Comparison of Surface Energy Fluxes from Global to Local Scale

JOHANNES MAYER, MICHAEL MAYER, LEOPOLD HAIMBERGER, AND CHUNLEI LIU

Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria
European Centre for Medium-Range Weather Forecasts, Bonn, Germany
South China Sea Institute of Marine Meteorology, Guangdong Ocean University, Zhanjiang, China
CMA-GDOU Joint Laboratory for Marine Meteorology, Guangdong Ocean University, Zhanjiang, China

ABSTRACT: This study uses the ECMWF ERA5 reanalysis and observationally constrained top-of-the-atmosphere radiative fluxes to infer net surface energy fluxes covering 1985–2018, which can be further adjusted to match the observed mean land heat uptake. Various diagnostics are applied to provide error estimates of inferred fluxes on different spatial scales. For this purpose, adjusted as well as unadjusted inferred surface fluxes are compared with other commonly used flux products. On a regional scale, the oceanic energy budget of the North Atlantic between the RAPID array at 26.5°N and moorings located farther north (e.g., at the Greenland–Scotland Ridge) is evaluated. On the station scale, a comprehensive comparison of inferred and buoy-based fluxes is presented. Results indicate that global land and ocean averages of unadjusted inferred surface fluxes agree with the observed heat uptake to within 1 W m⁻², while satellite-derived and model-based fluxes show large global mean biases. Furthermore, the oceanic energy budget of the North Atlantic is closed to within 2.7 (±0.2) W m⁻² for the period 2005–09 when unadjusted (adjusted) inferred surface fluxes are employed. Indirect estimates of the 2004–16 mean oceanic heat transport at 26.5°N are 1.09 PW (1.17 PW with adjusted fluxes), which agrees well with observed RAPID transports. On the station scale, inferred fluxes exhibit a mean bias of −20.1 W m⁻² when using buoy-based fluxes as reference, which confirms expectations that biases increase from global to local scales. However, buoy-based fluxes as reference are debatable, and are likely positively biased, suggesting that the station-scale bias of inferred fluxes is more likely on the order of −10 W m⁻².

KEYWORDS: Energy transport; Atmosphere-ocean interaction; Climate variability; Energy budget/balance; Heat budgets/fluxes; Surface fluxes; Buoy observations; Satellite observations; Reanalysis data; Climate variability

1. Introduction

The exchange of energy between atmosphere and the underlying Earth surface plays a pivotal role in climate dynamics and variability. Many physical processes of Earth’s climate system are associated with the exchange of energy through the surface (as well as the exchange of mass and momentum), such as changes of the ocean heat content, the global impact of El Niño events, and heating and cooling of Earth’s lower atmosphere, as well as the impact on the hydrological cycle through evaporation (Peixoto and Oort 1992; Trenberth et al. 2002b; Held and Soden 2006; Trenberth and Fasullo 2013). Furthermore, surface fluxes play an important role for the meridional energy transport, the formation of storm tracks, and the large-scale atmospheric circulation (Hoskins and Valdes 1990; Trenberth et al. 2002a; Trenberth and Fasullo 2017). Investigating surface energy fluxes is thus fundamental and can help to understand and interpret changes in Earth’s climate system.

A precise quantification of these surface fluxes is indispensable in order to determine long-term changes of the climate. However, this is difficult to accomplish as current observation systems do not achieve the required accuracy of less than 5 W m⁻² (Bourassa et al. 2013; Yu 2019; Cronin et al. 2019). Uncertainties in observation-based estimates of surface fluxes may be on the order of several tens of watts per square meter; in particular, satellite-derived estimates of individual surface flux components over the global ocean are assumed to have uncertainties of up to 20 W m⁻² (Brunke et al. 2011; Rhein et al. 2013; Kato et al. 2020; Tomita et al. 2021; Yu 2019). In situ–based estimates of net surface energy fluxes (i.e., turbulent plus radiative fluxes), on the other hand, are sparse over the global ocean (Cronin et al. 2019) and provide in most cases only a few years of adequate data, making them unsuitable for long-term climate change and variability studies.

While most satellite-derived surface flux estimates, such as from Japanese Ocean Flux Datasets with Use of Remote Sensing Observations (Tomita et al. 2019) or Objectively Analyzed Air–Sea Fluxes (OAFlux; Yu and Weller 2007), suggest unrealistically strong global mean ocean heating rates of up to 25 W m⁻², model-based surface fluxes from the latest generations of reanalyses usually have smaller ocean means that are in better agreement with observed heating rates (Brunke et al. 2011; Valdivieso et al. 2017; Tomita et al. 2021). However, they still feature similarly strong and likely spurious trends on annual to decadal time scales, making them also unsuitable for long-term climate studies (Robertson et al. 2020; Hersbach et al. 2020).

Another approach to estimate net surface energy fluxes is an indirect method combining atmospheric energy budget...
diagnostics evaluated with reanalysis data and independent observation-based top of the atmosphere (TOA) flux products (Mayer et al. 2017; Trenberth and Fasullo 2013, 2018). Reanalyses provide global gridded data with high spatial and temporal resolution, which are physically constrained by the model and thus in principle are optimally suited for global energy budget evaluations. The high spatiotemporal resolution is important for estimating horizontal eddy fluxes in atmosphere and ocean. Previous studies have shown that inferred surface energy fluxes indeed exhibit smaller biases than other estimates, but are still too large to adequately reproduce the long-term ocean and land heat uptake of less than 1 W m$^{-2}$ (Rhein et al. 2013; von Schuckmann et al. 2020). However, efforts have been made in recent years to further reduce the bias, and every new reanalysis may come with improvements over its predecessor toward achieving this goal.

In this study, we present indirectly estimated surface energy fluxes using the fifth major global reanalysis produced by ECMWF (ERA5; Hersbach et al. 2020), which are subsequently adjusted to the observed mean land heat uptake. Adjusted as well as unadjusted inferred surface fluxes are compared with buoy-based estimates, model-based fluxes from ERA5 forecasts, and satellite-derived surface flux estimates. We report improvements compared to previous estimates, and benefits and downsides of using ERA5. Furthermore, we use indirectly estimated fluxes to evaluate the regional oceanic energy budget of the North Atlantic Ocean, and budget closure using independent datasets is discussed. Additionally, the bias of indirectly estimated surface fluxes on global to local scale is provided.

This paper is structured as follows. The atmospheric and oceanic energy budget formalism is introduced in section 2. Section 3 describes the data used in this study. Results are presented in section 4. Section 5 is a summary and discussion.

2. Methodology

a. Atmospheric energy budget

We use the simplified formulation of the total atmospheric energy budget as derived by Mayer et al. (2017), where horizontal and vertical enthalpy fluxes of water are consistently removed. Thus, the sum of net turbulent plus radiative heat fluxes at Earth’s surface is defined as

$$F_S = F_{TOA} - \nabla \cdot \left( \frac{1}{g} \int_0^{r_{fs}} \left[ (1 - q) c_v T_a + L_v(T_a) q + \Phi + k \right] dv \right) - \text{AET},$$

(1)

where $F_{TOA}$ is the net energy flux at the TOA, $g$ is the gravitational acceleration, $p_S$ is the surface pressure, $q$ is the specific humidity, $c_v$ is the specific heat capacity of dry air, $T_a$ is the air temperature measured in Celsius, $L_v$ is the latent heat of vaporization, $\Phi$ is the geopotential, $k$ is the kinetic energy, $v$ is the horizontal wind vector, and AET is the vertically integrated atmospheric total energy tendency derived from analyzed state quantities. The divergence term on the right side is also referred to as the vertically integrated divergence of moist static plus kinetic energy flux (denoted as TDIV), which is mass-corrected as described in Mayer et al. (2021). That is, we combine TOA fluxes with the atmospheric divergence and tendency of energy to indirectly estimate net surface energy fluxes (denoted as inferred surface fluxes).

b. Oceanic energy and mass budget

The oceanic energy budget is evaluated for three closed domains within the North Atlantic Ocean (see Fig. 1): the southern domain (SD) covering the ocean area between the RAPID array at 26.5°N and the Greenland–Scotland Ridge (GSR) and Davis Strait (DS), the northern domain (ND) between the GSR in the south and Fram Strait (FRAM) and Barents Sea Opening (BSO) in the north, and the area covering both the southern and northern domains (SD+ND). The SD comprises the North Atlantic Ocean, the Labrador Sea, Hudson Bay, the Northwest Passage east of the Fury and Hecla Straits, the North Sea, and the Baltic Sea. The Mediterranean Sea is excluded. Energy transports through the Fury and Hecla Straits and the Straits of Gibraltar are neglected as they are on the order of <15 TW (Macdonald et al. 1994; Wu and Haines 1998; Straneo and Saucier 2008) and small compared to oceanic transports through the other sections. The ND covers the Norwegian Sea, Iceland Sea, and Greenland Sea.

We write the vertical integral of horizontal divergence of oceanic heat fluxes (Mayer et al. 2019) as

$$\nabla \cdot \int_0^Z \rho_o c_o \left( T_o(z) - T_{ref} \right) \rho_o c_o \frac{d}{dz} \left( T_o(z) - T_{ref} \right) dz = F_S - \text{MET} - \rho_o c_o \frac{d}{dz} \int_0^Z \left( T_o(z) - T_{ref} \right) dz,$$

(2)

where $\rho_o$ (1026 kg m$^{-3}$) is seawater density, $c_o$ (3990 J kg$^{-1}$ K$^{-1}$) is specific heat of seawater, $T_o$ is ocean temperature, $T_{ref}$ is the Celsius reference temperature of 275.15 K, $\mathbf{e}$ is the horizontal oceanic velocity vector, and $Z$ is ocean depth. The terms on the right side describe the net surface flux as obtained from Eq. (1), the sea ice melt energy tendency MET [i.e., the energy consumed/released during sea ice melt/freezing; computed following Mayer et al. (2019)], and the ocean heat content tendency (denoted as OHCT). That is, the divergence of oceanic heat transport balances surface energy fluxes over the ice-free ocean and changes in the ocean heat content. Ocean budget residuals are obtained by moving the divergence term to the right side of the equation. Note that we compute the divergence term from observed oceanic heat transports (OHT; according to the divergence theorem); for example, transports from the Davis Strait, Fram Strait, and Barents Sea Opening minus transports through the RAPID array yield the net transport into the SD. Alternatively, the RAPID transport at 26.5°N can be calculated indirectly by subtracting the right side of Eq. (2) from observed transports through the northern gateways.

The energy budget of an oceanic volume as formulated in Eq. (2) is unambiguous as long as its mass budget is closed (Schauer and Beszczynska-Möller 2009; Tsubouchi et al. 2020).
We write the divergence of total oceanic heat transports through the boundaries of an oceanic volume [equivalent to the left side of Eq. (2)] as

\[ \nabla \cdot \text{OHT} = v_{\text{in}}(T_{o,\text{in}} - T_{\text{ref}}) + v_{\text{out}}(T_{o,\text{out}} - T_{\text{ref}}), \]  

(3)

where \( v_{\text{in}} \) is the total volume transport entering the budget volume, \( v_{\text{out}} \) is the total volume transport leaving the budget volume, and \( T_o \) is the corresponding mean seawater temperature. If the sum of \( v_{\text{in}} \) and \( v_{\text{out}} \) is zero, the \( T_{\text{ref}} \) terms cancel out and the divergence of OHT becomes independent of the choice of reference temperature. In contrast, if the sum is nonzero, \( T_{\text{ref}} \) terms remain making the oceanic energy budget ambiguous. This motivates the following mass correction of the oceanic energy budget. We start with the definition of the oceanic mass budget, which reads as follows:

\[ \frac{\partial M}{\partial t} = J_{\text{ice}} + J_{\text{river}} + (P - E) - J_{\text{ocean}}. \]  

(4)

The term on the left side describes temporal mass changes in the budget volume. On the right, \( J_{\text{ice}} \) is the lateral mass transport owing to ocean circulation, \( J_{\text{ice}} \) is the lateral sea ice transport, \( J_{\text{river}} \) is the river discharge, and \( P \) and \( E \) are precipitation and evaporation, respectively (note that \( P - E \) is often referred to as the surface freshwater flux). The river discharge term for the North Atlantic Ocean north of \( 10^\circ \)N, including the American Mediterranean Sea (i.e., the combined Caribbean Sea and Gulf of Mexico), is on the order of \( 171 \times 10^3 \text{ m}^3 \text{ s}^{-1} \) (Dai and Trenberth 2002), which is equivalent to \( \sim 0.17 \text{ Sv} \) (1 Sv \( \equiv 10^6 \text{ m}^3 \text{ s}^{-1} \)). We thus assume a mean river discharge of \( \sim 0.13 \text{ Sv} \) in the SD and \( \sim 0.01 \text{ Sv} \) in the ND. Temporal mass changes are assumed to be zero when averaged over multiple years. We neglect the Greenland ice discharge in these computations as it is on the order of \( \sim 2 \text{ mSv} \) (60 Gt yr\(^{-1}\)) in the ND and roughly \( 7 \text{ mSv} \) (200 Gt yr\(^{-1}\)) in the SD (King et al. 2018). Consequently, the mass budget residual is equal to the right side of Eq. (4).

To estimate the effect of mass inconsistencies on the energy budget, mass budget residuals \( \Delta R_M \) are converted to an associated erroneous heat flux \( \Delta \text{OHT} = \Delta R_M c_o \rho_o \Delta T \), where \( \Delta T \) is the seawater temperature difference between southern and northern boundary of the volume, which is assumed to be approximately 15 K. The erroneous heat flux associated with mass inconsistencies (typically on the
order of ±1 W m\(^{-2}\)) is then subtracted from the energy budget residual, which can be considered as simple form of a mass correction analogous to the adjustments routinely performed for atmospheric budget diagnostics (see, e.g., Trenberth 1991; Fasullo and Trenberth 2008; Mayer and Haimberger 2012).

c. Terminology

The term ocean average, or ocean mean, refers to the global ocean area including regions covered by sea ice (in total 363.1 × 10\(^6\) km\(^2\) on the quasi-regular grid used for evaluations in this study; see below). Other water masses, such as inland waters (i.e., the Great Lakes or the Caspian Sea), are excluded. Land averages cover all areas (148.1 × 10\(^6\) km\(^2\)) at F90) that are excluded in the definition of ocean averages. Furthermore, we define the bias to be the difference between data under study minus a reference (usually observational products; e.g., buoy-based estimates). The term error, or mean absolute error, is defined as the absolute value of the bias. The RMS deviation is defined as

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\text{data}_i - \text{reference}_i)^2}.
\]

where \(N\) is the length of the data. Moreover, we define surface fluxes from the atmosphere to ocean as well as northward oceanic transports to be positive. The anomaly time series in section 4e are computed by subtracting the corresponding long-term mean of each calendar month (i.e., the climatology with respect to the times where buoy-based fluxes are available) from the original time series.

3. Data

a. Data sources

We employ the mass-balanced total atmospheric energy flux divergence and atmospheric energy tendency from Mayer et al. (2021), which are derived from ECMWF’s latest reanalysis dataset ERA5, in combination with net TOA fluxes from the DEEP-C dataset (Allan et al. 2014; Liu et al. 2020; publicly available at https://doi.org/10.17864/1947.271) to infer net surface energy fluxes [denoted as \(F_{\text{net,ERA5}}\); according to Eq. (1)] for the period 1985–2018. ERA5 provides global gridded data on 137 vertical model levels (up to 0.01 hPa) with 1-hourly temporal and ~0.28° spatial resolution (using a reduced Gaussian grid N320).

DEEP-C data are available as monthly averages on a full Gaussian grid F128 (corresponds to 0.7° spatial resolution) for the period 1985–2020. It is a backward extension of the net TOA fluxes from the Clouds and the Earth’s Radiant Energy System–Energy Balanced and Filled (CERES-EBAF) product in version 4.1 (Loeb et al. 2009, 2018). The TOA fluxes prior to CERES have been reconstructed by Liu et al. (2020) based on the procedure of Allan et al. (2014) but with some modifications. The TOA flux climatology is from CERES-EBAF and anomalies are from ERA5 constrained by ERBE WFOV (Earth Radiation Budget Experiment Satellite wide field of view, 72-day mean; Wong et al. 2006) anomalies at 10° × 10° resolution (covering 60°N–60°S) to keep the observed variability on a regional scale. Discontinuities in the reconstruction were dealt with using Atmospheric Model Intercomparison Project simulations and other high-resolution atmospheric model simulations. The global mean OHCT and net TOA flux have been compared and the general agreement in both the absolute value and the variability between them suggests robustness of the reconstruction over 1985–99 (Liu et al. 2020).

The unadjusted inferred net surface fluxes \((F_{\text{net,ERA5}})\) are compared with the following four products:

1) Adjusted inferred net surface fluxes (denoted as \(F_{\text{net,adj,ERA5}}\)) are computed in the same way as unadjusted inferred surface fluxes, but subsequently modified with the procedure described in Liu et al. (2017) [see section 2.4 therein and also Liu et al. (2020)], where unrealistically large land surface fluxes (in fact the contributing divergence of atmospheric energy transports) are zonally redistributed to the ocean. This product covers the period from January 1985 to November 2017 and is available at https://doi.org/10.17864/1947.000347.

2) Model-based net surface fluxes \((F_{\text{model,ERA5}})\) are turbulent plus radiative heat fluxes taken from monthly means of twice-daily 12-hourly ERA5 forecasts (i.e., the standard flux fields as available from the ERA5 archive) and are available from 1979 onward.

3) Satellite-derived net surface fluxes \((F_{\text{CERES,ERA5}})\) are combined from CERES-EBAF Ed4.1 (Kato et al. 2018) and OAFlux in version 3 (Yu and Weller 2007; Yu et al. 2008). CERES provides monthly means of net surface radiative (shortwave and longwave) fluxes at 1° spatial resolution covering the period from March 2000 onward. OAFlux provides monthly averages of turbulent (sensible + latent) heat fluxes for the period 1979 to 2018 also at 1° spatial resolution. CERES-EBAF surface fluxes cover the whole globe, while OAFlux data are available only for the ice-free ocean between roughly 60°N and 60°S. Consequently, analyses including \(F_{\text{CERES,ERA5}}\) are limited to this area and the period from 2001 onward.

4) Buoy-based net surface flux estimates \((F_{\text{Buoy}})\) were computed by several institutions using the COARE algorithm in version 3, and are used as such. NOAA’s Ocean Climate Stations project (OCS) provides data from the Kuroshio Extension Observatory (KEO) and Ocean Station Papa (PAPA) buoys (see www.pmel.noaa.gov/ocs/), the Woods Hole Oceanographic Institution (WHOI) provides Northwest Tropical Atlantic Station (NTAS), Stratus, and WHOI Hawaii Ocean Time-series Station (WHOTS) buoy data (http://uop.whoi.edu/), and NOAA’s Oceansites web page provides data from the TOA/TRITON, PIRATA, and RAMA buoy arrays (see https://www.pmel.noaa.gov/tao/drupal/fl ox/index.html). We excluded buoys with too short or incomplete time series, requiring time series to cover at least four years per calendar month to obtain reasonable estimate of the mean annual cycles. Buoy-based time series containing obviously spurious...
values (detected by visual inspection) have been discarded as well. In this way, we selected 14 buoys from the aforementioned sources: 7 in the Pacific Ocean, 5 in the Atlantic Ocean, and 2 in the Indian Ocean. All buoy-based net surface fluxes are used as provided by the institutions, no additional adjustments or computations were performed.

Hence, a total of five surface flux products are used in this study (see Table 1 for a summary, with corresponding names, acronyms, time periods, and adjustments).

The oceanic energy budget is evaluated using the ocean heat content tendency (OHCT) and sea ice melt energy tendency (MET) computed from the Ocean ReAnalysis Pilot system 6 (ORAP6.0; Zuo et al. 2021), which covers the period 1979–2019. ORAP6.0 is a successor to the Ocean Reanalysis System-5 (ORAS5; Zuo et al. 2019) and comes with several improvements, such as ERA5 atmospheric forcing and other updates to the assimilation system that will be reported on elsewhere. For regional studies (section 4b), the OHCT is integrated over the whole ocean depth and uniformly adjusted (OHCT adj hereafter) to the year-to-year variability of global net TOA fluxes, as done by Trenberth et al. (2019) and Liu et al. (2020). For global studies (section 4a), the unadjusted ocean heat content tendency (denoted as OHCT unadj) is employed. The divergence of oceanic energy transports is derived from mooring-derived and volume-conserving OHT estimates as provided by Tsubouchi et al. (2018) for various Arctic Gateways (available at https://doi.pangaea.de/10.1594/PANGAEA.909966), and Tsubouchi et al. (2020) for the Greenland-Scotland Ridge and Davis Strait (see http://metadata.nmdc.no/metadata-api/landingpage/0a2ae0e42ef7af767a92081e83784b1). Independent OHT estimates in the Atlantic Ocean at 26.5°N are provided by the RAPID-MOCHA project (Johns et al. 2011; McCarthy et al. 2015; Bryden et al. 2020; see https://mocha.rsmas.miami.edu/mocha/results/index.html). All these datasets contain corresponding volume transports needed to compute the oceanic mass budget. In addition, Tsubouchi et al. (2018) provide sea ice transports for DS, FRAM, and BSO (the sea ice transport through the GSR is assumed to be zero). Surface freshwater fluxes are taken from monthly means of twice-daily 12-hourly ERA5 forecasts. Other mass budget terms are long-term mean estimates from the literature, as described in section 2.

All data used in this study are aggregated to monthly means, if not provided as such. Reanalysis and DEEP-C data are also interpolated to a full Gaussian grid F90, which has a meridional spacing decreases slightly at high latitudes). Whenever possible, evaluations are performed for the period 1985–2018 as this is the period for which TOA fluxes are available allowing to infer surface energy fluxes.

### b. Data dependencies

Some products compared in this study (see Table 1) are based on the same input data and therefore cannot be considered as fully independent. Other datasets are constrained in a way such that additional data dependencies arise. For instance, the adjusted inferred surface flux is modified to match the observed mean land heat flux, while TOA fluxes are adjusted to the global long-term mean OHCT (Loeb et al. 2018). This leads to close agreement between oceanic and OHCT, as the land heat flux is comparatively small. Such data dependencies are visualized in Fig. 2, where we show employed input data and interdependencies among the products listed in Table 1. Fully independent products are OHT from RAPID and Arctic Gateways as they are not assimilated by any of the used data products. Buoy-based surface fluxes are solely computed from buoy observations using the COARE algorithm and are in principle also independent of

<table>
<thead>
<tr>
<th>Name</th>
<th>Acronym</th>
<th>Period (in this study)</th>
<th>Constraints/tuning</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOA net energy flux</td>
<td>$F_{\text{TOA}}$</td>
<td>1985–2018</td>
<td>To long-term global mean OHCT</td>
<td>Allan et al. (2014), Liu et al. (2020)</td>
</tr>
<tr>
<td>Unadjusted inferred net surface flux derived from ERA5</td>
<td>$F_{\text{inf,adj}}^\text{ERA5}$</td>
<td>1985–2017</td>
<td>Mass consistency†, land heat uptake†</td>
<td>Liu et al. (2020)</td>
</tr>
<tr>
<td>Adjusted inferred net surface flux</td>
<td>$F_{\text{inf,adj}}^\text{ERA5}$</td>
<td>1985–2017</td>
<td>Mass consistency†</td>
<td>Hersbach et al. (2020)</td>
</tr>
<tr>
<td>Model-based net surface flux from ERA5 forecast</td>
<td>$F_{\text{model}}^\text{ERA5}$</td>
<td>1985–2018</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Buoy-based net surface flux</td>
<td>$F_{\text{Buoy}}$</td>
<td>Variable</td>
<td>—</td>
<td>WHOI, NOAA</td>
</tr>
<tr>
<td>Satellite-derived turbulent net surface flux</td>
<td>$F_{\text{SOA}}$</td>
<td>2001–18</td>
<td>To buoy-based fluxes</td>
<td>Yu and Weller (2007)</td>
</tr>
<tr>
<td></td>
<td>$\text{OHT}_{\text{BSO}}$</td>
<td>2005–09</td>
<td>—</td>
<td>Tsubouchi et al. (2018)</td>
</tr>
<tr>
<td></td>
<td>$\text{OHT}_{\text{DS}}$</td>
<td>1993–2017</td>
<td>—</td>
<td>Tsubouchi et al. (2020)</td>
</tr>
<tr>
<td></td>
<td>$\text{OHT}_{\text{GSR}}$</td>
<td>1993–2017</td>
<td>—</td>
<td>Tsubouchi et al. (2020)</td>
</tr>
<tr>
<td></td>
<td>$\text{OHT}_{\text{RAPID}}$</td>
<td>2004–18</td>
<td>—</td>
<td>Johns et al. (2011)</td>
</tr>
<tr>
<td>Ocean heat content tendency</td>
<td>$\text{OHCT}^\text{adj}$</td>
<td>1985–2018</td>
<td>To annual global mean $F_{\text{TOA}}$†</td>
<td>Zuo et al. (2021)</td>
</tr>
</tbody>
</table>
other products. However, OAFlux data are tuned to match the buoy-based fluxes, making $F_{\text{S,CERES+OA}}$ strongly dependent on $F_{\text{S,Buoy}}$. ERA5 also assimilates buoy data (except for WHOI buoys) which may influence surface flux products derived from it. Josey et al. (2014) demonstrated such effects for ERA-Interim, but our evaluations using ERA5 (not shown) suggest that these effects are much smaller for this more recent reanalysis. WHOI buoys are not assimilated by ERA5 and are thus truly independent of any surface fluxes derived from it. ORAP6 uses atmospheric forcing based on ERA5, which is considered to have an impact on the correlation between $F_{\text{S,Buoy}}$ and $F_{\text{S,CERES+OA}}$. Other data dependencies are considered to be relatively weak.

4. Results

a. Global surface energy fluxes

In this section, we focus on global land and ocean averages. Some evaluations in this section are confined to 1985–2016 as the adjusted inferred flux data end with November 2017. We show 1985–2016 averages of global net surface flux fields in Fig. 3. In general, net surface fluxes over the global ocean are characterized by positive fluxes in the tropics, where the ocean efficiently absorbs incoming solar radiation, and negative fluxes at mid- and high latitudes (i.e., along the western boundary currents and over the Arctic Ocean; Sverdrup 1947; Stommel 1948; Seager and Simpson 2016), where warm water masses are transported poleward and large amounts of oceanic heat are lost to the atmosphere. Note that the pronounced regional features over the ocean are qualitatively similar to the global pattern of TEDIV from Eq. (1), as shown in Mayer et al. (2021), suggesting that strong air–sea fluxes along the equator and the western boundary currents are
largely balanced by atmospheric energy transports. As the ocean absorbs about 90% (Rhein et al. 2013) of the global energy imbalance at the TOA ($0.69 \pm 0.1$ W m$^{-2}$ over the period 1993–2018; von Schuckmann et al. 2020), long-term mean oceanic surface fluxes are expected to be on the order of 0.9 W m$^{-2}$ (i.e., $0.9 \times 0.69/0.71$, where the denominator is the relative ocean surface area). Over landmasses, mean surface fluxes should be on the order of $<0.1$ W m$^{-2}$ due to the small heat storage rate of soil masses (von Schuckmann et al. 2020).

The $F_{\text{inf, ERA5}}$ field in Fig. 3a agrees well with the described pattern, with a global 1985–2016 mean of 0.4 W m$^{-2}$ and
TEDIV, which is used to compute (e.g., along the Andes and Himalayas) is a remnant of of 30.7 W m
2021). Over the ocean, the unrealistically large global mean of 5.8 W m
needs to be balanced by analysis increments (Mayer et al. 2021). In Fig. 3b, we show the
earlier studies could not resolve in such detail (see, e.g., Trenberth and Fasullo 2008, 2017). In Fig. 3b, we show the
adjusted field \( F_{\text{inf,adj}}^{\text{ERA5}} \), which is qualitatively similar to the unadjusted field in Fig. 3a but exhibits a smaller RMS value as large spurious fluxes over land (e.g., the prominent peak over the Andes) are redistributed to relatively large ocean areas, while the global long-term mean is conserved.

The \( F_{\text{inf,adj}}^{\text{ERA5}} \) field in Fig. 3c exhibits a comparatively low RMS of 30.7 W m
2 as there is no artificial noise over land, but an unrealistically large global mean of 5.8 W m
indicating unrealistic heat loss of the atmosphere to the ocean in ERA5, which needs to be balanced by analysis increments (Mayer et al. 2021).

We present time series of global ocean (Figs. 4 and 5) and land averages (Fig. 6) of various \( F_S \) products. Over the ocean, the OHCT\textsuperscript{unadj} (integrated over the full ocean depth and not additionally adjusted to global \( F_{\text{TOA}} \); see section 3) is used as reference as it has the best correlation with oceanic surface fluxes. Its 1985–2016 mean is 0.8 W m
2, with a weakly positive but statistically insignificant trend of \(+4.2 \times 10^{-2} \text{ W m}^{-2} \text{ decade}^{-1} \). Furthermore, it features prominent intermittent cooling signals associated, for example, with the Mt. Pinatubo eruption in 1991 or strong El Niño events (e.g., in 1997/98, 2009/10, and 2015/16).

Ocean averages of \( F_{\text{inf,adj}}^{\text{ERA5}} \) are stable around the 1.7 W m
2 mean, with a weak but statistically insignificant negative trend owing to the rapid decrease between 1997 and 2001, which solely stems from the gradual change of TEDIV during that time [see Mayer et al. (2021) for discussion]. However, the long-term oceanic \( F_{\text{inf,adj}}^{\text{ERA5}} \) mean agrees with the mean OHCT\textsuperscript{unadj} and ocean heat warming rates from von Schuckmann et al. (2020) to within 1 W m
2, particularly after the decrease in the late 1990s where the 2001–16 mean \( F_{\text{inf,adj}}^{\text{ERA5}} \) is \( \sim 1.4 \text{ W m}^{-2} \) and the Pearson correlation coefficient with OHCT\textsuperscript{unadj} is \( \rho = 0.43 \) (compared to \( \rho = 0.40 \) for 1985–2016). This is significantly larger than the correlation between OHCT\textsuperscript{unadj} and global \( F_{\text{TOA}} \) alone (\( \rho = 0.23 \) for 2001–16); that is, the OHCT\textsuperscript{unadj} variability is more consistent with oceanic \( F_{\text{inf,adj}}^{\text{ERA5}} \) than with global \( F_{\text{TOA}} \), in agreement with the fact that all oceanic surface flux viability should show up in OHCT, while some of global TOA flux variability may also be redistributed to other parts of the system than the ocean (atmosphere, land, cryosphere). Consequently, the correlation between OHCT\textsuperscript{unadj} and ocean heating via surface fluxes is improved by the TEDIV term, which is used to compute \( F_{\text{inf,adj}}^{\text{ERA5}} \).

The oceanic \( F_{\text{inf,adj}}^{\text{ERA5}} \) is qualitatively similar to \( F_{\text{inf}}^{\text{ERA5}} \) but does not exhibit the spurious trend in the late 1990s as this is

RMS = 32.8 W m
2. The artificial noise over high topography (e.g., along the Andes and Himalayas) is a remnant of TEDIV, which is used to compute \( F_{\text{inf}}^{\text{ERA5}} \) (see Mayer et al. 2021). Over the ocean, the field is very smooth and features sharp gradients along coastal lines and ice edges, results that earlier studies could not resolve in such detail (see, e.g., Trenberth and Fasullo 2008, 2017). In Fig. 3b, we show the adjusted field \( F_{\text{inf,adj}}^{\text{ERA5}} \), which is qualitatively similar to the unadjusted field in Fig. 3a but exhibits a smaller RMS value as large spurious fluxes over land (e.g., the prominent peak over the Andes) are redistributed to relatively large ocean areas, while the global long-term mean is conserved.

The \( F_{\text{inf,adj}}^{\text{ERA5}} \) field in Fig. 3c exhibits a comparatively low RMS of 30.7 W m
2 as there is no artificial noise over land, but an unrealistically large global mean of 5.8 W m
indicating unrealistic heat loss of the atmosphere to the ocean in ERA5, which needs to be balanced by analysis increments (Mayer et al. 2021).

We present time series of global ocean (Figs. 4 and 5) and land averages (Fig. 6) of various \( F_S \) products. Over the ocean, the OHCT\textsuperscript{unadj} (integrated over the full ocean depth and not additionally adjusted to global \( F_{\text{TOA}} \); see section 3) is used as reference as it has the best correlation with oceanic surface fluxes. Its 1985–2016 mean is 0.8 W m
2, with a weakly positive but statistically insignificant trend of \(+4.2 \times 10^{-2} \text{ W m}^{-2} \text{ decade}^{-1} \). Furthermore, it features prominent intermittent cooling signals associated, for example, with the Mt. Pinatubo eruption in 1991 or strong El Niño events (e.g., in 1997/98, 2009/10, and 2015/16).

Ocean averages of \( F_{\text{inf,adj}}^{\text{ERA5}} \) are stable around the 1.7 W m
2 mean, with a weak but statistically insignificant negative trend owing to the rapid decrease between 1997 and 2001, which solely stems from the gradual change of TEDIV during that time [see Mayer et al. (2021) for discussion]. However, the long-term oceanic \( F_{\text{inf,adj}}^{\text{ERA5}} \) mean agrees with the mean OHCT\textsuperscript{unadj} and ocean heat warming rates from von Schuckmann et al. (2020) to within 1 W m
2, particularly after the decrease in the late 1990s where the 2001–16 mean \( F_{\text{inf,adj}}^{\text{ERA5}} \) is \( \sim 1.4 \text{ W m}^{-2} \) and the Pearson correlation coefficient with OHCT\textsuperscript{unadj} is \( \rho = 0.43 \) (compared to \( \rho = 0.40 \) for 1985–2016). This is significantly larger than the correlation between OHCT\textsuperscript{unadj} and global \( F_{\text{TOA}} \) alone (\( \rho = 0.23 \) for 2001–16); that is, the OHCT\textsuperscript{unadj} variability is more consistent with oceanic \( F_{\text{inf,adj}}^{\text{ERA5}} \) than with global \( F_{\text{TOA}} \), in agreement with the fact that all oceanic surface flux viability should show up in OHCT, while some of global TOA flux variability may also be redistributed to other parts of the system than the ocean (atmosphere, land, cryosphere). Consequently, the correlation between OHCT\textsuperscript{unadj} and ocean heating via surface fluxes is improved by the TEDIV term, which is used to compute \( F_{\text{inf,adj}}^{\text{ERA5}} \).

The oceanic \( F_{\text{inf,adj}}^{\text{ERA5}} \) is qualitatively similar to \( F_{\text{inf}}^{\text{ERA5}} \) but does not exhibit the spurious trend in the late 1990s as this is
corrected by the adjustment. This leads to a 1985–2016 oceanic \( F^\text{inf,adj}_{\text{ERA5}} \) mean of 0.5 W m\(^{-2}\), which is in fact closer to the long-term OHCT\(^{\text{unadj}} \) mean compared to any other surface flux product shown, but this matching by construction as \( F_{\text{TOA}} \) is adjusted to the OHCT\(^{\text{unadj}} \) mean and land heat uptake is small (see discussion in section 3b). This discrepancy is not surprising because the Scripps domain (see Fig. 1 therein). This discrepancy is not surprising because the Scripps domain (used, e.g., by Guinehut et al. (2012), Trenberth et al. (2016), Huybers (2016) for the Scripps domain (see Fig. 1 therein).

In Fig. 5 we compare satellite-derived surface fluxes (\( F_{\text{S,CERES+OA}} \)) with \( F^\text{inf}_{\text{ERA5}} \) confined to the period 2001–18 and global ocean area between 60°N–60°S as this is the time and region for which CERES+OAFlux data are available. This excludes the predominantly negative surface fluxes north of 60°N leading to a 2001–18 \( F^\text{inf}_{\text{ERA5}} \) mean of 4.9 W m\(^{-2}\) (compared to 1.5 W m\(^{-2}\) for the global ocean), whereas the \( F_{\text{S,CERES+OA}} \) mean is 28.6 W m\(^{-2}\) (135.2 W m\(^{-2}\) from CERES-based net radiative flux and –106.6 W m\(^{-2}\) from OAFlux-based turbulent flux) and additionally exhibits a strong positive trend of 3.3 W m\(^{-2}\) decade\(^{-1}\) during the whole period (see Fig. 5b). Nonetheless, the temporal correlation between \( F^\text{inf}_{\text{ERA5}} \) and \( F_{\text{S,CERES+OA}} \) is large (~0.6) as both depend strongly on TOA fluxes from CERES-EBAF (see section 3 and Fig. 2). The climatology of \( F_{\text{S,CERES+OA}} \) varies between –~42 W m\(^{-2}\) (February) and –~3 W m\(^{-2}\) (June), which is about 24 W m\(^{-2}\) higher compared to \( F^\text{inf}_{\text{ERA5}} \) confined to the same region and time. However, peak-to-peak amplitudes (~38 W m\(^{-2}\)) of the two estimates are similar.

To make the seasonal cycle of ocean \( F^\text{inf}_{\text{ERA5}} \) from Fig. 5 comparable to results from McKinnon and Huybers (2016), we cumulatively integrated the climatology, using their averaging period 2005–14 (the shorter period changes the annual cycle only weakly). The integrated climatology has a maximum of \( 9.4 \times 10^7 \) J m\(^{-2}\) in April and a minimum of \(-1.0 \times 10^7 \) J m\(^{-2}\) in August as these are the times when the sign of the fluxes changes (see Fig. 5). This is equivalent to an amplitude of 31.5 ZJ (1 ZJ = \( 10^{21} \) J) and thus about 5.5 ZJ smaller compared to results from McKinnon and Huybers (2016) for the Scripps domain (see Fig. 1 therein).

Von Schuckmann et al. (2020) estimated an average land heat uptake of <0.1 W m\(^{-2}\) based on recent studies using borehole temperature profiles. We use this value as reference for the average surface flux over global land area, where model-based net surface fluxes from ERA5 forecasts show benefits over inferred fluxes (Fig. 6). While \( F^\text{inf}_{\text{ERA5}} \) exhibits a relatively strong positive trend in the late 1990s that solely stems from the gradual change in TEDIV, \( F^\text{model}_{\text{ERA5}} \) is temporally stable around its long-term mean of 0.8 W m\(^{-2}\) (for comparison, the \( F^\text{inf}_{\text{ERA5}} \) land mean is ~2.7 W m\(^{-2}\)). This is remarkably close to the observed land heat uptake noted above. Land averages of \( F^\text{inf}_{\text{ERA5}} \) are tuned to the mean land surface flux from Beltrami et al. (2002) and thus exhibit a long-term mean of 0.09 W m\(^{-2}\) and good temporal stability with no significant trend. Please note that the difference between \( F^\text{inf}_{\text{ERA5}} \) and \( F^\text{model}_{\text{ERA5}} \) represents the magnitude of the adjustment from Liu et al. (2017); that is, the modifications that are made are stronger before 2000 than after. Also note
Table 2. Mean and RMS values of unfiltered time series of various surface flux products for global, ocean, and land averages. Averaging period is 1985–2016, except for the last two rows, which are based on 2001–18. Units are W m⁻² relative to the averaging area. To compute ocean and land averages relative to the global surface, multiply by 0.709 and 0.291, respectively.

<table>
<thead>
<tr>
<th>Term</th>
<th>Global</th>
<th>Ocean</th>
<th>Land</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>RMS</td>
<td>Mean</td>
</tr>
<tr>
<td>(F_{\text{ref}})(^{\text{ERA5}}) (\text{gfdl}^{\text{adj}})</td>
<td>0.39</td>
<td>7.11</td>
<td>1.68</td>
</tr>
<tr>
<td>(F_{\text{ref}})(^{\text{ERA5}})</td>
<td>0.39</td>
<td>7.10</td>
<td>0.53</td>
</tr>
<tr>
<td>(\text{OHCT}_{\text{ref}}^{\text{adj}})</td>
<td>5.80</td>
<td>9.98</td>
<td>7.85</td>
</tr>
<tr>
<td>((60^\circ N-60^\circ S) F_{\text{S,CERES-OA}})</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>((60^\circ N-60^\circ S) F_{\text{S,CERES}})</td>
<td>—</td>
<td>—</td>
<td>9.49</td>
</tr>
</tbody>
</table>

that climatological signals caused by ENSO phases and volcanic eruptions are barely visible in land averages as the enhanced ocean–land energy transport during these events (see Mayer et al. 2021) is mainly compensated by changes of \(F_{\text{TOA}}\) over land (not shown) rather than by net surface fluxes.

In summary, ocean averages of unadjusted inferred surface fluxes derived from ERA5 \(F_{\text{ref}}^{\text{ERA5}}\) have a mean bias of about 1 W m⁻² with respect to the observed ocean heat uptake, whereas model-based fluxes and satellite-derived estimates show substantially larger biases (~7–25 W m⁻²) and exhibit stronger trends and inconsistencies. Over land, however, model-based fluxes are superior to unadjusted inferred surface fluxes, with long-term averages close to observed values (see von Schuckmann et al. 2020) and good temporal stability. All long-term means and RMS values are summarized in Table 2.

b. Regional oceanic energy budgets

In this section, we use inferred surface fluxes as derived from the atmospheric energy budget [see Eq. (1)] to test their consistency with the oceanic energy budget [see Eq. (2)] in the Atlantic Ocean basin (Fig. 1); that is, we assess the degree of budget closure and estimate the bias of inferred surface fluxes on regional scale. Tsubouchi et al. (2017), Tsubouchi et al. (2020), and the RAPID project (Johns et al. 2011; McCarthy et al. 2015; Bryden et al. 2020) provide independent oceanic energy transports that allow us to evaluate the ocean energy budget for the three closed domains introduced in section 2 (see Fig. 1).

Figure 7 shows the temporal evolution of individual oceanic energy budget terms of the three domains SD, ND, and SD+ND, with a 12-month running average applied. While OHCT\(^{\text{adj}}\) (integrated over the full ocean depth and adjusted to global \(F_{\text{TOA}}\); see section 3) fluctuates around its relatively small mean, surface fluxes should in the long run balance the divergence of lateral oceanic energy transports. However, surface fluxes are too weak (or energy divergence is too strong) in the SD and SD+ND for an exact balance such that the budget residuals are positive, and vice versa in the ND. Nevertheless, transports through Arctic Gateways at high latitudes (e.g., through DS+GSR in Fig. 7b) are temporally more stable than other budget terms exhibiting a relatively large variance, which suggests that any energy input into the budget volume (i.e., anomalous surface fluxes or energy transport through the southern gateway) mainly changes the ocean heat content so that resulting energy transports through the northern gateways are only weakly affected. This is supported by a strong correlation between OHCT\(^{\text{adj}}\) and \(F_{\text{ref}}^{\text{ERA5}}\), which is \(\rho = 0.66\) and 0.46 for the ND and SD, respectively. Additionally, the OHCT\(^{\text{adj}}\) correlation with the transport through the southern gateway is \(\rho = 0.27\) in the ND and 0.64 in the SD, whereas the correlation with the northern gateway is in all cases \(\rho < 0.15\).

The OHCT\(^{\text{adj}}\) in the SD exhibits anomalous peaks in 1998/99 and 2011/12, which do not occur in the ND. Both peaks are caused by the northward advection of tropical oceanic heat (visible as anomalous peak in the northward transport at 26°N derived from ORAS5; not shown), which originated in the Pacific during the preceding El Niños and transported to the tropical Atlantic Ocean by the atmospheric bridge (see Mayer et al. 2014 and Fig. 1 therein).

The mean budget residual is \(-8.6 (4.6)\) W m⁻² in the ND (SD) and 4.1 W m⁻² in SD+ND (see Table 3 for long-term averages of individual budget terms). As outlined in section 2, inconsistencies in the mass budget inevitably project on the estimated energy budget [see Eq. (4)]. This effect can be estimated by converting the mass budget residual to corresponding spurious energy fluxes. We find an average mass budget residual of \(+0.09 (±0.47)\) Sv (i.e., mass excess within the domain) in ND (SD) and \(-0.06\) Sv in SD+ND, which corresponds to an average erroneous heat flux of \(2.4 (1.3)\) W m⁻² in ND (SD) and \(1.4\) W m⁻² in SD+ND when using \(\Delta T = 15\) K (see section 2). This erroneous heat flux is subtracted from the energy budget residual increasing the ND residual mean to \(-11.0\) W m⁻², whereas that of SD+ND (SD) is reduced to \(2.7 (3.2)\) W m⁻².

We repeated this procedure using \(F_{\text{ref}}^{\text{adj}}\) instead of \(F_{\text{ref}}^{\text{ERA5}}\), which is shown only for SD+ND (see Fig. 7c). \(F_{\text{ref}}^{\text{adj}}\) in the SD+ND has a 1985–2015 mean of \(-0.40\) W m⁻² (compared to \(-0.37\) W m⁻² of \(F_{\text{ref}}^{\text{ERA5}}\)), leading to a budget residual of \(1.2\) W m⁻², or \(-0.2\) W m⁻² after subtracting the erroneous heat flux associated with inconsistent mass budget. In the ND (SD), the budget residual is \(-13.6 (2.1)\) W m⁻², or \(-16.0 (0.8)\) W m⁻² in a mass-consistent budget. Hence, the energy budget residual is reduced in all but the ND when \(F_{\text{ref}}^{\text{adj}}\) is employed.

We find a pronounced annual cycle (not shown; note that time series in Fig. 7 are smoothed by a 12-month moving average) in the ocean budget residual of each domain. The northern (southern) domain has a minimum of \(-0.1 (±0.2)\) PW in March and a maximum of \(-0.1 (0.3)\) PW in September (October). Consequently, the annual cycle of the SD+ND residual looks very similar, with a minimum of \(-0.3\) PW in March and a maximum of \(-0.4\) PW in October, which indicates that the annual cycles of the RAPID transport, \(F_{\text{ref}}^{\text{adj}}\) and \(\text{OHCT}_{\text{ref}}^{\text{adj}}\) exhibit inconsistencies.

Figure 7 also shows indirectly estimated transports (indicated with superscript “ind”) through the southern gateway of each domain (GSR in the ND, and RAPID in the SD and SD+ND). The long continuous DS+GSR measurements used in the SD budget (Fig. 7b) allow us to indirectly estimate the transport at 26.5°N for 25 years, almost twice as long as...
currently available RAPID observations. Note that this approach is different from earlier attempts that inferred OHT at 26.5°N (Trenberth and Fasullo 2017; Liu et al. 2020) by integration of the budget terms between 26.5°N and the Bering Strait, using the latter as a choke point. The indirect estimate of OHTRAPID shows good temporal variability with no statistically significant trend, while RAPID observations clearly show a decrease between 2004 and 2009. The subsequent wind-driven decrease in 2009/10 (McCarthy et al. 2012) is longer lasting in the indirect estimate as the ocean cools (negative OHTCT) until the beginning of 2011. However, we find a correlation of 0.72 (or 0.74 with adjusted inferred surface fluxes) between observed and indirectly estimated transports for the period when both are available (2004–16; see Fig. 7b). This is slightly better than previous estimates from the literature; for example, Liu et al. (2020) obtained a correlation of 0.66 for the same period using inferred surface fluxes derived from ERA-Interim. Furthermore, the 2004–13 mean

Fig. 7. Temporal evolution of individual oceanic energy budget terms for the (a) northern domain (as defined in Fig. 1), (b) southern domain, and (c) combined northern plus southern domain. Shown are the full depth ocean heat content tendency (solid black line), unadjusted inferred surface fluxes derived from ERA5 (solid orange line), ocean energy transports through the northern (southern) gateway of each domain [solid blue (red) line], indirectly estimated transports through the southern gateways (dashed red line), and the residual of the oceanic energy budget (dotted gray line). Additionally, we show the adjusted inferred surface fluxes (solid yellow line) and the corresponding budget residual (dotted dark gray line) for the combined domain in (c). All lines are smoothed with a 12-month moving average. Temporal standard deviations are in parentheses. Units are PW (10^15 W).
indirect transport is 1.09 PW (or 1.17 PW with adjusted inferred surface fluxes), which is somewhat closer to the observed transport of 1.21 PW than values provided by Trenberth and Fasullo (2017).

In summary, based on our evaluations, the oceanic energy budget in the Atlantic Ocean is closed to within ~10 W m⁻². We find an average budget residual of 2.7 W m⁻² for the closed ocean domain between RAPID array and three northernmost Arctic Gateways (Davis Strait, Fram Strait, and the Barents Sea Opening) using $F_{\text{inf}}$ERA5 (or ~0.2 W m⁻² when using $F_{\text{inf}}$S,Era5). Nonetheless, the bias increases substantially with smaller averaging areas (as shown with the energy budget for ND and SD), which is most likely due to the transport imbalance divided by a relatively small area. It should also be noted that the necessary processing of the divergence (interpolation and truncation) can introduce uncertainties. It can change spatial averages of TEDIT, especially for small averaging areas. For instance, the long-term mean of TEDIT from Eq. (1) can differ up to 4 W m⁻² for the ND when interpolated from the native grid to the F90 grid used here, which consequently affects the inferred surface fluxes. Thus, although the interpolation error cannot fully explain the nonclosure of oceanic energy budgets, it can pose a significant contribution to it.

c. Comparison with buoy-based surface energy fluxes

In the following, we compare the three surface flux products $F_{\text{inf}}$ERA5, $F_{\text{model}}$ERA5, and $F_{\text{S,CERES}}$ERA5+OA with buoy-based surface flux estimates (denoted as $F_{\text{S,Buoy}}$) from 14 buoy locations (seven in the Pacific, five in the Atlantic, and two in the Indian Ocean; see Table 4). Please keep in mind that turbulent heat fluxes from OAFlux are tuned to these buoy-based fluxes (Yu et al. 2008). We do not discuss $F_{\text{inf}}$ERA5 in this section as differences from $F_{\text{inf}}$ERA5 are negligible on the station scale.

Figures 8 and 9 show anomaly time series and mean climatologies of surface fluxes at the 14 buoy locations. Surface fluxes near the equator are characterized by their weak seasonality and negative anomalies caused by El Niños (e.g., in 2009/10 and in some cases also in 2015/16). At higher

Table 3. Long-term means of oceanic energy budget terms for the northern domain (ND), southern domain (SD), and their combination (SD+ND) (see Fig. 1 for an overview map). Units are W m⁻², and PW in parentheses.

<table>
<thead>
<tr>
<th>Term</th>
<th>ND</th>
<th>SD</th>
<th>SD+ND</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{\text{inf}}$ERA5</td>
<td>-68.42 (−0.17)</td>
<td>-33.89 (−0.77)</td>
<td>-37.33 (−0.94)</td>
</tr>
<tr>
<td>OHCT(adj)</td>
<td>1.17 (0.003)</td>
<td>0.92 (0.02)</td>
<td>0.94 (0.02)</td>
</tr>
<tr>
<td>OHTRAPID</td>
<td>—</td>
<td>(1.20)</td>
<td>(1.20)</td>
</tr>
<tr>
<td>OHTFRAM+BSO</td>
<td>(0.12)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>OHTGSR</td>
<td>(0.28)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>OHTDIS+GSR</td>
<td>—</td>
<td>(0.28)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>Residual</td>
<td>-8.63 (−0.02)</td>
<td>4.58 (0.10)</td>
<td>4.13 (0.10)</td>
</tr>
<tr>
<td>Mass-corrected residual</td>
<td>−10.98</td>
<td>3.24</td>
<td>2.69</td>
</tr>
</tbody>
</table>

Table 4. List of buoys used in this study, with corresponding RMSE and bias for unadjusted inferred surface fluxes ($F_{\text{inf}}$ERA5), model-based surface fluxes from ERA5 forecasts ($F_{\text{model}}$ERA5), and satellite-derived surface fluxes ($F_{\text{S,CERES}}$ERA5+OA) based on original time series. As reference, buoy-based net surface fluxes are used. For each buoy location, the lower value of $\tilde{F}_{\text{bx,product}}$ (or $\tilde{F}_{\text{bx,ERA5}}$, $\tilde{F}_{\text{bx,ERA5}}$+GSR) is highlighted in bold. Mean values are averaged over all 14 buoys and weighted by the length of the time series. Units are W m⁻².

<table>
<thead>
<tr>
<th>No.</th>
<th>Project</th>
<th>Location</th>
<th>$F_{\text{inf}}$ERA5</th>
<th>$F_{\text{model}}$ERA5</th>
<th>$F_{\text{S,CERES}}$ERA5+OA</th>
<th>$F_{\text{inf}}$ERA5</th>
<th>$F_{\text{model}}$ERA5</th>
<th>$F_{\text{S,CERES}}$ERA5+OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TOA/TRITON</td>
<td>0°N 140°W</td>
<td>28.53</td>
<td>34.96</td>
<td>18.15</td>
<td>-21.03</td>
<td>-30.66</td>
<td>6.03</td>
</tr>
<tr>
<td>2</td>
<td>TOA/TRITON</td>
<td>0°N 165°E</td>
<td>18.42</td>
<td>19.44</td>
<td>22.14</td>
<td>-5.96</td>
<td>-8.24</td>
<td>14.97</td>
</tr>
<tr>
<td>4</td>
<td>PIRATA</td>
<td>0°N 23°W</td>
<td>23.55</td>
<td>32.48</td>
<td>17.49</td>
<td>-15.39</td>
<td>-28.55</td>
<td>0.19</td>
</tr>
<tr>
<td>5</td>
<td>PIRATA</td>
<td>10°S 10°W</td>
<td>30.43</td>
<td>37.34</td>
<td>20.71</td>
<td>-25.04</td>
<td>-27.23</td>
<td>14.13</td>
</tr>
<tr>
<td>6</td>
<td>PIRATA</td>
<td>12°S 23°W</td>
<td>22.89</td>
<td>29.27</td>
<td>39.96</td>
<td>-2.66</td>
<td>9.85</td>
<td>29.10</td>
</tr>
<tr>
<td>7</td>
<td>PIRATA</td>
<td>15°N 38°W</td>
<td>25.56</td>
<td>34.90</td>
<td>31.50</td>
<td>-13.05</td>
<td>-21.79</td>
<td>25.10</td>
</tr>
<tr>
<td>8</td>
<td>RAMA</td>
<td>15°N 90°E</td>
<td>26.69</td>
<td>24.69</td>
<td>22.22</td>
<td>-14.70</td>
<td>-16.92</td>
<td>-7.94</td>
</tr>
<tr>
<td>9</td>
<td>RAMA</td>
<td>8°S 67°E</td>
<td>31.32</td>
<td>27.84</td>
<td>17.08</td>
<td>-23.68</td>
<td>-23.14</td>
<td>-3.10</td>
</tr>
<tr>
<td>10</td>
<td>OCS KEO</td>
<td>32°N 145°E</td>
<td>38.03</td>
<td>27.89</td>
<td>21.61</td>
<td>-31.07</td>
<td>-18.19</td>
<td>-2.31</td>
</tr>
<tr>
<td>11</td>
<td>OCS PAPA</td>
<td>50°N 145°W</td>
<td>23.66</td>
<td>19.03</td>
<td>11.95</td>
<td>-19.07</td>
<td>-14.94</td>
<td>3.74</td>
</tr>
<tr>
<td>12</td>
<td>WHOI NTAS</td>
<td>14.7°N 51°W</td>
<td>31.83</td>
<td>35.05</td>
<td>19.65</td>
<td>-28.18</td>
<td>-32.02</td>
<td>14.52</td>
</tr>
<tr>
<td>13</td>
<td>WHOI Stratus</td>
<td>20°S 85.3°W</td>
<td>28.68</td>
<td>30.30</td>
<td>17.01</td>
<td>-24.31</td>
<td>-22.96</td>
<td>11.71</td>
</tr>
<tr>
<td>14</td>
<td>WHOI WHOTS</td>
<td>22.8°N 158°W</td>
<td>26.45</td>
<td>34.05</td>
<td>18.47</td>
<td>-20.52</td>
<td>-28.58</td>
<td>9.76</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td>27.78</td>
<td>29.80</td>
<td>20.80</td>
<td>-20.07</td>
<td>-21.47</td>
<td>9.21</td>
</tr>
</tbody>
</table>
latitudes, fluxes are dominated by their strong annual cycle, with positive (negative) values during summer (winter) months of the corresponding hemisphere, while anomalies are relatively small. At buoy locations between the tropics of Cancer and Capricorn, climatologies exhibit a double peak shifted by at least three months, which is caused by the seasonal shift of maximum solar insolation.

We find that at 8 of 14 locations (all five in the Atlantic Ocean and three in the Pacific Ocean, including all three WHOI buoys), the correlation between anomalies of $F_{\text{inf}}^\text{ERA5}$ and $F_{\text{S,Buoy}}$ (see $r$ values in the left panels of Figs. 8 and 9) is greater than that of $F_{\text{model}}^\text{ERA5}$ and $F_{\text{S,Buoy}}$. In only three cases it is also larger than the correlation between $F_{\text{S,CERES+OA}}$ and $F_{\text{S,Buoy}}$ anomalies. In 7 out of 14 cases, buoy-based flux anomalies correlate best with $F_{\text{S,CERES+OA}}$. In the remaining three cases, they show the largest correlation with model-based fluxes from ERA5 forecasts. This indicates that buoy-based surface flux anomalies are reproduced best by satellite-
derived fluxes from CERES+OAFlux, but differences among the four products are in general small.

Furthermore, at 10 of 14 locations unadjusted inferred surface fluxes are in better agreement with the $F_{S,Buoy}$ climatology than model-based fluxes from ERA5 forecasts (based on the RMS deviation of climatologies; not shown), and in four cases also better than $F_{S,CERES+OA}$. However, in the other 10 cases satellite-derived fluxes from CERES+OAFlux exhibit the best agreement with the buoy climatology, while $F_{S,ERA5}$ is not able to do that at any of these buoy locations.

In Fig. 10, we present bias and RMSE maps of the three surface flux products, where $F_{S,Buoy}$ is used as reference (note that these maps are based on original time series, i.e., climatology plus anomalies, whereas in Figs. 8 and 9 we showed them separately). The latter (Figs. 10a–c) reveals that ERA5-based products exhibit overall larger RMSE compared to
CERES+OAFlux. Nonetheless, in 9 out of 14 cases \( \frac{F_{\text{inf}, \text{ERA5}}}{F_{\text{model}, \text{ERA5}}} \) performs better than \( \frac{F_{\text{model}, \text{ERA5}}}{F_{\text{CERES+OA}}}; \) ERA5 (see Fig. 10d), with a mean RMSE averaged over all 14 buoys of 27.8 W m\(^{-2}\) compared to 29.8 W m\(^{-2}\) for \( F_{\text{model}, \text{ERA5}} \). In four cases (Fig. 10i), the RMSE of \( F_{\text{inf}, \text{ERA5}} \) is also smaller than that of \( F_{\text{CERES+OA}} \), because there CERES radiative fluxes do not agree well with those measured from buoys. The RMSE of \( F_{\text{CERES+OA}} \) is 20.8 W m\(^{-2}\) when averaged over all buoys, which is substantially smaller than that of ERA5-based fluxes (see Table 4). The same can be concluded from the bias metric (Fig. 10f–h). The bias averaged over all buoys is \(-20.1\) W m\(^{-2}\) for \( F_{\text{inf}, \text{ERA5}}; -21.5\) W m\(^{-2}\) for \( F_{\text{model}, \text{ERA5}} \), and \(-9.2\) W m\(^{-2}\) for \( F_{\text{CERES+OA}} \).
We conclude that unadjusted inferred surface fluxes are superior to model-based fluxes from ERA5 forecasts when using buoy-based fluxes as reference. Inferred fluxes exhibit a substantially smaller RMSE and bias, and also represent $F_{\text{TOA}}$ climatologies and anomalies reasonably well. However, both ERA5-based flux products show stronger deviations from buoy records than CERES+OAFlux, which can be attributed to the tuning of OAFlux to buoy-based fluxes.

5. Discussion and conclusions

We combine the atmospheric energy tendency and divergence of atmospheric energy fluxes from Mayer et al. (2021) with net TOA radiation from DEEP-C (Liu et al., 2017, 2020) to indirectly estimate turbulent plus radiative net surface energy fluxes (denoted as inferred fluxes) for the period 1985–2018, which are subsequently adjusted to observed global land averages as described in Liu et al. (2017). The adjusted as well as unadjusted inferred surface fluxes are compared with satellite-derived estimates from CERES+OAFlux, model-based fluxes from ERA5 forecasts, and buoy-based flux estimates (Table 1 lists all datasets used in this study).

We find a 1985–2018 mean global ocean surface heat flux of 1.7 W m$^{-2}$, which is smaller than that of other surface flux products that are constrained by observations. For example, surface fluxes from CERES+OAFlux have a long-term ocean mean of $\sim$28 W m$^{-2}$ for 60°N–60’S (see Fig. 5), and model-based fluxes from ERA5 forecasts exhibit an ocean mean of $\sim$6 W m$^{-2}$. Other widely used reanalysis-based fluxes show similar inconsistencies: ocean mean fluxes from the Japanese 55-year Reanalysis (Kobayashi et al., 2015) are at $-17$ W m$^{-2}$, and those from the Modern-Era Retrospective analysis for Research and Applications, version 2 (Gelaro et al., 2017) have a mean of $-5$ W m$^{-2}$ (Cronin et al., 2019). On the large scale, our inferred surface fluxes clearly benefit from the fact that the global mean divergence vanishes and the results are thus unbiased by construction (assuming the TOA flux product is unbiased) when considering global averages. However, they suffer from a spurious trend in the late 1990s, which likely stems from changes in the observing system [Robertson et al., 2020; see discussion in Mayer et al. (2021)]. Nevertheless, this spurious trend mainly dominates land averages, while it is only weakly noticeable in global ocean averages (because of the distribution over the much larger ocean area). It is evident that the spurious trend becomes less visible with smaller averaging areas, where interannual variability becomes the dominant signal (e.g., see Fig. 7).

Unlike the unadjusted inferred surface flux product, adjusted inferred surface fluxes presented in this study (see also Liu et al., 2017, 2020) retain some important physical properties: they conserve the global $F_{\text{TOA}}$ trend over the ocean (not shown) and agree by construction with the long-term OHCT mean, in agreement with the fact that $\sim$90% of the global TOA imbalance and its trends is stored in the ocean (von Schuckmann et al., 2020). Moreover, the adjusted inferred surface fluxes are fitted to the observed land heat uptake, which eliminates the spurious trend in the late 1990s present in the unadjusted inferred flux product. This makes the adjusted fluxes a good alternative to other commonly used surface flux products, which sometimes exhibit unrealistically strong trends or large long-term means. However, we would like to point out that only global land averages match the observed long-term mean land heat uptake; that is, regional land means may still contain unrealistic values and need further investigations, which will be done elsewhere.

The evaluation of the oceanic energy budget between the RAPID array and Arctic Gateways further suggests good accuracy and small bias of our inferred surface fluxes on the regional scale, with a budget residual of less than 3 W m$^{-2}$. This is remarkably small considering the fact that independent ocean heat transport datasets are used for this assessment. Additionally, the mean indirectly estimated transport at 26.5°N matches the observed transport from RAPID to within $\pm 0.1$ PW using unadjusted inferred surface fluxes, and even less when adjusted inferred fluxes are employed. We note that Mayer et al. (2019) found similarly good closure of the oceanic energy budget of the Arctic Ocean domain using inferred surface fluxes derived from ERA5, supporting the conclusion that we can generally achieve residuals $\leq 10$ W m$^{-2}$ for regional energy budgets with our data and methods.

The bias of inferred surface fluxes is largest at station scale. It should be noted that the majority of buoys used in this study are located in the tropical region between 30°N and 30°S, suggesting a bias of $-20$ W m$^{-2}$ in that particular region. Together with the global bias of roughly 1 W m$^{-2}$, this would suggest that inferred fluxes exhibit a bias on the order of $+20$ W m$^{-2}$ at higher latitudes (i.e., in regions not covered by buoys) to compensate the large station-scale bias in the tropics. However, this is contradicted by the results of our ocean budget evaluation in section 4b, which indicated good accuracy and small bias of inferred surface fluxes, even on relatively small scales and at high latitudes. In addition, the KEO (32°N, 145°E) and PAPA (50°N, 145°W) buoys located relatively far north do not point to a bias of opposite sign at high latitudes.

Most buoy-based latent and sensible heat fluxes are computed using the COARE algorithm in version 3b, and are not measured directly. In addition, it has been shown that different COARE versions lead to very different estimates of latent and sensible heat fluxes. Yu (2019) has demonstrated that turbulent heat fluxes computed with the COARE algorithm in version 4 (not yet released) tend to be stronger at all latitudes relative to those derived from older versions (differences of 5–20 W m$^{-2}$), which suggests that the station-scale bias of inferred fluxes would be smaller by this amount (more negative turbulent heat fluxes lead to less positive buoy-based heat fluxes and therefore smaller bias of our inferred fluxes). This makes buoy-based fluxes as a reference for other flux products debatable.

We further note that uncertainty assessments of measurements at individual buoy stations yielded bias estimates of similar order as the discrepancies between our flux estimates and buoy-based

estimates. For example, WHOI buoys outfitted with the Air-Sea Interaction Meteorology instrumentation are estimated to exhibit biases of up to 20% (equivalent to ~20 W m^−2 based on climatologies from Figs. 8 and 9) of the net ocean heat flux (Colbo and Weller 2009). Based on the unrealistically strong positive near-global mean of surface fluxes from CERES+OAFlux as well as results from Yu (2019), we deduce a positive bias (into the ocean) for buoy-based fluxes, most likely on the order of 10 W m^−2. This would suggest a smaller station-scale bias of our inferred surface fluxes, probably of −10 W m^−2 or less. Further research is needed to trace the sources of these discrepancies. It is critical to further reduce uncertainties, so that climate models can be validated with reliable benchmark datasets of air-sea fluxes.

Acknowledgments. JM and MM were financially supported by the Austrian Science Funds (FWF) project P33177. LH received support from the Austrian HRSM project GEOCLIM. CL was supported by the National Natural Science Foundation of China (42075036). The authors thank Hao Zuo (ECMWF) for production and provision of ORAP6.0 data. Computations were partly performed with routines provided through openIFS (https://confluence.ecmwf.int/display/OIFS/OpenIFS+Documentation). The use of openIFS was permitted by ECMWF. Data from the RAPID-MOCHA program are funded by the U.S. National Science Foundation and U.K. Natural Environment Research Council and are freely available at www.rapid.ac.uk/rapidmoc and mocha.rsmas.miami.edu/mocha. We acknowledge the GTMBA Project Office of NOAA/PMEL for providing Oceansites flux data at https://www.pmel.noaa.gov/tao/drupal/flux/index.html. The authors thank Aaron Donohoe and two anonymous reviewers for their insightful comments that helped to improve the manuscript.

Data availability statement. Data are accessible via the Copernicus Climate Data Store; see entry “Mass-consistent atmospheric energy and moisture budget data from 1979 to present derived from ERA5 reanalysis” (https://doi.org/10.24381/cds.c24516fb).

APPENDIX

Additional Table

Table A1 provides the same information as Table 4, but for the deseasonalized time series of \(F_{\text{ERAS}}\), model-based surface fluxes from ERA5 forecasts (\(F_{\text{model}}\)), and satellite-derived surface fluxes (\(F_{\text{OA}}\)). Note that the mean bias of the deseasonalized time series is zero.

REFERENCES


