A Mass and Energy Conservation Analysis of Drift in the CMIP6 Ensemble

March 19, 2021

Abstract

Coupled climate models are prone to ‘drift’ (long-term unforced trends in state variables) due to incomplete spin-up and non-closure of the global mass and energy budgets. Here we assess model drift and the associated conservation of energy, mass and salt in CMIP6 and CMIP5 models. For most models, drift in globally-integrated ocean mass and heat content represents a small but non-negligible fraction of recent historical trends, while drift in atmospheric water vapor is negligible. Model drift tends to be much larger in time-integrated ocean heat and freshwater flux, net top-of-the-atmosphere radiation (netTOA) and moisture flux into the atmosphere (evaporation minus precipitation), indicating a substantial leakage of mass and energy in the simulated climate system. Most models are able to achieve approximate energy budget closure after drift is removed, but ocean mass budget closure eludes a number of models even after de-drifting and none achieve closure of the atmospheric moisture budget. The magnitude of the drift in the CMIP6 ensemble represents an improvement over CMIP5 in some cases (salinity and time-integrated netTOA) but is worse (time-integrated ocean freshwater and atmospheric moisture fluxes) or little changed (ocean heat content, ocean mass and time-integrated ocean heat flux) for others, while closure of the ocean mass and energy budgets after drift removal has improved.

1 Introduction

In the climate modeling community, unforced trends in coupled model simulations are commonly referred to as model drift. Given the potential for drift to contaminate forced signals in climate simulations, it has been a topic of interest throughout the phases of the Coupled Model Intercomparison Project (CMIP). For CMIP2+ (?), CMIP3 (?) and CMIP5 (?), drift represented a non-negligible fraction of historical forced trends in global depth-integrated quantities such as ocean heat content (OHC) and steric sea level over recent decades. For surface and atmospheric variables such as global mean temperature or precipitation (i.e. variables that are less influenced by the slowly evolving deep ocean) drift is less important, but on regional scales it can still represent a substantial fraction of recent historical trends (?).

There are a number of causes of drift in coupled climate models. When a model simulation is initiated, an imbalance inevitably exists between the prescribed initial state (which is commonly estimated from observations) and the representation of physics in the model (i.e. the simulated ocean dynamics, advection and mixing). A coupling shock may also occur when the various model components (e.g. atmosphere, ocean, sea-ice) are first joined together, resulting in discontinuities in boundary fluxes (e.g. ?). In response, a model will typically drift from its initial state towards a quasi-steady state over time. The timescale over which the system reaches equilibrium depends on how long it takes anomalies to be advected or mixed through the deep ocean, which is typically many thousands of years (e.g. ?). The adjustment of the atmosphere and land surface is much faster. The most obvious solution to this issue would be to let the model run to equilibrium before performing any experiments of interest. The problem is that state-of-the-art coupled climate models are computationally expensive, which makes a 'spin-up' period of many thousands of years impractical. Instead, models are generally spun up for a few hundred years. Experiments will therefore exhibit changes/trends associated with incomplete model spin-up, as well as changes related to external forcing or internal climate variability. The overall reduction in drift from CMIP2+ to CMIP5 has been primarily attributed to longer spin-up times and more careful initialization of the coupled ocean-atmosphere system (**?**).

In addition to incomplete model spin-up, drift is also caused by spurious mass or energy 'leakage' into or out of the simulated climate system. This non-closure of the global mass and energy budgets arises due to small inconsistencies in the model treatment of energy (**??**) and/or water (**??**). In relation to the global energy budget, an essential characteristic is a close correspondence between the globally integrated net top-of-the-atmosphere radiation (netTOA) and OHC, because the latter represents Earth's primary energy store (**?**). In CMIP5 models, the difference between the time-integrated global netTOA and changes in OHC is overwhelmingly characterized by an approximately time-constant bias that is insensitive to changes in model forcing (i.e., it is the same for all experiments; **?**). This means it is generally possible to correct (or 'de-drift') output from a coupled model experiment by subtracting a drift signal taken from the corresponding control experiment. When calculating this drift signal there is the potential to over-fit and thus remove low-frequency signals associated with internal variability, so there are a number of (somewhat subjective) decisions to be made about fitting a linear or higher-order polynomial (or high-pass filter) to either the full length or a shorter segment the control time series (**?**). Once the data have been de-drifted, most CMIP5 models are approximately energy conserving (**?**).

The practice of de-drifting is commonplace in studies concerned with forced trends in model variables that have an obvious link to the slowly evolving deep ocean (e.g. OHC and steric sea level; **??**), but it is less well understood and applied in the context of net time-integrated heat and water fluxes into the atmosphere and ocean. For instance, changes in meridional transports of ocean heat and freshwater can be inferred from cumulative surface heat and freshwater fluxes (e.g. ???) and changes in ocean salinity can be used to infer global water cycle changes (?), but only if there is approximate closure of the relevant global budgets. If model leakage causes a substantial mismatch between changes in global OHC and the time-integrated net ocean heat flux, for instance, then any inferred change in meridional ocean heat transport is invalid. If de-drifting does not restore budget closure for any particular model, then that model may need to be excluded from the analysis ensemble (e.g. ??).

In this study, we extend the physically-based approach to drift analysis used by ? by considering both energy *and* mass conservation in the CMIP6 ensemble (?) before and after de-drifting. Relevant comparisons are made with the CMIP5 ensemble (?) in order to report on progress/improvements.

2 Methods

In order to assess drift in the CMIP6 and CMIP5 ensembles, we analyze data from the pre-industrial control (piControl) experiment. For each model, the drift in globally-integrated OHC is decomposed into a temperature and barystatic (mass-related; ?) component. This decomposition provides insights into the cause of the drift, and the temperature component can be compared against the time-integrated heat flux into the ocean to assess energy conservation. To assess ocean mass and salt conservation, we compare the global ocean mass to the time-integrated surface freshwater flux and global mean salinity, respectively (remembering that the ocean integrated salt content should be constant). Similarly, atmospheric mass conservation is assessed by comparing the global mass of water in the atmosphere to the time-integrated moisture flux into the atmosphere (i.e. evaporation minus precipitation). Each quantity in the OHC decomposition and mass and energy conservation analysis is derived/defined below.

2.1 Ocean heat decomposition

The amount of thermal energy stored in the global ocean is proportional to $c_p M T$, where c_p (units J(kg K)$^{-1}$) is the specific heat of seawater (a constant in the models), M (kg) the mass of the ocean and T (K) the average temperature of the ocean. The rate of heat gain or loss can therefore be represented as,

$$c_p \left[M \frac{dT}{dt} + T \frac{dM}{dt} \right]$$

where the left-hand term captures any gain or loss of heat related to a change in ocean temperature and the right-hand term represents any gain or loss of heat related to a change in ocean mass (the non-linear terms in the decomposition are negligible). For the purposes of this study, we therefore decompose the globally integrated OHC anomaly (H) into a temperature (H_T) and mass/barystatic (H_M) component,

$$H_T(t) = c_p \, M_0 \, \Delta T(t) \tag{1}$$

$$H_M(t) = c_p \, T_0 \, \Delta M(t) \tag{2}$$

where X_0 is the value of X at the first time step and ΔX the change in X since the first time step (i.e. $\Delta X = X(t) - X_0$).

2.2 Conservation

In an energy conserving coupled climate simulation, any change in the temperature component of global OHC should (on annual and longer timescales) be in response to a time-integrated net heat flux into the ocean,

$$\frac{dH_T}{dt} \approx \frac{dQ_h}{dt}, \qquad Q_h(t) = \int_0^t \int_{A_o} q_h(t,i,j)\ dA_o\ dt \tag{3}$$

where $A_o(i,j)$ (m^2) is the grid cell areas of the surface ocean and q_h (W m^{-2}) the net heat flux into the ocean; this net heat flux includes the net surface heat flux and for a small number of CMIP6 models an upward geothermal flux at the sea floor (Table S1). Similarly, any change in the mass of the global ocean should be in response to a time-integrated net freshwater flux,

$$\frac{dM}{dt} \approx \frac{dQ_m}{dt}, \qquad Q_m(t) = \int_0^t \int_{A_o} q_m(t,i,j)\ dA_o\ dt \tag{4}$$

where q_m (kg m^{-2} s^{-1}) is the net freshwater flux into the ocean (including runoff).

With respect to the atmosphere, the mass of global water vapor can be taken to represent the total mass of atmospheric water (M_a), since the globally-integrated mass of condensed water and ice in clouds is negligible (<1% of the total atmospheric water mass in the CMIP models). Any change in the mass of atmospheric water should be in response to a time-integrated net atmospheric moisture flux,

$$\frac{dM_a}{dt} \approx \frac{dQ_{ep}}{dt}, \qquad Q_{ep}(t) = \int_0^t \int_{A} q_{ep}(t,i,j)\ dA\ dt \tag{5}$$

where $A(i,j)$ (m^2) is the surface grid cell areas and q_{ep} (kg m^{-2} s^{-1}) the net atmospheric moisture flux (evaporation minus precipitation). There is a column-integrated but not global-integrated water vapor CMIP diagnostic, so it was necessary to calculate the global value as follows,

$$M_a(t) = \int_A w(t,i,j)\ dA \tag{6}$$

where w (kg m^{-2}) is the column-integrated atmospheric mass content of water vapor.

The drifts in global oceanic mass, atmospheric mass and OHC are approximately linear (e.g. Figures 1 and 2), so the time derivatives defined above were calculated as a simple linear trend (using Ordinary Least Squares regression) over the length of the control simulation. Any significant residual in Equations 3, 4 or 5 indicates a spurious source/sink of heat or mass in the simulated climate system, which we refer to as model leakage.

To put the magnitude of the model drifts into perspective, we compare them to estimates of current observed trends. For the global energy budget, we compare against estimates of the planetary energy imbalance, which range from 0.4–1.0 W m^{-2} for various estimation methods and time periods over the last couple of decades (**??**). This comparison is achieved by dividing the model energy drift by the planetary surface area of 5.1×10^{14} m^2. For the ocean mass budget, we compare against the current barystatic sea level rise of 1.8 mm/year (or approximately 6.6×10^{14} kg/year). This value represents 58% of the estimated total (i.e. steric plus barystatic) global sea level rise during the altimetry era (3.1 mm/year from 1993–present), as per the findings of **?**. Finally, for the atmospheric mass budget we compare against a constant relative-humidity warming rate of 1.68×10^{13} kg/year. This value represents the Clausius–Clapeyron response of 7% °C^{-1} to a trend in global average surface temperature of approximately 0.2 °C/decade over the 1990–2019 period (from the NOAA Merged Land Ocean Global Surface Temperature Analysis Version 5; **?**) for an approximate average mass of water vapor in the CMIP atmospheres of 1.2×10^{16} kg.

To compliment our analysis of energy and mass conservation in the CMIP oceans and atmospheres, we also consider energy conservation for the entire climate system by comparing the time-integrated global netTOA (Q_r) and ocean heat storage,

$$\frac{dH_T}{dt} \approx \frac{dQ_r}{dt}, \qquad Q_r(t) = \int_0^t \int_A q_r(t,i,j)\, dA\, dt \tag{7}$$

where q_r (W m^{-2}) is the netTOA. Since the global ocean is the main energy reservoir for the climate system, changes in OHC should approximately balance the time-integrated netTOA on annual and longer timescales (**??**). It is estimated that 89% of the current planetary energy imbalance is absorbed by the ocean, with the rest primarily partitioned into melting ice and warming the land (**?**). Since this melting is not completely captured by the CMIP5 and CMIP6 models (the models do not include dynamic ice sheets), a percentage even closer to 100% applies when comparing the model-derived netTOA and OHC.

Finally, the ocean should also conserve salt. In particular, any change in global-mean salinity (S) should be in response to a change in the global ocean mass (which in turn should be in response to a time-integrated net freshwater flux; Equation 4). In order assess budget closure, we relate a change in global-mean salinity (between time 0 and t) to an expected/equivalent change in ocean mass (ΔM) as follows:

$$\Delta M = M_0 \left(\frac{S_0}{S_t} - 1 \right) . \tag{8}$$

We note that while there is a net time-integrated salt flux into the ocean from rivers and/or sea ice in some models, its influence on global-mean salinity is negligible compared to the influence of ocean mass changes and is thus ignored in this study.

2.3 Model diagnostics

Each of the variables discussed in the equations above (Table 1) can be related to a CMIP diagnostic/s (Table 2). Detailed definitions for each diagnostic are available from the CMIP5 standard output (**?**) and CMIP6 data request (**?**) documentation, with additional information regarding ocean diagnostics provided by **?**. Tables S1-4 provide precise details of exactly which diagnostics and data file versions were used for each model in this study. We note that none of the models for which we present ocean surface heat or water flux results archived a heat flux correction (*hfcorr*) or water flux correction (*wfcorr*) diagnostic, respectively.

Identifying the correct diagnostics for use in this study was mostly straightforward, except in the case of the global ocean mass. Almost all of the CMIP6 and CMIP5 ocean models apply a Boussinesq approximation, which means volume is conserved rather than mass (and sea water density is only considered in so far as it influences ocean dynamics). As such, steric processes (i.e. contraction/expansion of sea water due to temperature and/or salinity change) are represented as a change in density, from which an implied change in mass is often inferred and reported by modeling groups (the so-called Boussinesq ocean mass), as opposed to the real world where temperature and/or fresh water input leads to direct changes in ocean volume. To avoid any confusion, Boussinesq models in CMIP6 were asked to archive a global ocean mass variable (*masso*) equal to the reference density (*rhozero*) multiplied by the ocean volume (*volo*), as opposed to the Boussinesq ocean mass (**?**). A small number of modeling groups did not follow this direction (and it was not a requirement for Boussinesq models in CMIP5), so for those models we performed the density-times-volume calculation in order to obtain the variable M used in the equations above. All models for which a global ocean mass time series could be constructed were included in the final ensemble (Table 3). The small number of (mostly CMIP5) models that archived a virtual salt flux diagnostic were left out of the ensemble, as it is not clear from the CMIP documentation how those fluxes impact/modify the global ocean mass, salinity and surface water flux diagnostics. Monthly mean data were converted to annual mean (accounting for the different number of days in each month) prior to analysis and results for only the first member from each model ensemble is presented, because all ensemble members from a given model tended to produce similar results.

The change in the definition of the global ocean mass diagnostic for CMIP6 means that for models that apply a Boussinesq approximation (which is almost all the models), neither the global mass nor volume diagnostics respond to steric processes – they both only respond to barystatic changes. It is possible to derive some steric information from the global average thermosteric sea level change diagnostic (*zostoga*), however another new development in CMIP6 is that the full steric sea level change (*zossga*) is not archived. That diagnostic would incorporate thermosteric changes, halosteric changes and the so-called non-Boussinesq steric effect, which relates to reorganization of ocean mass (**?**). In the absence of any diagnostic that fully captures steric changes, our analysis does not consider changes to the volume of the ocean.

3 Results

3.1 Example model

In order to illustrate the various aspects of our analysis, the results for a typical model (ACCESS-CM2; **?**) are shown in Figure 1. The first thing to note is the clear drift / non-zero trend in OHC (black curve, Figure 1a). If the model were energy conserving, the time series corresponding to the OHC temperature component anomaly (red curve), time-integrated ocean surface heat flux (orange curve) and time-integrated netTOA (gold curve) would approximately overlay one another, as per Equations 3 and 7. To put the magnitude of these drifts into perspective, the linear trend in those time series is 0.18 W m^{-2}, 0.02 W m^{-2} and 0.37 W m^{-2} respectively. These values (and the leakage of approximately 0.19 W m^{-2} between the TOA and ocean storage) are trivial compared to the corresponding climatological energy flows in the climate system, but are not an insignificant fraction of the anthropogenic signal (i.e. the current planetary energy imbalance of 0.4–1.0 W m^{-2}).

Similar principles apply for the ocean mass budget (Figure 1b). The time series corresponding to the ocean mass anomaly (blue curve) and time-integrated freshwater flux (grey curve) approximately overlay one another, indicating approximate water conservation. The corresponding linear trend is equivalent to a drop in global sea level of 0.2 mm/year, which is trivial compared to individual surface freshwater fluxes (e.g. the annual precipitation or evaporation flux) but is not an insignificant fraction of the estimated current rate of barystatic sea level rise (1.8 mm/year). Global mean salinity has been converted to an equivalent change in ocean mass (as per Equation 8; green curve) and it also approximately overlays the ocean mass time series, indicating approximate salt conservation. Finally, it is clear that the atmosphere does not conserve water (Figure 1c). The drift in the mass of atmospheric water vapor is negligible (linear trend of 3.2×10^{11} kg/year), but the drift in time-integrated water flux into the atmosphere (i.e. evaporation minus precipitation; -1.8×10^{14} kg/year) is not. While trivial compared to the individual annual fluxes of precipitation or evaporation, the magnitude of the drift in time-integrated atmospheric water flux is larger than our estimated observed trend in atmospheric water vapor ($+1.68 \times 10^{13}$ kg/year) and represents a loss of approximately 1.5% of total atmospheric water vapor every year.

Given that the ACCESS-CM2 model does not conserve energy and atmospheric mass, it is important for data users to know whether conservation can be achieved after de-drifting. To test this, we quantify the drift signal by fitting a cubic polynomial to the full-length of various time series shown in Figure 1a-c. That signal is then subtracted from the original time series in order to produce corresponding de-drifted time series (Figure 1d-f). Approximate conservation is achieved for the energy and ocean mass budget after drift removal, but the atmospheric moisture budget time series still do not overlay one another. In a practical sense, this means that after de-drifting the mass and heat content of the global ocean responds appropriately to time-integrated changes in surface heat and freshwater fluxes, whereas the mass of water vapor in the atmosphere does not respond in a physically consistent manner to time-integrated changes in precipitation and evaporation. This is problematic for data users looking to infer anomalous atmospheric moisture transports (for instance) from regional changes in water vapor and evaporation minus precipitation. With this description of an example model in mind, we can expand our analysis to the entire CMIP6 (and CMIP5) ensemble.

3.2 Drift and conservation

3.2.1 Temporal evolution

We begin our description of the CMIP6 ensemble by considering the temporal evolution of the drift in globally-integrated ocean mass and heat content (drift in atmospheric water vapor is negligible and thus not shown). Drift in both quantities is overwhelmingly characterized by linear trends that are relatively constant throughout the length of the control experiment (Figure 2a,c). To visualize any coherent drift signals other than the linear trends, detrended OHC and ocean mass time series were calculated (Figure 2b,d). The removed trend was estimated using Ordinary Least Squares regression on the annual mean time series. For most models, removal of the linear trends transforms the time series into stationary red noise, which is the expected regime under an equilibrium climate. However, some of the models show clear coherent signals, particularly in OHC. These signals could represent low frequency oscillations that are cut off by the control run length (i.e. multi-century variability in the models), but most appear to be an asymptotic progression to some stable 'red noise plus trend' state that is more indicative of incomplete spin-up (**?**). Of course, there is no way of testing this hypothesis unless the control simulation is run for long enough that either the second-order trend becomes zero (indicating the arrival at a stable state) or reverses (indicating oscillatory behaviour).

3.2.2 Energy budget

Since the ocean is the biggest energy reservoir in the climate system, we anchor our energy budget analysis around the drift in OHC. Similar to ACCESS-CM2 (Figure 1a), the drift in OHC is dominated by the temperature (as opposed to barystatic) component for essentially all models (Figure 3). The direction of that drift has a positive bias across the ensemble, which was also true for the CMIP3 ensemble (**?**). This is important because it means the drifts will not cancel in the calculation of an ensemble mean. While there are fewer outliers in CMIP6, the ensemble median magnitude of the drift in OHC is similar for CMIP5 and CMIP6 (Table 4).

Drift in OHC tends to be much smaller than for time-integrated netTOA, indicating a net leakage of energy in the simulated climate system (Figure 4a). In fact, while drift in OHC is typically a small but non-negligible fraction of the current planetary energy imbalance, the drift in time-integrated netTOA (and indeed the net system-wide energy leakage; Figure 5) is larger than the observed planetary imbalance for a number of models. Most of this leakage occurs somewhere between the TOA and ocean surface, as ocean energy leakage (i.e. the discrepancy between the time-integrated heat flux into the ocean and change in OHC temperature component; Figure 4b) is relatively modest. Similar to OHC, the ensemble median magnitude of the drift in time-integrated heat flux into the ocean has changed very little from CMIP5 to CMIP6. In contrast, the magnitude of the drift in time-integrated netTOA is substantially smaller in CMIP6, which explains the reduced total system energy leakage in CMIP6 (Table 4).

3.2.3 Mass budget

Drift in the ocean mass budget shares many similarities with the energy budget. Firstly, like drift in OHC, the magnitude of drift in global ocean mass typically represents a small but non-negligible fraction of observed trends (Figure 4c) and has changed very little from CMIP5 to CMIP6 (Table 4). Drift in time-integrated surface freshwater flux on the other hand is larger than observed sea level trends for a number of models (Figure 4c), indicating substantial non-closure of the ocean mass budget. The ensemble median magnitude of the drift in freshwater flux is larger/worse in CMIP6, due in part to a number of large outliers (Table 4). Many models do a relatively good job of conserving salt (Figure 4d) and the magnitude of the drift in ocean salinity has been reduced in CMIP6 (Table 4).

Given that atmospheric water vapor is not directly linked to the slowly evolving deep ocean, it is perhaps not surprising that the ensemble median drift magnitude (Table 4) represents a negligible fraction of estimated current trends (i.e. atmospheric variables tend not to exhibit much drift). The same cannot be said for the time-integrated moisture flux into the atmosphere (i.e. evaporation minus precipitation), which for most models is larger than estimated current trends in atmospheric water vapor (Figure 4e). In fact, for many models the gain or loss of water associated with the drift in time-integrated moisture flux represents an appreciable fraction of the total mass of atmospheric water vapor (1.2×10^{16} kg) every year. In the CMIP3 ensemble the drift in time-integrated atmospheric moisture flux was overwhelmingly negative (i.e. precipitation dominated over evaporation for most models; **?**) but the CMIP5 (**?**) and CMIP6 models are relatively evenly distributed between positive and negative drifts (Figure 4e). As was the case for the freshwater flux into the ocean, the ensemble median magnitude of the drift in time-integrated atmospheric moisture flux is larger in CMIP6 than it was in CMIP5 (Table 4).

3.3 De-drifting

With the exception of atmospheric water vapor, we've shown that the magnitude of the drift in various global energy and mass budget terms typically represents a non-negligible fraction of estimated current observed trends (OHC, ocean mass and time-integrated ocean heat flux) or approaches/exceeds the magnitude of those trends (time-integrated netTOA, ocean freshwater flux and atmospheric moisture flux). To avoid contamination of analyzed trends it is therefore important to quantify and remove this drift from forced experiments, particularly as the direction of the drift is biased for some variables (e.g. Figure 3) and thus will not cancel when calculating ensemble statistics. Since the temporal evolution of these drifts is quasi-linear (with slight curvature likely related to incomplete spinup; Figure 2) and insensitive to changes in model forcing (?), this can be achieved by fitting a simple polynomial (we fit a cubic, although a linear or quadratic fit yields similar results) to the control experiment and then subtracting the relevant segment of that polynomial from the forced data.

An additional motivation for de-drifting relates to budget closure. We saw earlier that approximate energy and ocean mass budget closure was achieved after de-drifting for the ACCESS-CM2 model (Figure 1d-e), but non-closure of the atmospheric water budget remained (Figure 1f). In order to extend this budget closure analysis to the entire ensemble, we regress the various (decadal mean) de-drifted time series against one another to test for corresponding variability (Figure 6). For reference, the ACCESS-CM2 linear regression coefficients were 0.99 (Q_r vs. Q_h), 0.98 (Q_r vs. H_T) and 0.98 (Q_h vs. H_T) for the de-drifted energy budget time series (Figure 1d); 0.89 (M vs. S), 1.02 (Q_m vs. M) and 0.91 (Q_m vs. S) for the ocean mass budget time series (Figure 1e); but only 0.39 (M_a vs. Q_{ep}) for the atmospheric water budget (Figure 1f).

Looking at the regression coefficients across the ensemble, it is clear that like ACCESS-CM2 most CMIP6 models show approximate energy budget closure after de-drifting (i.e. regression coefficients close to 1.0; Figure 6a). Most CMIP5 models also achieve approximate energy budget closure, but there has been a small improvement between CMIP5 and CMIP6.

Energy budget coefficients slightly less than 1.0 were common across the ensemble because the variance in the de-drifted time-integrated netTOA time series was typically marginally larger than the time-integrated heat flux into the ocean time series, which had a variance marginally larger than the OHC time series. The larger netTOA variance might be explained by the additional non-ocean heat stores represented in the models (e.g. continental energy storage; ?), but it is unclear why the time-integrated heat flux into the ocean would have a slightly larger variance than the OHC. In other words, while perfect/expected closure would normally be a regression coefficient of 1.0, for the comparisons against the time-integrated netTOA the expected coefficient might be slightly less than 1.0. See Figure S1 for energy budget regression coefficients for individual models.

In contrast to the energy budget, even after drift correction there remain large discrepancies between the time-integrated surface freshwater flux and both ocean mass and salinity for a number of models (Figure 6b). The ensemble median closure has improved from CMIP5 to CMIP6, but aside from the approximate closure between ocean mass and salinity the ocean mass budget closure across the CMIP6 ensemble falls short of that achieved for the energy budget (see Figure S2 for mass budget regression coefficients for individual models). Closure of the atmospheric mass budget after drift correction also eludes the models, with many showing essentially no meaningful relationship between variability in the de-drifted atmospheric water vapor and time-integrated moisture flux time series (Figure 6b).

4 Discussion

In the early coupled ocean-atmosphere models, drift was so large that it was necessary to constrain simulations via the use of offline flux adjustments (e.g. ?). This was still the case for most models participating in the first phase of CMIP (?), but with each subsequent CMIP iteration model drift improved to the point where no flux adjustment was required for most

CMIP5 models to achieve drifts in global, depth-integrated quantities (e.g. OHC or steric sea level) of magnitude less than corresponding forced historical trends (?). Our analysis suggests that when it comes to globally integrated OHC, there has been little improvement from CMIP5 to CMIP6 (fewer outliers, but a similar ensemble median magnitude). This indicates that model drift still represents a non-negligible fraction of historical forced trends in global, depth-integrated quantities and existing advice regarding the need to de-drift data from forced experiments still applies (?).

In order to better understand the component of model drift related to non-closure of the global energy and mass budgets, we compare drift in ocean state variables (global ocean mass, salinity and heat content) with time-integrated heat and freshwater fluxes at the ocean surface and TOA. We find that drift in OHC is typically much smaller than in time-integrated netTOA, indicating a leakage of energy in the simulated climate system. Most of this energy leakage occurs somewhere between the TOA and ocean surface and has improved (i.e. it has a reduced ensemble median magnitude) from CMIP5 to CMIP6 due to reduced drift in time-integrated netTOA. To put these drifts and leaks into perspective, the time-integrated netTOA and system-wide energy leakage approaches or exceeds the estimated current planetary imbalance for a number of models.

A similar story is true for ocean mass conservation. Drift in ocean mass is typically a small but non-negligible fraction of observed trends in barystatic sea level, while the time-integrated freshwater flux is typically much larger and approaches/exceeds the magnitude of recent observed trends for some models. Unlike the global energy budget, the ensemble median drift magnitude in time-integrated ocean freshwater flux is worse for CMIP6 than it was for CMIP5. In contrast, most models do a relatively good job of conserving salt and the drift in ocean salinity is reduced/improved in CMIP6. Given the importance of modeling and understanding changes in the global water cycle, we also consider the atmospheric mass budget. While drift in the global mass of atmospheric water vapor is negligible relative to estimated current trends, the drift in time-integrated moisture flux into the atmosphere (i.e. evaporation minus precipitation) and the consequent non-closure of the atmospheric moisture budget is relatively large (and worse for CMIP6), approaching/exceeding the magnitude of current trends for many models.

The causes of the energy and mass leaks we identify are many and varied, but must essentially belong to one of two categories. The first relates to deficiencies in model coupling, numerical schemes and/or physical processes. For example, the heat flux associated with water transport across the ocean boundary generally represents a global net heat loss for the ocean, because evaporation transfers water away at a temperature typically higher than precipitation adds water. The documented size of this global heat loss ranges from 0.15 W m^{-2} (**?**) to 0.30 W m^{-2} (**?**). In a steady state, this heat loss due to advective mass transfer is compensated by ocean mass and heat transport, which is in turn balanced by atmospheric transport. However, most atmospheric models do not account for the heat content of their moisture field, meaning they represent the moisture mass transport but not the heat content transport (**?**). Leakage in the simulated global heat budget therefore arises due to a basic limitation of the modeled atmospheric thermodynamics.

The second category has nothing to do with deficiencies of the model itself and instead relates to potential issues with the data that is archived and made available to the research community. For example, in discussions about ocean heat budget closure with people familiar with the ACCESS-CM2 model (Ryan Holmes, personal communication), it was discovered that the discrepancy between the OHC temperature component anomaly (Figure 1a, red curve) and time-integrated ocean surface heat flux (Figure 1a, orange curve) could be explained by a minor mistake in the construction of the ocean surface heat flux CMIP diagnostic (*hfds*; Table 2). In particular, the *hfds* diagnostic was missing contributions from the heat flux into the ocean associated with sea ice-ocean volume exchanges and frozen precipitation, as well as the effects of frazil ice formation below the surface layer of the model. When these terms are correctly included in *hfds*, there is closure between the OHC temperature component and time-integrated ocean surface heat flux. Given the high-level of model-specific knowledge (and access to data) required to precisely diagnose the cause of an apparent energy leak like this, a detailed examination of the underlying causes of non-conservation across the CMIP6 ensemble would be a difficult undertaking (and is beyond the scope of this study). A detailed assessment of energy and mass conservation is therefore best undertaken by the relevant modeling groups.

While the causes of non-conservation are of interest to model developers, for CMIP data users it is more important to know whether closure of global energy and mass budgets can be achieved after de-drifting. In other words, does the state of a reservoir like the ocean (i.e. its mass and heat content) respond appropriately to a time-integrated change in boundary heat and water fluxes once drift has been removed? In this regard, we find that almost all CMIP5 and CMIP6 models achieve approximate energy budget closure between the time-integrated netTOA flux, time-integrated ocean heat flux and OHC after de-drifting, whereas a number of models do not achieve ocean mass budget closure. The situation is even worse for the atmospheric water budget, with no models showing a strong relationship between variability in the global mass of atmospheric water vapor and time-integrated moisture fluxes into the atmosphere after de-drifting. In the case of the global energy and ocean mass budgets, CMIP6 closure represents an improvement over CMIP5. It appears that while progress in reducing the magnitude of global energy and ocean mass drifts is something of a mixed bag, the physical consistency between variations in surface fluxes and the ocean state after de-drifting has improved across the ensemble.

5 Data and code availability statement

The CMIP5 and CMIP6 model output used in this study is publicly available through a distributed data archive developed and operated by the Earth System Grid Federation (ESGF). The citation webpage for each unique model dataset (Table 3) provides a link to access the data from the relevant ESGF node. Following established best practices for reproducible computational research in the weather and climate sciences (?), the code we wrote to analyze those data has been uploaded to a Figshare repository (?) along with details of the associated software environment and data processing steps for each figure we present.

6 Acknowledgements

This study was supported by the Centre for Southern Hemisphere Oceans Research (CSHOR), jointly funded by the Qingdao National Laboratory for Marine Science and Technology (QNLM, China) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO, Australia), and the Australian Research Council's Discovery Project funding scheme (project DP190101173). We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP5 and CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP and ESGF. We also thank Ryan Holmes, Angeline Pendergrass and Martin Dix for their insightful comments on drafts of this work.

417 # Tables

variable	description	corresponding CMIP diagnostic/s
M	mass of global ocean	masso (or rhozero $\times$ volo)
T	average temperature of global ocean	thetaoga
S	average salinity of global ocean	soga
c_p	specific heat of sea water	cpocean
q_h	net heat flux into ocean	hfgeou + hfds
q_m	net freshwater flux into ocean	wfo
A_o	ocean surface grid cell area	areacello
w	column integrated mass of atmospheric water vapor	prw
q_r	net TOA radiative flux	rsdt - rsut - rlut
q_{ep}	net moisture flux into atmosphere	evspsbl - pr
A	surface grid cell area	areacella

Table 1: Variable definitions. For models where *cpocean* and/or *rhozero* were not provided, default values of 4000 J (kg K)$^{-1}$ and 1026 kg m^{-3} were used. See Table 2 for more details on the CMIP diagnostics.

variable	**name**	**units**	**time**	**shape**
areacella	grid cell area (atmosphere)	m^2	static	XY
areacello	grid cell area (ocean)	m^2	static	XY
cpocean	specific heat capacity of sea water	$J\ (kg\ K)^{-1}$	static	0
evspsbl	water evapotranspiration flux	$kg\ m^{-2}\ s^{-1}$	month	XY
hfds	net surface downward heat flux in sea water	$W\ m^{-2}$	month	XY
hfgeou	upward geothermal heat flux at sea floor	$W\ m^{-2}$	static	XY
masso	global sea water mass	kg	month	0
pr	precipitation flux	$kg\ m^{-2}\ s^{-1}$	month	XY
prw	atmosphere mass content of water vapor	$kg\ m^{-2}$	month	XY
rhozero	reference sea water density	$kg\ m^{-3}$	static	0
rlut	TOA outgoing longwave flux	$W\ m^{-2}$	month	XY
rsdt	TOA incoming shortwave flux	$W\ m^{-2}$	month	XY
rsut	TOA outgoing shortwave flux	$W\ m^{-2}$	month	XY
soga	global mean sea water salinity	g/kg	month	0
thetaoga	global mean sea water potential temperature	°C	month	0
volo	global sea water volume	m^3	month	0
wfo	net water flux into sea water	$kg\ m^{-2}\ s^{-1}$	month	XY

Table 2: CMIP diagnostics used in this study. The evaporation (*evspsbl*) diagnostic includes transpiration and sublimation, while precipitation (*pr*) includes liquid and solid phases from all types of clouds; TOA = top of atmosphere.

institution	**CMIP5 models**	**CMIP6 models**
BCC	BCC-CSM1.1 (?) BCC-CSM1.1(m) (?)	BCC-CSM2-MR (?) BCC-ESM1 (?)
BNU	BNU-ESM (?)	
CMCC	CMCC-CESM (?) CMCC-CM (?) CMCC-CMS (?)	
CNRM-CERFACS		CNRM-CM6-1* (?) CNRM-ESM2-1* (?)
CSIRO	ACCESS1-0 (?) ACCESS1-3 (?)	ACCESS-CM2 (?) ACCESS-ESM1-5 (?)
E3SM-Project		E3SM-1-0 (?) E3SM-1-1 (?)
EC-Earth-Consortium		EC-Earth (?) EC-Earth-Veg (?)
HAMMOZ-Consortium		MPI-ESM-1-2-HAM (?)
IPSL	IPSL-CM5A-LR (?) IPSL-CM5A-MR (?) IPSL-CM5B-LR (?)	IPSL-CM6A-LR (?)
MIROC	MIROC4h (?) MIROC-ESM (?) MIROC-ESM-CHEM (?)	
MOHC		HadGEM3-GC31-LL (?) UKESM1-0-LL* (?)
MPI-M	MPI-ESM-LR (?) MPI-ESM-MR (?) MPI-ESM-P (?)	MPI-ESM1-2-LR (?) MPI-ESM1-2-HR (?)
NASA-GISS		GISS-E2-1-G (?) GISS-E2-1-G-CC (?)
NCC	NorESM1-M (?)	

variable	CMIP5	CMIP6
dQ_r/dt (W m^{-2})	0.48 [0.20–1.63]	**0.24** [0.06–0.41]
dQ_h/dt (W m^{-2})	0.09 [0.07–0.17]	0.13 [0.09–0.34]
dH_T/dt (W m^{-2})	0.05 [0.01–0.21]	0.05 [0.01–0.09]
total leakage (W m^{-2})	0.47 [0.18–1.62]	**0.19** [0.05–0.42]
non-ocean leakage (W m^{-2})	0.57 [0.36–2.02]	0.46 [0.23–0.71]
ocean leakage (W m^{-2})	**0.06** [0.01–0.10]	0.17 [0.11–0.26]
dQ_m/dt (10^{15} kg yr^{-1})	**0.08** [0.00–0.60]	0.84 [0.19–122]
dM/dt (10^{15} kg yr^{-1})	0.02 [0.00–0.12]	0.04 [0.01–0.12]
dS/dt (10^{15} kg yr^{-1})	0.10 [0.04–0.16]	**0.03** [0.01–0.09]
dQ_{ep}/dt (10^{12} kg yr^{-1})	**368** [5.15–1030]	1008 [66.4–1479]
dM_a/dt (10^{12} kg yr^{-1})	0.10 [0.03–0.46]	0.19 [0.03–0.25]

Table 4: Drift in the CMIP5 and CMIP6 ensembles. The ensemble median [interquartile range] drift magnitude, calculated as the linear trend over the full length of the piControl experiment, is shown. Bold values indicate where drift in one of the CMIP projects is clearly smaller than the other (defined as a median drift magnitude at least 50% smaller). Drift in ocean salinity was calculated by first converting to an equivalent change in ocean mass (as per Equation 8).

Figure Captions

Figure 1. Annual-mean, globally-integrated energy and mass budget terms for the ACCESS-CM2 pre-industrial control experiment. The time series in panels (a), (b) and (c) represent the anomaly with respect to the first year, while the de-drifted time series in panels (d), (e) and (f) were calculated by fitting and then subtracting a cubic polynomial from the corresponding time series in panels (a), (b) and (c). Ocean salinity was converted to an equivalent change in ocean mass as per Equation 8 and a ten-year running mean was applied to the de-drifted time series.

Figure 2. Annual-mean, globally-integrated ocean heat content (OHC) and ocean mass for the CMIP6 pre-industrial control experiment. Each time series represents the anomaly with respect to the first year and a ten-year running mean has been applied. The thin black dashed lines correspond to a trend magnitude of 0.4, 0.2 and 0.1 Wm-2 respectively in panel (a) and 1.8, 0.9 and 0.45 mm/year in panel (c). For reference, 0.4 Wm-2 is the lower bound of current estimates of the planetary energy imbalance and 1.8 mm/year the estimated current rate of barystatic sea level rise.

Figure 3. Drift in globally-integrated ocean heat content (OHC; dH/dt) and its temperature (dHT/dt; Equation 1) and barystatic (dHm/dt; Equation 2) components. Values represent the linear trend over the entire length of the pre-industrial control experiment for CMIP5 (to the left of vertical dividing line) and CMIP6 (to the right). For comparison, the current planetary energy imbalance is shaded (estimates range from 0.4-1.0 Wm-2).

Figure 4. Drift in ocean and atmosphere state variables and boundary fluxes related to energy, mass and salt conservation. Each marker represents the linear trend over the full length of the pre-industrial control experiment, with CMIP5 and CMIP6 models designated with open and solid shapes, respectively. The colors represent models from the same institution (Table 3). Drift in ocean salinity was calculated by first converting to an equivalent change in ocean mass (as per Equation 8; see panel d) and the time-integrated moisture flux into the atmosphere (panel e) has not been plotted against the change in atmospheric water vapor because the water vapor trends are negligible in comparison. The thick dashed lines indicate a 1-to-1 relationship (i.e. conservation) and estimates of the magnitude of the current planetary energy imbalance (estimates range from 0.4-1.0 Wm-2; shading in panels a and b), barystatic sea level rise (1.8 mm/year; thin dashed lines in panels c and d) and trend in the global mass of atmospheric water vapor (thin dashed lines in panel e) are shown.

Figure 5. Energy leakage between the time-integrated netTOA and change in ocean heat content. Values represent the linear trend over the entire length of the pre-industrial control experiment for CMIP5 (to the left of vertical dividing line) and CMIP6 (to the right). For comparison, the magnitude of the current planetary energy imbalance is shaded (estimates range from 0.4-1.0 Wm-2). The MIROC models have a total leakage of approximately -3.5 W m-2, with offsetting ocean and non-ocean leakages of approximately -41.5 and 38.0 W m-2 respectively.

Figure 6. Mass and energy conservation after drift removal. For each model, linear regression coefficients were calculated between pairs of decadal-mean de-drifted time series of interest including the time-integrated netTOA (Qr), time-integrated heat flux into ocean (Qh), time-integrated moisture flux into atmosphere (Qep), time-integrated freshwater flux into

ocean (Qm), temperature component of globally integrated OHC (HT), ocean mass (M), ocean salinity (S) and mass of atmospheric water vapor (Ma). Each box shows the ensemble quartiles for the coefficients, while the whiskers extend to show the rest of the distribution, except for points determined to be outliers (values beyond 1.5 times the inter-quartile range). Values for each model (including the small number of outliers beyond the plot bounds) are given in Figures S1 and S2.

Figures

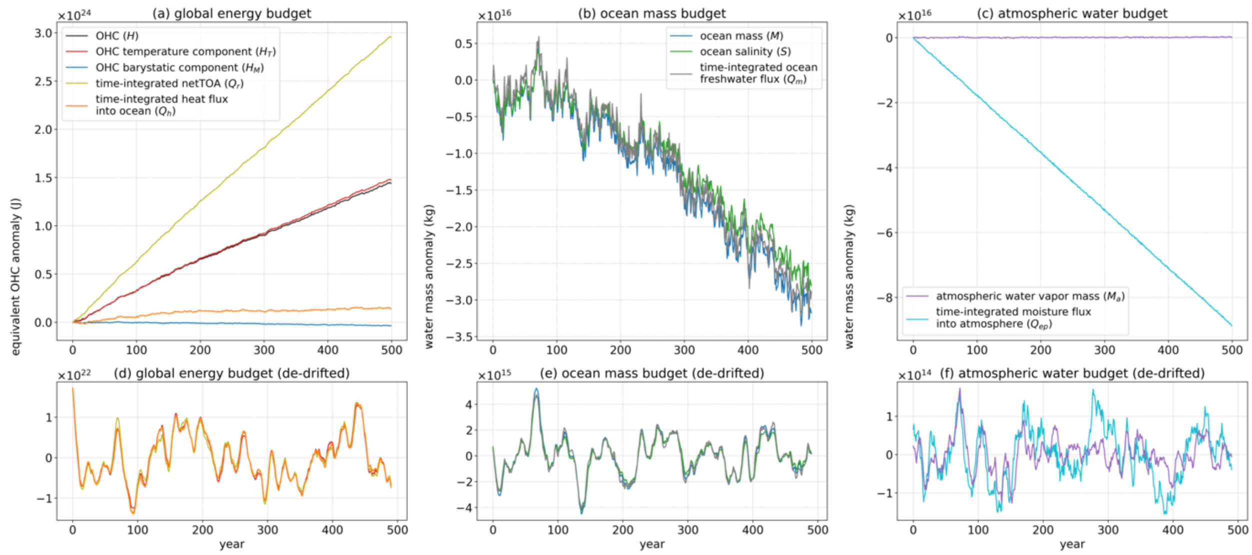

Figure 1: Annual-mean, globally-integrated energy and mass budget terms for the ACCESS-CM2 pre-industrial control experiment. The time series in panels (a), (b) and (c) represent the anomaly with respect to the first year, while the de-drifted time series in panels (d), (e) and (f) were calculated by fitting and then subtracting a cubic polynomial from the corresponding time series in panels (a), (b) and (c). Ocean salinity was converted to an equivalent change in ocean mass as per Equation 8 and a ten-year running mean was applied to the de-drifted time series.

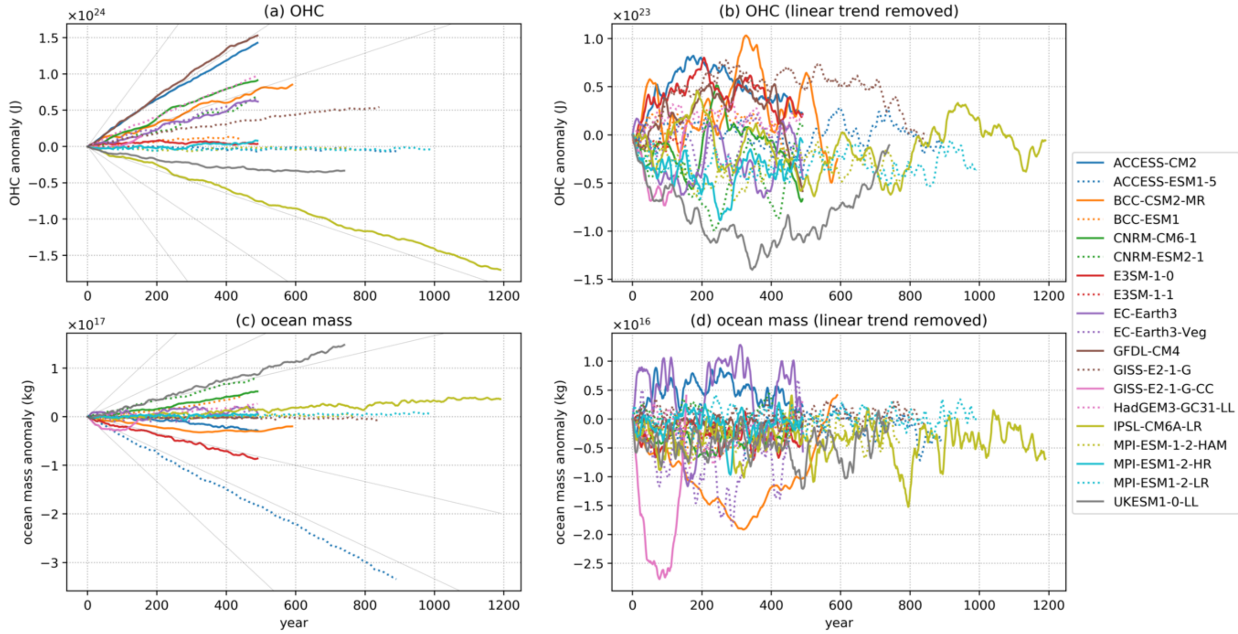

Figure 2: Annual-mean, globally-integrated ocean heat content (OHC) and ocean mass for the CMIP6 pre-industrial control experiment. Each time series represents the anomaly with respect to the first year and a ten-year running mean has been applied. The thin black dashed lines correspond to a trend magnitude of 0.4, 0.2 and 0.1 Wm-2 respectively in panel (a) and 1.8, 0.9 and 0.45 mm/year in panel (c). For reference, 0.4 Wm-2 is the lower bound of current estimates of the planetary energy imbalance and 1.8 mm/year the estimated current rate of barystatic sea level rise.

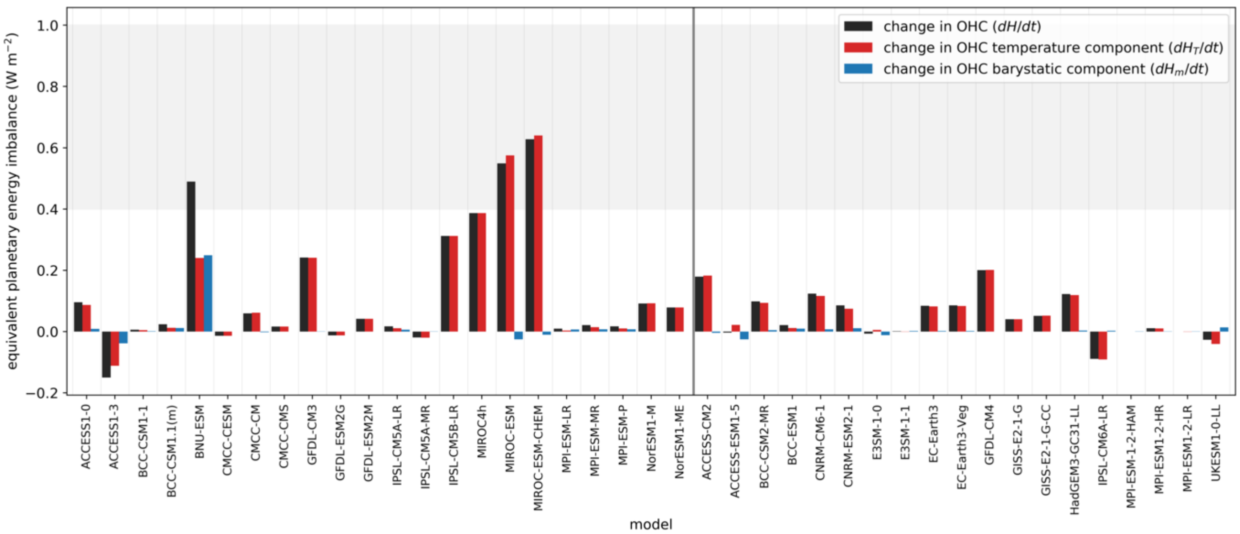

Figure 3: Drift in globally-integrated ocean heat content (OHC; dH/dt) and its temperature (dHT/dt; Equation 1) and barystatic (dHm/dt; Equation 2) components. Values represent the linear trend over the entire length of the pre-industrial control experiment for CMIP5 (to the left of vertical dividing line) and CMIP6 (to the right). For comparison, the current planetary energy imbalance is shaded (estimates range from 0.4-1.0 Wm-2).

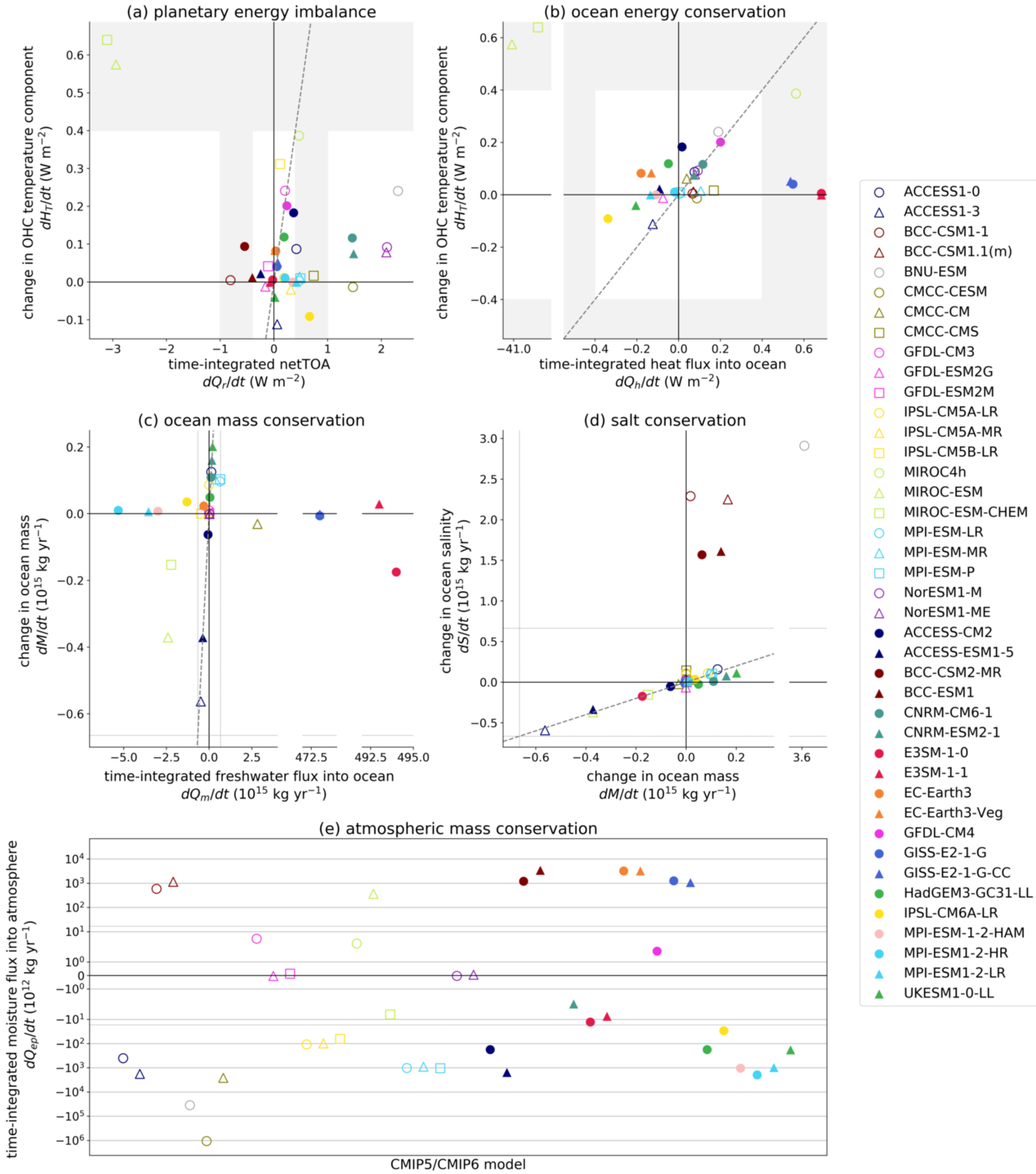

Figure 4: Drift in ocean and atmosphere state variables and boundary fluxes related to energy, mass and salt conservation. Each marker represents the linear trend over the full length of the pre-industrial control experiment, with CMIP5 and CMIP6 models designated with open and solid shapes, respectively. The colors represent models from the same institution (Table 3). Drift in ocean salinity was calculated by first converting to an equivalent change in ocean mass (as per Equation 8; see panel d) and the time-integrated moisture flux into the atmosphere (panel e) has not been plotted against the change in atmospheric water vapor

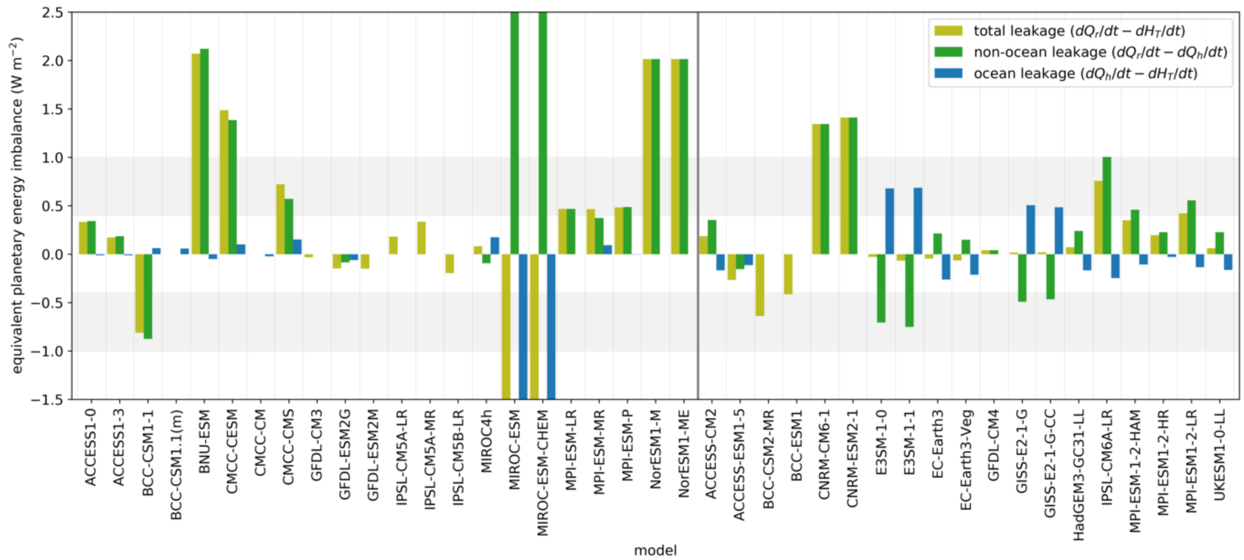

Figure 5: Energy leakage between the time-integrated netTOA and change in ocean heat content. Values represent the linear trend over the entire length of the pre-industrial control experiment for CMIP5 (to the left of vertical dividing line) and CMIP6 (to the right). For comparison, the magnitude of the current planetary energy imbalance is shaded (estimates range from 0.4-1.0 Wm-2). The MIROC models have a total leakage of approximately -3.5 W m-2, with offsetting ocean and non-ocean leakages of approximately -41.5 and 38.0 W m-2 respectively.

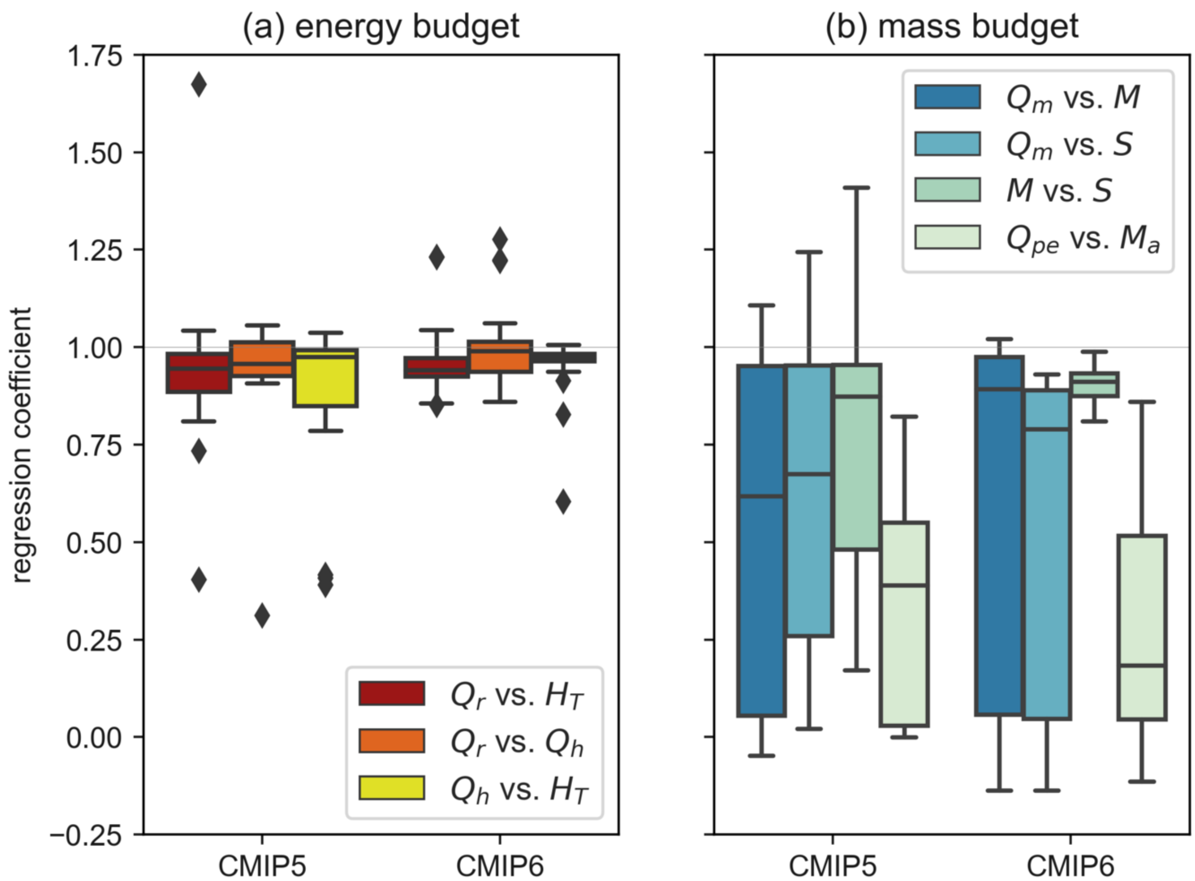

Figure 6: Mass and energy conservation after drift removal. For each model, linear regression coefficients were calculated between pairs of decadal-mean de-drifted time series of interest including the time-integrated netTOA (Qr), time-integrated heat flux into ocean (Qh), time-integrated moisture flux into atmosphere (Qep), time-integrated freshwater flux into ocean (Qm), temperature component of globally integrated OHC (HT), ocean mass (M), ocean salinity (S) and mass of atmospheric water vapor (Ma). Each box shows the ensemble quartiles for the coefficients, while the whiskers extend to show the rest of the distribution, except for points determined to be outliers (values beyond 1.5 times the inter-quartile range). Values for each model (including the small number of outliers beyond the plot bounds) are given in Figures S1 and S2.