Model biases in the atmosphere/ocean partitioning of poleward energy transport are persistent across three model generations.

A. Donohoe[1], R. Fajber[2], and K.C. Armour[2,3]and others

[1]Polar Science Center, Applied Physics Laboratory
University of Washington
Seattle, Washington 98195, USA.
[2] Department of Atmospheric Sciences, University of Washington
[3]School of Oceanography, University of Washington

Key Points:

- Observational estimates of poleward heat transport and its partitioning between the ocean and atmosphere are compared to that in three different generations of coupled climate modes. Model biases are found to be persistent across model generations.
- Models simulate too little poleward heat transport in the ocean in both hemispheres and too much poleward heat transport by the atmosphere in the Northern Hemisphere
- Stronger than observed evaporation in models removes energy from the tropical ocean and adds energy to the tropical atmosphere to drive atmospheric energy transport at the expense of ocean heat transport. This mechanism explains the majority of model biases in heat transport partitioning

Abstract

Observational estimates of the partitioning of poleward meridional heat transport (MHT) between the atmospheric heat transport (AHT) and ocean heat transport (OHT) are compared to that in coupled climate models across three generations of the coupled model inter-comparison project (CMIP3, CMIP5, CMIP6). Observational estimates of MHT are derived from satellite based top of atmosphere radiation data, AHT is calculated from high spatio-temporal resolution atmospheric reanalysis and OHT is derived as a residual. Model MHT is similarly constrained by top of atmosphere radiation, OHT is calculated from the net surface heat flux and AHT is derived as a residual. We demonstrate these contrasting methods of AHT/OHT partitioning give nearly identical results in a single model which justifies the comparison between AHT/OHT partioning in models and observations. Poleward OHT is biased low in both hemispheres in the models with largest magnitude biases in the Southern Hemisphere extratropics. Poleward AHT is biased high in the Northern Hemisphere especially in the vicinity of the peak poleward heat transport near 40°N. These model biases are significant and persistent across the three model generations and are consistent across observational MHT derived from three different estimates of satellite based TOA radiation and observational AHT derived from three different atmospheric reanalysis. The model biases in AHT and OHT are consistent with model biases in the spatial structure of processes that add and remove energy from the atmosphere. Specifically, larger then observed model evaporation in the tropics adds excess energy to the atmosphere to drive nearly hemispherically symmetric enhanced poleward AHT at the expense of weaker OHT whereas the weaker than observed equator-to-pole gradient in absorbed solar radiation in the models manifests as weaker OHT in the Southern Hemisphere.

Plain Language Summary

1 Introduction

The combined poleward meridional heat transport (MHT) by the ocean and atmosphere plays a fundamental role in moderating temperatures on Earth. In the absence of MHT, the equator-to-pole temperature gradient would be approximately three times greater than observed (Pierrehumbert, 2010) – based on radiative considerations alone – rendering the tropics uninhabitably warm and the high latitudes uninhabitably cold. Early observational estimates of the partitioning of MHT in the Northern Hemisphere (NH) into ocean heat transport (OHT) and atmospheric heat transport (AHT) found that poleward OHT was a factor of two greater than AHT equatorward of 20°N and that AHT comprises the vast majority of total MHT in the extratropics (Vonder Haar & Oort, 1973; Oort & Haar, 1976). More recent observational estimates of global MHT partitioning find that poleward OHT exceeds AHT in the deep tropics of both hemispheres (equatorward of 10°) and AHT dominates total poleward MHT in the mid and high latitudes of both hemispheres (Trenberth & Caron, 2001; J. Mayer et al., 2021). Additionally, the observed MHT partitioning features a modest but climatically impactful hemispheric asymmetry of OHT that moves energy northward across the equator and into the NH midlatitudes. This cross-equatorial OHT causes the NH to be warmer than the SH (Kang et al., 2014), the intertropical convergence zone to reside in the NH in the annual mean (D. Frierson et al., 2013; Marshall et al., 2013) and is generally understood to be associated with the Atlantic meridional overturning circulation (Broecker et al., 1985). Coupled climate model simulations of the present day climate system have significant (order 20%) inter-model spread in the total MHT (Donohoe & Battisti, 2012) but generally agree that MHT is dominated by AHT in the extratropics and is primarily due to OHT in the deep tropics (Yang et al., 2015; Wu et al., 2011; Donohoe et al., 2020). Model's are biased in their cross equatorial MHT and its partitioning between AHT and OHT (Hwang & Frierson, 2013) which impact the tropical mean state. Are model's also biased in the partitioning of MHT outside the deep tropics? A quantitative comparison

of the MHT/AHT/OHT partitioning between models and observations at the global scale is lacking in the literature and is the task of this manuscript.

We begin by discussing whether the partioning of MHT between AHT and OHT matters for the mean climate state and climate change. The same quantities of OHT and AHT potentially have unequal impacts on surface climate for the following reason: 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 upper atmosphere can be radiated to space with minimal impact on surface climate. Stated otherwise, while AHT and OHT are treated as equals in diagnostics of the vertically integrated energy budget of the climate system such as the commonly used climate feedback framework (Roe, 2009; Forster et al., 2021), the vertical structure of AHT fundamentally sets the climatological lapse rate (Hahn et al., 2020) and its changes with climate state (Fajber et al., 2021). Thus, AHT and OHT have disproportionate impacts on surface climate and its changes by way of their control on the vertical structure of atmospheric temperature. High latitude surface climate is particularly sensitive to the partitioning of MHT between OHT and AHT due to the substantial ($\approx 20\%$) contribution of the stratospheric circulation to AHT (Overland & Turet, 1994) that make a negligible contribution to the surface energy budget (Cardinale et al., 2020).

Previous work has demonstrated that (idealized and coupled) climate states with similar total MHT can have significantly different surface temperatures and sea ice extents due to the disproportionate impact of OHT and AHT on surface climate (Enderton & Marshall, 2009). Similarly, forced changes in surface temperature are sensitive to the spatial pattern of OHT (and ocean heat storage Rose & Ferreira, 2013) whereas changes in AHT tend to spatially homogenize the climate impact of radiative changes (Feldl & Roe, 2013).

The partitioning of MHT between AHT and OHT also has implications for the hydrological cycle due to the unequal impact of OHT and AHT on surface evaporation: OHT convergence is primarily balanced by evaporative heat loss whereas; AHT convergence above the surface stabilizes the atmospheric column and, thus, suppresses evaporation. As discussed previously, MHT/AHT/OHT biases in the deep tropics have been linked to biases in the ITCZ location (D. M. W. Frierson & Hwang, 2012) and its seasonal range (Kim et al., 2020) which impact the meridional width of the tropical precipitation (Donohoe et al., 2019). Similarly, the hemispheric asymmetry of OHT is argued to play an essential role in the magnitude and location of the mid-latitude storm tracks which are responsible for the majority of extratropical precipitation (Shaw et al., 2018).

Given the dependence of climate on the AHT/OHT partitioning we ask here: are climate models biased relative to observations in the partitioning of poleward heat transport between the atmosphere and ocean? Identification of mean state model biases in AHT/OHT partitioning likely has implications for future climate change as any model misrepresentation of energy exchange between the atmosphere and the ocean will impact both the mean state and future changes in temperature and hydrology. This question has not been previously addressed, in part, 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. Total MHT is constrained by top of atmosphere radiation at each latitude – as the combined AHT and OHT required to balance the local net radiative imbalance at equilibrium timescales. Thus satellite radiation data and standard climatological model output allow a like-for-like comparison of MHT between models and observations (Donohoe & Battisti, 2012). A similar approach is used to constrain OHT in models from the net surface heat flux whereas this approach leads to insurmountable uncertainty in observations (Wunsch, 2005) due to sparsity of observed surface turbulent energy flux estimates and the substantial global surface energy budget imbalance (Stephens et al., 2012).

Therefore, observational estimates of OHT are most commonly derived from the difference between total MHT inferred from TOA radiation and AHT calculated from the vertical and zonal integral of atmospheric (moist static) energy fluxes estimated from high spatio-temporal resolution atmospheric reanalysis as pioneered by Trenberth & Caron (2001). The disparate methodologies and underlying climate fields used to calculate the AHT/OHT in observations versus models prohibit and like-for-like comparison of MHT partitioning between models and observations and is likely the reason that the comparison has not been made in the literature to date.

The methodology for calculating AHT from observational reanalysis can be applied to high frequency three-dimensional atmospheric model output to facilitate a like-for-like comparison of observational and model AHT. However, this approach is both computationally expensive and requires careful treatment of atmospheric mass and moisture budgets(M. Mayer et al., 2017) to provide an AHT calculation that is self consistent with the coupled energy budget of the modeled climate system. Such an analysis across a multi-model ensemble is possible but would require substantial computational resources and an intricate knowledge of the model specific atmospheric energy budget or lack thereof (Lucarini & Ragone, 2011). Instead, the approach we use here is to export high frequency three-dimensional atmospheric data – akin to observational reanalysis– from a single model and then calculate the AHT using the same methodology that is used on the observational reanalysis. We demonstrate that the resultant AHT from this calculation is nearly identical to the AHT calculated from the standard model output (TOA and surface energy fluxes). This results suggest a comparison between model and observational estimates of MHT partitioning is physically meaningful despite the differing underlying methodologies and potentially will identify genuine model biases in the partioning of MHT.

This manuscript is organized as follows. In Section 2 we provide an overview of the observational and model methodologies for partitioning MHT into AHT and OHT, introduce the underlying data sets used for the calculations and demonstrate the near equivalence of these two approaches in a single coupled model. In section 3 we compare the observational and model MHT partitioning across three different model generation (CMIP3, CMIP5 and CMIP6) and additionally analyze the sensitivity of our finding to different underlying observational data sets used the partition MHT. In Section 4 we consider an alternative method for comparing AHT/OHT partioning in models and observations from the processes that contribute to spatial gradients in input to the atmopshere and ocean. A summary and discussion follows.

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

The methodology used here to partition MHT into AHT and OHT in coupled climate models and observations is described in detail in Donohoe et al. (2020) and here we summarize the conceptual approach. We emphasize that the AHT/OHT partitioning method differs between models and observations due to the availability and reliability of the climate fields required for the calculations and we highlight below reasons the contrasting approaches might yield different results. Additionally, we demonstrate that the two methods produce nearly identical results in a single climate model in which the climate fields used for the observational calculation are exported from the model.

2.1 MHT partitioning in coupled models

2.1.1 Method

The combined heat transport by the atmosphere and ocean (MHT) is calculated from the TOA radiation as the energy flux through a given latitude circle required to

balance the net radiative deficit spatially integrated over the polar cap bounded by that latitude:

$$MHT(\Theta) = -2\pi a^2 \int_{\Theta}^{90} RAD^*_{TOA}(\theta) cos(\theta) d\theta. \quad (1)$$

The * quantity represents an anomaly from the global mean. This approach removes any global mean radiative imbalance – which have a typical magnitude of order 1 W m^{-2} in pre-industrial model simulations (Lucarini & Ragone, 2011) – by using a spatially invariant additive adjustment, thus ensuring MHT is insensitive to whether the integral is calculated from south to north (as written) of from north to south (Donohoe & Battisti, 2012). The negative sign in front of the integral in Eq. 1 demands that MHT is positive (Northward) when there is a net radiative deficit ($\mathrm{RAD}^*_{TOA} < 0$) over the Northern polar cap.

The ocean heat transport (OHT) is similarly calculated from the net surface heat flux (SHF = radiative plus turbulent flux into the ocean) as the energy flux through a given latitude circle required to balance the surface heat loss integrated over the polar cap bounded by that latitude:

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

The removal of the global mean SHF from SHF* is equivalent to making a spatially invariant additive adjustment to conserve the global mean ocean energy budget. This approach assumes that, in the absence of any knowledge of the spatial structure of non energy conserving processes in the climate model, non-conservative processes are spatially homogeneous.

The atmospheric energy transport is calculated as the difference between MHT and OHT. Equivalently, AHT is calculated as the energy flux into the polar cap needed to balance spatial integral of the deficit in net atmospheric heating with the latter equal to $RAD^*_{TOA} - SHF^*$.

2.1.2 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.

2.2 Observational partitioning of MHT

2.2.1 Method for calculating MHT partitioning from observational data

Observational MHT is calculated from satellite derived TOA radiation using the same method that is used for the model calculation of MHT (Eq. 1). Removal of the global mean TOA radiative imbalance from RAD^*_{TOA} ensures that the global mean MHT divergence is zero and is equivalent to assuming that the measurement uncertainty and/or actual energy imbalances in the climate system (i.e. storage in the ocean and atmosphere)

are spatially homogeneous. We return to the impact of this assumption on observationally derived MHT in the results section.

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). The root cause and structural uncertainty in this imbalance is unclear and render the observational estimate of OHT via Eq. 2 untenable. Instead, the following conceptual approach is used following Vonder Haar & Oort (1973) and Trenberth & Caron (2001): AHT at each latitude is calculated from the time average of the vertically and zonally integrated meridional energy flux in the atmosphere derived from high temporal frequency atmospheric reanalysis and; OHT is then calculated as the residual of satellite derived MHT and reanalysis derived AHT. This method can equivalently be thought of as deriving the SHF as the residual of satellite estimated RAD_{TOA} and atmospheric heat flux divergence (Trenberth, 1997; Liu et al., 2015) and then calculating OHT via Eq. 2.

The mass budget of the atmosphere must be balanced prior to the calculation of AHT from atmospheric reanalysis as discussed at length in Trenberth & Stepaniak (2003). We implicitly balance the mass budget by removing the vertically averaged zonal mean MSE at each latitude following Donohoe & Battisti (2013), Cardinale et al. (2020) and Donohoe et al. (2020). This approach renders the AHT/OHT partitioning insensitive to the reference energy state used (M. Mayer et al., 2017) and has been argued to be physically interpretable as the heating of the atmospheric columns resulting from atmospheric transport (Liang et al., 2018).

2.2.2 Observational datasets used

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 $\pm$ 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.

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.

2.3 Comparison of the "observational" and "model" MHT partitioning methods in a single model

Given the disparate methods and climate fields that we use to partition MHT in models and observations, we now ask: do the differing MHT partitioning methods produce the same results? We ask this questions in a single model – a NCAR CESM coupled pre-industrial simulation– in which we partition the MHT into AHT/OHT using both the standard model approach (from TOA and SHF) and using the "observational approach" where AHT is calculated from high frequency 4-dimensional atmospheric data and OHT is derived as a residual. The two approaches give nearly identical partitioning of MHT into AHT and OHT in NCAR CESM (c.f the dashed and solid red and blue lines in Fig. 1) with a root mean squared difference AHT (and OHT) between the two methods of 0.07 PW.

The specifics of the NCAR CESM AHT calculation are discussed in Donohoe et al. (2020) and here we highlight the pertinent details for the purpose of comparing the observational and model approaches to MHT partitioning. AHT is calculated from the time-averaged product of instantaneous meridional velocity times temperature (VT), specific humidity (VQ) and geopotential (VZ). The time average of these products are used in conjunction with the time average of the fields independently ($\overline{V}$, $\overline{T}$, $\overline{Q}$ and $\overline{Z}$) to back out the temporal co-variances which can then be used (along with the time mean fields) to calculate AHT (Donohoe et al., 2020) in a mathematically equivalent calculation to the method used to calculate observational AHT. These products (VZ, VT, VQ) are interpolated from the native model grid to an output grid with 30 pressure levels and a horizontal resolution of 1.25° which is coarser than observational reanalysis we use to calculate observational AHT. We note that model output is the product of fields at the model time-step as opposed to instantaneous 6-hourly data used in the observational AHT calculation. However, we note that instantaneous 6-hourly data resolves co-variances at frequencies greater than $\frac{1}{6hours}$ with the exception of the discrete harmonics of the sampling period. We return to this point in Section 3.0.2.

We interpret the close correspondence of the two approaches to partitioning MHT into AHT and OHT in NCAR CESM as validation that the "observational" and "model" approaches we use here to partition MHT are directly comparable and use this result to justify the examination of potential model biases in MHT partitioning.

3 Results: model biases in MHT partitioning

Model biases in MHT partitioning are analyzed using three different generations of model ensembles and several different sets of underlying observational data sets. The presentation of our results is organized as follows: Section 3.1 uses a single observational

estimate of MHT partitioning – using the most contemporary and high resolution data– compared against three different generations of coupled climate model ensembles; Section 3.2 analyzes the sensitivity of our results to the observational data used by comparing 6 different observational estimates of MHT partitioning against the multi-generation model ensemble mean. The overarching conclusion is that the directionality and spatial structure of model biases in MHT partitioning are consistent across model generation and observational data sets used..

3.0.1 Consistent model biases in AHT/OHT partitioning across three generations of coupled model ensembles

The primary observational MHT partitioning used in this section is derived from the combination CERES EBAF TOA radiation and ERA5 atmospheric reanalysis. This observational estimate (solid line) is compared against a different generation of coupled climate model ensembles (dashed lines with ensemble average shown by the thick dashed line) in the different rows of Fig. 2.

The peak poleward MHT in both hemispheres is near 35° in both models and observations consistent with constraints due to Earth-Sun geometry whereby RAD^*_{TOA} is dominated by the second order Legendre polynomial (equator-to-pole scale) as discussed by Stone (1978). However, the amplitude of poleward MHT in models is biased low in the mid-latitudes of both hemispheres relative to observations across all three CMIP generations. The maximum poleward MHT in the SH is 5.7 PW in observations and exceeds the ensemble mean of CMIP3 (5.2 PW), CMIP5 (5.3 PW) and CMIP6 (5.4). The ensemble mean SH MHT is significantly different from the observational estimate (as measured by the 95% confidence of the ensemble mean spread) in all three model generations. The maximum poleward MHT in the NH is 5.8 PW in observations and exceeds the ensemble mean of CMIP3 (5.6 PW), CMIP5 (5.5 PW) and CMIP6 (5.7). The ensemble mean maximum NH MHT is only significantly different from the observational estimate in CMIP5. The inter-model spread in peak SH MHT (2 standard deviations) is as large as 23% of the ensemble mean and has values of 1.2 PW in CMIP3, 0.8 OW in CMIP5 and 0.8 PW in CMIP6. The inter-model spread in peak NH is smaller than its SH counterpart but is substantial in magnitude with values of 0.8 PW in CMIP3, 0.6 PW in CMIP5 and 0.6 PW in CMIP6. Trenberth & Fasullo (2010); 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).

In the NH, the model ensemble mean is significantly biased toward too little poleward OHT and too much poleward AHT (relative to our primary observational estimate) across all three model generations. The observational estimate of peak NH AHT is 4.4 PW as compared to in 4.7 PW in CMIP3, 4.7 PW in CMIP5 and 4.8 PW in CMIP6. The peak NH OHT occurs more equatorward in both observations and models but has significantly higher values in observations (2.0 PW) as compared to model ensemble means (1.7 PW in CMIP3, 1.8 PW in CMIP5 and 1.7 PW in CMIP6). The bias toward too little OHT extends poleward to the extratropics and the Arctic where OHT has been demonstrated to have large impacts on sea ice extent (Holland et al., 2006).

In the Southern Hemisphere, poleward OHT is significantly (at the 95% confidence interval) biased low relative to observations in all three model generations. The largest magnitude biases in OHT are found the vicinity of 40S where the equatorward transport of cold water in the oceanic Deacon cell (Marshall & Speer, 2012) manifests as surface energy fluxes into the ocean and thus a kink in the spatial structure of OHT that otherwise decreases in magnitude toward the pole. At 40S, the observational OHT is -

0.7 PW whereas the ensemble mean OHT at that latitude is -0.3 ± 0.2 PW , -0.2 ± 0.1 PW and -0.1 ± 0.1 PW in CMIP3, CMIP5 and CMIP6 respectively where the stated uncertainty is 2 standard deviations of the ensemble mean. The observational estimate of poleward OHT is only exceeded in 3 model simulations (2 in CMIP3 and 1 in CMIP5). In contrast, the poleward AHT in the SH is very similar between the models and observational estimates; the ensemble mean maximum poleward AHT in the SH is not significantly different from the observations in any model generation nor in all three model generations considered collectively.

These results suggest that the majority of the model biases in SH 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 fraction of MHT carried by AHT/OHT (i.e. normalizing each model by the model specific MHT) is biased toward too much poleward AHT and too little poleward OHT in both hemispheres with biases that are nearly hemispherically symmetric between the two hemispheres (not shown). Importantly, the directionality and spatial structure of model biases in MHT/AHT/OHT are remarkably consistent across three model generations spanning over 15 years of progress in climate modeling.

3.0.2 Sensitivity of results to observational data used

The previous section demonstrated that model are biases in the partitioning of AHT/OHT are consistent across three CMIP model generations when compared to our primary observational estimate (CERES EBAF and ERA5 based) of MHT, AHT and OHT. We now ask how sensitive these conclusions are to the choice of observational data sets of TOA radiation, atmospheric reanalysis and time period used. We use the ensemble model mean across all three CMIP generations (weighted by the number of models in each generation) as a reference for all analysis in this subsection.

We begin by analyzing the sensitivity of our results to choice of observational TOA radiation by analyzing the MHT/AHT/OHT derived from two additional satellite derived observational estimates of TOA radiation: the unadjusted CERES SSF data and the ERBE satellite data (left panels of Fig. 3). In these three panels, the varying choice of TOA radiation product alters the calculated observational MHT (solid black line) and, because the observational OHT is calculated from the difference of MHT and AHT, the observational OHT estimate (solid blue line) also varies across panels. In contrast, the observational AHT is derived from ERA5 reanalysis in all panels and, thus, is consistent between panels with the minor exception that the climatological average in the middle panel is calculated over the 1984-1990 ERBE period as opposed to the 2001-2018 CERES period for all other panels (which makes a negligible impact). Calculated observational poleward MHT is consistently larger than the model mean in both hemispheres for all three TOA radiation data sets. Observational poleward MHT – and, thus, the magnitude of model biases– is largest in the CERES SSF product in both hemispheres and smallest in the ERBE product. The model ensemble mean OHT is biased low as compared to that derived from all three TOA radiation datasets with largest magnitude biases in the CERES SSF based calculations especially in the Southern Hemisphere. These results suggest that the model biases in AHT/OHT are insensitive to observational TOA radiation data sets used.

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.

We next analyze the sensitivity of our results to the choice of atmospheric reanalysis used to calculate the AHT (Fig. 3 panels A, B and D). In these panels, the MHT is calculated from the CERES EBAF data consistently across all three panels, whereas the AHT is calculated from the ERA5, ERA interim and NCEP reanalysis respectively. Since OHT is calculated from the residual of MHT and AHT, the OHT difference between the three panels are equal and opposite the inter-panel differences in AHT. The model bias toward too much poleward AHT and too little poleward OHT is consistent across all three estimates of observational AHT. Poleward AHT is largest in the ERA5 based calculations followed by ERA interim whereas NCEP based calculations have the smallest poleward AHT with the most notable difference near the peak in the SH at 40S. Therefore, model biases in the AHT/OHT partitioning are smallest in magnitude in the ERA5 based estimates and largest in the NCEP based estimates. These results suggest that the directionality and spatial structure of model biases in AHT/OHT partitioning are consistent across atmospheric reanalysis datasets whereas the magnitude of the bias differs between the different reanalysis.

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).

Finally, we evaluate if energy accumulation due to the transient response to anthropogenic forcing impacts our observational estimates of OHT. Earth is not in equilibrium but, rather, is accumulating energy at a decadal and global mean rate of 0.71 W m^{-2} (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 since the latter is calculated from the spatial integral of inferred (TOA radiation minus AHT divergence) surface heat fluxes. We diagnose the impact of observed ocean heat uptake on implied OHT from the UK Hadley Center EN4 objective ocean analysis (Good et al., 2013) as follows: 1. the ocean column integral of potential temperature times density is computed at each (monthly) time step and grid-point; 2. de-seasonalized anomalies from the 2000-2018 period are computed; 3. the column energy tendency is calculated as the centered difference finite difference at each time step; 4. the time average tendency over the 2000-2018 time period is zonally averaged. The result is the rate of ocean heat uptake (STORAGE) in W m^{-2} in 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 over the polar cap:

$$OHT(\Theta)_{STORAGE} = 2\pi a^2 \int_{\Theta}^{90} -STORAGE*cos(\theta)d\theta. \tag{3}$$

$OHT_{STORAGE}$ is the 'implied' OHT that would be calculated from the surface heat fluxes needed to balance the local storage *in the absence of ocean transport*. If ocean heat uptake was preferentially in the high latitudes, 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. We remove the $OHT_{STOARGE}$ from the 'implied' observational OHT (calculated from Eq. 2) to isolate the 'dynamic' OHT that would need to be transported laterally in the ocean to balance the sum of STORAGE* and SHF*. $OHT_{STORAGE}$ is very small (< 0.1 PW) and, thus, the diagnosed 'dynamic' OHT (solid teal line in Fig. 3F) is visually indistinguishable from the observational 'implied' OHT (solid blue line). We note that, as described in the above example, high latitude ocean heat uptake would *reduce* our observational estimate of OHT and therefore model biases toward too little poleward OHT would be larger in magnitude than reported here even if ocean heat uptake was underestimated by EN4.

The negligible importance of ocean heat storage over the historical period are 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 simulations in the same models (Supplemental Fig. 3). The 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.

Collectively, these results suggest that the directionality of model biases in AHT/OHT partioning is robust to choice of observational estimates of satellite derived TOA radiation used to calculate MHT, atmospheric reanalysis used to calculated AHT and accounting for the impact of (the spatial pattern of) transient heat uptake by the ocean on inferred OHT. The magnitude of the model bias in AHT/OHT partitioning does vary with datasets used. In this regard, the use of CERES EBAF and ERA5 data for our primary analysis is a more conservative (only the ERBE and ERA5 data finds a smaller OHT bias) estimate of model biases in AHT/OHT partitioning.

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

Here we seek alternative estimates of model biases in AHT/OHT from the fields that contribute regional scale energy input into the ocean and atmosphere. Starting in the ocean, the SHF in Eq. 2 is decomposed into the individual fluxes that contribute to SHF:

$$SHF = RAD_{SURF} - SENS - L_vE \tag{4}$$

where RAD_{SURF} is the net surface radiation into the ocean, $SENS$ is the sensible heat flux from the ocean to the atmosphere and L_vE is the latent heat flux from the ocean to the atmosphere which is expressed here as the evaporation (E) times the latent heat of vaporization of water (L_v). Substitution of Eq. 4 into 2 gives:

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

where, similar to Eq. 1 and 2, the * represents the local anomaly from the global ocean domain average. Each term is presented here as the zonal average over the ocean grid-points at that latitude and the values at each latitude are weighted by the fractional area of the ocean at that latitude in order to calculate the OHT. Following Fajber et al. (2022), we decompose the OHT in Eq. 5 into the three processes that contribute to spatial gradients in energy input to the ocean:

$$OHT = OHT_{RAD,SURF} + OHT_{SENS} + OHT_E, \tag{6}$$

where, for example the OHT due to evaporation is

$$OHT(\Theta)_E = 2\pi a^2 \int_{\Theta}^{90} L_vE^* cos(\theta) d\theta. \tag{7}$$

Equivalent expressions for the $\mathrm{OHT}_{RAD,SURF}$ and OHT_{SENS} are defined by replacing L_vE^* in Eq. 7 with RAD^*_{SURF} and SENS^* receptively. These equations: 1. exactly replicate the OHT calculated in climate models via EQ. 2; 2) allow an alternative estimation of model biases in OHT from a comparison of the spatial structure of surface radiation and turbulent fluxes in models and observations and; 3) are independent of the global mean surface energy imbalance since only the global anomalies of each term contribute to OHT.

Observational estimates of E and $SENS$ are taken from the WHOI OA flux (Yu et al., 2004) and $SURF_{RAD}$ estimates from the CERES EBAF surface product (Kato & Coauthors, 2018). Zonal averages over the ocean domain in both observations and models are shown in Fig. 4C. In both models and observations, there is a net energy input to the ocean in the tropics and energy loss at the high latitudes (black lines in Fig. 4C) which demands poleward OHT. Energetically, the equator-to-pole gradient in energy input to the ocean is driven by surface radiation (primarily due to stronger solar fluxes in the low latitudes – Supplemental Fig. 2E,F) and is primarily damped by stronger evaporative energy loss in the tropics and subtropics. Thus, the climatological equator-to-pole gradient in evaporation reduces the demand for OHT (relative to that required in the absence of exchange with the atmosphere).

Evaporation is biased high in models (relative to the observational estimate) at all latitudes except the Arctic (Supplemental Fig. 1). Evaporation biases are largest (> 20

W m^{-2}) in the subtropics of both hemispheres and are much smaller in the high latitudes. Stated otherwise, the equator-to-pole gradient in evaporation is much stronger in the models compared to that in observations which manifests as stronger global anomalies of evaporative loss from the low latitude ocean in models (E^* is more negative in models in Fig. 4C) and stronger extratropical energy gain in the models due to global anomalies in evaporation (more positive E^*). Fig. 4D shows the biases in implied OHT due to evaporation (green line) defined as OHT_E calculated from the observational estimates of evaporation in Eq. 7 minus OHT_E calculated from the same equation using the model ensemble mean evaporation. Poleward OHT_E is stronger in the observations as compared to that in models by approximately 0.4 PW. This result is consistent with a weaker observational E^* gradients leaving more energy in the low-latitude ocean in the observations and, thus, demanding more OHT.

The observational RAD^*_{SURF} has a stronger equator-to-pole gradient than that in climate models (c.f. the solid and dashed orange lines in Fig. 4C) especially in the SH. Model biases in RAD^*_{SURF} are associated with greater than observed downwelling solar radiation into the extratropical Southern Ocean (50S to 80S) in models due to simulated clouds that are too shortwave transparent and weaker than observed surface longwave cooling in the model simulations of the Arctic (Supplemental Fig. 2E). As a result, observed poleward $OHT_{RAD,SURF}$ is larger than that in models with larger magnitude (0.4 PW) biases in the SH and a modest contribution to southward biased OHT at the equator. The model biases in $OHT_{RAD,SURF}$ mimic the impact of TOA radiation biases on MHT (left panels of Figure 2) including the partitioning between shortwave and longwave biases within each hemisphere (Donohoe & Battisti, 2012) which suggests 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 (Supplemental Fig. B and F). Sensible heat loss from the ocean is fairly spatially homogeneous in both models and observations (the red $SENS^*$ lines in Fig. 4C are small in magnitude) and, thus, OHT_{SENS} is small in magnitude. There is a modest model bias toward weaker (than the global mean) sensible heat loss from the ocean surface in the Arctic which results in stronger observational poleward OHT_{SENS} into the Arctic as compared to that in the models. The sum of model biases OHT_E, $OHT_{RAD,SURF}$ and OHT_{SENS} (solid black line in Fig. 4D) demonstrates that biases in spatial structure of energy input into the ocean demand stronger poleward OHT of 0.6 PW in the NH and 0.8 PW in the SH primarily due to nearly hemispherically symmetric (in the poleward sense) OHT_E which is enhanced by poleward $OHT_{RAD,SURF}$ in the SH. The bias in implied OHT (inferred from surface flux biases) matches the spatial structure but exceeds the magnitude of OHT biases identified in Section 2 from the residual of (CERES EBAF) TOA radiation constrained MHT and (ERA5) atmospheric reanalysis constrained AHT which is shown by the dashed black line in Fig. 4D. We note that there is no reason these two calculations of OHT biases should match as the two calculations use different conceptual approaches and rely on completely independent observational climate fields. Specifically, the turbulent energy fluxes and surface radiation used in the implied OHT are completely independent data sets to the atmospheric reanalysis used in Section 2. Nonetheless, the consistency of the direction, spatial pattern and magnitude of the identified biases in OHT approaches suggest that the model biases in surface energy fluxes are more than large enough to account for the AHT/OHT partitioning biases inferred from the residual TOA radiation and AHT estimates.

A similar (to the above OHT analysis) calculation of the model biases in implied AHT from the spatial structure of energy input to the atmosphere can be used to compute an alternative estimate of AHT biases – to be compared to the high frequency observational reanalysis based AHT calculation from Section 2. The AHT analog to Eq. 6 is:

$$AHT = AHT_{RAD,ATMOS} + AHT_{SENS} + AHT_E, \tag{8}$$

where the atmospheric analog to Eq. 7 for the AHT due to evaporation is

$$AHT(\Theta)_E = 2\pi a^2 \int_{\Theta}^{90} -L_v E^* cos(\theta) d\theta. \quad (9)$$

The zonal mean in the integrand (and the global mean that has been removed from) of Eq. 9 is computed over the combined ocean and land domains. Equivalent expressions for the $\mathrm{AHT}_{RAD,ATMOS}$ and AHT_{SENS} are defined by replacing -$\mathrm{L}_v E^*$ in Eq. 7 with RAD^*_{ATMOS} and -SENS* respectively. Here RAD_{ATMOS} is the net radiative heating of the atmospheric column which is equivalent to the net radiation at TOA minus RAD_{SURF} or, alternatively, the shortwave absorption within the atmospheric column minus the longwave radiative divergence integrated over the atmospheric column.

Fajber et al. (2022) demonstrated that poleward AHT is primarily determined by evaporation (AHT $\approx \mathrm{AHT}_E$) because E* dominates the spatial structure of energy input to the atmosphere. Furthermore, the AHT_E calculated by Eq. 9 is physically intepretable as the heat transport driven by the hydrologic cycle by way of E* directly demanding atmospheric moisture transport and the subsequent hand off of this energy to dry atmospheric heat transport where the atmosphere is condensationally heated due to moisture precipitating out of the atmospheric column (Fajber & Kushner, 2021). We note that E* (and SENS*) spatially integrated over the ocean domain has opposing impacts on AHT_E versus OHT_E (and AHT_{SENS} versus OHT_{SENS}) which can be understood as follows: 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. Eq. 9 is written from the perspective of the (Northern) polar cap such that a deficit of evaporation over the extratropics (E$^* < 0$) enhances the demand for poleward AHT via the minus sign in the integrand.

To more clearly see the compensation between biases in AHT/OHT due to model biases in E* (and SENS*) over the ocean domain we take the following approach to comparing models and observations of AHT via Eqs. 8,9: 1) 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 as was done to calculate model biases in OHT_E and OHT_{SENS} (green lines in Fig. 4A,B); 2) $\mathrm{AHT}_{RAD,ATMOS}$ 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. 4A,B) and; 3) 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 as compared to analogous calculation in the models. This strategy circumvents the lack of reliable observational estimates of turbulent energy fluxes over land by directly comparing model and observational estimates of turbulent energy fluxes over the ocean domain only and inferring the former from the assumption that surface energy balance over land requires that RAD_{SURF} is balanced by upward turbulent fluxes from the land to the atmosphere. The latter assumption is very nearly satisfied in all climate models considered here. We note that for all calculations in Section, only like climate fields (ocean domain turbulent fluxes and land domain RAD_{SURF}) in observations and models are compared. Additionally, by construction, biases in AHT_E and OHT_E oppose one another (the green lines in Fig. 4B and 4D are vertical mirror images of one another) since both are defined from the same observational and model data over the ocean domain only.

Model biases in AHT_E compose the vast majority biases of AHT biases diagnosed from Eq. 8 (c.f. the green and solid black lines in Fig. 4C) 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 com-

pared to models (due to stronger longwave cooling in the models– supplemental Fig. 2C) which contributes to weaker $\mathrm{AHT}_{RAD,ATMOS}$ export from the tropics in the models that generally opposes the low latitude biases in AHT_E (orange line in Fig. 4B). Weaker than observed atmospheric longwave cooling over the Antarctic continent results in more poleward $\mathrm{AHT}_{RAD,ATMOS}$ into the Antarctic continent in observations. 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 inf Supplemental Fig. 2D) but this model deficit in atmospheric heating of the tropics is nearly compensated for by weaker than observed longwave cooling of the atmosphere in models 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 have substantial latitudinal structure which contributes to poleward AHT. However, these physics are fairly consistent in models and observations and make little contribution to model biases in AHT (purple line in Fig. 4B).

These calculations demonstrate that the model biases in the partitioning of poleward heat transport between AHT and OHT that were inferred in Section 2 are consistent (in directionality, 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 are biased toward too little poleward OHT in both hemispheres and too much AHT in the Northern Hemisphere. These model biases are remarkably consistent across three generations of coupled model ensembles (CMIP3, CMIP5 and CMIP6). The model biases are also consistent across the observational TOA radiation dataset used to derive MHT and the atmospheric reanalysis used to calculate AHT. We have demonstrated that the differing methods used here to calculate AHT/OHT partitioning in models versus observations give nearly identical results when applied to a single model where AHT is calculated from high frequency atmospheric data on the model output grid – akin to the observational calculation of AHT from instantaneous reanalysis data. This result suggest that the comparison between observational and modeled AHT/OHT partioning is meaningful and is a statement that the methods we use for calculating MHT, AHT, OHT approximately balance the energy budget of the atmosphere and ocean.

We briefly reflect on the assumptions that we have made in our calculations of AHT and OHT to highlight places that our calculation may have gone wrong or otherwise differs from the methodology used elsewhere in the literature. Our calculation of MHT from satellite derived TOA radiation first removes a global mean energy imbalance via an additive adjustment at each latitude which is equivalent to assuming that the global energy imbalance (or lack of energy conservation) is spatially homogeneous. This choice is motivated in part because the largest uncertainty in the CERES and ERBE satellite radiation data is the absolute calibration of shortwave and longwave irradiances which would be represented by a global additive offset. The calculation of MHT (and OHT) in models also requires removal of the global mean radiative (and surface heat flux) imbalance even in the pre-industrial simulations where aphysical (i.e. non energy conserving) global mean energy imbalances of order 1 W m^{-2} would be associated with an MHT of order 0.5 PW through the South Pole via the conventions used in Eq. 1. Removal of a spatially invariant global mean imbalance from the model MHT and OHT calculation appears pervasive in the literature dating back to at least (Trenberth & Fasullo, 2010)

and it is unclear if this method has a formal justification or has been tested extensively in an ensemble of models. For lack of a better understanding of the spatial structure of non-conservative processes, the spatially homogenous adjustment seems like a good first step but this approach deserves more attention regarding the possible impact on derived MHT and OHT (Wunsch, 2005).

The method used here to calculate MHT, AHT and inferred OHT is inherently zonal mean and makes no distinction between energy balances separated into ocean basins and land domains. Two dimensional (lat, lon) based observational estimates (Trenberth & Caron, 2001) of AHT and OHT adjust the radiation over the land domain to balance the AHT divergence (assuming no energy storage and land), compute the OHT from the derived surface fluxes over the ocean only and adjust the OHT to match models in the high latitude Southern Ocean. Such approaches are motivated by known deficiencies and or inconsistencies (i.e. lack of energy and water mass conservation) across different observational datasets. The satisfaction of additional constraints physical constraints comes at the expense of: 1. more complicated methodologies for calculating MHT/AHT/OHT that may be difficult to replicate identically (i.e. depend on exactly how the additional constraints are satisfied) and 2. the additional complication of understanding whether model biases in, for example MHT, stem from biases in the TOA radiation or, alternatively, the adjustment of the satellite radiation over land needed to satisfy no energy flux into the land surface. Our approach here is to trust the most observationally constrained data, apply no adjustments beyond the removal of global means and compare the derived AHT/OHT to that in models. This approach is easily reproducible and we hope it will provide insights into the processes that control MHT/AHT/OHT including the inter-model spread and bias.

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. Mayer 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. The AHT associated with the surface pressure tendency vanishes on climatological time scales and has been shown to be aphysical on shorter timescales (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. Mayer 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 a mass flux into the extratropical ocean and subsequent return flow in the ocean. The energy flux of each of these mass fluxes is the product of mass flux and mean energy of the fluid and is of order 0.2 PW for each leg. 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" approaches where inclusion of a net mass flux 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/AHT/OHT despite the different climate fields that go into each calculation.

Remarkably, if we trust the validity of the OHT inferred from satellite TOA radiation and reanalysis based AHT, we infer a model OHT bias that is in descent agreement with model biases in the energy exchange between the ocean and atmosphere. The latter bias is due primarily to stronger than observed evaporation in the models. We note that the community has been reluctant to diagnose observational OHT and it biases from

the surface energy balance due to 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 and the bulk formula based estimates such as the WHOI OA flux (Yu et al., 2004) and SEAFLUX (Curry et al., 2004) having less evaporation. We chose to use the 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 (XX W m^{-2}) plus the ERA5 reanalysis land evaporation over land (YY W m^{-2} for a global total evaporation of ZZ W m^{-2}) nearly balancing the best observational estimate of global mean precipitation (WW W m^{-2}) from the NOAA GPCP (Adler et al., 2018). The lack of closure of the observed global mean surface energy budget suggest either 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 and suggests that equator-to-pole gradient of surface energy fluxes is consistent with those inferred from TOA radiation and AHT divergence. We caution the reader that if there are genuine uncertainties and/or structural biases on observational evaporation products then there is likely also uncertainty in the equator-to-pole gradient of evaporation and the implied OHT/AHT due to evaporation (AHT_E) that needs further examination. We hope that consideration of the meridional structure of surface energy fluxes constrained by TOA radiation and AHT can be used in conjunction with global mean imbalances to give an additional degree of freedom for reconciling which terms in the observed surface energy budget are most uncertain and/or biased.

Our work here has 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 stronger than observed dry (sensible) heat transport by transient eddies in the mid-latitudes of both hemisphere (there Fig. 4D) whereas AHT by stationary eddies in the NH is weaker than observed and moisture (latent heat) transport has negligible biases. Model biases in evaporation are expected to be manifested as biases in both moist and dry AHT for the following reason: dry AHT is set by the spatial pattern of condensational heating of the atmosphere which represents the portion of AHT_E that is not transported poleward as latent heat (Fajber et al., 2022). Thus, 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 speculation that model biases toward too much AHT result from stronger than observed evaporation is not inconsistent with the AHT biases primarily being sensible heat transport biases.

The analysis here has only diagnosed OHT from the spatial pattern of surface energy fluxes as opposed to an analysis of the energy transport in different oceanic circulations. An observational/model comparison of OHT by basin and decomposed into gyre and overturning circulations would help to elucidate if the model biases identified here are consistent with model biases in the ocean circulation.

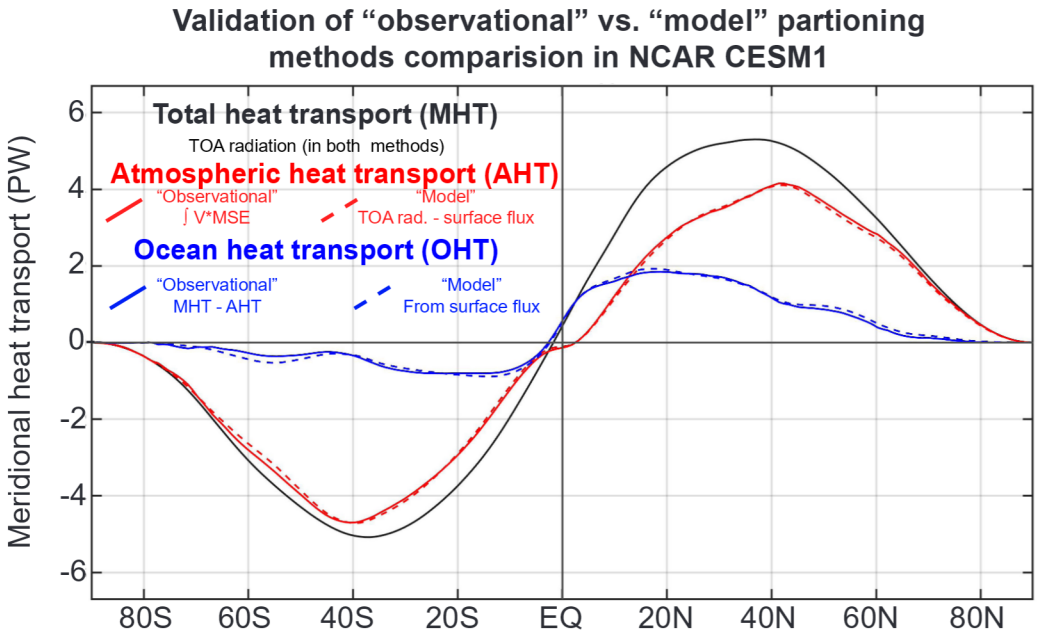

Figure 1. Comparison of the MHT/AHT/OHT partitioning method used for the observations versus that used for the models in an NCAR CESM1 simulation in which the atmospheric fields used to calculate AHT were exported akin to the atmospheric reanalysis. MHT (black) is calculated from the TOA radiation integrated 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).

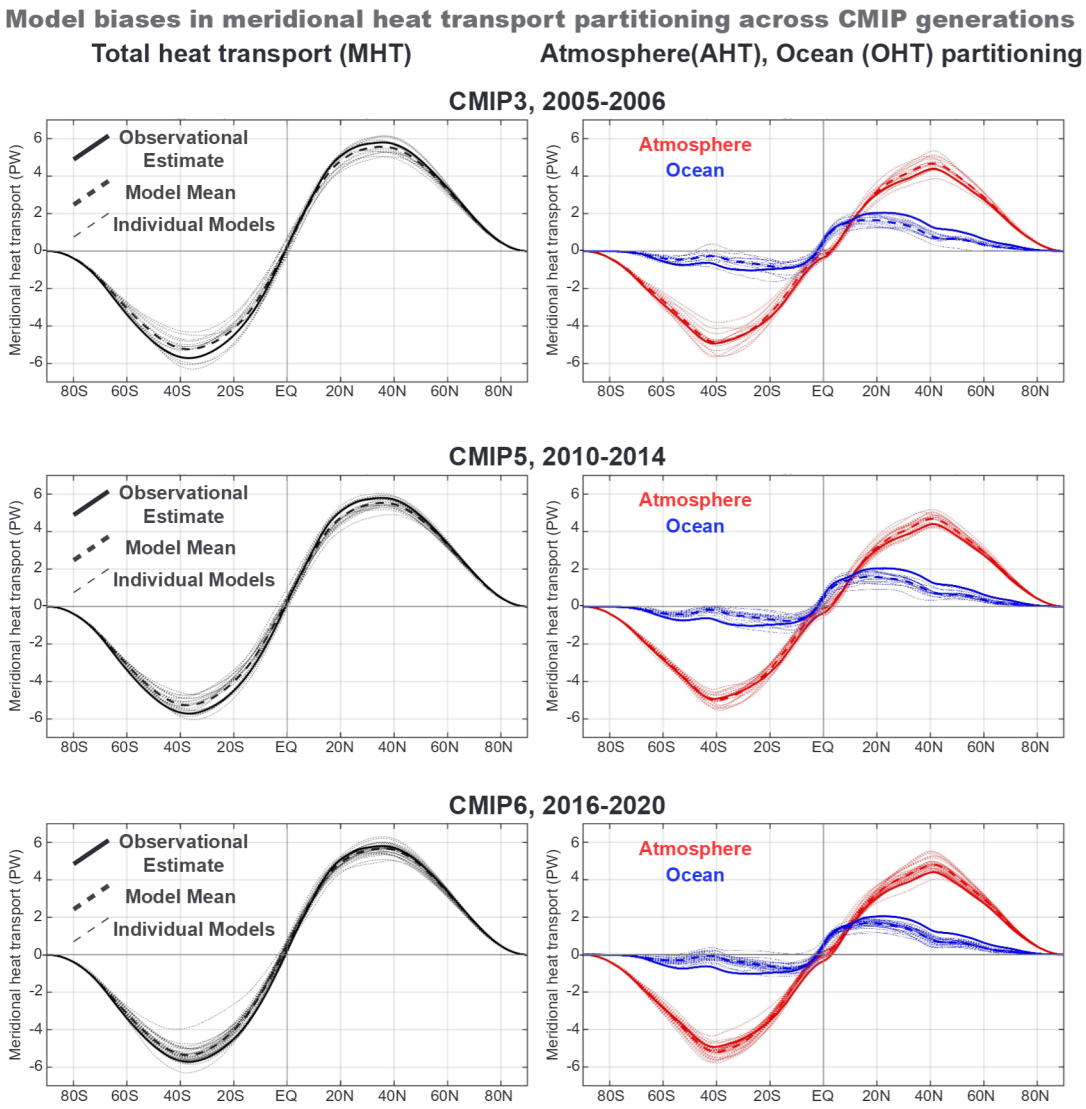

Figure 2. Comparison of meridional heat transport partioning in the atmosphere and ocean between observations and models. The left panels show the total (atmosphere plus ocean) meridional heat transport (MHT) in models (dashed lines with ensemble mean shown by the thicker dashed line) and observations (solid) line. The right panels show the atmospheric heat transport (AHT) in red and ocean heat transport (OHT) in blue. The top panel shows CMIP3 models. The middle panel shows CMIP5 models. The bottom panel shows CMIP6 models.

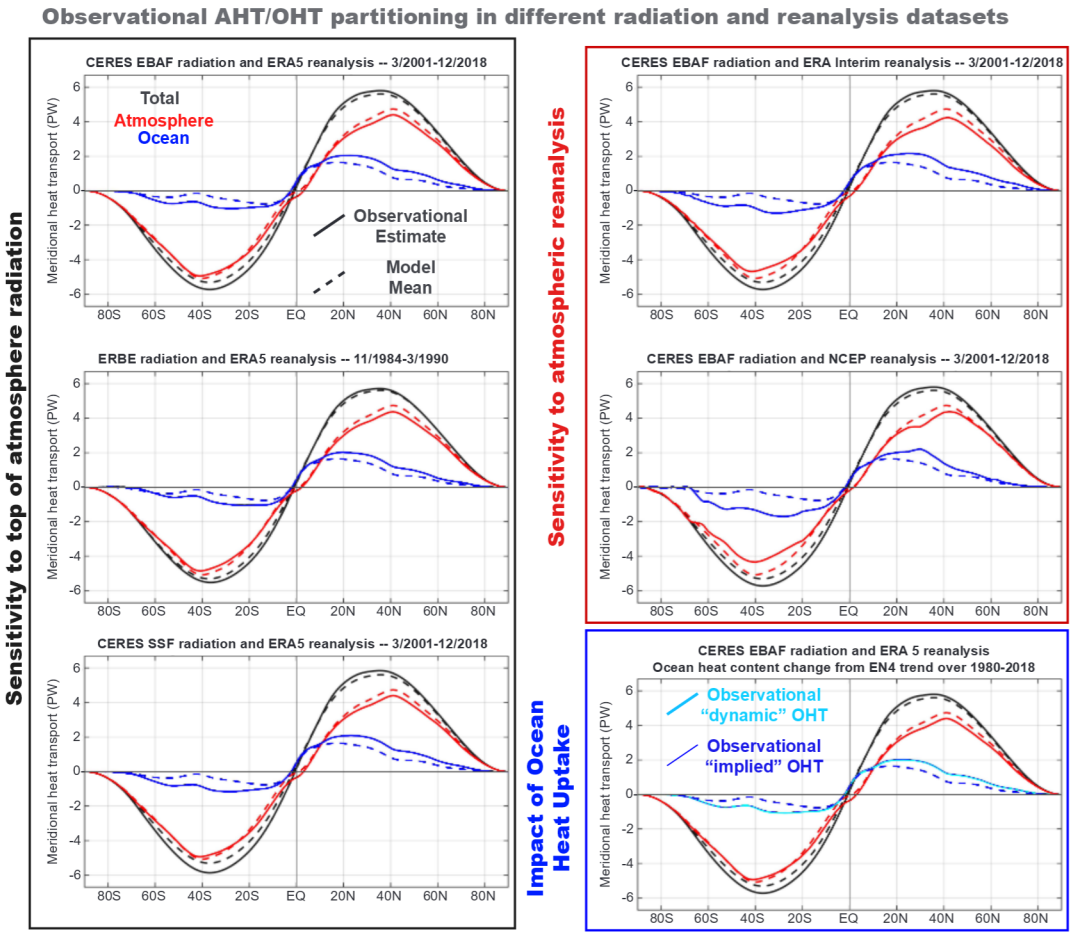

Figure 3. Six different observational estimates of meridional heat transport (black solid) partitioned into atmospheric (AHT, red) and oceanic (OHT, blue) contributions. The model mean from CMIP3, CMIP5 and CMIP6

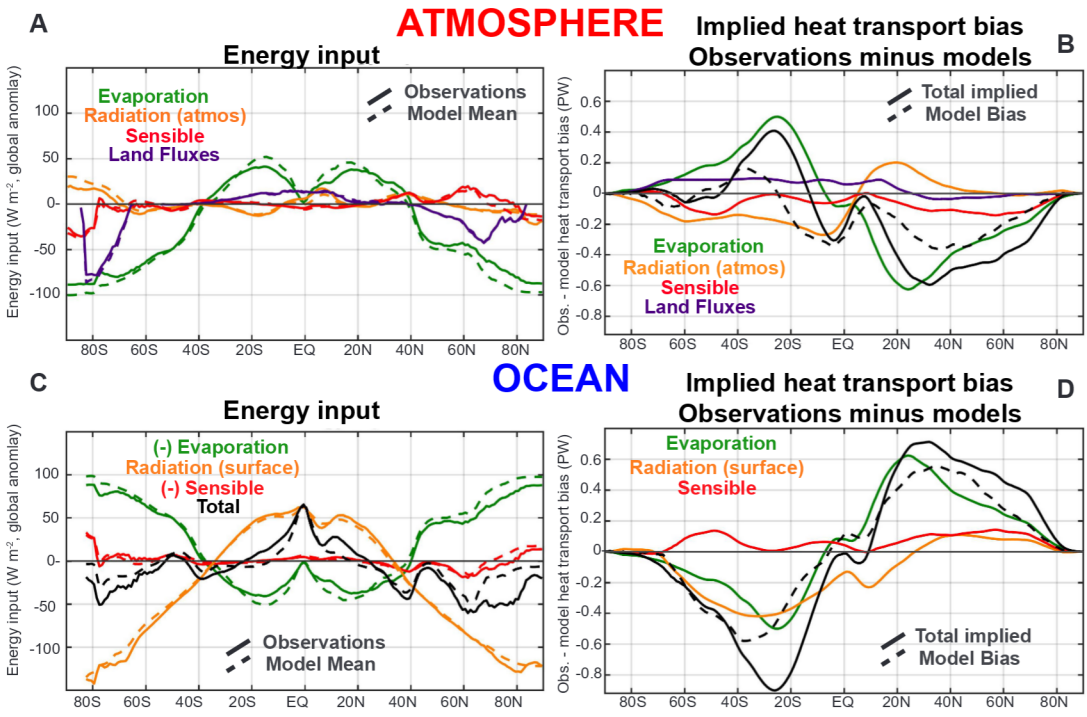

Figure 4. Model and observational estimates of the energy input into the atmopshere and ocean and the implied AHT and OHT biases resulting from each input. (A) Global anomaly energy input into the atmosphere in models (dashed) and observations (solid). See text for definition of terms. (B) Implied AHT bias (observations minus models) due to each energy input. The solid black line shows the sum of all terms and the dashed black line shows the bias 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.

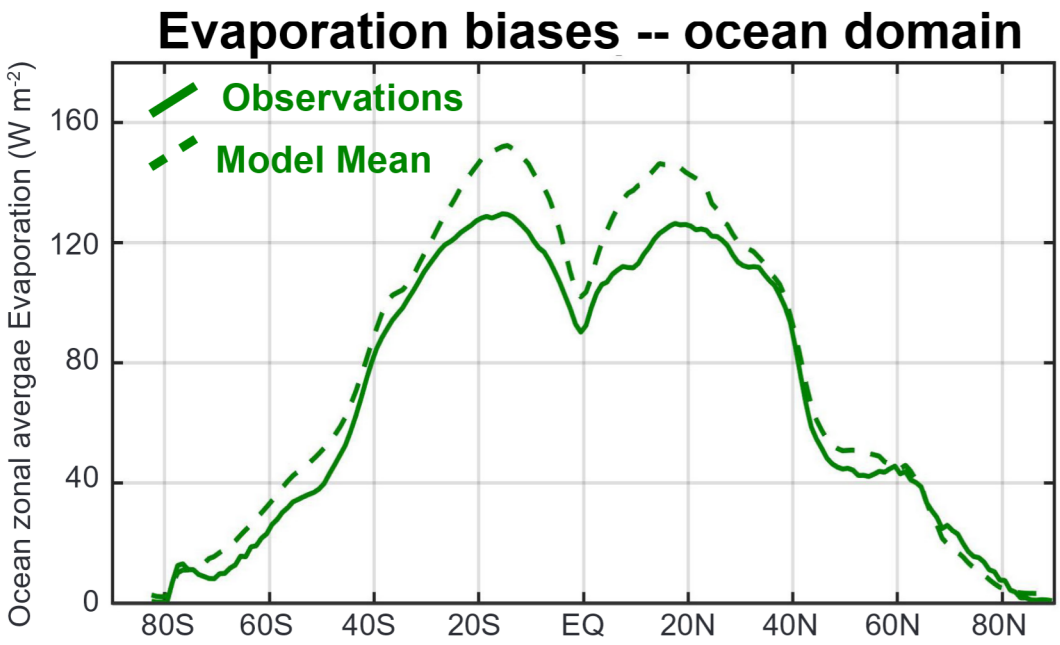

Figure S1. Comparison between evaporation over the ocean in models (ensemble mean) and observations (WHOI OA flux). All values show the annual mean average over the ocean domain and are expressed as latent heat fluxes in W m^{-2}.

Appendix A Open Research

AGU requires an Availability Statement for the underlying data needed to understand, evaluate, and build upon the reported research at the time of peer review and publication.

Authors should include an Availability Statement for the software that has a significant impact on the research. Details and templates are in the Availability Statement section of the Data and Software for Authors Guidance: `https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors#availability`

It is important to cite individual datasets in this section and, and they must be included in your bibliography. Please use the type field in your bibtex file to specify the type of data cited. Some options include Dataset, Software, Collection, ComputationalNotebook. Ex:

Acknowledgments

AD and KCA acknowledge support from National Science Foundation Award CLD-201964. RF was supported by the NOAA Climate and Global Change Postdoctoral Fellowship programs for the Advancement of Earth System Science (CPAESS) under award NA18NWS4620043B.

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).

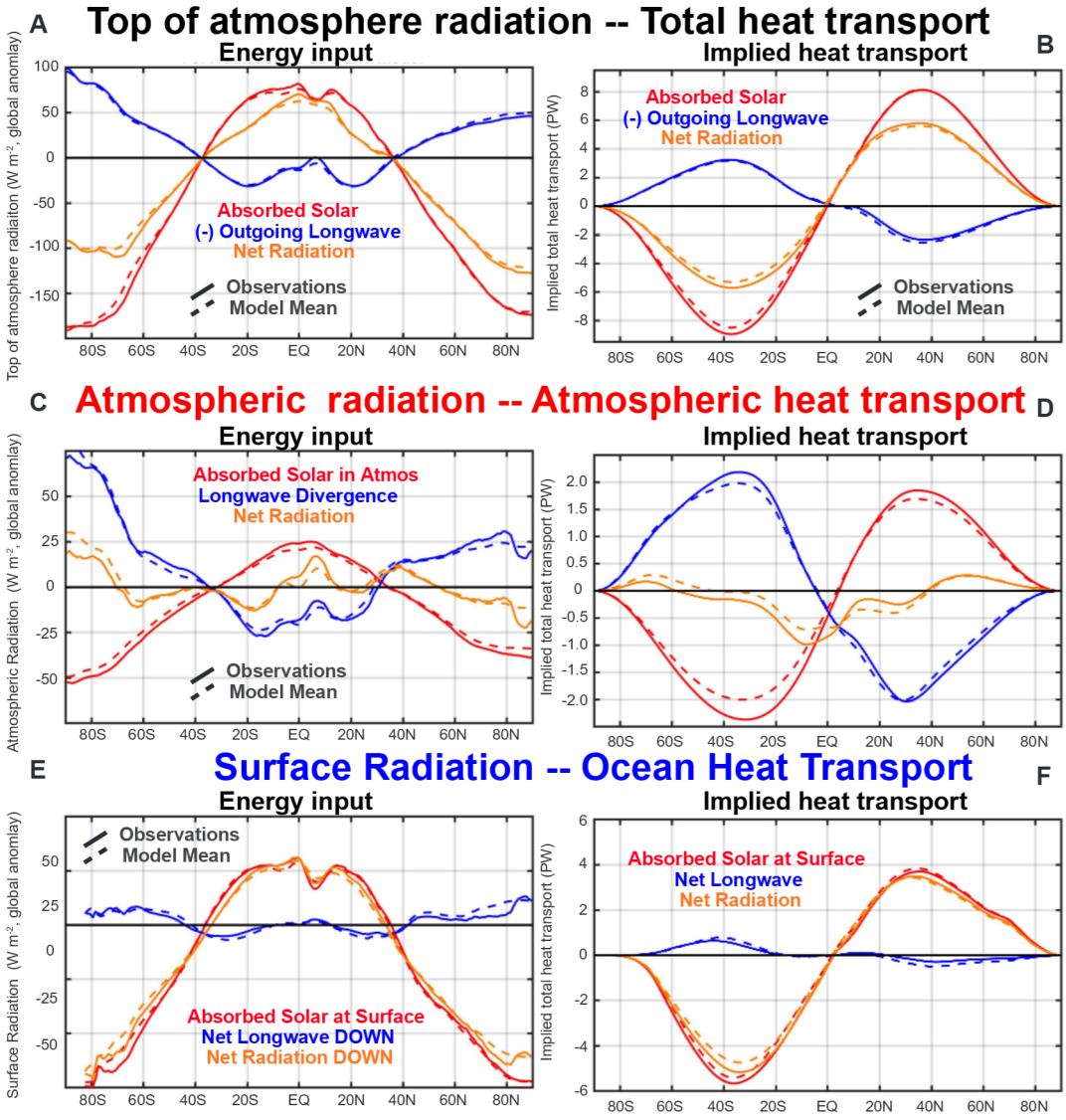

Figure S2. Comparison of radiation in observations (solid lines) and model ensemble mean (dashed lines) at the top of atmosphere (A), in the atmospheric column (C) and at the surface (E). Shortwave fluxes are shown in red, longwave fluxes are shown in blue and the net radiation is shown in orange with 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 between panels.

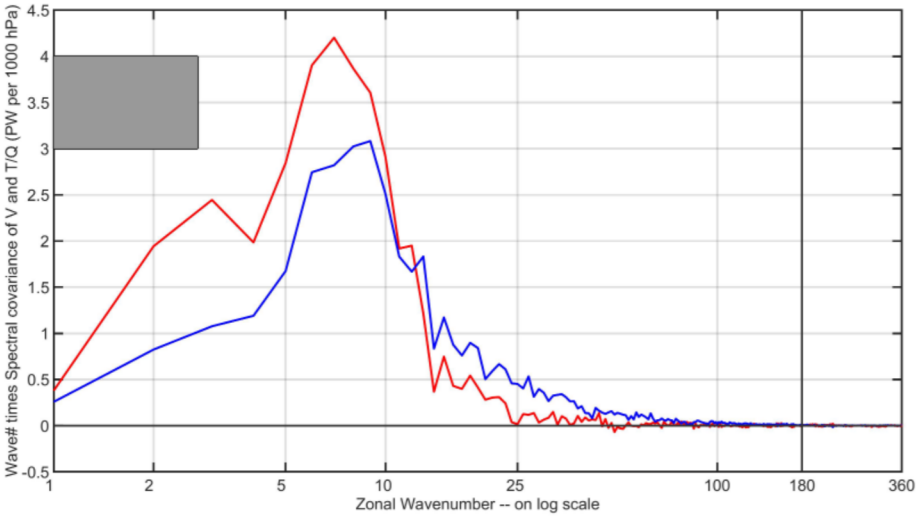

Figure S3. The spectra of atmospheric heat transport at 40N and 700 hPa. The red line shows the spectral co-variance of meridional velocity and temperature (time the specific heat of dry air) and the blue line shows the spectral covariance of meridional velocity and specific humidity (times the latent 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.

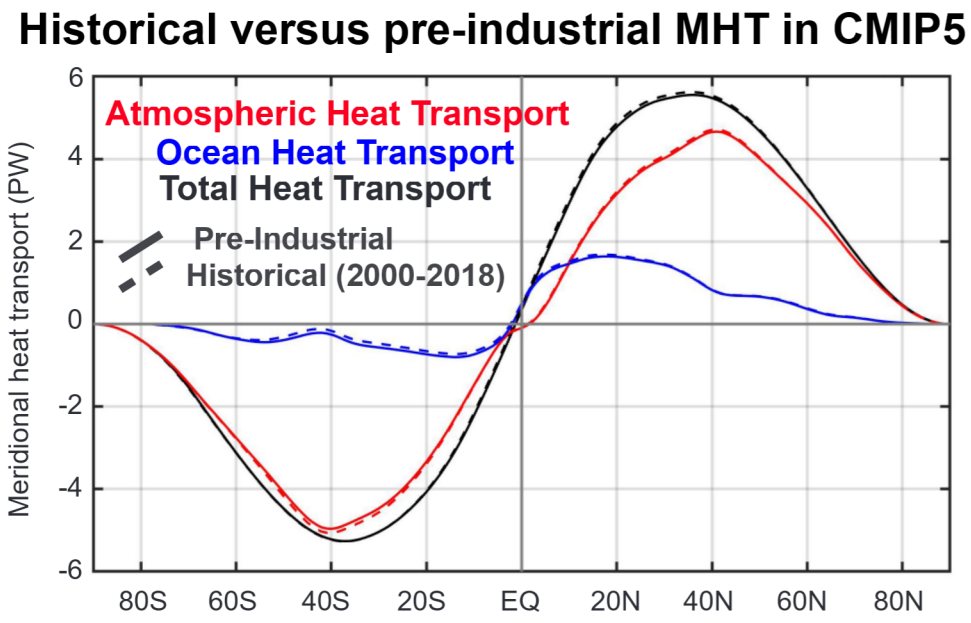

Figure S4. Comparison of the MHT/AHT/OHT partitioning between CMIP historical simulations (dashed lines) and pre-industrial simulation (solid lines). The MHT is shown in black. The AHT is shown in red. The OHT is shown in blue. Both lines are the ensemble mean of the 12 models that have sufficient output 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.

doi: 10.3390/atmos9040138

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.

Broecker, W., Peteet, D., & Rind, D. (1985). Does the ocean–atmosphere system have more than one stable mode of operation? *Nature*, *315*(6014), 21-26.

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

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])

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., Atwood, A. R., & Byrne, M. P. (2019). Controls on the width of tropical precipitation and its contraction under global warming. *Geophys. Res. Lett.*, *46*, 9958-9967. doi: 10.1029/2019GL082969

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.

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

Fajber, R., Donohoe, A., & Armour, K. (2021). *Investigating the impact of the vertical structure of atmospheric heat transport on arctic climate feedbacks using a single column model.* Retrieved from `https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/970063` (AGU Fall meeting)

Fajber, R., Donohoe, A., Ragen, S., Armour, K., & Kushner, P. (2022). The hydrologic cycle drive atmospheric heat transport. *Proc. Nat. Acad. Sci.*, *338*(6108), 1-7. doi: Submitted

Fajber, R., & Kushner, P. (2021). The hydrologic cycle drive atmospheric heat transport. *J. Atmos. Sci.*, *78*, 2161-2176.

Feldl, N., & Roe, G. (2013). The nonlinear and nonlocal nature of climate feedbacks. *J. Climate*, *26*, 8289-8304.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., ... Zhang, H. (2021). The earth's energy budget, climate feedbacks, and climate sensitivity [Book Section]. In M.-D. V. et al. (Eds.), *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (chap. 7). Cambridge

884 University Press.
885 Frierson, D., Hwang, Y., Fuckar, N., Seager, R., Kang, S., Donohoe, A., ... Battisti,
886 D. (2013). Why does tropical rainfall peak in the northern hemisphere? the role
887 of the oceans meridional overturning circulation. *Nat. Geo. Sci.*, *6*, 940-944.
888 Frierson, D. M. W., & Hwang, Y. T. (2012). Extratropical influence on ITCZ shifts
889 in slab ocean simulations of global warming. *J. Climate*, *25*, 720-733.
890 Good, S. A., Martin, M., & Rayner, N. (2013). En4: quality controlled ocean tem-
891 perature and salinity profiles and monthly objective analyses with uncertainty
892 estimates. *J. Geophys. Res.-Oceans*, *118*, 6704-6716.
893 Hahn, L., Armour, K., Battisti, D., Donohoe, A., Pauling, A., & Bitz, C. (2020).
894 Antarctic elevation drives hemispheric asymmetry in polar lapse rate climatology
895 and feedback. *grl*, *47*(16), e2020GL088965.
896 Hersbach, H. B. B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz?Sabater, J.,
897 Nicolas, J., & Peubey, C. (2020). The era5 global reanalysis. *Quart. J. Roy.*
898 *Meteor. Soc.*. doi: 10.1002/qj.3803
899 Holland, M. M., Bitz, C., & Tremblay, B. (2006). Future abrupt reductions in the
900 summer arctic sea ice. *Geophys. Res. Lett.*, *33*(23). doi: 10.1029/2006GL028024
901 Hwang, Y., & Frierson, D. (2013). Link between the double-intertropical conver-
902 gence zone problem and cloud bias over southern ocean. *Proc. Nat. Acad. Sci.*,
903 *110*, 4935-4940.
904 Johnson, G., Lyman, J., & Loeb, N. (2016). Improving estimates of earth's energy
905 imbalance. *Nat. Clim. Chang.*, *6*(7), 639-640.
906 Kang, S., Seager, R., Frierson, D., & Liu, X. (2014). Croll revisited: Why is the
907 northern hemisphere warmer than the southern hemisphere? *Climate Dyn.*, 16
908 pages. doi: 10.1007/s00382-014-2147-z
909 Kato, S., & Coauthors. (2018). Surface irradiances of edition 4.0 clouds and the
910 Earth's radiant energy system (CERES) energy balanced and filled (EBAF) data
911 product. *J. Climate*, *31*(11), 4501-4527.
912 Kim, H., Kang, S., Takahashi, K., Donohoe, A., & Pendergrass, A. (2020). Mecha-
913 nisms of tropical precipitation biases in climate models. *Climate Dyn.*, In Press.
914 Liang, M., Czaja, A., Graversen, R., & Tailleux, R. (2018). Poleward energy trans-
915 port: is the standard definition physically relevant at all time scales? *Climate*
916 *Dyn.*, *50*, 1785-1797.
917 Liu, C., Allan, R., Berrisford, P., Mayer, M., Hyder, P., Loeb, N., & Smith, D.
918 (2015). Combining satellite observations and reanalysis energy transports to
919 estimate global net surface energy fluxes 1985-2012. *jgr*, *18*, 9374-9389.
920 Loeb, N. G., & Coauthors. (2018). Clouds and the Earth's radiant energy system
921 (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition 4.0
922 data product. *J. Climate*, *31*(2), 895-918.
923 Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S.,
924 ... Wong, T. (2009). Towards optimal closure of the earth's top-of-atmosphere
925 radiation budget. *J. Climate*, *22*, 748-766.
926 Lucarini, V., & Ragone, F. (2011). Energetics of IPCC4AR4 climate models: energy
927 balance and meridional enthalpy transports. *Rev. Geophys.*, *49*, RG1001.
928 Marshall, J., Donohoe, A., Ferreira, D., & McGee, D. (2013). The oceans role in
929 setting the mean position of the inter-tropical convergence zone. *Climate Dyn.*, 14.
930 doi: 10.1007/s00382-013-1767-z
931 Marshall, J., & Speer, K. (2012). Closure of the meridional overturning circulation
932 through southern ocean upwelling. *Nat. Geo. Sci.*, *5*(3), 171-180.
933 Mayer, J., Mayer, M., & Haimberger, L. (2021). Consistency and homogeneity of at-
934 mospheric energy, moisture, and mass budgets in ERA5. *jcli*, *34*(10), 3955-3974.
935 Mayer, M., Haimberger, L., Edwards, J., & Hyder, P. (2017). Toward consistent di-
936 agnostics of the coupled atmosphere and ocean energy budgets. *jcli*, *22*(30), 9225-
937 9246.

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.

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.

Overland, J., & Turet, P. (1994). Variability of the atmospheric energy flux across 70°n computed from the gfdl data set. *Centennial Volume, Geophys. Monogr.*, *84*, 313-325.

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

Roe, G. (2009). Feedbacks, timescales, and seeiing red. *Annu. Rev. Earth Planet. Sci.*, *37*, 930115.

Rose, B., & Ferreira, D. (2013). Ocean heat transport and water vapor greenhouse in a warm equable climate: a new look at the low gradient paradox. *J. Climate*, in press.

Shaw, T., Barpanda, P., & Donohoe, A. (2018). A moist static energy framework for zonal-mean storm-track intensity. *J. Atmos. Sci.*, *75*(6), 1979-1994.

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.

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.

Trenberth, K. E. (1997). Using atmospheric budgets as a constraint on surface fluxes. *J. Climate*, *10*, 2796-2809.

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. (2010). Simulation of present day and 21st century energy budgets of the southern oceans. *J. Climate*, *23*, 440-454.

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. , *6*, 138-144. doi: 10.1038/nclimate2876

Wu, Y., Ting, M., Seager, R., Huang, H., & Cane, M. (2011). Changes in storm tracks and energy transports in a warmer climate simulated by the gfdl cm2.1 model. *Climate Dyn.*, *37*, 53-72.

Wunsch, C. (2005). The total meridional heat flux and its oceanic and atmospheric partition. *J. Climate*, *18*, 4374-4380.

Yang, H., Li, Q., Wang, K., Sun, Y., & Sun, D. (2015). Decomposing the meridional heat transport in the climate system. *Climate Dyn.*, *44*(9), 2751-2768.

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.