction needed to make the AHT calculation physically meaningful (Trenberth & Stepaniak, 2003; Liang et al., 2018).

To show that the "observational" and "model" methods are comparable, we partition MHT into AHT and OHT using both methods in a NCAR CESM1 coupled pre-industrial control simulation (see Donohoe et al., 2020, for details). The two approaches give nearly identical partitioning of MHT into AHT and OHT (cf. the dashed and solid red and blue lines in Supporting Information Fig. S1) with a root mean squared difference AHT (and OHT) between the two methods of 0.07 PW. The close correspondence of the two calculations of MHT partitioning suggests that the "observational" and "model" approaches we use here to partition MHT are directly comparable. We use this result to justify the examination of potential model biases in MHT partitioning using these two methodologies.

3 Results: model biases in MHT partitioning

Climate model biases in MHT partitioning are analyzed using pre-industrial control simulations from three different CMIP generations and several different sets of observational products (see Supporting Information for details). The presentation of our results is organized as follows. Section 3.1 presents the observational estimate of MHT partitioning using the most contemporary and high resolution data available, which is compared against the MHT partitioning in the three CMIP ensembles. Section 3.2 analyzes the sensitivity of our results to the observational data used by comparing eight different observational estimates of MHT partitioning against the multi-generation CMIP ensemble mean. The results show that the sign and spatial structure of model biases in MHT partitioning are consistent across model generation and observational data sets used.

3.1 Consistent model biases in AHT-OHT partitioning across three CMIP generations

In this section, we use CERES Energy Balanced and Filled (EBAF) TOA radiation (Loeb et al., 2009) and the ERA5 atmospheric reanalysis (Hersbach et al., 2020) to calculate an observational estimate of MHT and its partitioning over the period 2001-2020. This observational estimate (solid line) is compared against each of the three CMIP ensembles (in each row of Fig. 1; with dashed lines showing individual models and the thick dashed lines showing CMIP ensemble averages).

Poleward MHT peaks near 35° in both hemispheres in both models and observations (Fig. 1), consistent with constraints due to Earth-Sun geometry whereby the meridional distribution of net TOA radiation ($\mathrm{RAD^*_{TOA}}$) is dominated by the second order Legendre polynomial (equator-to-pole scale) as discussed by Stone (1978). However, across all three CMIP generations, the amplitude of poleward MHT in models is biased low in the mid-latitudes of both hemispheres relative to the observational estimate. In the Southern Hemisphere (SH), the observational estimate of maximum poleward MHT is 5.7 PW, which is significantly larger (95% confidence interval of t-test) than the ensemble means of CMIP3 (5.2 PW), CMIP5 (5.3 PW), and CMIP6 (5.4 PW). In the Northern Hemisphere (NH) the observational estimate of maximum poleward MHT is 5.8 PW, exceeding the ensemble mean of CMIP3 (5.6 PW), CMIP5 (5.5 PW), and CMIP6 (5.7 PW), but only for CMIP5 is the difference statistically significant. In the SH, the inter-model spread in peak MHT (2 standard deviations) is as large as 23% of the ensemble mean and has values of 1.2 PW in CMIP3, 0.8 PW in CMIP5, and 0.8 PW in CMIP6. The inter-model spread in peak NH MHT is smaller than its SH counterpart with values of 0.8 PW in CMIP3, 0.6 PW in CMIP5, and 0.6 PW in CMIP6. Donohoe & Battisti (2011) demonstrated that the the inter-model spread and bias in MHT in CMIP3 results from biases and spread in the albedo of clouds which impact the equator-to-pole gradient of absorbed solar radiation. The bias and spread in MHT is only slightly reduced in CMIP5 and CMIP6, and also results primarily from model differences in mean-state shortwave cloud radiative effects (not shown).

We next analyze the partitioning of MHT between OHT and AHT. In the NH, the model ensemble mean is significantly biased toward too little poleward OHT and too much poleward AHT in all three CMIP generations. The observational estimate of peak NH AHT is 4.4 PW as compared to 4.7 $\pm$ 0.2 PW in CMIP3, 4.7 $\pm$ 0.1 PW in CMIP5, and 4.8 $\pm$ 0.1 PW in CMIP6 where the stated uncertainty is two standard deviations of the ensemble mean. The peak in NH OHT is robustly equatorward of the peak AHT, but has significantly larger values for the observational estimate (2.0 PW) than in the model ensemble means (1.7 PW in CMIP3, 1.8 PW in CMIP5, and 1.7 PW in CMIP6). The model bias toward smaller than observed OHT extends poleward to the Arctic where OHT has been demonstrated to have large impacts on sea ice extent (Holland et al., 2006; Seager et al., 2002).

In the SH, poleward OHT in the models is biased low relative to the observational estimate in all three CMIP generations. The largest biases in OHT are found the vicinity of 40°S where the observational OHT is -0.7 PW compared to the ensemble mean OHT at that latitude is -0.3 $\pm$ 0.2 PW in CMIP3, -0.2 $\pm$ 0.1 PW in CMIP5, and -0.1 $\pm$ 0.1 PW in CMIP6. The observational estimate of poleward OHT is only exceeded in three model simulations (two in CMIP3 and one in CMIP5). In contrast, the poleward AHT in the SH is not significantly different between the models and observational estimates.

These results suggest that in the SH, the majority of the model biases in MHT are a result in biases in OHT, whereas in the NH the models generally simulate too much poleward AHT and too little poleward OHT. Alternatively, the fractional contribution of AHT-OHT to total MHT (i.e., normalizing each model by the model specific MHT) is biased toward too much poleward AHT and too little poleward OHT with biases that are nearly hemispherically symmetric between the two hemispheres (not shown). Importantly, the sign and spatial structure of model biases in MHT and AHT-OHT partitioning are remarkably consistent across the three CMIP generations spanning over 20 years of progress in climate modeling.

3.2 Sensitivity of results to observational data sets used

We next consider whether the identified model biases in AHT-OHT partitioning are sensitive to the choice of observational data sets (TOA radiation and atmospheric reanalysis) used to partition MHT. We use the mean of all ensemble members across all three CMIP generations, referred to as the CMIP-mean, as a reference for all analyses in this subsection.

We begin by analyzing the MHT and AHT/OHT partitioning estimated using two additional satellite-derived observational estimates of TOA radiation (see Supporting Information for details): the unadjusted CERES single scanner footprint (SSF) data and the ERBE satellite data which spans the 1984-1990 (left panels of Fig. 2 bordered by the black box). In these three panels, the choice of TOA radiation product alters the calculated observational MHT (solid black line) whereas the AHT is unchanged between panels (ERA5 is used in each). Because the observational OHT is calculated from the difference of MHT and AHT, the observational OHT estimate (solid blue line) also varies between panels. Observational MHT calculated from the three different TOA radiation products is consistently larger than the CMIP-mean in both hemispheres. Model biases in MHT are largest when the CERES SSF product is used (Fig. 2E) and smallest when the ERBE product is used (Fig. 2C). The CMIP-mean OHT is biased low compared to that derived from all three TOA radiation datasets with largest magnitude biases when CERES SSF is used, especially in the SH. Model biases in AHT/OHT partitioning are insensitive to observational TOA radiation data set used which give a consistent estimate of MHT despite their substantial (≈ 5 W m^{-2} = 2.5 PW globally integrated) differences in global mean TOA radiative balance associated with absolute calibration uncertainty (Loeb et al., 2009).

We next analyze the sensitivity of our results to the choice of atmospheric reanalysis used to calculate the AHT (Fig. 2 panels A, B, D, F and H). In these five panels, the MHT is identical (calculated using the CERES EBAF product) whereas the AHT is calculated from the ERA5, ERA-interim, NCEP, MERRA2, and JRA reanalyses. Since OHT is calculated from the residual of MHT and AHT, the OHT difference between the three panels are equal and opposite to the the inter-panel differences in AHT. The CMIP-mean bias toward too much poleward AHT and too little poleward OHT is apparent using all five observational estimates of AHT. Poleward AHT is largest when using ERA5 followed closely by JRA, MERRA2 and then ERA interim, whereas using NCEP produces the smallest poleward AHT with the most notable difference near the peak in the SH at 40°S. Therefore, model biases in the AHT-OHT partitioning are smallest using ERA5 and largest using NCEP. These results suggest that the sign and spatial structure of model biases in MHT partitioning are consistent across atmospheric reanalysis datasets, whereas the magnitude of the bias depends on the reanalyses product used. Differences in AHT calculated between the different reanalyses are not impacted by differences in the spatial resolution (see analysis and Fig. S2 in the Supporting Information) as even the coarsest product (NCEP) resolves the spatial scales responsible for the vast majority of AHT.

Finally, we evaluate whether heat storage due to the transient response to anthropogenic forcing impacts our observational estimates of OHT. The Earth is not in equilibrium but, rather, is accumulating energy at an average rate of 0.7 W m^{-2} globally (Johnson et al., 2016). The vast majority of this energy accumulation is stored in the ocean (Von Schuckmann et al., 2016) and it is possible that the spatial structure of this energy storage projects onto our diagnoses of observational OHT for the following reason: observed 'implied' OHT is calculated from the spatial integral of inferred surface heat fluxes (TOA radiation plus AHT convergence) and the latter is balanced by the sum of OHT divergence and ocean heat storage in a transient system. We diagnose the impact of observed ocean heat storage on the implied OHT ($\mathrm{OHT}_{\mathrm{STORAGE}}$) from the trend in ocean heat content, derived from UK Hadley Center EN4 objective ocean analysis (Good et al., 2013) over the CERES period (see Supporting Information for details). $\mathrm{OHT}_{\mathrm{STORAGE}}$ is removed from the 'implied' OHT to estimate the 'true' OHT (solid teal line in Fig. 1F) that must be transported laterally in the ocean to close the ocean energy budget. $\mathrm{OHT}_{\mathrm{STORAGE}}$ is very small (< 0.1 PW in magnitude) and, thus, the diagnosed 'true' OHT is visually indistinguishable from the observational 'implied' OHT (solid blue line in Fig. 1F). The global mean ocean heat uptake of 0.7 W m^{-2} translates to 0.4 PW of global energy input to the ocean but the implied OHT of ocean heat storage is significantly smaller in magnitudes due to ocean heat uptake being more globally uniform than regionally isolated. The negligible impact of ocean heat storage on 'implied' OHT over the historical period is consistent with the small (< 0.1 PW) differences between OHT in the ensemble mean of historical CMIP5 simulations averaged over the 2000-2018 time period as compared the pre-industrial control simulations using the same models (Supporting Information Fig. S3).

Collectively, these results suggest that the sign of model biases in AHT-OHT partitioning is robust to different observational products (satellite TOA radiation and atmospheric reanalysis) used to partition MHT. Additionally, the spatial pattern of transient heat uptake by the ocean makes a negligible impact on estimated OHT. However, the magnitude of the model bias in AHT-OHT partitioning does vary with observational datasets used. In this regard, the use of CERES EBAF and ERA5 data for our primary analysis (Fig. 1) is a conservative estimate of model biases in AHT-OHT partitioning (a smaller OHT bias is found only when using the combination of ERBE and ERA5 products).

4 Biases in energy input to the atmosphere and ocean and inferred AHT and OHT biases

Here we evaluate potential causes of the persistent model biases in AHT and OHT in terms of model biases in the spatial structure of energy input into the ocean and atmosphere. Starting in the ocean, energy conservation demands that OHT across a latitude band balances the net surface heat flux *out* of the ocean (-SHF by our sign convention) integrated over the polar cap bounded by that latitude, which from Eq. 1 is represented by:

$$\mathrm{OHT}(\Theta) = 2\pi a^2 \int_{\Theta}^{90} (-\mathrm{SHF}^*)\, cos(\theta) d\theta. \tag{2}$$

SHF is equal to the net downward surface radiation ($\mathrm{RAD_{SURF}}$) into the ocean minus the upward turbulent energy fluxes of sensible (SENS) and latent heat ($\mathrm{L_vE}$):

$$\mathrm{SHF} = \mathrm{RAD_{SURF}} - \mathrm{SENS} - \mathrm{L_vE}. \tag{3}$$

Substitution of Eq. (3) into Eq. (2) allows the OHT to be decomposed into the implied transports of each term contributing to SHF:

$$\mathrm{OHT} = \mathrm{OHT_{RAD,SURF}} + \mathrm{OHT_{SENS}} + \mathrm{OHT_E}, \tag{4}$$

where, for example, the OHT implied by evaporation ($\mathrm{OHT_E}$) is:

$$\mathrm{OHT}(\Theta)_\mathrm{E} = 2\pi a^2 \int_{\Theta}^{90} \mathrm{L_vE}^* cos(\theta) d\theta, \tag{5}$$

where, as in Eq. 1 and 2, the * indicates that the global (ocean domain) mean has been removed from the term. Because SENS^* is small compared to the other terms (Fig. 3C) and $\mathrm{RAD_{SURF}}$ is dominated by solar input to the surface (Supporting Information Figs. S5E,F), the predominant energy balance in this framework can be summarized as follows: the magnitude of OHT (black line in Fig. 3D) is governed by the imbalance between excess (relative to the global mean) solar radiation entering the tropical ocean (orange line) and excess evaporative loss (green). Perfect local compensation between surface solar input and evaporation implies zero OHT whereas weaker evaporative loss demands a larger fraction of solar input be realized as OHT. We use this framework to understand model biases in OHT in terms of biases in the meridional structure of terms contributing to SHF.

The latitudinal structure of CMIP-mean $\mathrm{L_vE}$, SENS and $\mathrm{SURF_{RAD}}$ over the ocean domain is compared to observational estimates of the same quantities with $\mathrm{L_vE}$ and SENS taken from the WHOI Objectively Analyzed (OA) Air-Sea Flux product (Yu et al., 2004) and $\mathrm{SURF_{RAD}}$ estimates from the CERES EBAF surface product (Kato & Coauthors,

2018) in Fig. 3C. Evaporation is biased high in models (relative to the observational estimate) at all latitudes except the Arctic (Supporting Information Fig. S4). Evaporation biases are largest (> 20 W m^{-2}) in the subtropics of both hemispheres and are much smaller in the high latitudes. These evaporation biases manifest as enhanced subtropical ocean energy loss by E^* in the models (cf. the dashed and solid green lines in Fig. 3C) and an implied model bias toward too little (by approximately 0.4 PW) poleward OHT due to evaporation in each hemisphere ($\mathrm{OHT_E}$, green line in Fig. 3D). Thus, evaporation biases alone explain the majority of the model bias in OHT identified in Section 3 (compare green and dashed black lines in Fig. 3D).

The observational $\mathrm{RAD^*_{SURF}}$ has a stronger equator-to-pole gradient than that in climate models (cf. the solid and dashed orange lines in Fig. 3C) especially in the SH. Model biases in $\mathrm{RAD^*_{SURF}}$ are associated with larger than observed downwelling solar radiation into the extratropical Southern Ocean (Supporting Information Fig. S5E) due to clouds that are optically thinner than observed (Donohoe & Battisti, 2012). As a result, observed poleward $\mathrm{OHT_{RAD,SURF}}$ is larger than that in models with larger magnitude (0.4 PW) biases in the SH. The model biases in $\mathrm{OHT_{RAD,SURF}}$ mirror the impact of TOA radiation biases on MHT (left panels of Figure 1) including the partitioning between shortwave and longwave biases within each hemisphere, suggesting that model biases in MHT and OHT in the SH are due to biases in shortwave absorption whereas those in the NH are due to biases in OLR and net surface longwave (Supporting Information Figs. S5B,F).

The sum of model biases in $\mathrm{OHT_E}$, $\mathrm{OHT_{RAD,SURF}}$ and $\mathrm{OHT_{SENS}}$ (solid black line in Fig. 3D) finds that models would have weaker than observed poleward OHT of 0.6 PW in the NH and 0.8 PW in the SH based on biases in energy input to the ocean. This overall inferred OHT bias is primarily due to a nearly hemispherically mirror-imaged bias in $\mathrm{OHT_E}$ which is enhanced by poleward $\mathrm{OHT_{RAD,SURF}}$ in the SH. The bias in OHT inferred from surface flux biases matches the spatial structure but exceeds in magnitude the OHT biases calculated in Section 3 from TOA radiation and atmospheric reanalysis (dashed black line in Fig. 3D). These two calculations of model OHT biases do not have to match as they use different conceptual approaches and rely on completely independent observational climate fields. Nonetheless, the consistency of the sign, spatial pattern, and magnitude of the OHT biases calculated using the two different approaches suggest that the model biases in surface energy fluxes are large enough to account for the AHT-OHT partitioning biases inferred from the residual TOA radiation and AHT estimates.

We use a similar calculation of the model biases in implied AHT from the spatial structure of energy input to the atmosphere to compute an alternative estimate of AHT biases to those calculated in Section 3. The AHT analog to Eq. 4 is:

$$\mathrm{AHT} = \mathrm{AHT_{RAD,ATMOS}} + \mathrm{AHT_{SENS}} + \mathrm{AHT_E}, \tag{6}$$

where the atmospheric analog to Eq. 5 for the AHT due to evaporation ($\mathrm{AHT_E}$) is:

$$\mathrm{AHT}(\Theta)_E = 2\pi a^2 \int_{\Theta}^{90} -\mathrm{L_v E^*} cos(\theta) d\theta. \tag{7}$$

The spatial integral is over a global (land plus ocean) domain. Here $\mathrm{RAD_{ATMOS}}$ is the net radiative heating of the atmospheric column which is equivalent to the net radiation at TOA minus $\mathrm{RAD_{SURF}}$. Fajber et al. (2023) demonstrated that poleward AHT is primarily determined by evaporation ($\mathrm{AHT} \approx \mathrm{AHT_E}$) because $\mathrm{L_v E^*}$ dominates the spatial structure of energy input to the atmosphere. We note that $\mathrm{L_v E^*}$ spatially integrated over the ocean domain has opposing impacts on $\mathrm{AHT_E}$ versus $\mathrm{OHT_E}$ (and likewise for $\mathrm{SENS^*}$ and $\mathrm{AHT_{SENS}}$ versus $\mathrm{OHT_{SENS}}$). This arises because excess evaporation over the low latitudes ($E^* > 0$) adds energy to the atmosphere to enhance the demand for poleward AHT at the expense of removing energy from the low-latitude ocean to reduce the demand for poleward OHT.

To more clearly see the compensation between biases in AHT-OHT due to model biases in L_vE^* (and SENS*) over the ocean domain, we take the following approach to compare models and observations of AHT via Eqs. 6 and 7. First, AHT_E and AHT_{SENS} are calculated from the observational WHOI OA evaporation and sensible heat flux data *over the ocean domain only*, and are compared to analogous model calculations over the ocean domain. Then, the contribution of turbulent energy fluxes over land to the combined AHT_E and AHT_{SENS} is estimated from the CERES EBAF net surface radiation spatially integrated over land. This approach assumes that (via surface energy balance) surface radiative gain is balanced by turbulent loss. These calculations are compared to analogous calculations in the models. Finally, $AHT_{RAD,ATMOS}$ is calculated from the CERES EBAF TOA and surface data *over the global domain* and is compared to the analogous global domain calculation in models (orange lines in Fig. 3A,B). This strategy circumvents the lack of reliable observational estimates of turbulent energy fluxes over land – instead inferring them from a like-with-like observational-to-model comparison of surface radiation over land and assuming that RAD_{SURF} is balanced by upward turbulent fluxes from the land to the atmosphere (the latter assumption has been validated in models).

Model biases in AHT_E compose the vast majority of AHT biases diagnosed from Eq. 6 (cf. the green and solid black lines in Fig. 3B) and suggest that the stronger than observed poleward AHT in models is driven by an enhanced equator-to-pole gradient in evaporation. Model RAD^*_{ATMOS} is more negative in the deep tropics as compared to observations (due to stronger longwave cooling in the models– Supporting Information Fig. S5) which contributes to smaller $AHT_{RAD,ATMOS}$ export from the tropics in the models that generally opposes the low latitude biases in AHT_E (orange line in Fig. 3B). Interestingly, shortwave absorption in the atmosphere is biased low in the models, which reduces the demand for poleward AHT by nearly 0.4 PW in both hemispheres (red line in Supporting Information Fig. S5D). However, this model deficit in atmospheric heating of the tropics is nearly compensated for by weaker than observed longwave cooling of the atmosphere such that there is almost no bias in $AHT_{RAD,ATMOS}$ at the equator-to-pole scale. Turbulent energy fluxes over the land inferred from net surface radiation are nearly identical in models and observations and make a negligible impact on AHT biases (cf. purple dashed and solid lines in Fig. 3C,D).

These calculations demonstrate that the model biases in the partitioning of poleward heat transport between AHT and OHT that were inferred in Section 3 are consistent (in sign, spatial structure, and magnitude) with the model biases in energy input into the atmosphere and ocean by radiative fluxes and turbulent exchange between the atmosphere and ocean. Stronger than observed evaporation in the models contributes to enhanced poleward AHT at the expense of reduced OHT that is nearly hemispherically symmetric whereas radiative biases due to thinner than observed clouds in the extratropical Southern Ocean results in too weak poleward MHT that is primarily manifested in the surface energy budget and implied OHT bias.

5 Summary and discussion

Coupled climate models have too little poleward OHT in both hemispheres and too much AHT in the NH, compared to observational estimates. These model biases are remarkably consistent across three generations of coupled model ensembles (CMIP3, CMIP5, and CMIP6) and across different sets of observational TOA radiation and atmospheric reanalysis data. These conclusions are not impacted by observed transient energy accumulation in the ocean.

The method used here to balance the mass budget of the atmospheric reanalysis differs from that used in the work of Trenberth & Stepaniak (2004) and M. et al. (2017). Specifically, we implicitly assume zero net atmospheric mass flux through a given latitude circle whereas other works adjust the mass flux to balance the polar cap spatial integral of the surface pressure tendency and evaporation minus precipitation. Our choice stems from defining the energy budget with respect to a fixed mass of atmosphere (Donohoe & Battisti, 2013; Liang et al., 2018). The AHT associated with the mass flux due to evaporation minus precipitation is primarily compensated for a return flow of mass and energy in the ocean and requires a consistent treatment of the energy fluxes through the atmosphere, surface and ocean (M. et al., 2017) that depends on the choice of zero point energy (e.g., the units used for temperature). Physically, a poleward (water) mass flux in the atmosphere is balanced by the mass flux of precipitation minus evaporation and an equivalent equatorward mass flux in the ocean. The energy flux of each of these mass fluxes is the product of mass flux and mean energy of the fluid, has a minimal net (AHT+OHT) poleward energy transport but is of order 0.2 PW in magnitude for each the compensating AHT and OHT. The standard definition of SHF in climate models does not include the sensible heat of this net (water) mass flux across the air/sea interface and we believe including this term would create an inconsistency between the model derived and observationally inferred OHT. Our interpretation is supported by the near equivalence of the AHT calculated in CESM via the "observational" and "model" partitioning calculations using our method of calculating AHT from reanalysis data whereas including the net mass flux of water in the AHT creates a substantial mismatch between the two calculations (not shown). We emphasize that all choices made here were aimed at creating a consistent way to compare observational and model MHT and AHT-OHT partitioning despite the different climate fields that go into each calculation.

This work focused on model biases in the vertical zonal and time integral of atmospheric moist static energy fluxes that comprise AHT without regard for biases in the underlying atmospheric circulations and associated temperature and humidity structures of the atmosphere. Donohoe et al. (2020) demonstrated that model biases in poleward AHT primarily result from larger than observed dry (sensible) heat transport by transient eddies in the mid-latitudes of both hemisphere (their Fig. 4D) and in the NH smaller than observed dry heat transport by stationary eddies; the moisture (latent heat) transport has negligible biases. Model biases in evaporation are expected to be manifested as biases in both moist and dry AHT because dry AHT is set by the spatial pattern of condensational heating of the atmosphere which represents the portion of $\mathrm{AHT_E}$ that is not transported poleward as latent heat (Fajber et al., 2023); while spatial patterns of evaporation directly demand poleward moist AHT, the energy input to the atmosphere via evaporation is handed off to dry AHT where precipitation forms and the atmosphere is heated condensationally. Therefore, our finding that model biases toward too much AHT result from stronger than observed evaporation is consistent with the finding that excess poleward AHT in the models is expressed as a bias toward too much dry heat transport.

Remarkably, the model OHT bias inferred from observational estimates from satellite TOA radiation and atmospheric reanalyses is in descent agreement with model biases in the energy exchange between the ocean and atmosphere calculated from independent observational estimates of surface heat fluxes. The latter bias is due primarily to stronger than observed low-latitude evaporation in the models. We note that the community has been reluctant to diagnose OHT from the observed surface energy balance because of uncertainty in the turbulent energy fluxes. Yet, our analysis paints a consistent picture of the model biases in turbulent energy fluxes – whether these are inferred from the residual of TOA radiation and AHT or from bulk formula. We also note that observational estimates of global mean evaporation and its equator-to-pole gradient vary substantially (Stephens et al., 2012) with reanalysis products generally having more evaporation than the bulk formula based estimates such as WHOI OA flux (Yu et al., 2004) and SEAFLUX (Curry et al., 2004). We chose to use WHOI OA flux for the analysis in Section 4 because the bulk formula in this product are optimized to match buoy observations – making it the most observationally constrained estimate of evaporation. Additionally, the global constraint of evaporation balancing precipitation is nearly satisfied from the combination of the WHOI OA FLUX evaporation over the ocean (62.8 W m^{-2} contribution to global mean) plus the ERA5 reanalysis evaporation over land (12.9 W m^{-2} for a global total evaporation of 75.7 W m^{-2}) nearly balancing the best observational estimate of global mean precipitation (77.9 W m^{-2}) from the NOAA GPCP (Adler et al., 2018). The lack of closure of the observed global mean surface energy budget suggests that observational surface radiation and/or turbulent energy fluxes are poorly constrained and one hypothesized solution is that both global mean evaporation and precipitation are substantially underestimated (Stephens et al., 2012). Our analysis circumvents this debate by removing global mean quantities, showing that the equator-to-pole gradient of surface energy fluxes is consistent with that inferred from TOA radiation and AHT divergence. This suggests that the meridional structure of surface energy fluxes constrained by TOA radiation and AHT could be used in conjunction with global mean imbalances to give an additional constraint for reconciling which terms in the observed surface energy budget are most uncertain and/or biased.

6 Open Research

All underlying observational data sets are publicly available. The CMIP data analyzed in this study can be found in the Earth System Grid392 Federation (ESGF) repository at `https://esgf-node.llnl.gov/projects/esgf-llnl`. Observational calculations of AHT from the vertical integral of 6 hourly reanalysis data (MERRA, ERA5, NCEP and JRA) are available at `https://atmos.uw.edu/~aaron/cmip_AHT_partition/`.

Acknowledgments

AD, KCA, TC, and DSB acknowledge support from National Science Foundation Award AGS-2019647. RF was supported by the NOAA Climate and Global Change Postdoctoral Fellowship programs for the Advancement of Earth System Science (CPAESS) under award NA18NWS4620043B.

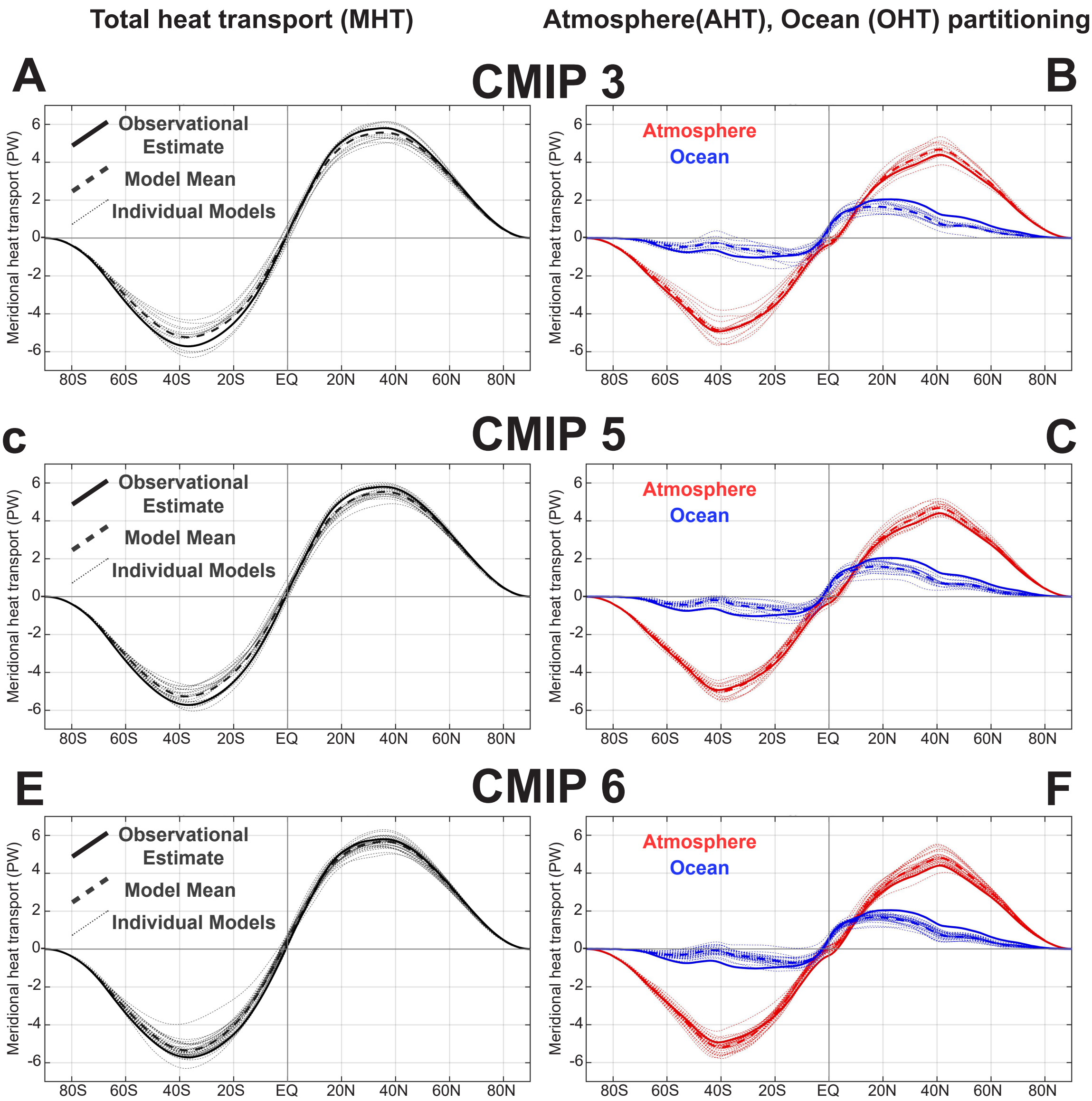

Figure 1. Observational and model (left panels) total meridional heat transport (MHT) and (right panels) its partitioning between the atmosphere (AHT, red) and ocean (OHT, blue). Results from the CMIP3, CMIP5, and CMIP6 models are shown in the top, middle and bottom panels respectively. The observational estimates are shown by the heavy solid line, individual coupled models are shown by the dotted lines and the model ensemble mean is shown by the heavy dashed line.

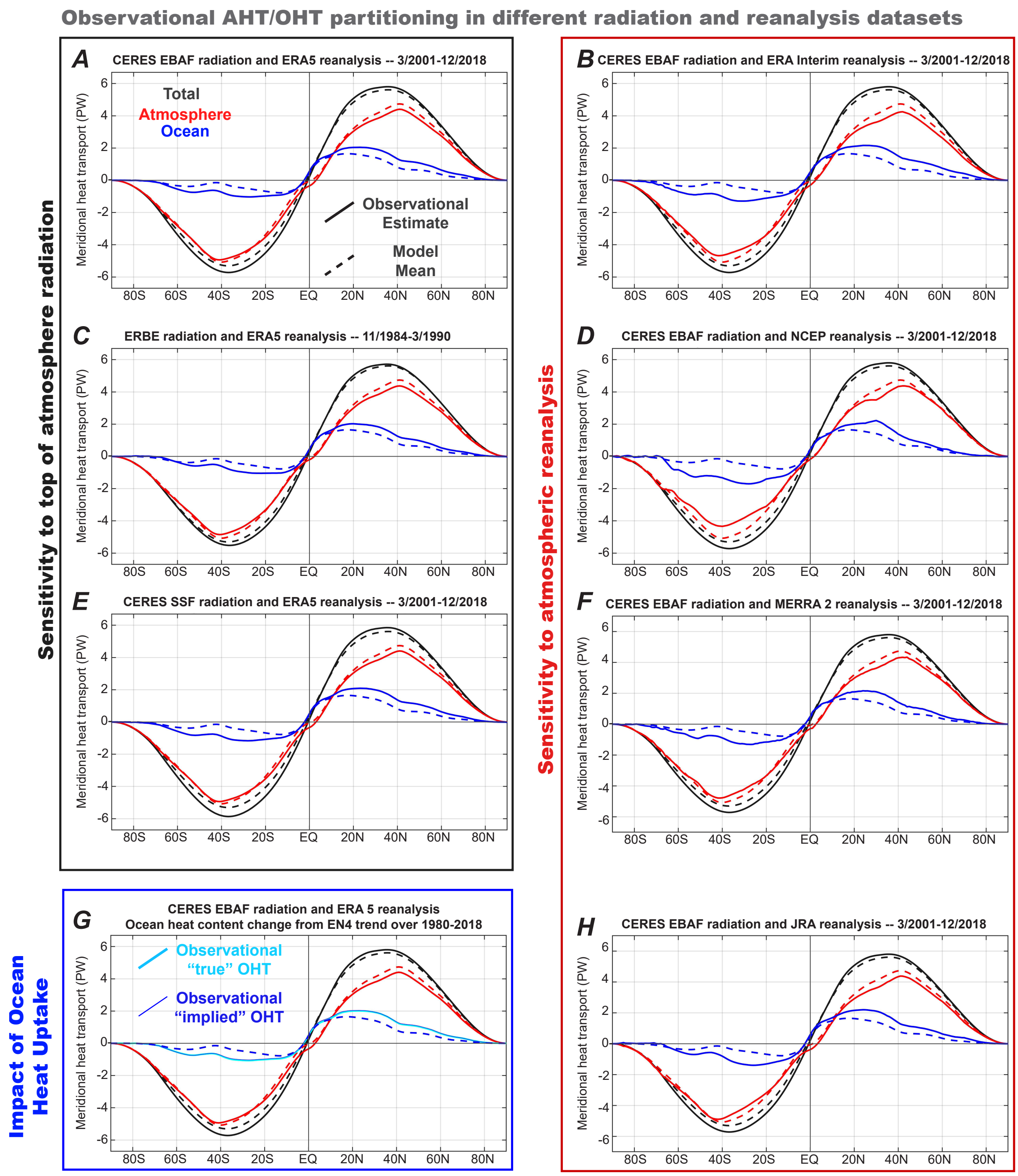

Figure 2. Comparison of MHT, OHT and AHT in models and observations using eight different observational estimates of MHT (black solid), AHT (red), and OHT (blue). The left panels show the sensitivity of the transports to TOA radiation product used with CERES EBAF on the top panel, ERBE in the second panel, and the unadjusted CERES SSF on the bottom and with the ERA5 AHT estimate across all panels. The right panels show the observational transports calculated using CERES EBAF TOA radiation in all panels but using different atmospheric reanalysis products in each panel: (B) ERA Interim; (D) NCEP; (F) MERRA2 and; (H) JRA. Panel (G) shows the impact of observed spatial patterns in ocean heat storage on implied OHT using EN4 ocean heat content changes over 2000-2018. The model mean is the average over all models in CMIP3, CMIP5, and CMIP6 (CMIP-mean).

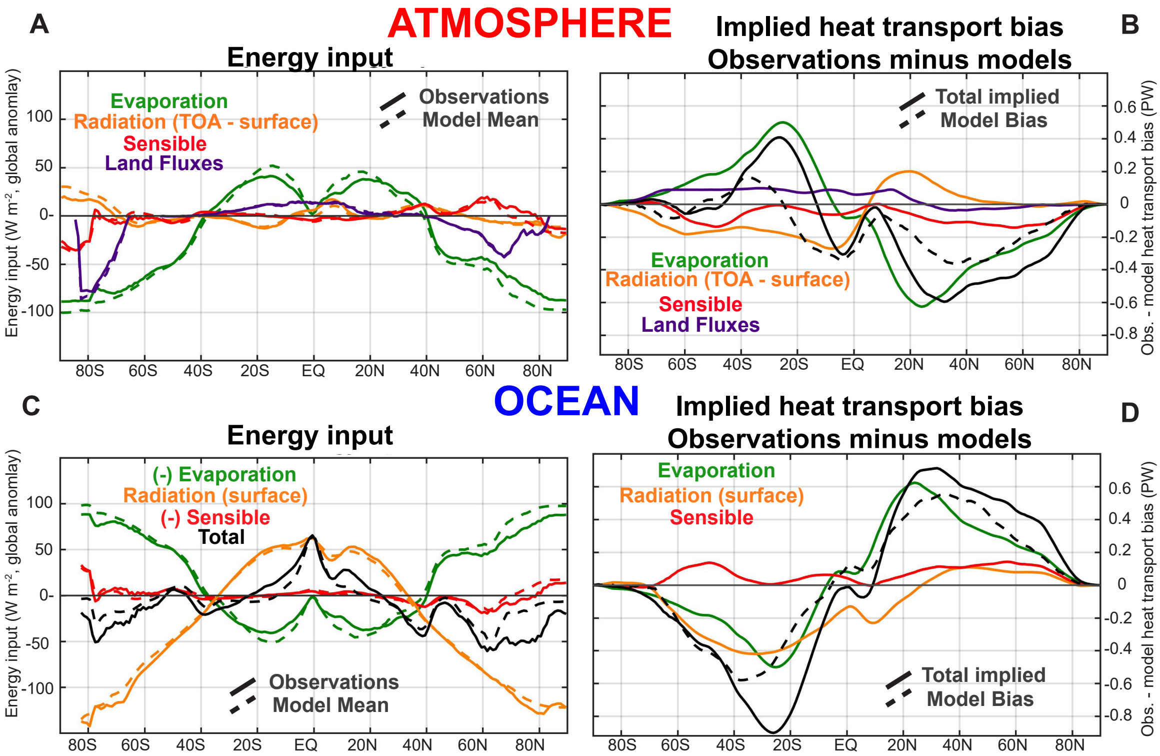

Figure 3. Model and observational estimates of the energy input into the atmosphere and ocean and the implied AHT and OHT biases resulting from each input. (A) Global anomaly energy input into the atmosphere in models (dashed) and observations (solid). See text for definition of terms. (B) Implied AHT bias (observations minus models) due to each energy input. The solid black line shows the sum of all terms. The dashed black line shows the bias in heat transport inferred from CERES and ERA5 data. (C) As in A but for the energy input to the ocean. (D) As in B but for the implied OHT bias.

References

Adler, R., Sapiano, M., Huffman, G., Wang, J., Gu, G., Bolvin, D., ... Shin, D. (2018). The global precipitation climatology project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, *9*(4). doi: 10.3390/atmos9040138

Cardinale, C., Rose, B., Lang, A., & Donohoe, A. (2020). Stratospheric and tropospheric flux contributions to the polar cap energy budgets. *J. Climate*, *34*(11), 4261-4278.

Curry, J., Bentamy, A., Bourassa, M., Bourras, D., Bradley, E. F., Brunke, M., ... Zeng, X. (2004). Seaflux. *Bull. Amer. Meteor. Soc.*, 490-424. doi: 10.1175/BAMS-85-3-409

Donohoe, A., Armour, K., Roe, G., & Battisti, D. (2020). The partitioning of meridional heat transport from the Last Glacial Maximum to CO_2 quadrupling in coupled climate models. *J. Climate*, *33*(10), 4141-4165.

Donohoe, A., & Battisti, D. (2011). Atmospheric and surface contributions to planetary albedo. *J. Climate*, *24*(16), 4401-4417.

Donohoe, A., & Battisti, D. (2012). What determines meridional heat transport in climate models? *J. Climate*, *25*, 3832-3850.

Donohoe, A., & Battisti, D. (2013). The seasonal cycle of atmospheric heating and temperature. *J. Climate*, *26*(14), 4962-4980.

Enderton, D., & Marshall, J. (2009). Controls on the total dynamical heat transport of the atmosphere and oceans. *J. Atmos. Sci.*, *66*, 1593-1611.

Fajber, R., Donohoe, A., Ragen, S., Armour, K., & Kushner, P. (2023). The hydrologic cycle drive atmospheric heat transport. *Proc. Nat. Acad. Sci.*, *120*(25), e2217202120. doi: 10.1073/pnas.2217202120

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., ... Rummukainen, M. (2013). Evaluation of climate models [Book Section]. In T. Stocker et al. (Eds.), *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (p. 659Â–740). Cambridge University Press. doi: 10.1017/CBO9781107415324.018

Good, S. A., Martin, M., & Rayner, N. (2013). En4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res.-Oceans*, *118*, 6704-6716.

Hersbach, H. B. B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz?Sabater, J., Nicolas, J., & Peubey, C. (2020). The era5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*. doi: 10.1002/qj.3803

Holland, M. M., Bitz, C., & Tremblay, B. (2006). Future abrupt reductions in the summer arctic sea ice. *Geophys. Res. Lett.*, *33*(23). doi: 10.1029/2006GL028024

Johnson, G., Lyman, J., & Loeb, N. (2016). Improving estimates of Earth's energy imbalance. *Nat. Clim. Chang.*, *6*(7), 639-640.

Kato, S., & Coauthors. (2018). Surface irradiances of edition 4.0 clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) data product. *J. Climate*, *31*(11), 4501-4527.

Liang, M., Czaja, A., Graversen, R., & Tailleux, R. (2018). Poleward energy transport: is the standard definition physically relevant at all time scales? *Climate Dyn.*, *50*, 1785-1797.

Loeb, N. G., & Coauthors. (2018). Clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition 4.0 data product. *J. Climate*, *31*(2), 895-918.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., ... Wong, T. (2009). Towards optimal closure of the Earth's top-of-atmosphere radiation budget. *J. Climate*, *22*, 748-766.

Lucarini, V., & Ragone, F. (2011). Energetics of IPCC4AR4 climate models: energy balance and meridional enthalpy transports. *Rev. Geophys.*, *49*, RG1001.
M., M., Haimberger, L., Edwards, J., & Hyder, P. (2017). Toward consistent diagnostics of the coupled atmosphere and ocean energy budgets. *J. Climate*, *22*(30), 9225-9246.
Mayer, J., Mayer, M., & Haimberger, L. (2021). Consistency and homogeneity of atmopsheric energy, moisture and mass budgets in ERA5. *J. Climate*, *34*(10), 3955-3974.
Oort, A., & Haar, T. V. (1976). On the observed annual cycle in the ocean–atmosphere heat balance over the Northern Hemisphere. *J. Phys. Oceanogr.*, *6*, 781-800.
Pierrehumbert, R. (2010). *Principles of planetary change.* Cambridge University Press.
Seager, R., Battisti, D., Yin, J., Naik, N., Gordon, N., Clement, A., & Cane, M. (2002). Is the gulf stream responsible for europe's mild winters? *Quart. J. Roy. Meteor. Soc.*, *128*, 2563-86.
Stephens, G., Li, J., Wild, M., Clayson, C., Loeb, N., Kato, S., … Andrews, T. (2012). An update on Earth's energy balance in light of the latest global observations. *Nat. Geo. Sci.*, *5*, 691-696. doi: 10.1038/ngeo15802
Stone, P. (1978). Constraints on dynamical transports of energy on a spherical planet. *Dynam. Atmos. Oceans*, *2*, 123-139.
Trenberth, K. E., & Caron, J. M. (2001). Estimates of meridional atmosphere and ocean heat transports. *J. Climate*, *14*, 3433-3443.
Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bull. Amer. Meteor. Soc.*, *90*(3), 311-324.
Trenberth, K. E., & Stepaniak, D. P. (2003). Co-variability of components of poleward atmospheric energy transports on seasonal and interannual timescales. *J. Climate*, *16*, 3691-3705.
Trenberth, K. E., & Stepaniak, D. P. (2004). The flow of energy through the earth's climate system. *Quart. J. Roy. Meteor. Soc.*, *130*, 2677-2701.
Vonder Haar, T., & Oort, A. (1973). New estimate of annual poleward energy transport by Northern Hemisphere oceans. *J. Phys. Oceanogr.*, *2*, 169-172.
Von Schuckmann, K., Palmer, M., Trenberth, K., Cazenave, A., Chambers, D., Champollion, N., … Wild, M. (2016). An imperative to monitor earth's energy imbalance. *Nat. Clim. Change.*, *6*, 138-144. doi: 10.1038/nclimate2876
Yu, L., Weller, R., & Sun, B. (2004). Improving latent and sensible heat flux estimates for the Atlantic Ocean (1988-1999) by a synthesis approach. *J. Climate*, *17*, 373-393.

JOURNAL OF GEOPHYSICAL RESEARCH

Supporting Information for "Model biases in the atmosphere-ocean partitioning of poleward heat transport are persistent across three CMIP generations"

A. Donohoe[1,3], R. Fajber[2], T. Cox [3], K.C. Armour[3,4], D.S. Battisti [3]and

G.H. Roe [4]

[1]Polar Science Center, Applied Physics Laboratory

University of Washington

Seattle, Washington 98195, USA.

[2] Department of Atmospheric and Oceanic Sciences, McGill University

[3] Department of Atmospheric and Oceanic Sciences, University of Washington

[4]School of Oceanography, University of Washington

[5]Department of Earth and Space Sciences, University of Washington

1. Coupled models analyzed

We analyze pre-industrial (PI) control simulations in coupled climate models that represent the equilibrium response to fixed green house gas concentrations. We analyze 66 model simulations from three different generations of the coupled climate model intercomparison project (CMIP): CMIP3 (Meehl et al., 2007) which ran from 2005-2006 (14 simulations); CMIP5 (Taylor et al., 2012) which ran from 2010-2014 (20 simulations) and; CMIP6 (Eyring et al., 2016) which ran from 2014-2020 (32 simulations). All calculations discussed here use annual mean long term climatologies calculated from the last 50 of available years of the PI simulation. We additionally analyze 12 CMIP5 historical simulations to evaluate the differences between the MHT/AHT/OHT in the PI simulations and historical era which may impact the observational-model comparison.

2. Observational datasets used

2.1. Top of atmosphere radiation

Observational MHT is primarily calculated using satellite derived RAD_{TOA} from the Clouds and Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) product version 4.0 (Loeb & Coauthors, 2018). This product is a gridded retrieval of net longwave and shortwave radiation at the TOA derived from instruments on the Aqua and Terra satellites. The retrieved RAD_{TOA} is subsequently adjusted to satisfy Earth's global energy imbalance of 0.71 ± 0.10 W m^{-2} constrained by long-term changes in global ocean heat content changes (Johnson et al., 2016). This adjustment is accomplished via modification of uncertain parameters in the retrieval algorithm (e.g. radiative transfer model) used to produce the gridded product and primarily involves adjustment of the absolute calibration of the shortwave and longwave fluxes which have a combined uncertainty (95% confidence interval) of 4.2 W m^{-2} (Loeb et al., 2009). We also analyze unadjusted gridded CERES data from single scanner footprints (SSF) to diagnose the impact of the EBAF adjustment on MHT. The average of four (FM1 and FM2 on Terra and FM3 and FM4 and Aqua) SSF RAD_{TOA} data sets is analyzed. The climatological average RAD_{TOA} over the 3/2001-12/2018 period is used to calculate MHT from all CERES products with the exception of the Aqua SSF data which begin in 7/2002. We also use RAD_{TOA} from the Earth Radiation Budget Experiment (ERBE Barkstrom & Hall, 1982). Climatological ERBE RAD_{TOA} over the 11/1984-3/1990 period is used to calculate an additional observational estimate of MHT.

Given that the global mean net TOA radiative imbalance ranges from 7.0 W m^{-2} (3.6 PW globally) in the unadjusted CERES dataset to 4.9 W m^{-2} (2.5 PW) in ERBE dataset (see table 1 of Loeb et al., 2009) to 0.7 W m^{-2} (0.4 PW) in the CERES EBAF dataset (Johnson et al., 2016), it is perhaps surprising that the calculated MHT only differs by of order 0.1 PW across these data sets. We interpret this result to imply that the largest differences between the TOA radiation data sets is the absolute calibration (addition of a spatially invariant constant) of the shortwave and longwave fluxes which are the stated largest source of uncertainty in the data sets (Loeb & Coauthors, 2018) and make no impact on the derived MHT calculated here via removal of the global mean value. Stated otherwise, the spatial gradients in net TOA radiation are less uncertain (or at least consistent between datasets) as compared to the global means.

2.2. Atmospheric reanalysis

AHT is derived from the time average of the vertical and zonal integral of the meridional flux of moist static energy calculated from high spatial-temporal resolution atmospheric reanalysis. Our analysis primarily focuses on AHT estimates calculated from the European Center for Medium Range Forecasting's (ECMWF) ERA5 reanalysis (Hersbach et al., 2020). We use instantaneous 6-hourly ERA5 data on 37 pressure levels and a horizontal resolution of 0.5°. Additional AHT calculations are performed and analyzed using two other sets of 6-hourly instantaneous atmospheric reanalysis: 1. ECMWF's ERA-interim reanalysis which has 37 vertical levels and horizontal resolution of 1.5° (Dee et al., 2011) and; 2. the National Center for Atmospheric Research's (NCEP) reanalysis which has 17 vertical levels and a horizontal spectral resolution of T62.

The following four-dimensional (pressure level, latitude,longitude, time) atmospheric fields are used to calculate AHT; meridional velocity (V), temperature (T), specific humidity (Q) and geopotential height (Z). The climatological surface pressure is used to set the bounds of the vertical integration. AHT calculations are preformed for each month then the results are averaged to produce a long-term average climatology. AHT climatologies are computed over the corresponding time period of the radiation data: 3/2001-12/2018 when used in conjunction with CERES data and 11/1984-2/1990 when used in conjunction with ERBE data.

3. Estimating the impact of ocean energy storage on 'implied' OHT

We first calculate the latitudinal structure of the observed long-term trend in ocean heat storage (STORAGE) over the CERES period (2000-2018) from potential temperature data in the UK Hadley Center EN4 objective ocean analysis (Good et al., 2013).

STORAGE is equal to the linear trend in zonal-mean, vertically (mass-weighted) integrated (deseasonalized) ocean potential temperature. The result is the rate of ocean heat uptake (STORAGE) in W m^{-2} at each latitude averaged over the CERES era. We convert this to an implied OHT due to ocean heat storage (OHT$_{STORAGE}$) by spatially integrating the local departure STORAGE from the global mean (indicated by an *) over the polar cap:

$$OHT(\Theta)_{STORAGE} = 2\pi a^2 \int_{\Theta}^{90} -STORAGE^{*} cos(\theta) d\theta. \tag{1}$$

OHT$_{STORAGE}$ is the 'implied' OHT that would be calculated from the surface heat fluxes needed to balance the local storage *in the absence of lateral ocean transport.* We remove the OHT$_{STOARGE}$ from the 'implied' observational OHT (=MHT-AHT) to isolate the 'dynamic' OHT that would need to be transported laterally in the ocean to balance the ocean energy budget (the sum of STORAGE and energy lost from the surface of the ocean to the atmosphere). If ocean heat uptake is preferentially in the high latitudes (as is observed), the associated downward extratropical surface fluxes would be diagnosed as an equatorward 'implied' OHT and our observational based estimate of poleward OHT from the inferred surface fluxes would be biased *low* relative to an equilibrium climate system with no STORAGE. Thus, the observed high latitude ocean heat uptake *reduces* our observational estimate of OHT and therefore the model biases toward too little poleward OHT are larger in magnitude than reported here even if the magnitude of ocean heat uptake was underestimated by EN4.

Consistent with the reasoning above, model biases toward too little OHT (relative to observations) are stronger in magnitude when comparing historical simulations to (historical) observations than found in the present work which compares pre-industrial (PI) simulations with (historical) observations. Historical simulations have slightly weaker poleward OHT into the Southern Ocean compared to their PI counterparts (c.f. the dashed and solid lines in Supplemental Fig. 3) – which is consistent with the expectations discussed above based on preferential STORAGE in the Southern Ocean – and enhanced poleward AHT in the SH as one would expect from down-gradient energy transport under delayed Southern Ocean warming (Armour et al., 2019). In addition to the differences in the AHT/OHT partitioning between the pre-industrial and historical simulations being small in magnitude (relative to the model biases) these results suggest that the model bias toward too much poleward AHT and too little poleward OHT in the SH would be larger in magnitude if observations over the historical period were compared to the historical (as opposed to PI) simulations.

4. Impact of spatial resolution on calculated AHT

Given that the ERA5 reanalysis is the highest spatial resolution considered here and produces the largest poleward AHT, the reader may be suspicious of whether the reanalysis are of sufficient spatial and temporal resolution (on the model output grid) to capture the processes responsible for AHT. We address the potential limitation of the 6-hourly instantaneous temporal resolution of the data first. Instantaneous data does *not* alias the variance (or co-variance) at any frequency with the exception of the discrete harmonics of the sampling period (periods of 6 hours, 3 hours, 1.5 hours, etc) which should be negligible in a continuous spectra. To test this conclusion, we sub-sampled random (white noise) 1 minute data at 6 hourly intervals and found the variance was reduced by less than 0.01% over 100,000 Monte-Carlo realizations. To evaluate the potential limitation of the horizontal resolution of the reanalysis, we calculate the cross-spectra of meridional velocity and temperature/humidity at 40N, 700 hPa during DJF, the location and season of global maximum climatological poleward AHT (supplemental Fig. 3). Both moist and dry AHT are primarily accomplished by wavenumbers less than 15 with negligible contributions from wavenumbers greater than 90 (corresponding to the smallest resolved wave at 2° longitude grid spacing). Therefore, reducing the resolution of the reanalysis from 0.5 degrees to 2 degrees is equivalent to spectrally truncating the co-spectra at wavenumber 90 which results in a loss of covariance (AHT) of 0.009 % for the dry AHT and 0.021 % for the moisture transport. Stated otherwise, the enhanced horizontal resolution of the ERA5 reanalysis (relative to the resolution of the NCEP reanalysis) makes a negligible contribution to the derived AHT. This analysis does not preclude the possibility that spatial structures smaller than the 0.5° resolution of the ERA5 reanalysis contribute to AHT but does suggest that the enhanced resolution of the ERA5 reanalysis relative to the NCEP reanalysis makes a negligible contribution to the calculated AHT. This conclusion is consistent with the near equivalence of two different AHT calculations in the NCEP CESM simulation shown in Section 2.3; the AHT calculated (dynamically) from the vertical and zonal integral of the product of meridional velocity and temperature/humidity on the 1.25° and 30 vertical level output grid matches that inferred from (energy conservation) of TOA radiation and surface fluxes (Fig. 1 of main text).

References

Armour, K., Siler, N., Donohoe, A., & Roe, G. (2019). Meridional atmospheric heat transport constrained by energetics and mediated by large-scale diffusion. *J. Climate*, *32*(12), 3655-3680. doi: 10.1175/JCLI-D-18-0563.1

Barkstrom, B. R., & Hall, J. B. (1982). Earth radiation budget experiment (erbe)-an overview. *J. Energy*, *6*, 141-146.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., P., K., … van de Berg, A. B. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, *137*, 553-597. Retrieved from `https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim` ([accessed 05-January-2018, https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim])

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development (Online)*, *9*(5). doi: 10.5194/gmd-9-1937-2016

Good, S. A., Martin, M., & Rayner, N. (2013). En4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res.-Oceans*, *118*, 6704-6716.

Hersbach, H. B. B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz?Sabater, J., Nicolas, J., & Peubey, C. (2020). The era5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*. doi: 10.1002/qj.3803

Johnson, G., Lyman, J., & Loeb, N. (2016). Improving estimates of Earth's energy imbalance. *Nat. Clim. Chang.*, *6*(7), 639-640.

Loeb, N. G., & Coauthors. (2018). Clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition 4.0 data product. *J. Climate*, *31*(2), 895-918.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., …

Wong, T. (2009). Towards optimal closure of the Earth's top-of-atmosphere radiation budget. *J. Climate*, *22*, 748-766.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., ... Taylor, K. E. (2007). The WCRP CMIP3 multi-model dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, *88*, 1383-1394.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, *93*, 485-498.

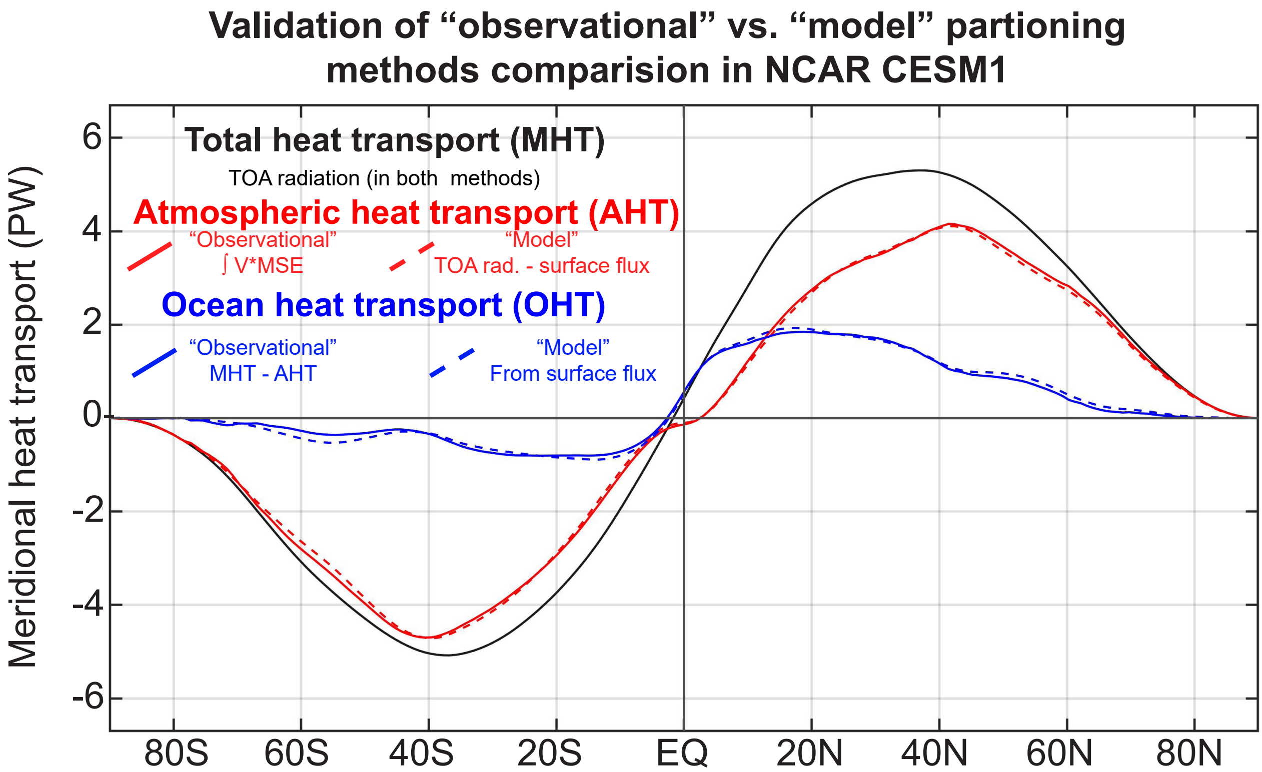

Figure S1. Comparison of the MHT/AHT/OHT partitioning method used for the observations versus that used for the models in an NCAR CESM1 simulation in which the atmospheric fields used to calculate AHT were exported akin to the atmospheric reanalysis. MHT (black) is calculated from the TOA radiation integrated over the polar cap in both methods. AHT (red) is calculated from the time averaged vertical and zonal integral of the product of atmospheric MSE and meridional velocity in the observational approach (solid) and from the spatial integral over the polar cap of TOa radiation minus the surface flux in the model approach (dashed). OHT (blue) is calculated from the residual of MHT and AHT in the observational methodology (solid) and from the spatial integral over the polar cap of the surface heat flux in the model methodology (dashed).

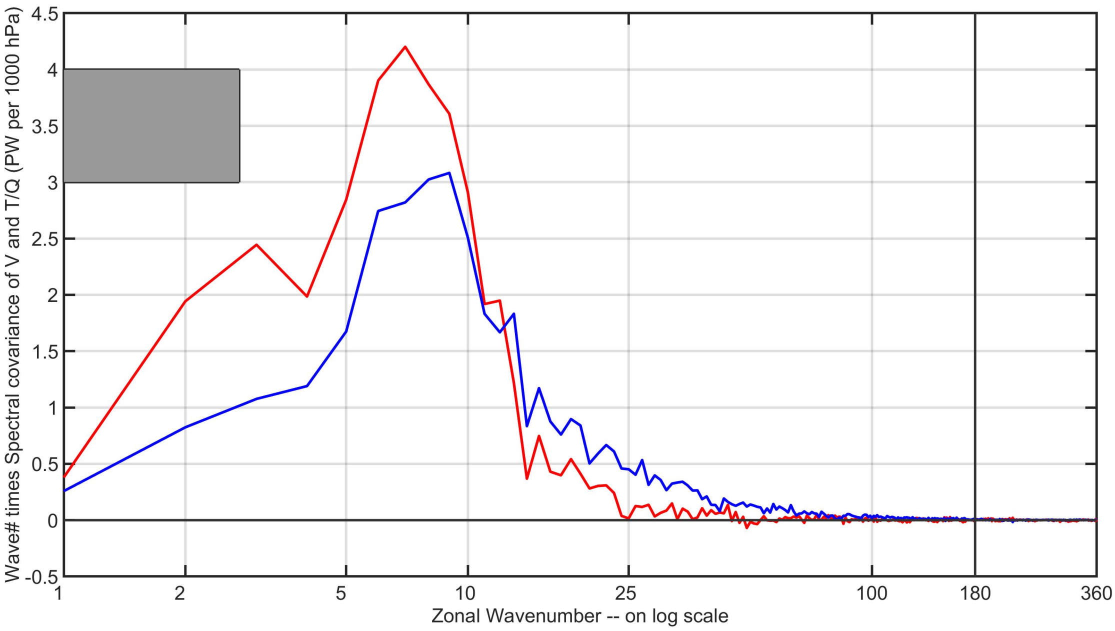

Figure S2. The spectra of atmospheric heat transport at 40N and 700 hPa. The red line shows the spectral co-variance of meridional velocity and temperature (time the specific heat of dry air) and the blue line shows the spectral covariance of meridional velocity and specific humidity (times the latent heat of vaporization of water). The co-spectra are calculated from the product of the spectral power of meridional velocity and temperature/humidity at each instant times the cosine of the spatial phase (wavenumber specific) then time averaged. The wavenumbers on the x-axis are presented on a log scale such that the independent spectral realizations are more densely packed on the right hand side of the plot and the spectral co-variances on the y-axis are multiplied by wavenumber in order to preserve the interpretation of the area under the curve representing the heat transport. The gray shaded box shown an area equal to one PW of zonally and vertically integrated AHT if the spectral co-variance at 700 hPa was realized throughout the atmospheric column. The vertical black line shows the spectral truncation of 4 degrees longitude grid spacing.

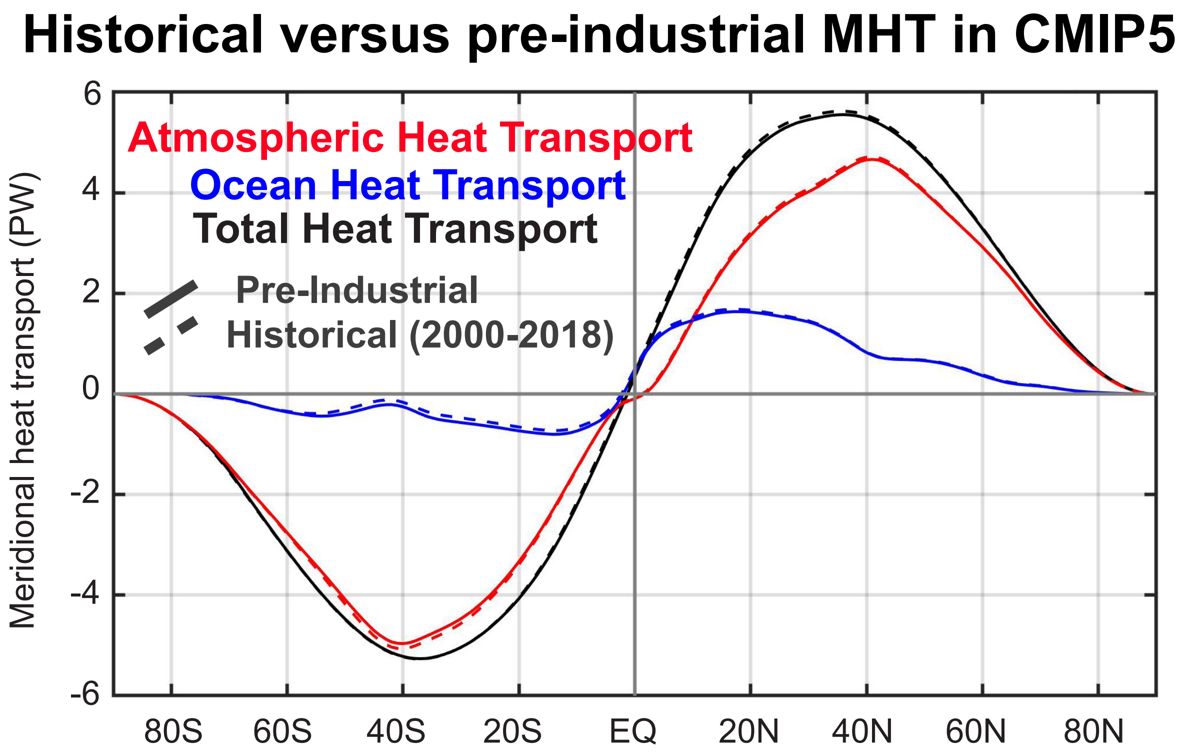

Figure S3. Comparison of the MHT/AHT/OHT partitioning between CMIP historical simulations (dashed lines) and pre-industrial simulation (solid lines). The MHT is shown in black. The AHT is shown in red. The OHT is shown in blue. Both lines are the ensemble mean of the 12 models that have sufficient output for the historical simulations. Historical simulations are averaged over the 2000-2018 period with no adjustment made for ocean heat storage to mimic the observational methodology.

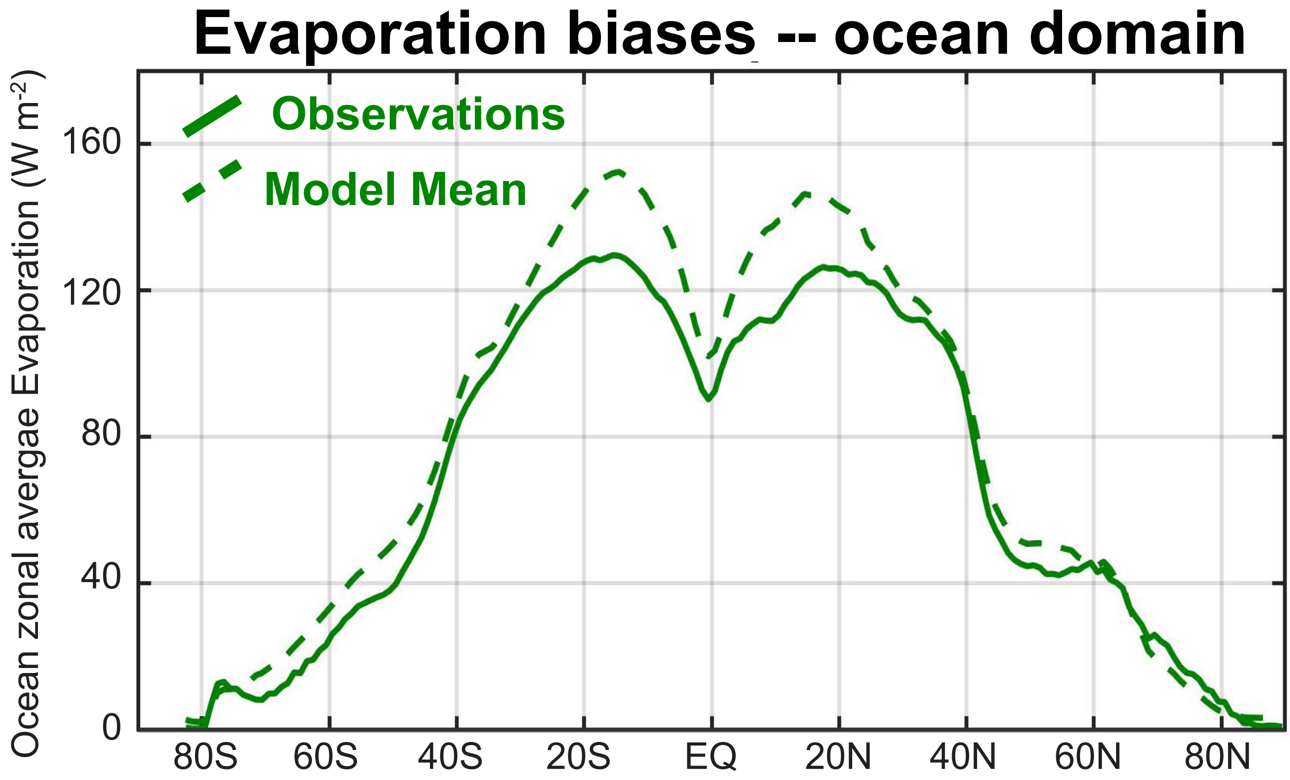

Figure S4. Comparison between evaporation over the ocean in models (ensemble mean) and observations (WHOI OA flux). All values show the annual mean average over the ocean domain and are expressed as latent heat fluxes in W m^{-2}.

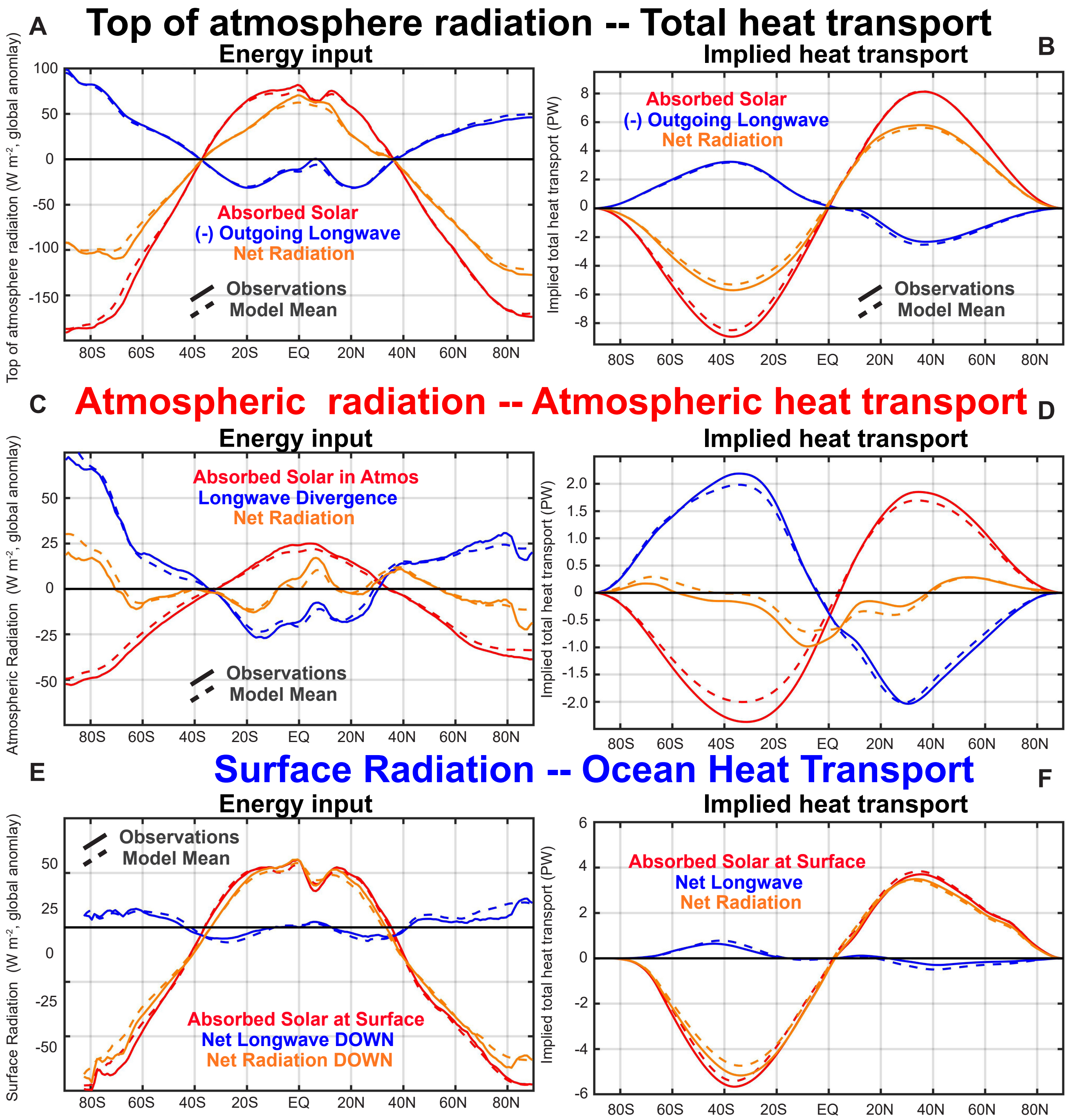

Figure S5. Comparison of radiation in observations (solid lines) and model ensemble mean (dashed lines) at the top of atmosphere (A), in the atmospheric column (C) and at the surface (E). Shortwave fluxes are shown in red, longwave fluxes are shown in blue and the net radiation is shown in orange with positive values defined as a heating tendency on the climate system, atmosphere and surface respectively. The global mean of each term has been removed to emphasize the contribution to the spatial gradients in heating.

The right panels show the implied heat transport of the radiative components for the total (atmosphere plus ocean) meridional heat transport (B, MHT), atmospheric heat transport (D, AHT) and ocean heat transport (F, OHT) in PW. Note that the y-axis range differs