



Diagnostic evaluation of river discharge into the Arctic Ocean and its impact on oceanic volume transports

Susanna Winkelbauer¹, Michael Mayer^{1,2}, Vanessa Seitner¹, Ervin Zsoter², Hao Zuo², and Leopold Haimberger¹

¹Department of Meteorology and Geophysics, University of Vienna, Vienna Austria

²European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

Correspondence: Susanna Winkelbauer (susanna.winkelbauer@univie.ac.at)

Abstract. This study analyses river discharge into the Arctic Ocean using state-of-the-art reanalyses such as the fifth-generation European Reanalysis (ERA5) and the reanalysis component from the Global Flood Awareness System (GloFAS). GloFAS, in its operational version 2.1, combines the land surface model (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land, HTESSEL) from ECMWF's ERA5 with a hydrological and channel routing model (LISFLOOD). Further we analyse

5 GloFAS most recent version 3.1, which is not coupled to HTESSEL but uses the full configuration of LISFLOOD. Seasonal cycles, as well as annual runoff trends are analysed for the major Arctic watersheds - Yenisei, Ob, Lena and Mackenzie - where reanalysis-based runoff can be compared to available observed river discharge records. Further we calculate river discharge over the whole Pan-Arctic region and, by combination with atmospheric inputs, storage changes from the Gravity Recovery and Climate Experiment (GRACE) and oceanic volume transports from ocean reanalyses, try to close the non-steric

10 water volume budget. Finally we provide best estimates for every budget equation term using a variational adjustment scheme. Seasonal river discharge peaks are underestimated in ERA5 and GloFAS v2.1 by up to 50%, caused by pronounced declining trends due to spurious signals in ERA5s data assimilation system. The new GloFAS v3.1 product exhibits distinct improvements and performs best in terms of seasonality and long term means, however opposing to gauge observations it also features declining trends. Calculating runoff indirectly through the divergence of moisture flux is the only reanalyses based estimate

15 that is able to reproduce the river discharge increases measured by gauge observations (Pan-Arctic increase of 2% per decade). In addition we look into Greenlandic discharge, which makes out about 10% of of the total Pan-Arctic discharge and features strong increases mainly due to glacial melting.

The variational adjustment brought reliable estimates of the volume budget terms on an annual scale, requiring only moderate adjustments of less than 1% for each individual term. Approximately $6584 \pm 84 \text{ km}^3$ freshwater leave the Arctic Ocean per

20 year through its boundaries. About two thirds of this are recovered through runoff from the surrounding land areas to the Arctic Ocean ($4379 \pm 25 \text{ km}^3$ per year) and about one third is supplied by the atmosphere. On a seasonal scale however the variational approach demonstrated that there are systematical errors present in the data-sets, that are not considered in their uncertainty estimation. Hence the budget residuals of some month were too large to be eliminated within the a priori spreads of the individual terms.



1 Introduction

Rapid surface warming in the Arctic region has strong impacts on the Arctic water balance and its individual hydrological components, almost certainly leading to an amplification in runoff, evapotranspiration and precipitation (Rawlins et al., 2010; Collins et al., 2014). Increasing river discharge and precipitation trends and intensified sea ice melt coupled with an increase of freshwater inflow through Bering Strait lead to an increase of liquid freshwater stored in the Arctic Ocean (Morison et al., 2012; Haine et al., 2015; Haine, 2020). Ultimately, this could result in enhanced southward exports of low-density waters (Lin et al., 2021) into the Atlantic Ocean, impacting the oceanic circulation also on a global scale. Altogether the hydrological cycle is a complex process with tight coupling between the individual components, having impacts on energy and mass budgets and eventually sea level rise. Therefore the quantification of the individual hydrological components and their changes is of great importance.

With the Arctic Ocean being almost entirely surrounded by landmasses and some of the world's largest rivers draining into it, the link between ocean and surrounding land is remarkably strong. Hence runoff forms one of the key variables in the Arctic freshwater budget. However, direct quantification of river discharge into the Arctic Ocean is aggravated by the fact that about 30-40% of the Pan-Arctic drainage area is now unmonitored (Shiklomanov et al., 2002). Hydrological monitoring suffered a widespread decline from 74% in 1986 to 67% by 1999 - in Siberia even 73% of river gauges were closed between 1986 and 1999 (Shiklomanov et al., 2002; Shiklomanov and Vuglinsky, 2008). In addition discharge may bypass gauging stations through braided channels or as submarine groundwater and also climatological conditions pose a hindrance to gauge measurements, as they lead to river freeze up in late autumn and flooding in spring due to river-ice break up (Syed et al., 2007).

Atmospheric reanalyses produce gridded estimates of atmospheric and land components, providing spatially continuous estimates of variables such as runoff. They represent a huge advancement in climate monitoring and we want to use this study to evaluate runoff from state-of-the art reanalyses and GloFAS to find a best estimate of Pan-Arctic river discharge and to incorporate it into the Arctic freshwater and volume budget. However data assimilation systems can introduce biases and also inhomogenities due to changes in the observing system are often inevitable (Hersbach et al., 2020). This can lead to spurious jumps in the time series of reanalyzed quantities. This is the case for ERA5s runoff component which features spurious trends in higher latitudes.

The first complete freshwater budget for the Arctic Ocean is proposed by Aagaard and Carmack (1989) and updated by Serreze et al. (2006) and Dickson et al. (2007). Since then, the amount of available data, in particular of atmospheric and oceanic reanalyses, opened new possibilities for evaluation of the coupled oceanic and atmospheric energy and hydrological cycles. For example, Mayer et al. (2019) have presented a greatly improved depiction of the Arctic energy cycle. With regard to freshwater or the water volume budget, the data situation has improved as well. New collections of hydrological data of the far north have been published (Shiklomanov et al., 2021b). Tsubouchi et al. (2012, 2018) presented observation-based estimates of volume fluxes through Arctic gateways. This opens the opportunity to overspecify the Arctic volume budgets and thus also to give residual and bias estimates.

This paper is structured as follows. The next section describes the used data and presents the study domain, followed by the



methodology. The results are presented in Sect. 4 and are subdivided into seasonal cycles and trends for the four major Arctic watersheds (Sect. 4.1), Pan-Arctic seasonalities and trends (Sect. 4.2) and an assessment of budget closure by comparison with oceanic fluxes (Sect. 4.3). Section 5 presents conclusions and in the appendix a list of acronyms used throughout the text can be found.

2 Data and study domain

Runoff is taken from the European Centre for Medium-Range Weather Forecast's (ECMWF) 5th generation global climate reanalysis ERA5 (Hersbach et al., 2020), as well as from its offline simulation ERA5-Land and is downloaded through the Copernicus Climate Change Service (C3S) Climate Data Store (Hersbach et al., 2019; Muñoz Sabater, 2019). Runoff from ERA5 and ERA5-Land are both produced by the land-surface model Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL, Balsamo et al. (2009)) of the ECMWF Integrated Forecasting System (IFS). However contrary to ERA5, ERA5-Land is not coupled to the atmospheric model of the IFS and no data assimilation is used. An advantage of ERA5-Land is the enhanced global resolution of 9 km (31 km for ERA5) (Muñoz Sabater et al., 2021). Runoff data are converted into river discharge, by integration over the associated domain.

Furthermore we consider the Global Flood Awareness System (GloFAS) river discharge reanalysis. GloFAS is developed by ECMWF and the Joint Research Centre (JRC), as part of the Copernicus Emergency Management Service (CEMS). GloFAS is publically available with data accessible from the Copernicus Climate Change Service Climate Data Store (CDS). Its operational version, GloFAS 2.1 (Harrigan et al. (2019), hereafter denoted $GloFAS_{E5}$), combines a simplified version of the hydrological river routing model LISFLOOD (Knijff et al., 2010), to simulate groundwater processes and river routing, with runoff data from HTESSEL, the land surface model used in ERA5 (Harrigan et al., 2020). In addition we examine an experimental GloFAS version, that also uses LISFLOODs channel routing, but forces it with runoff from ERA5-Land - hereafter denoted $GLOFAS_{E5L}$ - and GloFAS version 3.1 ($GLOFAS_{E5new}$), which uses the full configuration of the LISFLOOD model and is not coupled to HTESSEL but rather produces its own runoff by using directly precipitation, evaporation and temperature from ERA5. While $GLOFAS_{E5}$ and $GLOFAS_{E5new}$ are available from 1979 to near real time, the experimental version $GLOFAS_{E5L}$ is only available from 1999 to 2018.

Data from ERA5 and GloFAS are compared to available observed river discharge records. Observing records vary among the various countries and rivers, with the longest time series coming from Russia, where discharge monitoring began in the mid 1930s. In contrast discharge measurement in North America did not begin until the 1970s (Holmes et al., 2018). The data used in this study comes from Roshydromet (Ob, Yenisei and Lena) and from the Water Survey of Canada (Mackenzie) and was downloaded through the Arctic Great Rivers Observatory (Shiklomanov et al., 2021b). Table 1 shows coordinates of gauge observations and GloFAS sampling locations for Yenisei, Ob, Lena and Mackenzie. For the Pan-Arctic approach river discharge from additional 20 rivers was taken. Gauging records for Kolyma, Severna Dvina, Pechora and Yukon are available for our period of interest 1981-2019 and are also downloaded through the Arctic Great Rivers Observatory (Shiklomanov et al., 2021b), while records for 16 further rivers are taken from Regional Arctic Hydrographic Network data set (R-ArcticNET, Lammers



	Gauges	GloFAS
Yenisei	67.48°N; 86.50°E	67.45°N; 86.45°E
Ob	66.57°N; 66.53°E	66.55°N; 66.45°E
Lena	70.70°N; 127.65°E	72.25°N; 126.75°E
Mackenzie	67.45°N; -133.75°E	67.45°N; -133.75°E

Table 1. Positions of gauge observations and GloFAS locations for Ob, Yenisei, Lena and Mackenzie.

Straits	ORCA
Fram	78.80°-78.80°N, 20.60°W-11.50°E
Davis	66.60°-67.30°N, 61.20°-54.00°W
Bering	65.90°-65.70°N, 170.00°-168.30°W
BSO	77.40°-69.70°N, 18.00°-20.40°E
Hecla and Fury	69.85°-70.00°N, 84.50°-84.32°W

Table 2. Start and end points of the sections used for lateral flux calculations on the native ORCA grid.

et al., 2001) for the period 1981-1999.

Atmospheric components like precipitation, evaporation, atmospheric storage change and the divergence of moisture flux (VI-
 95 WVD) are taken from ERA5 and in Sect. 4.3 we additionally use VIWVD data from the Japanese 55-year Reanalysis JRA55
 (Kobayashi et al., 2015) and JRA55 Conventional (Kobayashi et al., 2014). Land storage is derived from snow depth (given as
 water equivalent) and soil water changes from ERA5.

Oceanic volume fluxes through the main Arctic Gateways are calculated by integrating the cross-sectional velocity component
 around the Pan-Arctic boundary from the Copernicus Marine Environment Monitoring Service (CMEMS) Global ocean Re-
 100 analysis Ensemble Product (GREP, Desportes et al., 2017; Storto et al., 2019), an ensemble of four global ocean reanalyses
 for the period from 1993 to present. GREP consists of current ocean reanalysis efforts from the Centro Euro-Mediterraneo sui
 Cambiamenti Climatici (CGLORS, Storto and Masina, 2016), the UK Met Office (FOAM, Blockley et al., 2014), Mercator
 Ocean (GLORYS, Garric et al., 2017) and the ECMWF (ORAS5, Zuo et al., 2018, 2019). The four ocean reanalyses are all
 forced by the ERA-Interim (Dee et al., 2011) reanalysis with different bulk formulas and differences in the data assimilation
 105 schemes and observational datasets. For more information see Storto et al. (2019). We also look into the runoff climatology
 Bt06 (Bourdalle-Badie and Treguier, 2006), that is used in the global ocean-ice model ORCA025. In addition volumetric fluxes
 are derived from moorings within the so called ARCGATE project (Tsubouchi et al., 2019), covering the period from October
 2004 to May 2010. These observation based estimates of volume fluxes come from a mass-consistent framework that views
 the Arctic Ocean as a closed box surrounded by landmasses and hydrographic observation lines placed in the four major Arctic
 110 gateways. The hydrographic lines consist of arrays of moored instruments measuring variables like temperature, salinity and
 velocity, making it possible to calculate fluxes of volume, heat and freshwater. For more details about the framework see Tsub-
 ouchi et al. (2012, 2018).



Figure 1. Map of the main study area, consisting of the oceanic area bounded by moorings in Bering Strait (BS), Davis Strait (DS), Fram Strait (FS), the Barents Sea Opening (BSO) as well as Hecla and Fury Strait (HS & FuS) (indicated by grey shading; corresponds to 11.3×10^6 km²) and the land area draining into it (blue shading; corresponds to 18.2×10^6 km²). Further the four largest river catchments - Ob, Yenisei, Lena and Mackenzie - are displayed.

Additionally storage terms for land and ocean are calculated using ocean bottom pressure changes and land water storage from GRACE (Gravity Recovery & Climate Experiment) Release 6.0 version 03 from April 2002 to June 2017. Monthly
 115 ocean bottom pressure anomalies and land mass anomalies are derived from time-variable gravity observations and are given as equivalent water thickness changes (Landerer, 2020a, b). Satellite observations are processed by three different centers, the Center for Space Research at University of Texas, Austin (CSR), the Jet Propulsion Laboratory (JPL) and the Geoforschungs Zentrum Potsdam (GFZ). We estimate the optimal ocean bottom pressure and land water storage components by taking the mean of those three solutions. Land storage includes the total terrestrial water storage anomalies from soil moisture, snow,
 120 surface water, groundwater and aquifers.

Figure 1 presents the study domain. The Arctic Ocean is bounded by moorings in the main gateways. Bering Strait forms the only passage to the Pacific Ocean and delivers low salinity waters into the Arctic, while liquid freshwater and sea ice leave the Arctic Ocean mainly through Fram and Davis Strait. The fourth major strait is the Barents Sea Opening (BSO), where high salinity waters from the Atlantic Ocean are imported into the Arctic. Furthermore there are two small passages, Fury and Hecla
 125 Strait, that connect the Arctic Ocean with Hudson Bay.

The terrestrial domain consists of land areas draining into the Arctic Ocean, excluding Greenland but including the Canadian Arctic Archipelago (CAA) as well as islands along the Eurasian coast. At the Pacific passage Yukon and Anadyr rivers are



considered, as they build important sources for inflow of low salinity waters into the Arctic Ocean through Bering Strait. The total oceanic and terrestrial areas correspond to $11.3 \times 10^6 \text{ km}^2$ and $18.2 \times 10^6 \text{ km}^2$ respectively. For volume budget analyses we further incorporate Greenlandic discharge and storage change north of Davis and Fram Strait, which adds an additional area of $0.95 \times 10^6 \text{ km}^2$.

3 Methods

3.1 Budget equations

A popular way to calculate the oceanic freshwater budget is through the assumption of a reference salinity. However the outcome is dependent of those reference salinities in a nonlinear way, so that slight differences in the choice of the reference value lead to very different estimates of freshwater transports, both temporally and spatially. Hence Schauer and Losch (2019) declared freshwater fractions not useful for the analysis of oceanic regions and rather recommend the usage of salt budgets for salinity assessments. In this paper we do not calculate salt budgets, but we estimate volume budgets and hence also avoid the usage of a reference salinity. Hereinafter the volumetric budget equations for atmosphere, land and ocean are formulated.

140 a) Atmosphere

The change of water storage in the atmosphere - here expressed as water vapor integrated from the earth's surface to the top of the atmosphere (i.e. total column water vapor; hereafter denoted S_A ; atmospheric liquid water and ice are neglected) - denotes the left side of the volumetric budget equation for the atmosphere and is balanced by the surface freshwater fluxes evapotranspiration ET and precipitation P (both in SI units ms^{-1}) and the vertically integrated horizontal moisture flux divergence (last term; hereafter denoted as VIWVD):

$$\frac{\partial}{\partial t} \left(\frac{1}{g\rho_w} \int_0^{p_s} q dp \right) = ET - P - \left(\nabla \cdot \left(\frac{1}{g\rho_w} \int_0^{p_s} q \mathbf{v} dp \right) \right) \quad (1)$$

with the gravitational constant g , surface pressure p_s , specific humidity q , density of freshwater ρ_w and the horizontal wind vector \mathbf{v} . The equation above is probably more familiar to the reader as a mass budget equation, without ρ_w in the denominator. In order to get volumetric water fluxes, we divided by the density of freshwater $\rho_w = 1000 \text{ kg m}^{-3}$, which is assumed constant in this paper, neglecting dependence on temperature and soluble substances. Further all terms are integrated over the Arctic area A_{total} to obtain SI units of $\text{m}^3 \text{s}^{-1}$. For presentation of results, we will often use $Sv (=10^6 \text{ m}^3 \text{s}^{-1})$ or $\text{km}^3 \text{a}^{-1}$ as more convenient units.

150 b) Land

The change in land water storage (S_L) can be expressed as sum of changes of volumetric soil water $SWVL_n$ integrated over the corresponding soil depth l_n and changes in snow depth SD - given as snow water equivalent. Changes in land water storage are balanced through precipitation P_L and evapotranspiration ET_L over land and runoff R (in SI units ms^{-1}). To obtain



volumetric fluxes we again perform areal integration over the corresponding area (here the land area A_l).

$$\frac{\partial}{\partial t} \left(SD + \int_{l_n} SWV L_n dl_n \right) = P_L - ET_L - R \quad (2)$$

Rearranging Eq. (2) we obtain an estimate of the water discharging into the Arctic Ocean independent of runoff itself. As we
 160 prefer analyzed quantities we can further insert Eq. (1) to substitute P and ET , which are derived from short-term forecasts in
 ERA5, through the analyzed quantities $VIWVD$ and atmospheric storage change:

$$R = -\frac{\partial S_L}{\partial t} + P_L - ET_L = -\frac{\partial S_L}{\partial t} - \frac{\partial S_A}{\partial t} - VIWVD \quad (3)$$

c) Ocean

As per Bacon et al. (2015) the oceanic mass budget equation can be expressed as

$$165 \quad \iiint_V \frac{\partial \rho}{\partial t} dV = F^{surf} - \iint_{A_{sz}} \rho \mathbf{v} \cdot \mathbf{n} ds dz \quad (4)$$

The left side of Eq. (4) denotes the change in mass over a closed volume, F^{surf} describes any surface mass input/output in the
 form of precipitation, evaporation and runoff. The last term of Eq. (4) expresses ice and ocean side-boundary fluxes from or into
 the volume (integration is performed along the depth z and the along-boundary coordinate s). Further we apply the Boussinesq
 approximation and assume ρ as constant. We adopt the reference salinity used in the Nucleus for European Modelling of the
 170 Ocean (NEMO) ocean model of $\rho_O = 1035 \text{ kg/m}^3$ (Madec and Team, 2019) and divide Eq. (4) by ρ_O . This yields an expression
 where steric effects are ignored and only volume, also considered as Boussinesq mass ($M_O = \rho_O V$), is conserved (Madec and
 Team, 2019; Bacon et al., 2015):

$$\frac{\partial}{\partial t} \left(\iiint_V dV \right) = P_O - ET_O + R - \iint_{A_{sz}} \mathbf{v} \cdot \mathbf{n} ds dz \quad (5)$$

Hence the change of oceanic volume (derived from bottom pressure changes; hereafter denoted S_O) is balanced by precipitation
 175 and evapotranspiration over the oceanic domain (P_O and ET_O), runoff from the land domain R and further volume can leave
 and enter the ocean laterally over it's vertical boundaries, described by the last term of the equation (oceanic lateral transport,
 denoted F). F is calculated by integrating the cross-sectional velocity component along the side areas of the Arctic boundary.
 Volume exchange between liquid ocean and sea-ice is conserved in the NEMO model, so we assume sea-ice to be transported
 by the ocean currents and do not explicitly add ice volume transports to prevent ice transports from being counted twice.

180 Changes in ocean density do not affect the volume as the steric effect is missing due to the Boussinesq approximation.

In this paper we mostly present monthly means, derived by averaging the corresponding fields from reanalyses in their native
 temporal resolution. Horizontal interpolation and vertical interpolation has been avoided by using all reanalysis products in
 their native grid representation. Care has been taken also to average over the same area for all products as far as this is possible.
 The lateral fluxes through the ocean gateways were evaluated along paths on the native ORCA grid that followed the ARCGATE



185 mooring arrays as closely as possible - ORCA coordinates are given in table 2. This is essential, since the net lateral volume
 fluxes in and out of the Arctic are very small (~ 0.2 Sv in the annual mean) compared to the fluxes through individual straits
 (e.g. ~ 2.3 Sv for Davis Strait, Curry et al., 2011).

3.2 Variational approach for budget closure

Given all the various, largely independent data sets we use, closure between the budget terms will not be perfect, resulting in a
 190 budget residual. In order to get rid of any residual and obtain a closed budget with physical terms only, we follow Mayer et al.
 (2019), L'Ecuyer et al. (2015) and Rodell et al. (2015) and use a variational Lagrange multiplier approach to enforce budget
 closure on annual and monthly scales. Therefore the following cost function J is minimized:

$$J = \sum_i \frac{(F_i - F'_i)^2}{\sigma_i'^2} + \lambda \sum_i F'_i \quad (6)$$

With the Lagrange multiplier λ , the a priori estimates of the budget terms F'_i , the adjusted budget terms F_i , the uncertainty of
 195 the respective budget term $\sigma_i'^2$ and the budget residual $\sum_i F'_i$.

a) Annual optimization

Inserting the annual means of the individual budget terms into Eq. (6) and differentiation in respect to λ and the a priori esti-
 mates of the budget terms yields eight equations with eight unknowns. Solving the system of equations results in an expression
 for the adjusted budget terms F_k :

$$200 \quad F_k = F'_k + \frac{\sigma_k'^2}{\sum_i \sigma_i'^2} \sum_i F'_i \quad (7)$$

Hence the budget residual is distributed across the budget terms according to their relative uncertainty. The a priori uncertain-
 ties are derived from the standard deviations of the mean annual budget terms. The a posteriori uncertainties are calculated
 following Mayer et al. (2018):

$$\sigma_k^2 = \left(\frac{1}{\sigma_k'^2} + \frac{1}{\sum_i \sigma_i'^2 - \sigma_k'^2} \right) \quad (8)$$

205 b) Monthly optimization

Monthly optimization is performed in two steps. First adjusted fluxes are calculated for each month separately following Eq.
 (7), whereat the a priori uncertainty is estimated by taking the maximum of the seasonal standard deviations and is kept fixed
 throughout all months. However the annual means of the resulting monthly fluxes do not coincide with the annually optimized
 fluxes. Therefore we follow Rodell et al. (2015) and apply a second Lagrangian optimization, where the adjusted monthly
 210 fluxes from the first step (F_k) are adjusted in relation to their uncertainty, so that their annual mean is equal to the annual
 optimized fluxes F_m :

$$FO_k = F_k + \frac{12\sigma_k'^2}{\sum_i \sigma_i'^2} \left(F_m - \frac{1}{12} \sum_i F_i \right) \quad (9)$$



However the second step again generates small monthly residuals. Therefore the whole procedure is performed iteratively a second time, using the a posteriori uncertainties gained through Eq. (8). This results in the desired monthly fluxes, that satisfy both, a closed budget and consistency with the annually optimized fluxes.

3.3 Trend and relative error calculation

We calculate trends following Zsótér et al. (2020) and Stahl et al. (2012) by applying a linear regression to the annual mean time series and calculating trends over a fixed 10 year period:

$$t = \frac{10 * s}{m} \quad (10)$$

With the long term annual means m and the slope s - defined as annual change from the linear regression. Hence t determines the change over a decade relative to the sample mean value. Trends are calculated over the common period of all datasets 1981-2019, except for $GloFAS_{E5L}$ which is calculated over 1999-2018, using annual means and not considering auto-correlation. Significance is determined by using the Wald Test with a t-distribution, where p-values smaller than 0.05 are considered as significant.

To compare river discharge estimates from the various reanalyses to river discharge observations we use the Pearson's correlation coefficient r and a normalised root mean square error (NRMSE), which is calculated by dividing the RMSE through the RMS of the observed values ($NRMSE(x) = RMS(x - obs) / RMS(obs)$).

4 Results

We first discuss seasonal cycles and trends for the four major Arctic catchments - Yenisei, Ob, Lena and Mackenzie. Then we extend our assessments to the Pan-Arctic region, where we compare the total terrestrial Arctic runoff with oceanic freshwater fluxes through the main gateways.

4.1 Analysis of major catchments

4.1.1 Seasonal cycles

Figure 2 shows seasonal cycles of various hydrological components for the Yenisei, Ob, Lena and Mackenzie catchments. The top panels compare runoff from ERA5 and ERA5-Land with river discharge from $GLOFAS_{E5}$, $GLOFAS_{E5L}$ and $GLOFAS_{E5new}$ as well as with observed discharge values.

Observations show a distinct runoff peak in June due to snow melt and river ice breakup and weak runoff through winter. The spring flood season of Eurasian rivers depends on the basin size and usually ends by the end of June at small size rivers and by the end of July or beginning of August at large rivers like Ob, Yenisei and Lena (Yang et al., 2007). While smaller rivers usually exhibit a low-flow season in the summer to fall period, discharge from larger rivers mostly shows a slower decrease, also because of summer rainfalls providing additional discharge water. Especially at rivers of Eastern Siberia (e.g. Lena) intense



rainfalls in the summer-fall season may occasionally cause rainfall floods (Shiklomanov et al., 2021a). In late winter, with the maximum of river freeze-up, discharge reaches its minimum.

Runoff data from ERA5 and consequently also from GLOFAS_{ERA5} clearly underestimate the summer peaks recorded by gauges
 245 and reach only about 25 to 50% of the observed peak discharge values. In the low flow season the reanalyses slightly exceed
 observed discharge, however this does not alter the annual means considerably resulting in a clear underestimation of discharge
 by ERA5 and GLOFAS_{ERA5} also in annual terms. The difference between ERA5 runoff and GLOFAS_{ERA5} discharge is expected to
 be caused by two sink terms, a groundwater loss component, calibrated in LISFLOOD, that removes water that is lost to deep
 groundwater systems, and the open water evaporation component, which also removes water through evaporation over water
 250 surfaces in LISFLOOD. The negative contribution can reach up to 20-40% of the average flow by any of the two terms in cer-
 tain regions (described in the LISFLOOD model documentation, Burek et al., 2013). Better accordance in terms of peak height
 and annual means are achieved by runoff from ERA5-Land - except for the Lena basin - while the sink terms in GloFAS again
 causes an underestimation by GLOFAS_{ERA5L}. At the Ob basin the runoff peak in ERA5-Land, and also ERA5, occurs in May,
 thus a month earlier than the observed peak. GloFAS discharge is not in phase with ERA5 runoff, but reaches its peak in June,
 255 presumably due to the delay by the river routing component. In contrast to GLOFAS_{ERA5} and GLOFAS_{ERA5L}, GLOFAS_{ERA5new}
 reaches values similar to observations and agrees best with the observed values in terms of annual means, peak heights and
 seasonality. Cucchi et al. (2020) run the hydrological model WaterGAP (Müller Schmied et al., 2016) with precipitation, tem-
 perature and radiation forcing from ERA5 and from the bias corrected version of ERA5 WFDE5 (Cucchi et al., 2020). Model
 runs with ERA5 forcing show similar river discharge seasonalities at the Lena catchment as GloFAS_{ERA5new}.
 260 The middle panels of Fig. 2 show snowfall and snow melt as well as the atmospheric components net precipitation (P-E) and
 divergence of moisture flux (VIWVD). Seasonal cycles of P-E and VIWVD minus storage change agree quite well in terms
 of peak heights and timing, with low moisture inflow and low net precipitation in summer and higher values in autumn and
 winter. Annual values of P-E and VIWVD- ΔS differ by 2-16% depending on the catchment. Seasonal cycles of ERA5 snow
 melt show that there is a lag of one month between the peak in snow melt and river discharge. This can partly be explained by
 265 the time it takes for the water to reach the river mouth and by water resource management effects, but mostly is caused due to
 delayed river ice break up in the lower parts of the basins. For example at the upper part of Ob river ice breaks up around April
 to May, while the lower part breaks up between May and June (Yang et al., 2004b). While human impacts through water with-
 drawals for agricultural use are rather limited compared to rivers in lower latitudes, water resource management via dams and
 reservoirs can significantly alter the seasonal discharge cycle of the larger Arctic rivers (Shiklomanov et al., 2021a). Especially
 270 the Ob and Yenisei basins, but also Lena and Kolyma, are affected by multiple reservoirs using water for hydroelectric power
 generation and delaying discharge from high-flow periods to the low-flow season (Lammers et al., 2001; Ye et al., 2003; Yang
 et al., 2004b, a; Shiklomanov et al., 2021a).

The lower panels of Fig. 2 show land storage change from ERA5 and from GRACE. Additionally ERA5 storage is separated
 into its components of soil water change ΔS_{SWVL} and snow depth change ΔS_{SD} . In autumn and winter land water is accu-
 275 mulated through a rise in snow depth, in summer storage is lost through snow melt and consequent runoff, and to a smaller
 part also through evaporation. GRACE shows a significant annual loss of water storage over the past decades, while storage

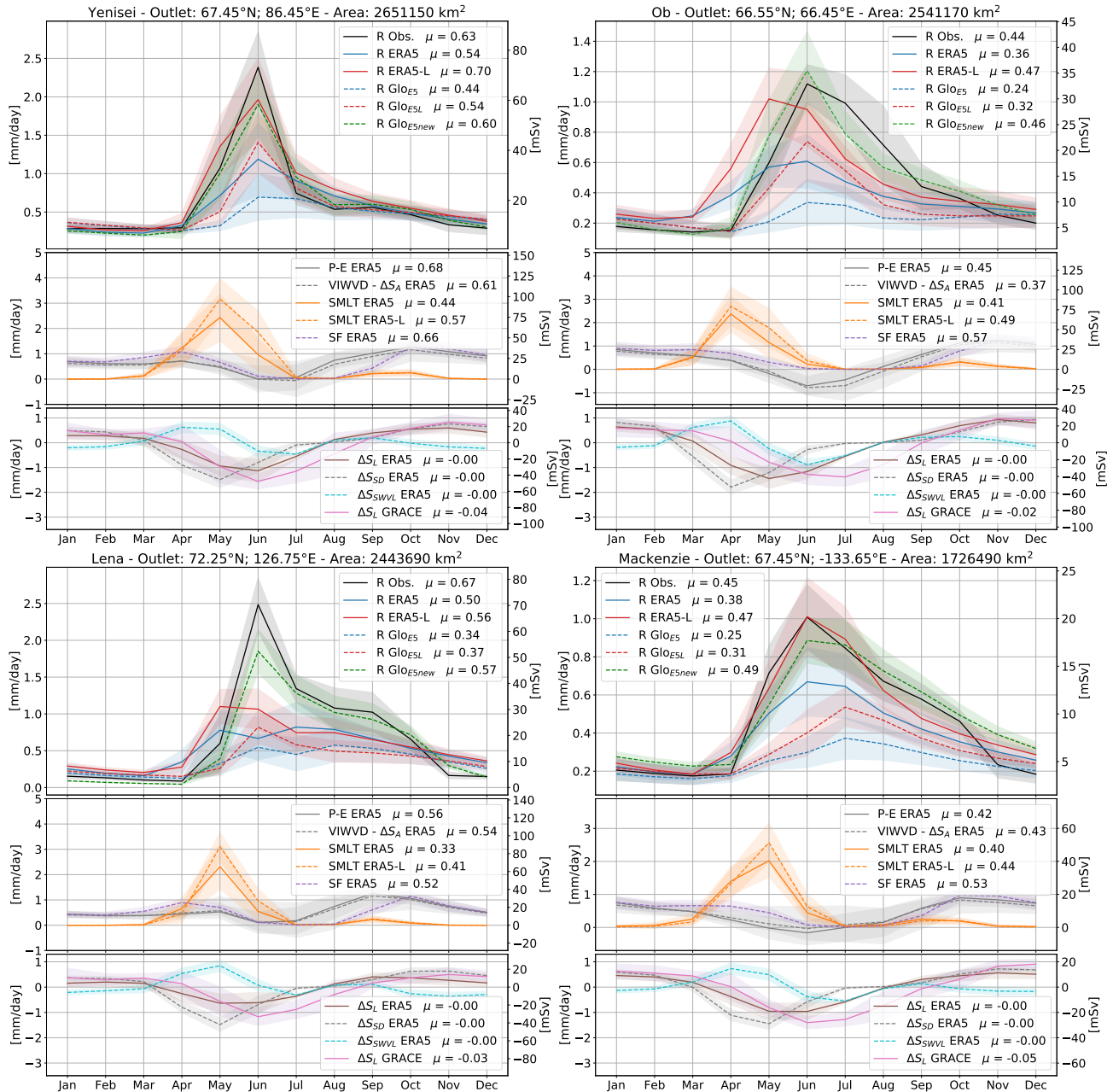


Figure 2. Yenisei, Ob, Lena and Mackenzie seasonal cycles of runoff and river discharge values in the top panel, terrestrial net precipitation (P-E), divergence of moisture flux (VIWVD), snow melt (SMLT) and snowfall (SF) in the middle and land storage change (ΔS_L), snow depth change (ΔS_{SD}) and soil water change (ΔS_{SWVL}) in the bottom panel. Shading denotes variance as the \pm one standard deviation. Furthermore locations of gauge observations and catchment areas are given. Long term annual means in mm/day are given in the legend.



change in ERA5 exhibits only a slight annual decline. In terms of seasonality GRACE features the largest storage changes in June and July, while ΔS from ERA5 tends to peak 1-2 month earlier. Again this could be caused by delayed river ice breakup and backwater that is considered in GRACE, but not in ERA5.

280 4.1.2 Trends

Figure 3 shows annual river discharge values for Yenisei, Ob, Lena and Mackenzie rivers. The corresponding temporal means, standard deviations, decadal trends, normalised RMSEs and correlation coefficients are given in Table 3. Observations indicate a slight increase in river discharge over the past decades. The largest changes occurred at the Ob and Lena basins with a rise of about 4% per decade. Yenisei shows no significant long term trend, as, after a rise until the early 2000s, discharge values seem to have slightly decreased over the past decade. In contrast to observations, runoff from ERA5 and discharge from GLOFAS_{E5} show distinct negative trends of 11 to 16% per decade. These strong decreases are a result of changing biases in the snow assimilation in ERA5 and are discussed further in Sect. 3.3.3. The effect of the sink terms removing water in GLOFAS_{E5}, and hence producing a negative shift, results in a strong underestimation towards the end of the time series with values reaching only slightly more than 50 % of the observed values. Runoff from ERA5-Land shows clearly better results, with long term means generally deviating only about 5% from observations (10% for Lena) and quite low normalised RMSE values, however the sink terms in GLOFAS_{E5L} again leads to a clear underestimation of discharge. Calculating runoff indirectly through net precipitation minus land storage change and VIWVD minus atmospheric storage change (Eq. (3)) yield results mostly within 10% of the observed discharge values.

In contrast to GLOFAS_{E5} and GLOFAS_{E5L}, river discharge from GLOFAS_{E5new} represents observed discharge from gauges very well and concerning the normalised RMSEs it provides the best results when compared to observations. Both discharge from GLOFAS_{E5new} and estimation through P-E also feature slightly negative trends, only calculation through VIWVD exhibits no, or even slightly positive trends.

4.2 Pan-Arctic approach

To get a complete estimate of the total amount of freshwater entering the Arctic Ocean via land, which is needed for subsequent budget calculation, river discharge is calculated over the whole Pan-Arctic area. First we look into the Pan-Arctic seasonal cycle and afterwards at annual means and long-term trends over the whole Arctic drainage area.

Pan-Arctic freshwater discharge estimates vary significantly between different studies. This comes to a big part from different definitions of the geographic area (Prowse and Flegg, 2000; Shiklomanov and Shiklomanov, 2003) as there is no strict boundary to the south that defines the Arctic and past studies disagree whether to include e.g. Greenland and the Hudson Bay or not. Another reason for discrepancies between different studies comes due to different approaches of discharge evaluation (Shiklomanov et al., 2021a) - some approaches are based on model simulations and others are derived from river discharge measurements. Using only in-situ measurements poses several challenges - especially the handling of the large unmonitored areas, that account for about 30–40% of the total drainage area (Shiklomanov et al., 2002). Most studies adopt the method of hydrological analogy (Shiklomanov and Shiklomanov, 2003) and calculate total discharge by expanding gauged runoff over the

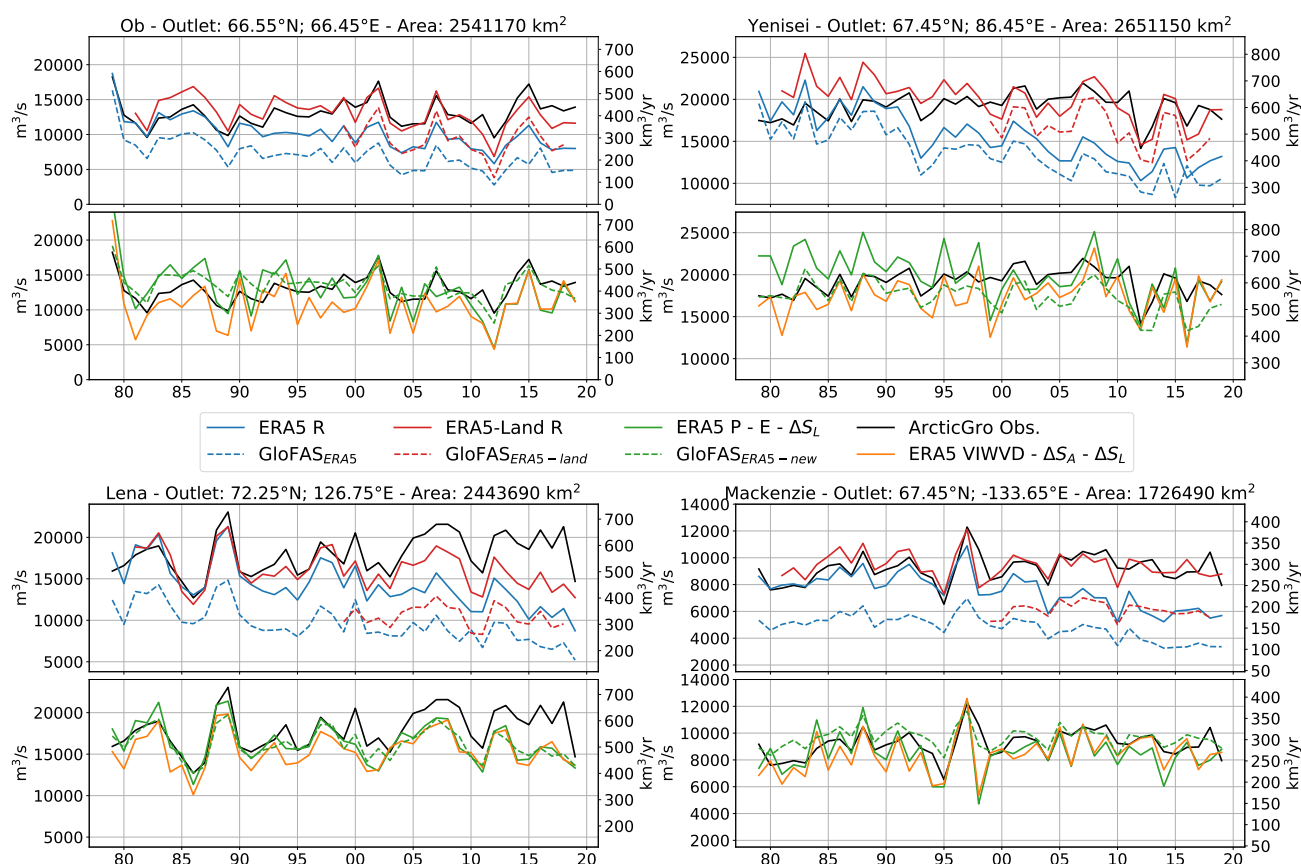


Figure 3. Annual means of observed river discharge, ERA5 runoff, ERA5 net precipitation, ERA5 VIWVD and ERA5-Land runoff as well as GLOFAS_{E5}, GLOFAS_{E5L} and GLOFAS_{E5new}.

unmonitored area. However hydrological, climatic and land cover conditions between gauged and ungauged areas can differ quite a lot, resulting in inaccurate estimates (Shiklomanov et al., 2021a). Hence we compare two different estimation methods for observed discharge that are displayed in Fig. 4. For that purpose we consider gauging records from the 24 largest Arctic rivers, that amount for roughly 70% of the total drainage area. Pan-Arctic river discharge for the ungauged areas for each month was estimated as follows: Firstly following e.g. Serreze et al. (2006) and applying the method of hydrological analogy for each calendar month by transforming discharge into runoff and, with the assumption that runoff from the ungauged area is the same as runoff from the gauged area, multiplication with the whole drainage area to transform back to river discharge - here denoted as *ae* (area estimate). However taking runoff from reanalysis (e.g. ERA5-Land) for the same 24 drainage areas and expanding it over the whole area yields different results, with smaller summer peaks, than taking runoff directly from the total drainage area. Hence monthly correction factors are calculated from our most reliable products ERA5-Land and GLOFAS_{ERA5new} and by multiplication with those, a more accurate estimate of observed Pan-Arctic river discharge should be possible - denoted as



		Obs.	E5 R	E5 P-E	E5 VIWVD	E5-L R	Glo _{E5}	Glo _{E5L}	Glo _{E5new}
Yenisei	μ	19.2	15.5	19.8	17.4	20.0	13.4	16.4*	17.3
	σ	1.5	3.0	2.9	2.4	2.4	2.9	2.3*	1.9
	Trend	0.00±0.01 ⁿ	-0.14±0.02	-0.06±0.02	0.01±0.02 ⁿ	-0.06±0.01	-0.16±0.02	-0.10±0.05* ⁿ	-0.05±0.01
	NRMSE		0.25	0.15	0.14	0.11	0.34	0.17*	0.10
	r		0.24	0.37	0.49	0.53	0.21	0.81*	0.64
Ob	μ	12.9	9.9	12.6	10.6	12.5	6.8	9.3*	13.4
	σ	1.8	1.8	2.9	3.3	2.1	1.7	2.3*	1.7
	Trend	0.04±0.02	-0.11±0.02	-0.08±0.03	0.02±0.04 ⁿ	-0.06±0.02	-0.15±0.03	-0.10±0.99* ⁿ	-0.03±0.02 ⁿ
	NRMSE		0.28	0.21	0.25	0.13	0.45	0.32*	0.11
	r		0.39	0.40	0.58	0.66	0.32	0.89*	0.68
Lena	μ	18.0	14.2	16.4	15.6	16.1	9.6	10.5*	16.2
	σ	2.5	2.8	2.4	2.2	2.3	2.2	1.3*	1.8
	Trend	0.04±0.02	-0.12±0.02	-0.03±0.02 ⁿ	0.00±0.02	-0.05±0.02	-0.15±0.02	-0.03±0.05* ⁿ	-0.02±0.02 ⁿ
	NRMSE		0.27	0.14	0.17	0.15	0.50	0.45*	0.14
	r		0.33	0.65	0.74	0.66	0.27	0.73*	0.76
Mackenzie	μ	9.2	7.5	8.6	8.5	9.4	4.8	6.1*	9.7
	σ	1.0	1.3	1.5	1.5	1.0	0.9	0.5*	0.7
	Trend	0.02±0.02	-0.11±0.02	-0.01±0.03 ⁿ	0.03±0.02	-0.02±0.01 ⁿ	-0.13±0.02	-0.00±0.04* ⁿ	-0.01±0.01 ⁿ
	NRMSE		0.24	0.17	0.16	0.09	0.49	0.36*	0.09
	r		0.29	0.49	0.55	0.65	0.35	0.65*	0.73

Table 3. Mean values and standard deviations of river discharge in units $10^3 \text{ m}^3 \text{ s}^{-1}$ as well as relative decadal trends and their standard errors (see Eq. (10)), calculated over the period 1981–2019 (1999–2018 for GLO_{E5L}, indicated by the superscript *). Trends that are not significant are marked with an superscript ⁿ. Additionally normalised RMSE values and correlation coefficients for annual means are given. Bold values identify the best correlation coefficients and NRMSE values.

Ee (ERA estimate) and *Ge* (GloFAS estimate), leading to more plausible high flow peaks. Following the results of the previous sections we trust most in GloFAS_{ERA5new} as it featured the smallest NRMSEs and highest correlation coefficients and hence use the GloFAS estimate *Ge* for the following Pan-Arctic analysis.

4.2.1 Seasonal Cycle

Figure 5 shows the mean seasonal cycles of the various Pan-Arctic river discharge estimates for the period 1981 to 2019 (1999 to 2018 for GLOFAS_{E5L}).

Results are similar as for the individual catchments, with large amounts of water entering the Arctic Ocean in June. Observations (*Ge*) show summer peaks of about $4.8 \times 10^5 \text{ m}^3 \text{ s}^{-1}$ and annual means of $4031 \text{ km}^3 \text{ yr}^{-1}$. ERA5 runoff and discharge from GLOFAS_{E5} and GLOFAS_{E5L} clearly underestimate the summer peaks and annual means, however they slightly overestimate runoff in the low flow winter months. ERA5-Land runoff performs quite well in terms of annual means, but the June

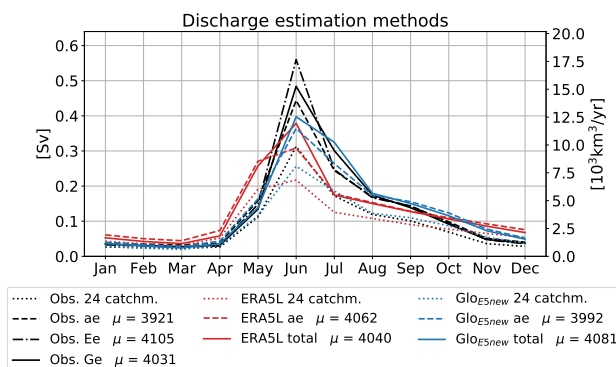


Figure 4. Pan-Arctic discharge estimates using hydrological analogy (ae) and monthly correction factors from ERA5-Land (Ee) and GloFAS_{ERA5new} (Ge). Long term annual means in $\text{km}^3\text{yr}^{-1}$ are given in the legend.

maximum is still about 25-35% too low. *GLOFAS_{ERA5new}* performs best in terms of seasonality, but still slightly underestimates the summer peak. Indirectly calculated discharge through Eq. (3) does not reach those distinct high June peaks. Annual values from calculation of P-E reach values similar to ERA5-Land runoff, while the VIWVD estimate is about 5% too low. The runoff climatology Bt06 exhibits a similar seasonal cycle as our observation based estimate, however it is roughly 9% lower than our estimate.

The bottom panel of Fig. 5 shows land storage change components. The time lag between the observed storage change through GRACE and ERA5 is clearly evident again. And just as for the four major basins, also the Pan-Arctic area shows a major decline of land water storage over the past decades, reaching -132 km^3 per year for our area of interest, while land storage from ERA5 shows considerably smaller declines of -34 km^3 per year. The largest changes for GRACE water storage occur over the Canadian Arctic Archipelago and the mountainous areas of Mackenzie and Yukon basin, suggesting a tight linkage to glacial melting. Additionally Greenland features a storage decline of -134 km^3 per year, accounting for roughly 50% of the total storage change. Wouters et al. (2019) use monthly GRACE Stokes coefficients to examine global glacier mass losses for regions defined in the Randolph Glacier Inventory (RGI), excluding Greenland. Building the sum over all regions that roughly resemble our area of interest, yields a glacial mass change of about -100 Gt (gigatons), or roughly -109 km^3 per year. Hence the sum of glacial mass change and storage change from ERA5 resembles the land storage change from GRACE within 10%.

4.2.2 Trends

Long term means of annual discharge over the coinciding period of all datasets 1999-2018 (1981-2018 in brackets) and the distribution to Eurasia, North America and the CAA can be seen on the left panel of Fig. 6 and long term annual means, relative trends, correlation coefficients and NRMSE values are provided in Table 4. About 75 % of Pan-Arctic runoff come from the Eurasian watersheds, followed by North America with about 17 % and the smallest contribution coming from the CAA. Total Pan-Arctic runoff values, calculated over 1999-2018, range between $2625 \text{ km}^3\text{yr}^{-1}$ for *GLOFAS_{E5}* and $3952 \text{ km}^3\text{yr}^{-1}$ for

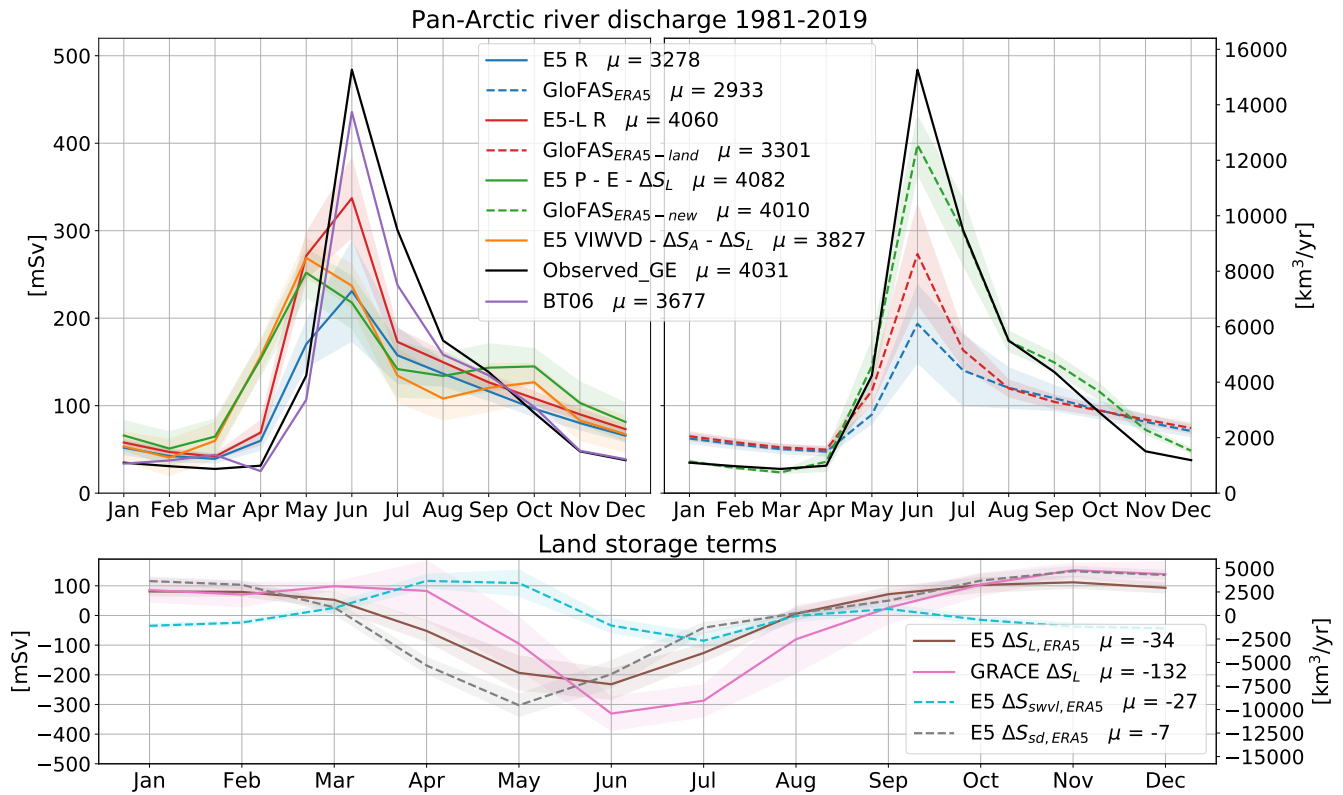


Figure 5. Top panel: Seasonal Cycles of Pan-Arctic runoff for ERA5, ERA5-Land and various GloFAS versions, as well as indirectly calculated runoff through P-E and VIWVD. Additionally an estimate of observed river discharge is given. Seasonal Cycles are calculated over the period 1981-2018 (1999-2018 for *GLOFAS_{ERA5}*). Bottom panel: storage change components from GRACE and ERA5. Shading denotes variance as the \pm one standard deviation. Additionally long term annual means in $\text{km}^3\text{yr}^{-1}$ are given and appropriately the right axis is scaled in $\text{km}^3\text{yr}^{-1}$.

indirectly calculated runoff through $\text{P-E} - \Delta S_L$ from ERA5. Our observation based estimates are considerably higher, reaching $4117 \text{ km}^3\text{yr}^{-1}$. Using observations of river discharge only and omitting the CAA, Serreze et al. (2006) obtain values of $3200 \text{ km}^3\text{yr}^{-1}$. Including the CAA, but excluding Yukon river, Haine et al. (2015) combine runoff from ERA-Interim with river discharge observations and report annual values of $3900 \pm 390 \text{ km}^3\text{yr}^{-1}$ for the period 1980-2000 and $4200 \pm 420 \text{ km}^3\text{yr}^{-1}$ for the period of 2000-2010. Our observation based estimates, including Yukon river, reach values of $3971 \pm 31 \text{ km}^3\text{yr}^{-1}$ for 1980-2000 and $4155 \pm 42 \text{ km}^3\text{yr}^{-1}$ for 2000-2010. Subtracting the contribution of Yukon River (about $200 \text{ km}^3\text{yr}^{-1}$), our estimates are about 5% lower than those from Haine et al. (2015). Excluding Yukon river for ERA5-Land runoff and indirectly calculated runoff through ERA5 P-E, both estimates are quite close to the estimates made by Haine et al. (2015) for the period of 1981-2000, however they are substantially too low in the 2000-2010 period. ERA5 and *GLOFAS_{ERA5}* feature far too low values for both periods.

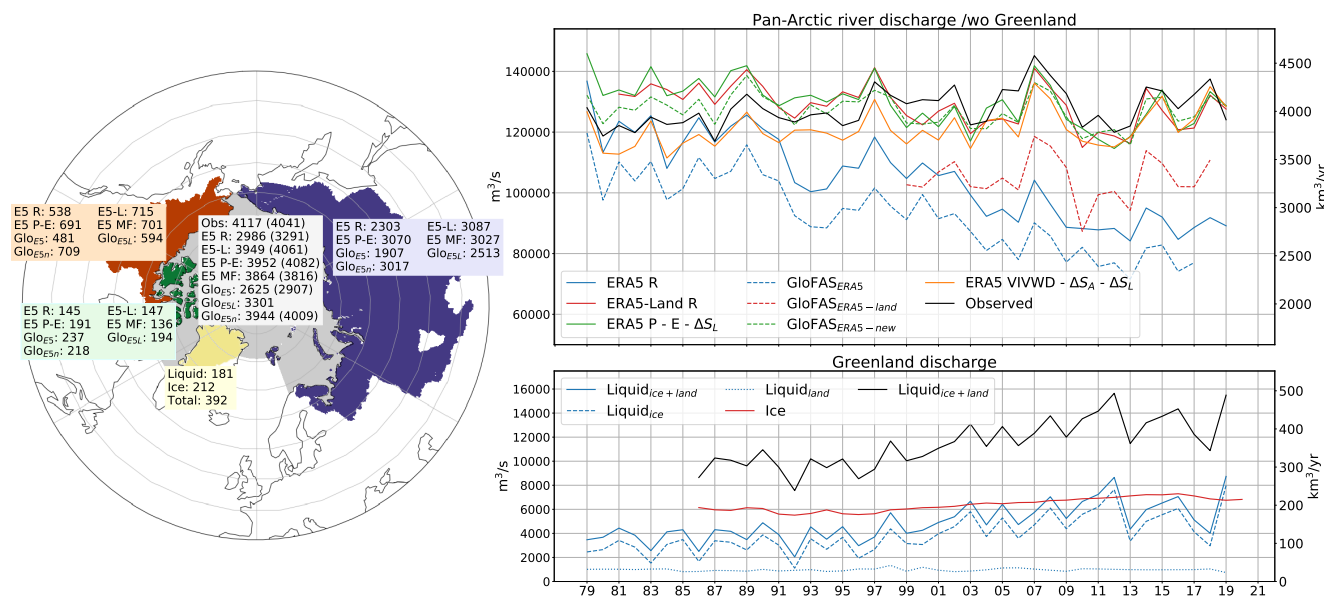


Figure 6. The left panel shows the drainage area and mean annual discharge values in $\text{km}^3 \text{yr}^{-1}$ for the whole area, as well as for Eurasia, North America and the CAA individually. Values are calculated over the coinciding period of all datasets 1999-2018 (values in brackets are calculated over the period 1981-2018). On the right hand side annual means of our observation based estimate, ERA5, ERA5-Land, $GLOFAS_{ERA5}$, $GLOFAS_{ERA5L}$ and $GLOFAS_{ERA5new}$ runoff, as well as indirectly calculated runoff through P-E and VIWVD are shown. Additionally Greenlandic liquid (from land and ice) and solid discharges are displayed in the bottom right panel.

For the whole Pan-Arctic region, both ERA5 and $GLOFAS_{ERA5}$ feature runoff decreases of 10-11 % per decade, while ERA5-Land and indirectly calculated runoff through ERA5 P-E exhibit negative trends of about 2-3 % per decade. In contrast the runoff estimate through ERA5's VIWVD features a slight increase of 2%, identically to the trend present in our observation based estimate. $GLOFAS_{ERA5new}$ comes closest to gauge observations concerning annual means and NRMSE values, while indirectly calculated runoff through VIWVD features the same trend as observations and also the highest correlation concerning annual means. However the VIWVD estimate generally yields a roughly 5% lower discharge.

The strong decreases in ERA5 runoff are reinforced through two discontinuities in the dataset, one around 1992 and the second one around 2004, that lead to a significant drop and hence a clear underestimation of runoff over the past decades. As $GLOFAS_{ERA5}$ takes ERA5 as input, it also adopts those discontinuities. In contrast ERA5-Land does not exhibit such breaks, suggesting that the error may come from the data assimilation system in ERA5. While the discontinuity around 2004 was traced back to the introduction of the IMS (Interactive Multisensor Snow and Ice Mapping System) snow product into the assimilation system (Zsótér et al., 2020), the reason for the break around 1992 could not be identified yet and is discussed further in the following section.

Contrary to discharge estimates from reanalyses, observations show opposing trends and a rise in river discharge over the past decades. In NOAA's Arctic Report Card 2018 Holmes et al. (2018) stated that river discharge from Eurasians largest rivers (Ob,



	Observed G_e	E5 R	E5 (P-E)	E5 VIWVD	E5-L R	G_{loE5}	G_{loE5L}	$G_{loE5_{new}}$
μ	127.8	104.4	129.4	121.0	128.8	92.2	104.7*	127.1
σ	6.1	13.0	7.1	5.9	6.6	11.3	7.0*	5.1
Pan-Arc. Trend	0.02±0.01	-0.10±0.01	-0.03±0.01	0.02±0.01	-0.02±0.01	-0.11±0.01	0.00±0.03* ⁿ	-0.01±0.01
NRMSE	-	0.22	0.06	0.06	0.06	0.30	0.20*	0.04
r	-	-0.14	0.29	0.78	0.38	-0.15	0.85*	0.53

Table 4. Mean values and standard deviations of Pan-Arctic river discharge in $m^3 s^{-1} * 10^{-3}$ as well as decadal relative trends, calculated over the period 1981-2019 (1999-2018 for $G_{loFAS_{E5L}}$, indicated by *). Trends that are not significant are indicated by the superscript ⁿ. Bold values identify the best estimates in respect of long term means, trends, correlation coefficients, NRMSE values

Yenisei, Lena, Kolyma, Pechora, Severnaya Dvina) has increased by 3.3 ± 1.6 % per decade since 1976 and by 2.0 ± 1.8 % per decade for the largest North American rivers (Mackenzie and Yukon). Our observation based estimates show an increase of 2 ± 1 % for the whole Pan-Arctic region.

380 While runoff from Greenland was left out so far, contributions of land and ice discharge are not negligible. Additionally to Eurasian and North American discharge, Fig. 6 shows discharge from the Greenlandic ice and land areas north of our boundaries in Davis and Fram Strait. Greenlandic liquid discharge from land and ice is taken from Mankoff et al. (2020a), who calculate discharge from daily runoff estimates of the Modèle Atmosphérique Régional (MAR) and the Regional Atmospheric Climate Model (RACMO). We estimate the optimal discharge by taking the mean of MAR and RACMO discharge. Solid

385 ice discharge through calving of marine-terminating glaciers and melt-water from ice–ocean boundary melting at submarine glaciers is taken from Mankoff et al. (2020b) over the regions central west (CW), north west (NW) and north (NO) (see Mankoff et al. (2020b), their Fig. 1) to roughly account for our region of interest. With a total contribution of $392 \text{ km}^3 \text{ yr}^{-1}$ over the period 1999-2018, discharge from Greenland accounts for roughly 10% of total Pan-Arctic discharge. Most of the freshwater supplied to the Arctic Ocean comes from solid ice discharge, followed by liquid discharge from glaciers and only a

390 small contribution from land runoff. Liquid discharge from land and ice show pronounced seasonalities with peaks in June and July, while solid ice discharge stays continuous throughout the year. The bottom figure of the right panel of Fig. 6 shows annual means for the individual Greenlandic discharge components and Table 5 displays mean values and trends. Liquid ice discharge exhibits a vast positive trend of 26% per decade and also solid ice discharge shows a clear rise of about 8% per decade. It can be expected that Greenland discharge will further increase in the future (Muntjewerf et al., 2020; Church et al., 2013; Vaughan

395 et al., 2013, e.g.)

Adding Greenlandic discharge to our G_e observation estimate, yields a total discharge of $4423 \text{ km}^3 \text{ yr}^{-1}$. A recent assessment (Shiklomanov and Lammers, 2013; Shiklomanov et al., 2021a) estimates the total discharge to the Arctic Ocean at approximately $4300 \text{ km}^3 \text{ yr}^{-1}$ for the period 1936-2006. Differences may stem from slight variations in discharge area - Shiklomanov and Lammers (2013) use an area of approximately $19 \times 10^6 \text{ km}^2$, while our area together with the considered Greenlandic area

400 ($0.95 \times 10^6 \text{ km}^2$) sums up to $19.15 \times 10^6 \text{ km}^2$ - and from different calculation periods. Discharge was very likely smaller in the mid-20th century.



		Liquid _{Ice}	Liquid _{Land}	Liquid _{total}	Ice	Total
	μ	4.2	1.0	5.2	6.7	11.9
Greenland	σ	1.5	0.1	1.5	0.5	2.0
	Trend	0.25±0.05	0.01±0.02 ⁿ	0.20±0.04	0.08±0.01	0.13±0.02

Table 5. Mean values, standard deviations and trends for liquid discharge from ice and land as well as solid ice discharge for Greenland north of Davis and Fram Strait over the period 1987-2019. Trends that are not significant are indicated by the superscript ⁿ.

4.2.3 ERA5 Runoff discontinuities

As mentioned above, a possible reason for the negative shifts in ERA5 runoff lies in the data assimilation system and its removal of soil moisture. Zsótér et al. (2020) assess the *GLOFAS_{E5}* river discharge reanalysis as well as ERA5 and ERA5-Land runoff and compare them with available river discharge observations. Similarly to our diagnostics for Arctic rivers, Zsótér et al. (2020) find river discharge decreases in GloFAS-ERA5 for several rivers of the world, that are not supported by observations. They find trends in tropical and subtropical areas being driven by changes in precipitation, while changes in snow melt have a very strong influence on river discharge trends in the northern latitudes. Thus the runoff decreases in the northern latitudes are likely linked to snow assimilation and other processes related to snow melt. Figure 7 shows ERA5 snowfall, snow melt and the sum of snow melt and snow evaporation as well as the corresponding parameters from ERA5-Land for the whole Pan-Arctic region, Eurasia, North America and the CAA. ERA5 shows great differences between snow gain and snow loss and Pan-Arctic snow melt exhibits the same discontinuities around 1992 and 2004 as runoff from ERA5. In contrast snow gain and snow loss are in balance for ERA5-Land. Zsótér et al. (2020) find similar results and attribute the differences to the land data assimilation, that has impacts on snow and soil moisture. The discontinuity in 2004 was traced back to a change in operational snow analysis, through the introduction of the 24-km Interactive Multi-Sensor Snow and Ice Mapping System (IMS) snow cover information to the snow assimilation system in 2004 (Zsótér et al., 2020). While the discontinuity around 2004 is present in the Eurasian watershed as well as in North America, the discontinuity around 1992 is only present in Eurasia. Figure 8 shows the spatial distribution of the snow melt discontinuities, calculated as differences between snow melt climatologies from 1979-1991 and 1992-2003, as well as 1992-2003 and 2004-2019. The latter difference exhibits spurious signals at the coastal mountain range of Alaska, the Rocky Mountains and at mountainous regions in Siberia, the prior discontinuity spots rather random signals only over Eurasia.

4.3 Comparison with oceanic fluxes

In this section we look at the oceanic volume budget as described in Eq. (5). For this purpose we calculate volumetric fluxes through the oceanic gateways - BSO, Fram, Davis, Bering, Hecla and Fury Strait - using ocean reanalyses and compare them to our various freshwater input estimates. Oceanic storage change is derived from ocean reanalyses and from GRACE. To close the budget and get rid of any residuals we use a variational approach (see Sect. 2.1.2).

Figure 9 shows seasonal cycles of lateral net volume fluxes from ORAS5 and from the GREP ensemble for 1993-2018

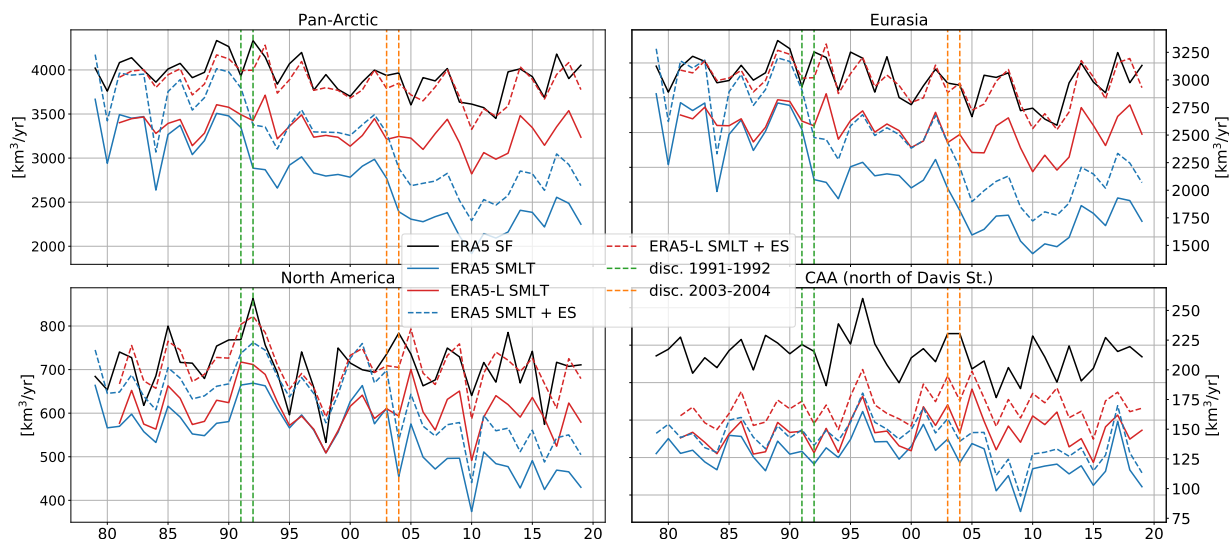


Figure 7. Annual means of ERA5 snowfall (SF, black), ERA5 snow melt (SMLT, blue) and the sum of ERA5 snow melt and snow evaporation (SMLT+ES, blue dashed) as well as ERA5-Land snow melt (red) and ERA5-Land SMLT+ES (red, dashed) for the whole Pan-Arctic region (top left), Eurasia (top right), North America (bottom left) and the CAA (bottom right).

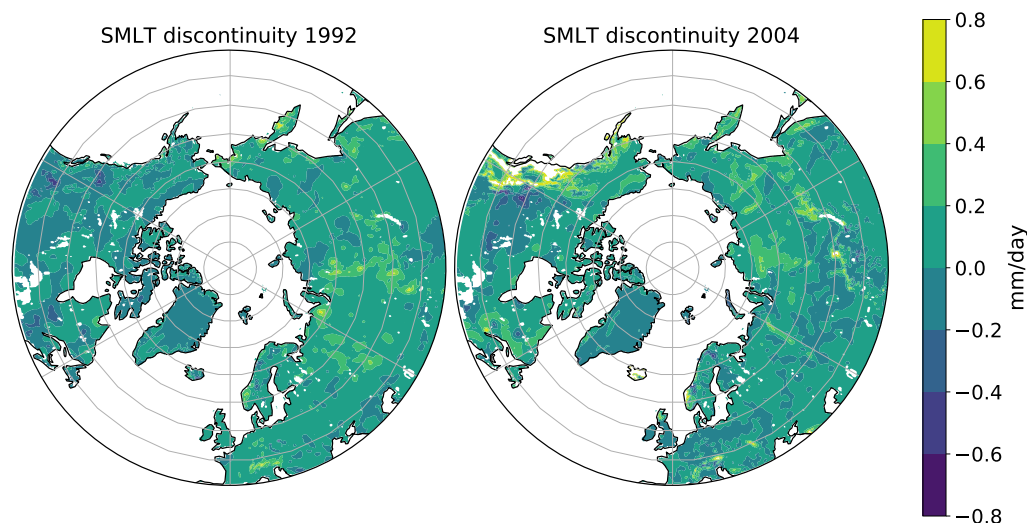


Figure 8. ERA5 snow melt discontinuities, calculated by subtracting the long term SMLT means from 1979-1991 and 1992-2003 and respectively 1992-2003 and 2004-2019.



The GREP ensemble shows an annual mean lateral transport out of the Arctic region of $0.207 \pm 0.048 Sv$, with ORAS5 being on the higher end of the ensemble with an annual outflow of $0.238 Sv$ and a summer peak of $0.578 Sv$. The seasonal cycles of the oceanic volume transports resemble the seasonal cycles of the freshwater input and peak in June, as the ocean reacts almost instantly to surface freshwater input and generates barotropic waves that lead to mass-adjustment in about a week (Bacon et al., 2015).

In addition to the oceanic transports the left panel of Fig. 9 shows the forcing terms involved in the generation of the ORAS5 fluxes:

$$F_{ORAS5} = (P_O - ET_O)_{ERAInterim} + R_{BT06} - \frac{\partial S_{ORAS5}}{\partial t} + damp_{ORAS5} + FW_{adj} \quad (11)$$

Next to net precipitation from ERA-Interim and river discharge from the BT06 climatology, we need to add a surface salinity damping term ($damp_{ORAS5}$), which represents an additional non-physical surface freshwater forcing in the ocean reanalyses (last term of Eq. 6.1 in Madec and Team, 2019). Additionally a freshwater adjustment term (FW_{adj}) should be added due to the assimilation of global mean-sea-level changes. This term is not separately saved as model output in ORAS5 and hence is not considered in our analysis (therefore it is marked grey in Eq. (11)). However we deem the freshwater adjustment term as a rather small source of error and nevertheless would expect good closure concerning Eq. (11). With an annual deviation of $0.005 Sv$ and an NRMSE value of 0.19 the accordance between ORAS5 and its freshwater input estimate is fairly reasonable. Next to the FW adjustment term, another cause for the discrepancies between forcing and computed volume fluxes is that the ocean reanalyses compute their own turbulent air-sea fluxes (REF) and do not use that from ERA-Interim. However, given the generally low values of sensible and latent heat fluxes in the Arctic we consider this a moderate source of error.

The right panel of Fig. 9 shows various estimates of freshwater input minus oceanic storage change. To estimate the atmospheric freshwater input over the ocean we take net precipitation from ERA5 as well as VIWVD from ERA5, ERA-Interim, the Japanese 55-year Reanalysis JRA55 and JRA55 Conventional. For the reanalysis-based estimates we use the river discharge reanalyses we have most confidence in, namely $GloFAS_{ERA5new}$, ERA5-Land runoff and the indirect estimates through P-ET and VIWVD minus storage change. Land storage is only derived from GRACE, while oceanic storage is taken from GRACE and from ORAS5. GRACE and ORAS5 show quite similar seasonal cycles, with mass being accumulated by the ocean in summer and released in winter, most likely caused by the seasonal variation of wind stress curl and seasonal changes in Ekman pumping (Bacon et al., 2015), however they disagree in terms of annual trends, as ORAS5 indicates a slight increase of mass over the past decades, while GRACE points to a decrease. The observation based estimates are calculated using river discharge from our observation estimates (ae , Ee , Ge) and oceanic storage change from GRACE. The atmospheric components VIWVD and atmospheric storage change over the ocean are still taken from reanalyses (ERA5, ERA-Interim, JRA55, JRA55c), as we lack the corresponding observations. Greenlandic discharge is taken from Mankoff et al. (2020a, b). Calculating the mean over our reanalyses and observation based estimates we obtain annual values of $0.208 \pm 0.004 Sv$ and $0.210 \pm 0.003 Sv$, respectively. Thus, compared to the GREP volume flux we get an imbalance of $0.001 - 0.003 Sv$, accounting for roughly 1% of the fluxes. Figure 10 shows the budget residuals, calculated as GREP ensemble average volume transport minus the various freshwater estimates. From January to April residuals are generally positive, as the freshwater input from atmosphere and land is low due

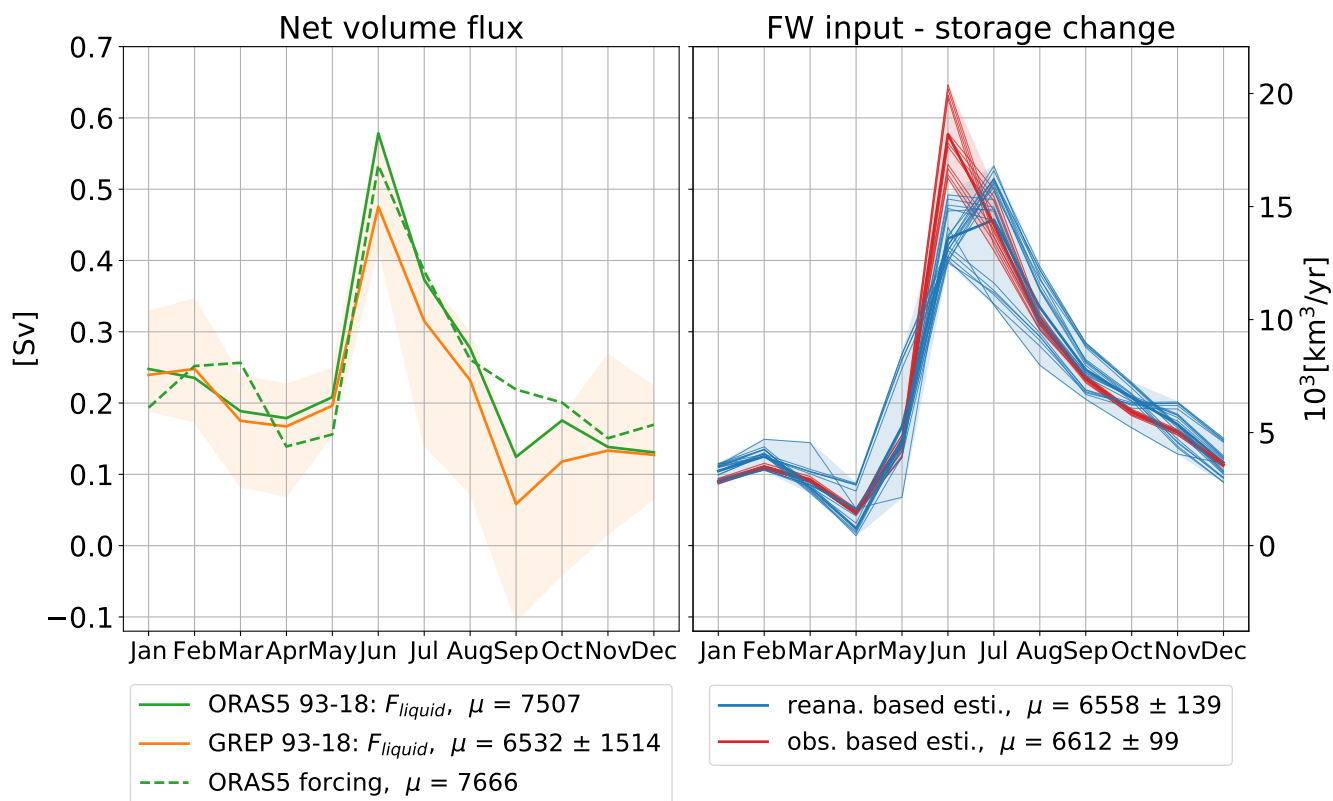


Figure 9. Left panel: Mean seasonal cycles of lateral net volume fluxes (defined positive out of the Arctic) from ORAS5 (1993–2018) and GREP (1993–2018), as well as the sum of forcing terms generating ORAS5. Right panel: Observation and reanalyses based estimates of freshwater input minus oceanic storage.

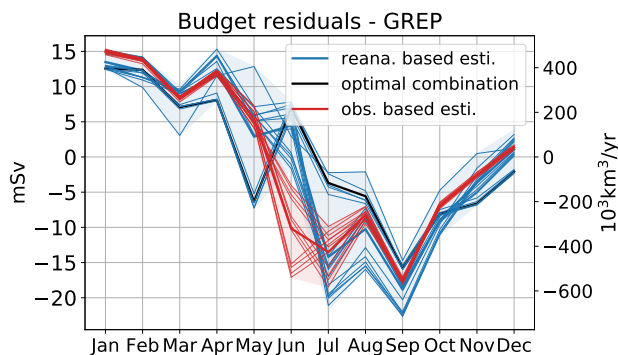


Figure 10. Mean annual cycles of different realisations of the budget residuals (reanalyses based: blue lines, observation based: red lines) and the optimal combination yielding the smallest NRMSE value (black line).



	GREP	ORAS5	ORAS5 forcing	ARCGATE	reana. based estimate	obs. based estimate
μ	0.207/0.203*	0.238/0.227*	0.243	0.151*	0.208	0.210
NRMSE	-	0.218 ¹	0.327 ¹ /0.195 ²	0.813 ^{1*} /0.857 ^{2*}	0.457 ¹ /0.376 ²	0.491 ¹ /0.347 ²
r	-	0.971 ¹	0.856 ¹ /0.928 ²	-0.313 ^{1*} /-0.522 ^{2*}	0.605 ¹ /0.739 ²	0.732 ¹ /0.855 ²

Table 6. 1993-2018 (*2004-2010) long term means, NRMSEs and correlation coefficients of oceanic volume transports from GREP, ORAS5 and ARCGATE, as well as estimates derived from the forcing terms used in ORAS5 (as described in Eq. (11)). Additionally our reanalyses and observation based estimates are given. NRMSE values and correlation coefficients are calculated using monthly means and in respect to GREP⁽¹⁾ and ORAS5⁽²⁾. Units are Sverdrup (Sv).

to river icing and low precipitation values, however freshwater that entered the Arctic in the past season and moves slowly with the ocean currents still find it's way through the Arctic gateways. In June, the large freshwater peak generates barotropic waves that are also seen in the reanalysis volume transports, residuals arise from different river discharge estimates and are generally positive for reanalyses and negative for observation based estimates. In late summer river discharge tends to decline and precipitation adopts the main role of delivering freshwater to the ocean. Oceanic reanalyses show a fast decline after the June peak and exhibit a volume transport minimum in September leading to a negative residual peak in all estimates. We assume that volume that enters the Arctic as freshwater through precipitation is slowly transported with the ocean currents and modified by wind stress, partly freezes on the way and takes weeks up to month to leave the Arctic area, explaining the elevated winter and spring transports.

Table 6 shows annual means of the various transport estimates, standard deviations and NRMSEs. The normalised RMSE values are calculated with respect to GREP⁽¹⁾ and ORAS5⁽²⁾, indicated by the superscripts. In addition to the estimates from Fig. 9, also the observation-based mean volume transport from the ARCGATE project (Tsubouchi et al., 2019; Tsubouchi et al., 2012, 2018) is given. To compare the ARCGATE value to ORAS5 and GREP, mean values and NRMSEs are calculated over the ARCGATE period (10/2004 - 05/2010, indicated by a *). With an annual mean volume transport of 0.151Sv the ARCGATE estimate is about 25% lower than the GREP estimate. The high NRMSE values emerge as the observation based ARCGATE flux does not show any peaks in June but rather stays low throughout the summer (not shown). This is probably a result from the mooring arrays in the gateways being too sparse and hence the velocity field not being measured accurately enough to dissolve the barotropic wave signal. Instead ARCGATE shows the export of freshwater that travels with the oceanic currents, dilutes on the way and remains around the continental shelves for some time. Hence the oceanic reanalyses and the observation based ARCGATE estimate should only be compared in terms of annual means and not seasonalities.

4.3.1 Volumetric budget closure

To close the volumetric budget and get rid of the small residual we use a variational adjustment procedure (see Sect. 2.1.2). A priori estimates are calculated by taking the mean over all trustworthy estimates of the individual budget terms - given in Table



7. We perform the adjustment upon long term annual means to close the annual budget and upon the monthly climatologies of the individual terms to close the budget on a monthly scale. The uncertainties for the annual optimization are estimated as the standard deviations between the various estimates (see Table 7) of the mean annual budget terms. The uncertainties for the monthly optimization are estimated by taking the annual maximum of the monthly standard deviations, indicated by the dashed lines in Fig. 11. Oceanic lateral transport F features the largest uncertainties and hence also experiences the largest adjustments amongst all budget terms.

Figure 12 shows the adjusted long term annual volume budget over the oceanic and land domains. About $6584 \pm 84 \text{ km}^3$

F	GREP
R	Obs _{Ge} , GloFAS _{ERA5new} , ERA5-Land
$R_{Greenland}$	Liquid and solid from Mankoff et al. (2020a, b)
$Ocean_{Atm,in}$	VIWVD from ERA5, ERA-Interim, JRA55, JRA55c
$Ocean_{stor}$	GRACE (CSR, GFZ, JPL)
$Land_{Atm,in}$	VIWVD from ERA5, ERA-Interim, JRA55, JRA55c
$Land_{stor}$	GRACE (CSR, GFZ, JPL)

Table 7. List of employed datasets in the variational adjustment procedure.

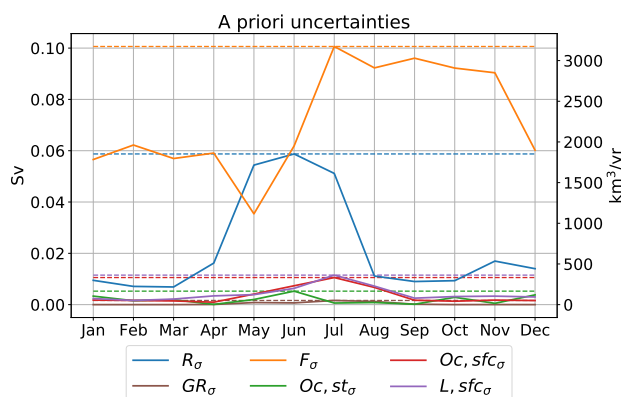


Figure 11. Mean annual cycles (1993-2018) of spread for runoff (R), Greenland runoff (GR), lateral ocean fluxes (F), oceanic storage (Oc, st) and atmospheric input over ocean (Oc, sfc) and land (L, sfc). Spread is calculated as standard deviate between various estimates (see table 7) of the individual terms and as RMS error estimate for GRACE.

leave the Arctic Ocean through the main gateways on an annual basis, whereof approximately $23 \pm 6 \text{ km}^3$ are the result of a slight decline in the oceans volume storage. The remaining 6561 km^3 enter the ocean via the surface, where roughly two thirds ($4379 \pm 25 \text{ km}^3$) are supplied by runoff and about one third ($2182 \pm 80 \text{ km}^3$) is delivered through the atmosphere. Oceanic transports out of the Arctic domain exceed the atmospheric moisture entering the Arctic ($6295 \pm 121 \text{ km}^3$) by nearly 5%, indicating an annual loss of water volume of roughly 300 km^3 . The bulk part of this loss is generated through glacial melt and to a smaller part by surface drying of the Arctic land areas due to snow and permafrost degradation. Arctic land areas feature a

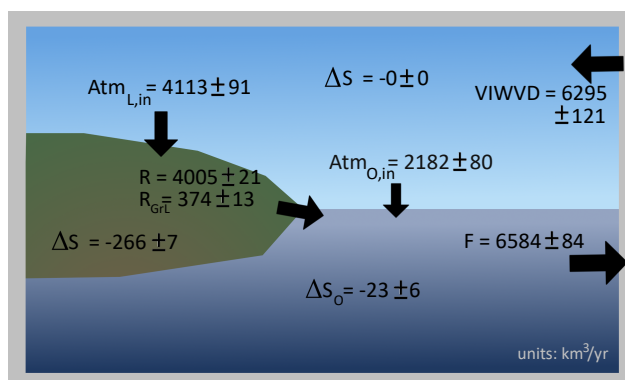


Figure 12. 1993-2018 adjusted long term means of the Arctic hydrological cycle. Units are km^3 per year; arrow areas are scaled by their magnitude.

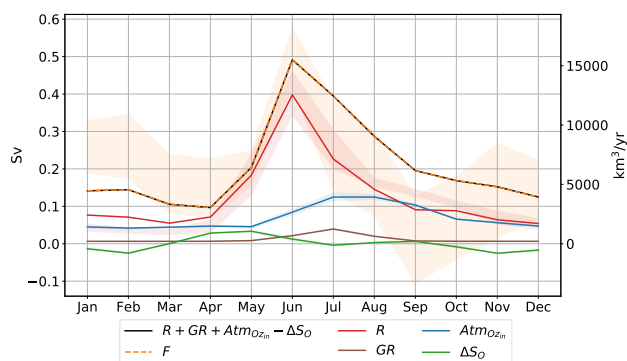


Figure 13. 1993-2018 adjusted mean annual cycle of the Arctic Oceans hydrological cycle. Shading represents the uncertainties of the a priori estimates.

decline of $266 km^3$ per year, wherefrom approximately 50% come from Greenland, about 40% from Arctic glaciers (excluding Greenland) and the remaining 10% are the result of a decline in land water storage and snow cover. The reported slight decline in oceanic storage is probably caused due to variations generated by barotropic fluctuations forced by decadal wind variations/trends, while contributions from fresh water fluxes are assumed negligible (Volkov and Landerer, 2013; Fukumori et al., 2015). Volkov and Landerer (2013) find that an intensification of the westerly winds over the North Atlantic and over the Russian Arctic continental shelf lead to a decrease of ocean mass in the central Arctic. Further they show the correlation between Arctic Ocean mass fluctuations and northward wind anomalies over the Bering Seas and the northeastern North Atlantic. They reveal that cyclonic/anticyclonic anomalies of the large-scale ocean circulation lead to negative/positive Arctic mass anomalies.

Table 8 provides adjusted monthly and annual climatologies and Fig. 13 shows the adjusted seasonal cycles of the oceanic budget terms. A metric to identify whether the variational adjustment is successful is the comparison of the adjusted terms to



	R	GR	Atm _{O,in}	ΔS _O	F
Jan	2415	204	1420	-421	4460
Feb	2238	204	1320	-793	4554
Mar	1735	203	1396	9	3325
Apr	2250	207	1488	898	3047
May	5780	264	1438	1053	6428
Jun	12549	679	2660	393	15495
Jul	7137	1242	3935	-128	12441
Aug	4585	628	3933	101	9045
Sep	2868	234	3259	199	6163
Oct	2778	210	2073	-257	5317
Nov	2018	207	1773	-801	4800
Dec	1706	205	1491	-534	3937
Mean	4005	374	2182	-23	6584

Table 8. 1993-2018 adjusted mean annual cycle of the Arctic Oceans hydrological cycle; units are km³ per year.

510 their a priori uncertainty (L'Ecuyer et al., 2015), hence Fig. 13 also shows the spreads of the a priori estimates (shaded areas). Due to the large differences between oceanic volume transports and freshwater input terms in late winter to early spring and in September, both F and runoff R feature adjustments beyond their a priori spreads, demonstrating that the a priori uncertainties are larger than indicated as systematic biases are not incorporated.



5 Conclusions

- 515 We analysed and compared various estimates of runoff into the Arctic Ocean on a seasonal and annual basis and find considerable differences in terms of seasonalities, mean values and trends. Further we used a non-steric formulation of the Arctic Ocean volume budget equation and compare freshwater input into the ocean to lateral volume transports over the Arctic boundaries. To close the budget and get best estimates of all budget terms we applied a variational adjustment procedure. The main outcomes of this study are the following:
- 520 – River discharge observations for the major Arctic catchments show distinct runoff peaks in June due to snow melt and river ice breakup and low runoff values in winter. Annual trends indicate slight discharge increases over the past decades - the largest increases evident at the Lena and Ob basins with a rise of about $4 \pm 2\%$ per decade. The total Pan-Arctic area (excluding Greenland) exhibits an upward trend of $2 \pm 1\%$ per decade. Holmes et al. (2018) show that the six largest Eurasian rivers (Ob, Yenisei, Lena, Kolyma, Pechora, Severnaya Dvina) exhibit an increase of $3.3 \pm 1.6\%$ and the two
 - 525 largest North American rivers (Mackenzie and Yukon) an increase of $2.0 \pm 1.8\%$ per decade.
 - We estimate Pan-Arctic river discharge from gauge observations using monthly correction factors from GloFAS_{ERA5new}, as the popular method of hydrological analogy tends to underestimate the high flow summer peaks (see Fig. 4), and obtain a long term annual flux of $4031 \text{ km}^3 \pm 203$ (excluding Greenland).
 - Runoff from ERA5 is substantially too low, with the largest errors occurring in the high flow summer months, and
 - 530 features strong declines of $10 \pm 1\%$ per decade on a Pan-Arctic basis and even stronger declines at the major catchments. These strong and unrealistic declines can be traced back to two inhomogeneities (1992 and 2004) present in ERA5s snow melt and are caused by the data assimilation system removing snow. While the discontinuity around 2004 was traced back to the introduction of the IMS snow cover information to the snow assimilation system, the 1992 inhomogeneity still needs further investigation.
 - 535 ERA5-Land does not feature these inhomogeneities and exhibits more moderate runoff declines of $5\text{--}6 \pm 2\%$ for the Eurasian major watersheds and $2 \pm 1\%$ for Mackenzie river and on the Pan-Arctic scale.
 - Calculating runoff through the divergence of moisture flux is the only reanalyses based estimate that exhibits a slightly positive trend of $2 \pm 1\%$ per decade and thus features the best agreement with observations in terms of trends and the highest correlation coefficient. However VIWVD tends to underestimate runoff by roughly 5%.
 - 540 – River discharge from GloFAS_{ERA5} reflects the variation and the inhomogeneities from ERA5 and shows an additional negative shift due to two sink terms removing water in LISFLOOD that leads to a discharge underestimation of up to 50% towards the end of the time series. In contrast GloFAS_{ERA5new} entails a vast improvement and reproduces the observed values best in terms of annual means and NRMSEs.
 - Liquid and solid discharge from Greenland account for roughly 10% of total Pan-Arctic discharge and hence should not
 - 545 be neglected in assessments of the Arctic freshwater budget. Liquid discharge features a vast increase of $20 \pm 4\%$ per



decade and solid discharge an increase of $8 \pm 1\%$. Due to glacial melting also land storage changes are particularly strong over the Greenlandic ice sheet, accounting for roughly half of total Pan-Arctic land storage change.

- Comparing all trustworthy estimates of freshwater input into the Arctic Ocean to volume transports from ocean reanalysis yields only a small imbalance of 1% in terms of annual means.

550 – The variational adjustment procedure provides us with one best estimate of every budget term for every calendar month - listed in Tab. 8. About two thirds of the freshwater provided for oceanic volume transports come from runoff ($4379 \pm 25 \text{ km}^3$ per year) and about one third is provided by the atmosphere. With oceanic and land storage declining - especially the land areas suffer a strong decline - there's a surplus of roughly 300 km^3 per year of more water leaving the Arctic area than entering.

555 Thus the variational adjustment brought reliable estimates of the volume budget terms on an annual scale, requiring only moderate adjustments of less than 1% for each individual term. With an annual value of $4379 \pm 25 \text{ km}^3$ (calculated over 1993-2018), our adjusted runoff estimate is slightly higher than estimates made by Shiklomanov and Lammers (2013) and Shiklomanov et al. (2021a) for the period 1936-2006. However considering the different calculation periods and assuming a decadal rise of roughly 2%, the estimates come quite close. On a seasonal scale however, the adjustment process was not a full success, as
 560 the budget residuals of some month are too large to be eliminated within the a priori spreads of the individual terms. This is very likely caused by systematic errors being present in the data-sets, or at least in their seasonal cycles, that are not taken into account in our a priori uncertainty estimates.

A possible reason for inconsistencies between runoff and ocean reanalyses is the usage of climatological river discharge data to
 565 specify land freshwater input in ocean reanalyses. Our analyses show that the seasonal cycle of the ORAS5 runoff climatology Bt06 fits quite well to our observation based estimates (see Fig. 5), however the lack of inter-annual variability in the freshwater input alters i.a. oceanic volume transports. In their BRONCO (Benefits of dynamically modelled river discharge input for ocean and coupled systems) project Zuo et al. (in preparation) work on implementing a time-varying land freshwater input, derived from discharge data from GloFAS version 2.1, into ORAS5. This should further reduce the inconsistencies between runoff and
 570 oceanic fluxes from ocean reanalyses.

In conclusion our results show that seasonal peaks of river discharge are underestimated in most reanalyses, mostly due to spurious signals in the data assimilation system. We find distinct improvements in the new GloFAS_{ERA5new} product, especially when investigating seasonal cycles and long term means it features vast enhancements compared to its precursors.
 575 However for interannual variability and trend analysis we would recommend the use of the VIWVD estimate, as it reproduces trends from gauge observations quite accurately. In order to expand gauge observations onto the total Pan-Arctic area we would advise to use reliable river discharge estimates like GloFAS_{ERA5new} to calculate discharge over the ungauged areas, as this results in more plausible seasonal cycles than the method of hydrological analogy. However in terms of trends and inter-annual



variability the various observation based estimates (ae, Ge, Ee) show almost identical results, as the general trend is shaped by
 580 discharge from the major Arctic catchments.

Code and data availability. ERA5 and ERA5-Land data are available on the Copernicus Climate Change Service (C3S) Climate Data Store (Hersbach et al., 2019; Muñoz Sabater, 2019). GloFAS river discharge data are also available on the Copernicus Climate Change Service (C3S) Climate Data Store (Harrigan et al., 2019). GRACE monthly ocean bottom pressure anomalies and land water-equivalent-thickness anomalies are available through the Physical Oceanography Distributed Active Archive Center (Landerer, 2020b, a). River gauge observations
 585 are downloadable through the Arctic Great Rivers Observatory (Shiklomanov et al., 2021b) and the Regional Arctic Hydrographic Network data set (Lammers et al., 2001).

Appendix A: List of Acronyms

ARCGATE Mooring-derived data of oceanic fluxes through the Arctic gateways

BT06 runoff climatology used in ORAS5

590 **BRONCO** Benefits of dynamically modelled river discharge input for ocean and coupled atmosphere-land-ocean systems.

C3S Copernicus Climate Change Service

CAA Canadian Arctic Archipelago

CDS Climate Data Store

CGLORS ocean reanalyses from the Euro-Mediterranean Center on Climate Change

595 **CMEMS** Copernicus Marine Environment Monitoring Service

CEMS Copernicus Emergency Management Service

ECMWF European Centre for Medium-Range Weather Forecasts

ERA5 ECMWF's fifth atmospheric reanalysis

ERA5-Land offline simulation of ERA5

600 **ERA-Interim** ECMWF interim reanalysis

FOAM ocean reanalyses from the UK Met Office

GloFAS Global Flood Awareness System

GLORYS ocean reanalyses from Mercator Ocean



- GRACE** Gravity Recovery and Climate Experiment
- 605 **GREP** Global ocean Reanalysis Ensemble Product
- hbr0** experimental ORAS5 run using river discharge from GloFAS_{ERA5}
- HTESSEL** Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land
- JRA55** Japanese 55-year Reanalysis
- JRC** Joint Research Centre
- 610 **LISFLOOD** model for river basin scale water balance and flood simulation
- NEMO** Nucleus for European Modelling of the Ocean
- NRMSE** normalized root mean square error
- Oras5** ECMWF's Ocean Reanalysis System 5
- r** Pearsons correlation coefficient
- 615 **ET** evapotranspiration
- F** oceanic lateral transport
- F^{surf}** freshwater surface fluxes (precipitation, evaporation, runoff)
- g** gravitational constant
- P** precipitation
- 620 **p_s** surface pressure
- q** specific humidity
- R** runoff or river discharge
- ρ_{ow}** density of freshwater
- S_A** atmospheric storage
- 625 **S_L** land storage
- S_O** ocean storage
- SD** snow depth



SWVL Volumetric soil water per layer

v horizontal wind vector

630 **VIWVD** Vertical integral of divergence of moisture flux

Author contributions. S.W. performed the data analysis, including the production of the figures in the paper and prepared the manuscript. M.M., V.S., E.Z., H.Z. and L.H. contributed to the interpretation of results and the writing of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

635 *Acknowledgements.* The authors thank Takamasa Tsubouchi for the helpful discussions on ARCGATE data. S.W., V.S., and M.M. were supported by Austrian Science Fund project P33177. V.S. was additionally supported by Copernicus Marine environment Service contract 114- R&D-GLO-RAN-CMEMS Lot 8.



References

- Aagaard, K. and Carmack, E. C.: The role of sea ice and other fresh water in the Arctic circulation, *Journal of Geophysical Research: Oceans*, 94, 14 485–14 498, <https://doi.org/10.1029/JC094iC10p14485>, 1989.
- Bacon, S., Aksenov, Y., Fawcett, S., and Madec, G.: Arctic mass, freshwater and heat fluxes: methods and modelled seasonal variability, *Phil. Trans. R. Soc. A*, 373, <https://doi.org/10.1098/rsta.2014.0169>, 2015.
- Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, *Journal of Hydrometeorology*, 10, 623 – 643, <https://doi.org/10.1175/2008JHM1068.1>, 2009.
- Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., Mirouze, I., Peterson, K. A., Sellar, A., and Storkey, D.: Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts, *Geoscientific Model Development*, 7, 2613–2638, <https://doi.org/10.5194/gmd-7-2613-2014>, 2014.
- Bourdalle-Badie, R. and Treguier, A.-M.: A climatology of runoff for the global ocean-ice model ORCA025, 2006.
- Burek, P., Van der Knijff, J., and De Roo, A.: LISFLOOD - Distributed Water Balance and Flood Simulation Model - Revised User Manual 2013, <https://doi.org/10.2788/24982>, 2013.
- Church, J., Clark, P., Cazenave, A., Gregory, J., Jevrejeva, S., Levermann, A., Merrifield, M., Milne, G., Nerem, R., Nunn, P., Payne, A., Pfeffer, W., Stammer, D., and Unnikrishnan, A.: Sea Level Change. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)], Cambridge University Press, 2013.
- Collins, M., Knutti, R., Arblaser, J., Dufresne, J.-L., Fichet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A., and Wehner, M.: Long-term Climate Change: Projections, Commitments and Irreversibility, pp. 1029–1136, 2014.
- Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., and Buontempo, C.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies, *Earth System Science Data*, 12, 2097–2120, <https://doi.org/10.5194/essd-12-2097-2020>, 2020.
- Curry, B., Lee, C. M., and Petrie, B.: Volume, Freshwater, and Heat Fluxes through Davis Strait, 2004–05, *Journal of Physical Oceanography*, 41, 429 – 436, <https://doi.org/10.1175/2010JPO4536.1>, 2011.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- Desportes, C., Garric, G., Régnier, C., Drévillon, M., Parent, L., Drillet, Y., Masina, S., Storto, A., Mirouze, I., Cipollone, A., Zuo, H., Balmaseda, M., Peterson, D., Wood, R., Jackson, L., Mulet, S., and Greiner, E.: CMEMS-GLO-QUID-001-026, E.U. Copernicus Marine Service Information, 2017.
- Dickson, R., Rudels, B., Dye, S., Karcher, M., Meincke, J., and Yashayev, I.: Current estimates of freshwater flux through Arctic and subarctic seas, *Progress in Oceanography*, 73, 210–230, <https://doi.org/10.1016/j.pocean.2006.12.003>, 2007.



- Fukumori, I., Wang, O., Llovel, W., Fenty, I., and Forget, G.: A near-uniform fluctuation of ocean bottom pressure and sea level across the deep ocean basins of the Arctic Ocean and the Nordic Seas, *Progress in Oceanography*, 134, 152–172, <https://doi.org/10.1016/j.pocean.2015.01.013>, 2015.
- Garric, G., Parent, L., Greiner, E., Drévilion, M., Hamon, M., Lellouche, J.-M., Régner, C., Desportes, C., Le Galloudec, O., Bricaud, C., Drillet, Y., Hernandez, F., and Le Traon, P.-Y.: Performance and quality assessment of the global ocean eddy-permitting physical reanalysis GLORYS2V4., in: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, p. 18776, 2017.
- Haine, T. W., Curry, B., Gerdes, R., Hansen, E., Karcher, M., Lee, C., Rudels, B., Spreen, G., de Steur, L., Stewart, K. D., and Woodgate, R.: Arctic freshwater export: Status, mechanisms, and prospects, *Global and Planetary Change*, 125, 13–35, <https://doi.org/10.1016/j.gloplacha.2014.11.013>, 2015.
- Haine, T. W. N.: Arctic Ocean Freshening Linked to Anthropogenic Climate Change: All Hands on Deck, *Geophysical Research Letters*, 47, <https://doi.org/10.1029/2020GL090678>, 2020.
- Harrigan, S., Zsoter, E., Barnard, C., F., W., Salamon, P., and Prudhomme, C.: River discharge and related historical data from the Global Flood Awareness System, v2.1, <https://doi.org/10.24381/cds.a4fdd6b9>, 2019.
- Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., and Pappenberger, F.: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, *Earth System Science Data*, 12, 2043–2060, <https://doi.org/10.5194/essd-12-2043-2020>, 2020.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 monthly averaged data on single levels from 1979 to present, <https://doi.org/10.24381/cds.f17050d7>, 2019.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Holmes, R., Shiklomanov, A., Suslova, A., Tretiakov, M., McClelland, J., Spencer, R., and Tank, S.: River Discharge, Arctic Report Card: Update for 2018, <https://arctic.noaa.gov/Report-Card/Report-Card-2018/ArtMID/7878/ArticleID/786/River-Discharge>, 2018.
- Knijff, J. M. V. D., Younis, J., and Roo, A. P. J. D.: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, *International Journal of Geographical Information Science*, 24, 189–212, <https://doi.org/10.1080/13658810802549154>, 2010.
- Kobayashi, C., Endo, H., Ota, Y., Kobayashi, S., Onoda, H., Harada, Y., Onogi, K., and Kamahori, H.: Preliminary Results of the JRA-55C, an Atmospheric Reanalysis Assimilating Conventional Observations Only, *SOLA*, 10, 78–82, <https://doi.org/10.2151/sola.2014-016>, 2014.
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, *Journal of the Meteorological Society of Japan*. Ser. II, 93, 5–48, <https://doi.org/10.2151/jmsj.2015-001>, 2015.
- Lammers, R. B., Shiklomanov, A. I., Vörösmarty, C. J., Fekete, B. M., and Peterson, B. J.: Assessment of contemporary Arctic river runoff based on observational discharge records, *Journal of Geophysical Research: Atmospheres*, 106, 3321–3334, <https://doi.org/10.1029/2000JD900444>, 2001.



- Landerer, F.: CSR TELLUS GRACE Level-3 Monthly Ocean Bottom Pressure Anomaly RL06 v03, <https://doi.org/10.5067/TEOCN-3AC63>, 2020a.
- Landerer, F.: CSR TELLUS GRACE Level-3 Monthly Land Water-Equivalent-Thickness Surface Mass Anomaly RL06 v03, <https://doi.org/10.5067/TELND-3AC63>, 2020b.
- 715 Lin, X., Massonnet, F., Yang, C., Artale, V., de Toma, V., and Rana, A.: Arctic freshwater cycle and the interaction with the North Atlantic, <https://doi.org/10.5194/egusphere-egu21-436>, eGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-436, 2021.
- L'Ecuyer, T. S., Beaudoin, H. K., Rodell, M., Olson, W., Lin, B., Kato, S., Clayson, C. A., Wood, E., Sheffield, J., Adler, R., Huffman, G., Bosilovich, M., Gu, G., Robertson, F., Houser, P. R., Chambers, D., Famiglietti, J. S., Fetzer, E., Liu, W. T., Gao, X., Schlosser, C. A., Clark, E., Lettenmaier, D. P., and Hilburn, K.: The Observed State of the Energy Budget in the Early Twenty-First Century, *Journal of*
- 720 *Climate*, 28, 8319 – 8346, <https://doi.org/10.1175/JCLI-D-14-00556.1>, 2015.
- Madec, G. and Team, N. S.: NEMO ocean engine, <https://doi.org/10.5281/zenodo.1464816>, 2019.
- Mankoff, K. D., Noël, B., Fettweis, X., Ahlstrøm, A. P., Colgan, W., Kondo, K., Langley, K., Sugiyama, S., van As, D., and Fausto, R. S.: Greenland liquid water discharge from 1958 through 2019, *Earth System Science Data*, 12, 2811–2841, <https://doi.org/10.5194/essd-12-2811-2020>, 2020a.
- 725 Mankoff, K. D., Solgaard, A., Colgan, W., Ahlstrøm, A. P., Khan, S. A., and Fausto, R. S.: Greenland Ice Sheet solid ice discharge from 1986 through March 2020, *Earth System Science Data*, 12, 1367 – 1383, <https://doi.org/10.5194/essd-12-1367-2020>, 2020b.
- Mayer, M., Alonso Balmaseda, M., and Haimberger, L.: Unprecedented 2015/2016 Indo-Pacific Heat Transfer Speeds Up Tropical Pacific Heat Recharge, *Geophysical Research Letters*, 45, 3274–3284, <https://doi.org/10.1002/2018GL077106>, 2018.
- Mayer, M., Tietsche, S., Haimberger, L., Tsubouchi, T., Mayer, J., and Zuo, H.: An Improved Estimate of the Coupled Arctic Energy Budget, *Journal of Climate*, 32, 7915 – 7934, <https://doi.org/10.1175/JCLI-D-19-0233.1>, 2019.
- 730 Morison, J., Kwok, R., Peralta Ferriz, C., Alkire, M., Rigor, I., Andersen, R., and Steele, M.: Changing Arctic Ocean freshwater pathways, *Nature*, 481, 66–70, <https://doi.org/10.1038/nature10705>, 2012.
- Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, *Earth System Science Data Discussions*, 2021, 1–50, <https://doi.org/10.5194/essd-2021-82>, 2021.
- 735 Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J., and Döll, P.: Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use, *Hydrology and Earth System Sciences*, 20, 2877–2898, <https://doi.org/10.5194/hess-20-2877-2016>, 2016.
- 740 Muntjewerf, L., Petrini, M., Vizcaino, M., Ernani da Silva, C., Sellevold, R., Scherrenberg, M. D. W., Thayer-Calder, K., Bradley, S. L., Lenaerts, J. T. M., Lipscomb, W. H., and Lofverstrom, M.: Greenland Ice Sheet Contribution to 21st Century Sea Level Rise as Simulated by the Coupled CESM2.1-CISM2.1, *Geophysical Research Letters*, 47, <https://doi.org/10.1029/2019GL086836>, 2020.
- Muñoz Sabater, J.: ERA5-Land monthly averaged data from 1981 to present, <https://doi.org/10.24381/cds.68d2bb3>, 2019.
- Prowse, T. D. and Flegg, P. O.: The magnitude of river flow to the Arctic Ocean: dependence on contributing area, *Hydrological Processes*, 14, 3185–3188, [https://doi.org/10.1002/1099-1085\(200011/12\)14:16/17<3185::AID-HYP170>3.0.CO;2-S](https://doi.org/10.1002/1099-1085(200011/12)14:16/17<3185::AID-HYP170>3.0.CO;2-S), 2000.
- 745 Rawlins, M. A., Steele, M., Holland, M. M., Adam, J. C., Cherry, J. E., Francis, J. A., Groisman, P. Y., Hinzman, L. D., Huntington, T. G., Kane, D. L., Kimball, J. S., Kwok, R., Lammers, R. B., Lee, C. M., Lettenmaier, D. P., McDonald, K. C., Podest, E., Pundsack, J. W., Rudels, B., Serreze, M. C., Shiklomanov, A., Øystein Skagseth, Troy, T. J., Vörösmarty, C. J., Wensnahan, M., Wood, E. F., Woodgate, R.,



- Yang, D., Zhang, K., and Zhang, T.: Analysis of the Arctic System for Freshwater Cycle Intensification: Observations and Expectations, *Journal of Climate*, 23, 5715 – 5737, <https://doi.org/10.1175/2010JCLI3421.1>, 2010.
- Rodell, M., Beaudoin, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J., and Wood, E. F.: The Observed State of the Water Cycle in the Early Twenty-First Century, *Journal of Climate*, 28, 8289 – 8318, <https://doi.org/10.1175/JCLI-D-14-00555.1>, 2015.
- Schauer, U. and Losch, M.: “Freshwater” in the Ocean is Not a Useful Parameter in Climate Research, *Journal of Physical Oceanography*, 49, 2309 – 2321, <https://doi.org/10.1175/JPO-D-19-0102.1>, 2019.
- Serreze, M. C., Barrett, A. P., Slater, A. G., Woodgate, R. A., Aagaard, K., Lammers, R. B., Steele, M., Moritz, R., Meredith, M., and Lee, C. M.: The large-scale freshwater cycle of the Arctic, *Journal of Geophysical Research: Oceans*, 111, <https://doi.org/10.1029/2005JC003424>, 2006.
- Shiklomanov, A. and Lammers, R.: 5.11 - Changing Discharge Patterns of High-Latitude Rivers, in: *Climate Vulnerability*, edited by Pielke, R. A., pp. 161–175, Academic Press, Oxford, <https://doi.org/10.1016/B978-0-12-384703-4.00526-8>, 2013.
- Shiklomanov, A., Déry, S., Tretiakov, M., Yang, D., Magritsky, D., Georgiadi, A., and Tang, W.: River Freshwater Flux to the Arctic Ocean, pp. 703–738, https://doi.org/10.1007/978-3-030-50930-9_24, 2021a.
- Shiklomanov, A., Holmes, R., McClelland, J., Tank, S., and R.G.M., S.: Arctic Great Rivers Observatory. Discharge Dataset, Version YYYYMMDD, <https://www.arcticgreatrivers.org/data>, 2021b.
- Shiklomanov, A. I. and Vuglinsky, V. S.: Hydrological networks, https://www.arcticobserving.org/images/pdf/Workshops/4th_stpetersburg/slideshow/shiklomanov_ppt.pdf, 2008.
- Shiklomanov, A. I., Lammers, R. B., and Vörösmarty, C. J.: Widespread decline in hydrological monitoring threatens Pan-Arctic Research, *Eos, Transactions American Geophysical Union*, 83, 13–17, <https://doi.org/10.1029/2002EO000007>, 2002.
- Shiklomanov, I. and Shiklomanov, A.: Climatic Change and the Dynamics of River Runoff into the Arctic Ocean, *Water Resources*, 30, 593–601, <https://doi.org/10.1023/B:WARE.0000007584.73692.ca>, 2003.
- Stahl, K., Tallaksen, L. M., Hannaford, J., and van Lanen, H. A. J.: Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble, *Hydrology and Earth System Sciences*, 16, 2035–2047, <https://doi.org/10.5194/hess-16-2035-2012>, 2012.
- Storto, A. and Masina, S.: C-GLORSv5: an improved multipurpose global ocean eddy-permitting physical reanalysis, *Earth System Science Data*, 8, 679–696, <https://doi.org/10.5194/essd-8-679-2016>, 2016.
- Storto, A., Masina, S., Simoncelli, S., Iovino, D., Cipollone, A., Drevillon, M., Drillet, Y., Schuckmann, K., Parent, L., Garric, G., Greiner, E., Desportes, C., Zuo, H., Balmaseda, M., and Peterson, K.: The added value of the multi-system spread information for ocean heat content and steric sea level investigations in the CMEMS GREP ensemble reanalysis product, *Climate Dynamics*, 53, <https://doi.org/10.1007/s00382-018-4585-5>, 2019.
- Syed, T. H., Famiglietti, J. S., Zlotnicki, V., and Rodell, M.: Contemporary estimates of Pan-Arctic freshwater discharge from GRACE and reanalysis, *Geophysical Research Letters*, 34, <https://doi.org/10.1029/2007GL031254>, 2007.
- Tsubouchi, T., Bacon, S., Naveira Garabato, A. C., Aksenov, Y., Laxon, S. W., Fahrbach, E., Beszczynska-Möller, A., Hansen, E., Lee, C. M., and Ingvaldsen, R. B.: The Arctic Ocean in summer: A quasi-synoptic inverse estimate of boundary fluxes and water mass transformation, *Journal of Geophysical Research: Oceans*, 117, <https://doi.org/10.1029/2011JC007174>, 2012.



- 785 Tsubouchi, T., Bacon, S., Aksenov, Y., Garabato, A. C. N., Beszczynska-Möller, A., Hansen, E., de Steur, L., Curry, B., and Lee, C. M.: The Arctic Ocean Seasonal Cycles of Heat and Freshwater Fluxes: Observation-Based Inverse Estimates, *Journal of Physical Oceanography*, 48, 2029 – 2055, <https://doi.org/10.1175/JPO-D-17-0239.1>, 2018.
- Tsubouchi, T., von Appen, W.-J., Schauer, U., Kanzow, T., Lee, C., Curry, B., de Steur, L., Ingvaldsen, R., and Woodgate, R. A.: The Arctic Ocean volume, heat and fresh water transports time series from October 2004 to May 2010, <https://doi.org/10.1594/PANGAEA.909966>, 2019.
- 790 Vaughan, D., Comiso, J., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., Ren, J., Rignot, E., Solomina, O., Steffen, K., and Zhang, T.: Observations: Cryosphere. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)], Cambridge University Press, 2013.
- 795 Volkov, D. L. and Landerer, F. W.: Nonseasonal fluctuations of the Arctic Ocean mass observed by the GRACE satellites, *Journal of Geophysical Research: Oceans*, 118, 6451–6460, <https://doi.org/10.1002/2013JC009341>, 2013.
- Wouters, B., Gardner, A., and Moholdt, G.: Global Glacier Mass Loss During the GRACE Satellite Mission (2002–2016), *Frontiers in Earth Science*, 7, <https://doi.org/10.3389/feart.2019.00096>, 2019.
- Yang, D., Ye, B., and L. Kane, D.: Streamflow changes over Siberian Yenisei River Basin, *Journal of Hydrology*, 296, 59–80, <https://doi.org/10.1016/j.jhydrol.2004.03.017>, 2004a.
- 800 Yang, D., Ye, B., and Shiklomanov, A.: Discharge Characteristics and Changes over the Ob River Watershed in Siberia, *Journal of Hydrometeorology*, 5, 595 – 610, [https://doi.org/10.1175/1525-7541\(2004\)005<0595:DCACOT>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0595:DCACOT>2.0.CO;2), 2004b.
- Yang, D., Zhao, Y., Armstrong, R., Robinson, D., and Brodzik, M.-J.: Streamflow response to seasonal snow cover mass changes over large Siberian watersheds, *Journal of Geophysical Research: Earth Surface*, 112, <https://doi.org/10.1029/2006JF000518>, 2007.
- 805 Ye, B., Yang, D., and Kane, D. L.: Changes in Lena River streamflow hydrology: Human impacts versus natural variations, *Water Resources Research*, 39, <https://doi.org/10.1029/2003WR001991>, 2003.
- Zsótér, E., Cloke, H. L., Prudhomme, C., Harrigan, S., de Rosnay, P., Munoz-Sabater, J., and Stephens, E.: Trends in the GloFAS-ERA5 river discharge reanalysis, <https://doi.org/10.21957/p9jrh0xp>, 2020.
- Zuo, H., Alonso-Balmaseda, M., Mogensen, K., and Tietsche, S.: OCEAN5: The ECMWF Ocean Reanalysis System and its Real-Time analysis component, <https://doi.org/10.21957/la2v0442>, 2018.
- 810 Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, *Ocean Science*, 15, 779–808, <https://doi.org/10.5194/os-15-779-2019>, 2019.
- Zuo, H., de Boisseson E., Zsoter, E., S., H., de Rosnay, P., and C., P.: Benefits of dynamically modelled river discharge input for ocean and coupled atmosphere-land-ocean systems, *Journal of advances in modelling earth systems*, in preparation.
- 815



Validation of key Arctic energy and water budget components in CMIP6

Susanna Winkelbauer^{1,2,4} · Michael Mayer^{1,2,3,4} · Leopold Haimberger^{1,4}

Received: 10 August 2023 / Accepted: 4 January 2024 / Published online: 3 February 2024
© The Author(s) 2024

Abstract

We investigate historical simulations of relevant components of the Arctic energy and water budgets for 39 Coupled Model Intercomparison Project Phase 6 (CMIP6) models and validate them against observation-based estimates. We look at simulated seasonal cycles, long-term averages and trends of lateral transports and storage rates in atmosphere and ocean as well as vertical fluxes at top-of-atmosphere and the surface. We find large inter-model spreads and systematic biases in the representation of annual cycles and long-term averages. Surface freshwater fluxes associated with precipitation and evaporation as well as runoff from Arctic lands tend to be overestimated by most CMIP6 models and about two thirds of the analysed models feature an early timing bias of one month in the runoff cycle phase, related to an early snow melt bias and the lack of realistic river routing schemes. Further, large biases are found for oceanic volume transports, partly because data required for accurate oceanic transport computations has not been archived. Biases are also present in the simulated energy budget components. The net vertical energy flux out of the ocean at the Arctic surface as well as poleward oceanic heat transports are systematically underestimated by all models. We find strong anti-correlation between average oceanic heat transports and mean sea ice cover, atmospheric heat transports, and also the long-term ocean warming rate. The latter strongly suggests that accurate depiction of the mean state is a prerequisite for realistic projections of future warming of the Arctic. Our diagnostics also provide useful process-based metrics for model selection to constrain projections.

Keywords Arctic · Energy budget · Water budget · CMIP6 · Validation · Bias

1 Introduction

The Arctic has undergone major changes in recent decades due to climate warming, with important implications not only for the Arctic itself, but also for the global climate system. Various feedback mechanisms like the ice-albedo feedback or the Planck feedback (Goosse et al. 2018) lead to a faster warming of the Polar regions compared to the globe. While earlier studies report a warming about twice the global average (Serreze et al. 2009; Walsh 2014), more

recent observational datasets suggest an even stronger warming about 4 times the global average (Rantanen et al. 2022). Rising temperatures provoke the degradation of permafrost (Rowland et al. 2010), thawing of the Greenland ice sheet (Mouginot et al. 2019) and sea ice melt (Stroeve and Notz 2018). The decline in sea ice area and thickness has been particularly prominent in recent decades (Kwok 2018) and is caused by both atmospheric and oceanic processes (Docquier and Koenigk 2021). Northward heat transports in both the atmosphere and ocean counterbalance an average net loss of energy to space in the Arctic. Variability and trends in those transports have major impacts on the state and change of the Arctic system, including sea ice, the atmosphere and the ocean (Docquier and Koenigk 2021).

Arctic warming also has a strong impact on the Arctic water balance, leading to an increase of runoff from land areas and the Greenland ice sheet as well as increases in precipitation. The reasons for enhanced Arctic precipitation changes are still under debate. While earlier studies attribute increases in area-integrated evaporation due to increased

✉ Susanna Winkelbauer
susanna.winkelbauer@univie.ac.at

¹ Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

² b.geos, Korneuburg, Austria

³ European Centre for Medium-Range Weather Forecasts, Bonn, Germany

⁴ Austrian Polar Research Institute (APRI), Vienna, Austria

open water areas together with enhanced moisture transports from lower latitudes (Bintanja and Selten 2014), more recent studies argue that the changes are consequences of the Planck feedback and therefore energetically driven (Pithan and Jung 2021; Bonan et al. 2023)

The effects of Arctic warming are not only limited to the Arctic—the melting of glaciers and the Greenland ice sheet contribute to rapid sea-level rise around the globe (e.g., Moon et al. 2018; Box et al. 2022) and the release of larger amounts of freshwater to the Atlantic Ocean (Haine et al. 2015) could have major implications for the oceanic circulation at a global scale. Further, there is contrasting evidence regarding the hypothesis that a reduction in the meridional temperature gradient might affect weather and climate in the mid-latitudes (e.g., Blackport and Screen 2020; Coumou et al. 2018; Francis and Vavrus 2012; Screen and Simmonds 2013)

Thus, the Arctic represents a complex system marked by tight couplings between atmosphere, ocean and sea ice, encompassing processes on various spatial and temporal scales. Analyzing the Arctic energy and water budgets is crucial to understand the physical processes of the system as well as the couplings between its components and to comprehend the pronounced warming trend and the resulting impacts on the Arctic system itself and globally. Further, improved process understanding and accurate validation data is needed to develop and enhance climate models and subsequently improve our knowledge of future Arctic change.

The development of the Coupled Model Intercomparison Project, a global collaborative initiative with its latest generation CMIP6 (Eyring et al. 2016), whose data are used to i.a. underpin the 6th Assessment Report of the Intergovernmental Panel on Climate Change (e.g., Fox-Kemper et al. 2021), helps to assess projected future changes under various greenhouse gas emission scenarios and is essential in understanding and quantifying the strength and the effects of climate change. However, the complex interactions between atmosphere, ocean and sea ice pose a major challenge to Arctic climate simulations and introduce large uncertainties and biases (Cai et al. 2021; Knutti 2008). This raises the need for a thorough evaluation of historical climate model simulations against observations in order to detect model biases, find potential shortcomings and improve our confidence in future projections.

However, due to the harsh environmental conditions and sheer remoteness, measurements in the polar regions are relatively sparse (Khosravi et al. 2022), complicating especially ocean and sea ice diagnostics. Satellite observations help in the quantification of surface properties, however in-situ data to assess subsurface properties, like vertically resolved temperatures in the ocean, are limited.

Over the past years, the usage of ocean reanalyses (ORAs) proved to be useful to study past ocean states, long-term

climate trends and investigate ocean variability (Storto et al. 2019b; von Schuckmann et al. 2020; Mayer et al. 2021c, 2022). However, as their reliability depends i.a. on the quality and quantity of observational data assimilated into the models, the reanalyses are affected by data paucity in the Arctic. Nevertheless, Mayer et al. (2021c) show that ORAs realistically represents observed trends and temporal variabilities of ocean heat content (OHC) in the Norwegian Sea. Cheng et al. (2022) find that the uncertainty of Arctic OHC is larger than for the other world basins, however they still find consistent trends for Arctic OHC between observations and a reanalysis product. Mayer et al. (2022) find a good agreement between ORAs and observations of the variability of ocean heat transport (OHT) anomalies into the Arctic Mediterranean, but they find OHT to be biased small by about 14%. In general, OHC is more strongly constrained in ORAs than oceanic transports and hence are deemed to be more reliable. A largely observation-based estimate of OHTs is provided by Tsubouchi et al. (2018), who derive transport estimates from moorings in a mass-consistent way, creating a largely model-independent estimate of Arctic OHTs.

Serreze et al. (2009) provide holistic estimates of annual cycles and long-term means of the coupled Arctic energy and water budget. However, their results contained inconsistencies of the various terms as indicated by large budget residuals, which is likely related to inaccurate data and suboptimal diagnostic methods (such as a biased atmospheric budget framework, see Mayer et al. 2017). Therefore, Mayer et al. (2019) combine transports from Tsubouchi et al. (2018) with state-of-the-art reanalyses and other observational products and provide updated and improved, consistent estimates of the coupled Arctic energy budget for the period 2005–2009. Similarly Winkelbauer et al. (2022) provide observationally constrained estimates of the key components of the Arctic water budget using observational datasets as well as reanalyses for 1993–2019.

In this study, we will use the observationally constrained estimates from Mayer et al. (2019) and Winkelbauer et al. (2022) as well as updated estimates from observations and reanalyses to evaluate a large ensemble of CMIP6 models. We aim to analyse the models' ability to accurately simulate some of the key components of the Arctic energy and water budgets and analyse the simulated long-term averages and seasonal cycles of the various energy and water cycle variables and their connections to understand typical model biases.

The paper is structured as follows. Section 2 introduces the main energy and water budget equations and describes the numerical methods used for calculating them, and Sect. 3 describes the data sets analysed and the study area. The results are presented in Sect. 4 and are divided into water (Sect. 4.1) and energy (Sect. 4.2) budget analyses. Conclusions and discussions follow in Sect. 5.

2 Methods

In this section we formulate the vertically integrated energy and water balance equations for the Arctic and describe the analytical methods used.

2.1 Energy and water budgets

For the Arctic energy cycle, we follow Mayer et al. (2019) and define the equation for the total energy budget of the atmosphere as

$$F_S = F_{TOA} - AET - \nabla \cdot F_A - L_f(T_p)P_{snow} \quad (1)$$

with the net (turbulent plus net radiative) vertical energy flux at the surface F_S , the net radiation at the top of the atmosphere F_{TOA} , the atmospheric total energy tendency AET and the divergence of vertically integrated lateral atmospheric energy transports $\nabla \cdot F_A$, which is equal to atmospheric energy transports over the lateral boundaries (AHT). The last term represents the cooling of the surface due to falling snow and consists of the latent heat of fusion L_f ($-0.3337 \times 10^6 \text{ J kg}^{-1}$) and the snowfall rate P_{snow} . Vertical fluxes are defined as positive downwards. The energy budget equation for an ocean-sea ice column reads as follows:

$$F_S = OHCT + \nabla \cdot F_O + MET + IHCT + \nabla \cdot F_I - L_f(T_p)P_{snow} + L_f\rho_{snow} \frac{\partial d_{snow}}{\partial t} \quad (2)$$

with the temporal tendency of ocean heat content OHCT, the divergence of vertically integrated ocean heat transport $\nabla \cdot F_O$, the sea ice melt energy tendency MET (i.e. the energy absorbed or released during melt and freeze, respectively, computed as the product of monthly sea ice thickness change and L_f), the sea ice sensible heat content tendency IHCT, the divergence of latent heat transport associated with sea ice transports $\nabla \cdot F_I$ and the snowfall term. The last term describes latent heat changes in conjunction with changes in grid-point-averaged snow thickness (d_{snow}).

For the oceanic water budget equation we follow Winkelbauer et al. (2022) and formulate it in its volumetric form:

$$\Delta S_O = P + ET + R - \nabla \cdot F_{vol} \quad (3)$$

with the change of ocean volume denoted as ΔS_O , the surface water fluxes precipitation P and evapotranspiration ET (counted positive downward), runoff from surrounding land areas R and the divergence of lateral oceanic volume fluxes $\nabla \cdot F_{vol}$.

Furthermore, following Gauss's divergence theorem the divergence terms in equations 2 and 3 can be replaced by transports of energy and volume across the lateral boundaries when considering closed oceanic regions.

2.2 Oceanic transports

Oceanic transports of volume (OVT), heat (OHT) and ice (OIT) through a given strait are defined as follows:

$$OVT = \int_{x_s}^{x_e} \int_0^{z(x)} \vec{v}_o(x, z) \cdot \vec{n} dz dx \quad (4)$$

$$OIT = \int_{x_s}^{x_e} d(x) \vec{v}_i(x) \cdot \vec{n} dx \quad (5)$$

$$OHT = c_p \rho \int_{x_s}^{x_e} \int_0^{z(x)} (\theta(x, z) - \theta_{ref}) \vec{v}_o(x, z) \cdot \vec{n} dz dx \quad (6)$$

where \vec{v}_o is the velocity vector of the oceanic flow and \vec{n} is the vector normal to the strait. Furthermore, x defines the width along the strait, with the straits' starting point x_s and the end point x_e . The straits' depth is given by z , where x and z together form the cross sectional area of the strait. Ice transports are calculated by integrating the cross-sectional ice velocity \vec{v}_i over the grid point average ice depth (d) and integrating over the section. Latent heat transports into the study area through ice exports (IHT) are then estimated by multiplying OIT with the sea ice density (assumed constant at 928 kg m^{-3}) and the latent heat of fusion L_f ($-0.3337 \times 10^6 \text{ J kg}^{-1}$). Computation of heat transports requires potential temperature θ , the specific heat of seawater c_p and the density of seawater ρ . Throughout this study, c_p and ρ are kept constant at $3996 \text{ J kg}^{-1} \text{ K}^{-1}$ and 1026 kg m^{-3} , respectively, because variations in c_p and ρ tend to compensate each other and together lead to only small changes in the computed heat transports (Fasullo and Trenberth 2008) which are neglected in the context of this study.

As discussed by Schauer and Beszczynska-Möller (2009), unambiguous heat transports would actually demand closed volume transports through the examined straits, which is not the case for the single straits considered here, and only approximately satisfied for the total oceanic transport through all straits. As a result, heat transports have to be calculated relative to a reference temperature θ_{ref} , which should represent the mean temperature of the assessed flow. Strictly speaking this reference temperature should vary spatially and temporally according to the investigated flow (Bacon et al. 2015). While changes in the reference temperature have only minor effects on the net Arctic transports (not shown), they are larger for transports through individual straits and may become significant the stronger θ_{ref} changes. However, to simplify the analysis we follow e.g. Tsubouchi et al. (2012), Tsubouchi et al. (2018), Muilwijk et al. (2018), Shu et al. (2022), Heuzé et al. (2023) and calculate all heat transports relative to a 0°C reference. Usage of the same reference temperature for all models and straits also allows

Table 1 List of models included in the analysis, their modelling components and links to relevant references

i	Model name	Atmosphere	Land	Ocean	Ice	References
1	ACCESS-CM2	UM10.6 GA7.1	CABLE2.5	MOM5.1	CICE5.1.2	Bi et al. (2020)
2	ACCESS-ESM1-5	UM7.3 GA1	CABLE2.4	MOM5.1	CICE4.1	Ziehn et al. (2020)
3	BCC-CSM2-MR	BCC-AGCM3-MR	BCC AVIM2	MOM4	SIS2	Wu et al. (2019)
4	BCC-ESM1	BCC-AGCM3-Chem	BCC AVIM2	MOM4	SIS2	Wu et al. (2019)
5	CAMS-CSM1-0	ECHAM5-CAMS	CoLM 1.0	MOM4	SIS1	Chen et al. (2019)
6	CanESM5	CanAM5	CLASS3.6-CTEM	NEMO3.4	LIM2	Swart et al. (2019)
7	CAS-ESM2-0	IAP AGCM5.0	CoLM	LICOM2.0	CICE4.0E	Zhang et al. (2020)
8	CESM2	CAM6	CLM5	POP2	CICE5.1	Danabasoglu et al. (2020)
9	CESM2-WACCM	WACCM6	CLM5	POP2	CICE5.1	Danabasoglu et al. (2020)
10	CMCC-CM2-HR4	CAM4	CLM4.5	NEMO3.6	CICE4.0	N/A
11	CMCC-CM2-SR5	CAM5.3	CLM4.5	NEMO3.6	CICE4.0	N/A
12	CNRM-CM6-1	ARPEGE-Clim6.3	Surfex 8.0c	NEMO3.6	Gelato 6.1	Voltaire et al. (2019)
13	CNRM-CM6-1-HR	ARPEGE-Clim6.3	Surfex 8.0c	NEMO3.6	Gelato 6.1	Voltaire et al. (2019)
14	EC-Earth3P-HR	IFS cy36r4	HTESSEL	NEMO3.6	LIM3	Haarsma et al. (2020)
15	EC-Earth3	IFS cy36r4	HTESSEL	NEMO3.6	LIM3	Döscher et al. (2022)
16	FGOALS-f3-L	FAMIL2.2	CLM4.0	LICOM3.0	CICE4.0	He et al. (2019)
17	FGOALS-g3	GAMIL3	CAS-LSM	LICOM3.0	CICE4.0	Li et al. (2020)
18	FIO-ESM2-0	CAM5	CLM4.0	POP2	CICE4	Bao et al. (2020)
19	GFDL-CM4	AM4.0	LM4.0	MOM6	SIS2	Held et al. (2019)
20	GFDL-ESM4	AM4.1	LM4.1	MOM6	SIS2	Dunne et al. (2020)
21	GISS-E2-1-G	GISS-E2.1	GISS LSM	GISS Ocean v1	GISS SI	Kelley et al. (2020)
22	HadGEM3-GC31-LL	UM-HG3-GA7.1	JULES-HG3-GL7.1	NEMO-HG3-GO6	CICE-HG3-GS18	Williams et al. (2018)
23	HadGEM3-GC31-MM	UM-HG3-GA7.1	JULES-HG3-GL7.1	NEMO-HG3-GO6	CICE-HG3-GS18	Williams et al. (2018)
24	INM-CM5-0	INM-AM5-0	INM-LND1	INM-OM5	INM-ICE1	N/A
25	IPSL-CM6A-LR	LMDZ	ORCHIDEE	NEMO-OPA	NEMO-LIM3	Boucher et al. (2020)
26	IPSL-CM6A-LR-INCA	LMDZ	ORCHIDEE	NEMO-OPA	NEMO-LIM3	Boucher et al. (2020)
27	KACE-1-0-G	UM-HG3-GA7.1	JULES-HG3-GL7.1	MOM4p1	CICE-HG3-GS18	Lee et al. (2020)
28	KIOST-ESM	AM2	LM3.0	MOM5	SIS	Pak et al. (2021)
29	MIROC-ES2L	CCSR-NIES AGCM	MATSIRO6	COCO4.9	COCO4.9	Hajima et al. (2020)
30	MIROC6	CCSR AGCM	MATSIRO6	COCO4.9	COCO4.9	Tatebe et al. (2019)
31	MPI-ESM1-2-HR	ECHAM6.3	JSBACH3.20	MPIOM1.63	MPIOM1.63	Mauritsen et al. (2019)
32	MPI-ESM1-2-LR	ECHAM6.3	JSBACH3.21	MPIOM1.63	MPIOM1.63	Mauritsen et al. (2019)
33	MRI-ESM2-0	MRI-AGCM3.5	HAL 1.0	MRI.COMv4	MRI.COMv4	Yukimoto et al. (2019)
34	NESM3	ECHAM6.3	JSBACH	NEMO3.4	CICE4.1	Cao et al. (2018)
35	NorCPM1	CAM(OSLO4.1)	CLM4	MICOM1.1	CICE4	Bethke et al. (2021)
36	NorESM2-MM	CAM6	CLM5	BLOM	CICE5.1.2	Seland et al. (2020)
37	SAM0-UNICON	CAM5-UNICON	CLM4.0	POP2	CICE4.0	Park et al. (2019)
38	TaiESM1	TaiAM1	CLM4.0	POP2	CICE4	Wang et al. (2021)
39	UKESM1-0-LL	UM-HG3-GA7.1	JULES-ES- 1.0	NEMO-HG3-GO6	CICE-HG3-GS18	Sellar et al. (2019)

for better inter-comparisons with one another (Muilwijk et al. 2018).

Transports must be calculated on the native grids of the models to maintain the conservation properties of the models. However, ocean models often use curvilinear grids where the North Pole is placed over land areas to avoid singularities over the ocean. The number of poles (tri- vs. dipolar), the exact location of the poles, and the Arakawa partition vary between models, resulting in a large number

of different grid types, making it difficult to compare models and with observations. We have developed two methods for calculating accurate ocean transports on different CMIP6 model grids, which are described in Winkelbauer et al. (2023) and are available via the Python package StraitFlux (Winkelbauer 2023).

Net Arctic transports are calculated as the sum of transports through Fram Strait, Davis Strait, the Barents Sea

Table 2 List of all CMIP6 variables used through this study, including their units, number of available models *n* and the indices of missing models

Variable	Description	Unit	n	Missing models
rsus, rsds, rlus, rlds	Surface up-/downward, short-/longwave raditations	Wm ⁻²	39	–
rsut, rsdt, rlut	Toa outgoing short-/longwave and incident shortwave raditations	Wm ⁻²	39	–
hfss, hfls	Surface sensible and latent heat flux	Wm ⁻²	38	28
pr	Precipitation flux	kg m ⁻² s ⁻¹	39	–
evspsbl	Evaporation (incl. sublimation and transpiration)	kg m ⁻² s ⁻¹	38	28
mrro	Runoff flux	kg m ⁻² s ⁻¹	36	5, 34, 35
thetao	Sea water potential temperature	degC	38	27
uo,vo	Sea water x/y velocity	ms ⁻¹	36	20,27,35
thkcello	Ocean cell thickness	m	26	7, 8, 9, 16, 17, 18, 21, 24, 27, 28, 33, 34, 35
sithick / sivol	Actual thickness of sea ice / total volume of sea ice divided by grid-cell area (=sithick*siconc)	m	35	17, 21, 24, 27
siconc	Percentage of grid cell covered by sea ice	%	36	20, 21, 27
siu, siv	Sea ice x/y velocity	ms ⁻¹	20	7, 8, 9, 13, 16, 17, 18, 20, 21, 24, 27, 28, 29, 30, 33, 34, 35, 36, 38

All variables where downloaded through ESGF

Opening and Bering Strait (see Fig. 1 below for the location of the cross-sections).

2.3 Metrics

To validate CMIP6 output against observations, scalar quantities are regridded to regular grids in the resolution of the available observation-based data (0.25° × 0.25° and 1° × 1° grids). However, quantities using vector-based components are computed on the respective native grids of the models to avoid any errors associated with the interpolation of vector quantities. Spatial averages are calculated over the Arctic areas as defined in Fig. 1 and long-term average seasonal cycles are determined over the 1993–2014 period.

We calculate decadal trends by applying a linear regression to the monthly anomaly (i.e., deseasonalized) time series. Significance is determined by the Wald test with a t-distribution, with p-values less than 0.05 considered significant. Inter-model correlations are calculated using Pearson's correlation coefficient *r* and to assess seasonal model performance we use normalised mean errors (nME). The normalisation for each variable is done using the largest error of all models for the variable in question to facilitate inter-model comparisons. For instance, the nME for model *j* over *N* years (whereby annual averages are calculated using only the assessed season, e.g. DJF for winter) is calculated as follows:

$$nME_j = \frac{\sum_{i=1}^N (data_{j,i} - reference_{j,i})}{MAX_k(nME_k)} \quad (7)$$

To determine sampling errors of long-term averages that can arise, e.g., from different states of natural variability modes in the model runs compared to observations, we use a bootstrapping approach of random sampling with replacement. Thus, for every model and variable we calculate 1000 long-term averages of the desired period (e.g. 22 years for the 1993–2014 period) out of randomly drawn annual averages within the most recent decades (1980–2014). The sampling error is then estimated as 2-sigma standard deviation from the distribution of the randomly sampled long-term averages.

Confidence ellipses for two-dimensional datasets (see all scatter-diagrams in Sec. 4 and the supplementary material) are calculated using the Pearson correlation coefficient as described at https://carstenshulp.github.io/2018/09/14/Plot_Confidence_Ellipse_001.html. They are determined for the 2-sigma standard deviation and therefore encompass about 95% of all values in the 2D space.

3 Data and study domain

3.1 CMIP6 models

We use monthly output from 39 models that participated in the Climate Model Intercomparison Project Phase 6 [CMIP6, Eyring et al. (2016)]. Table 1 lists all the models used in this study, including their modelling components, and provides links to key references. We use historical model runs for 37 models and the hist-1950 model run for EC-Earth3P-HR and HadGEM3-GC31-MM. We use one

ensemble member per model and choose the first available member per model, r1i1p1f2 for CNRM-CM6-1, CNRM-CM6-1-HR, MIROC-ES2L and UKESM1-0-LL, r1i1p1f3 for HadGEM3-GC31-LL and HadGEM3-GC31-MM, r1i1p2f1 for EC-Earth3P-HR and r1i1p1f1 for the remaining 32 models. The models have different horizontal and vertical resolutions (please refer to the individual model documentation listed in Tab. 1) and differ in their modelling components for atmosphere, land, ocean and sea ice. However, the models are not completely independent and often overlap in one or more modelling components. Therefore, when calculating the multi-model mean (MMM), models should ideally be preselected to avoid overlapping components or weighted with respect to their independence and performance (Brunner et al. 2020). Hence, results might differ when compiling a model ensemble that maximizes independence of its members, but this is not the focus of this study.

All data are obtained from the Earth System Grid Federation (ESGF) website (<https://esgf-node.llnl.gov/search/cmip6/>). We assess different components of the energy and water budgets. Table 2 lists the variables used in this study, not all variables are available for all models, therefore the number of available models (n) and a list of missing models (numbers correspond to indices in Table 1) are also given in Table 2. The variables listed in Table 2 are used to derive the main budget components represented by Eqs. 1 to 3, such as F_s , F_{TOA} , AET, OHCT, MET, OHT, OVT and OIT. F_{TOA} and F_s are calculated directly using

all available radiative and turbulent heat flux components (see Table 2) and oceanic transports are calculated using StraitFlux (Winkelbauer et al. 2023). Heat content tendencies in the ocean (OHCT) and sea ice (MET) are calculated from sea water potential temperature and sea ice thickness respectively, using a Theil-Sen trend estimator. Atmospheric energy tendencies (AET) are calculated on temperature and humidity levels using central differences of monthly mean values and the atmospheric heat transport AHT, which is equal to the divergence term $\nabla \cdot F_A$, is estimated indirectly using equation 1.

Sea ice extent was calculated similarly to Shu et al. (2020) as the area of all grid cells with sea ice concentration (*siconc*) greater than 15%. For sea ice thickness we either use the variable *sivol* or multiply *sithick* with *siconc*, depending on the availability of the variable through ESGF.

3.2 Observational data

To quantify the representation of the energy and water budget components in CMIP, we compare the modelled seasonal cycles and long-term averages with observation based estimates.

Winkelbauer et al. (2022) provide observationally constrained estimates of the key components of the Arctic water budget using in-situ and satellite observations as well as reanalyses, and enforcing budget closure with a variational

Table 3 List of datasets used to calculate the energy and water budget variables

Variable	Data	Time period
R	Adapted from Winkelbauer et al. (2022)	1993–2014
P-E	Adapted from Winkelbauer et al. (2022)	1993–2014
OVT	GREP	1993–2014
	ArcGate	10/2004–05/2010
OIT	GREP	1993–2014
	ArcGate	10/2004–05/2010
F_s	Mayer et al. (2021b)	1993–2014
	Mayer et al. (2019)	2005–2009
F_{TOA}	DEEPC	1993–2014
	Mayer et al. (2019)	2005–2009
OHCT	GREP	1993–2014
	Mayer et al. (2019)	2005–2009
MET	GREP	1993–2014
	CS2SMOS	October–March 2011–2014
	Mayer et al. (2019)	2005–2009
AET	ERA5	1993–2014
	Mayer et al. (2019)	2005–2009
AHT	ERA5	1993–2014
	Mayer et al. (2019)	2005–2009
OHT	GREP	1993–2014
	ArcGate	10/2004–05/2010

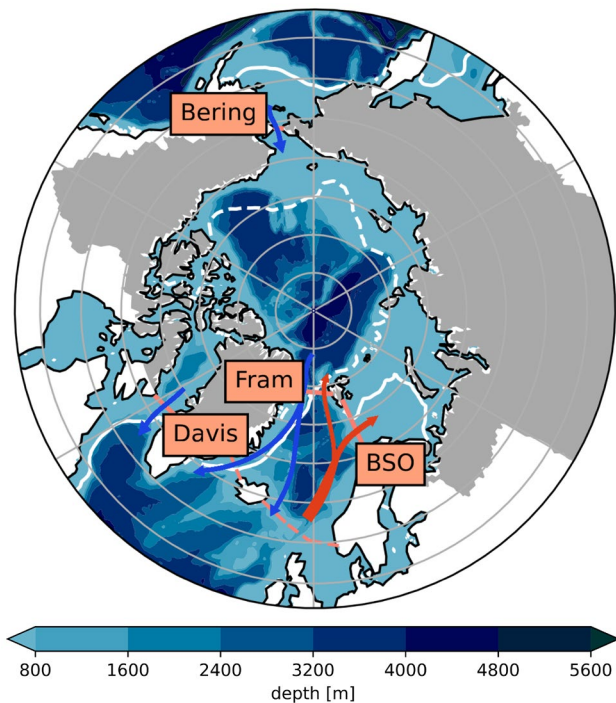


Fig. 1 Map of the main study area, consisting of the oceanic area bounded by the main Arctic gateways (indicated by solid orange lines; corresponds to $11.3 \times 10^6 \text{ km}^2$) and the terrestrial drainage area (grey shading; corresponds to $18.2 \times 10^6 \text{ km}^2$ for mainlands and islands and additional $0.95 \times 10^6 \text{ km}^2$ for Greenland). The orange dashed line indicates the position of the Greenland-Scotland Ridge, which bounds, together with Fram Strait and BSO, the region of the Nordic Seas. Additionally, the main currents flowing in and out of the Arctic (red and blue arrows for warm inflow and cold outflow, respectively) and 1993–2014 mean March (white solid) and September (white dashed) 30% sea ice concentration lines (taken from the GREP reanalyses ensemble) are shown. Shading in the oceanic areas indicates the bathymetry

approach. To avoid use of fluxes based on short-term forecasts from reanalyses, which are known to be biased (Trenberth et al. 2011), the net surface water flux (P-E) was derived from moisture flux divergence, which can be computed from analysed state quantities and thus is more strongly constrained by observations. We adapt results from Winkelbauer et al. (2022) to the 1993–2014 period and use them to validate seasonal cycles and trends of the freshwater input components R and P-E into the Arctic Ocean simulated by the CMIP6 models. As we also want to assess lateral oceanic transports through individual straits and for liquid water and sea ice separately, we additionally calculate oceanic transports directly from the Copernicus Marine Environment Monitoring Service (CMEMS) Global ocean Reanalysis Ensemble Product (GREP, Desportes et al. 2017; Storto et al. 2019a), an ensemble of four global ocean reanalyses: the CMCC Global Ocean Physical Reanalysis System (CGLORS, Storto and Masina 2016), the Forecasting Ocean

Assimilation Model (FOAM, MacLachlan et al. 2015), Global Ocean Reanalysis and Simulation Version 4 (GLO-RYS2V4, Garric et al. 2017) and Ocean Reanalysis System 5 (ORAS5, Zuo et al. 2015). The GREP ensemble members use the NEMO ocean model and are all run at $1/4^\circ$ horizontal resolution with 75 vertical levels. They all use the same atmospheric forcing (ERA-Interim Dee et al. 2011), however there are differences in the data assimilation methods, used observational products, the reanalysis initial states, NEMO versions, the sea ice models, physical and numerical parameterizations, and air-sea flux formulations. For further details, we refer to the individual data documentations and Storto et al. (2019a). Additionally, we look into an improved version of FOAM (GloRanV14, hereinafter called FOAMv2). Unlike the other reanalyses, FOAMv2 uses a non-linear free surface scheme (NLFS), which introduces some differences when looking into seasonal cycles of volume transports (see Section 4.1.1). Further, we use mooring-derived transports from the so-called ArcGate project (Tsubouchi et al. 2012, 2018), which are available from October 2004 to May 2010.

For the energy budget, we compare the CMIP6 output to results from Mayer et al. (2019), who provide a consistent, closed estimate of the seasonal cycle of the Arctic energy budget for the period 2005–2009 using observations and reanalyses and also a variational optimization approach. They calculate energy budget terms from Eqs. 1 and 2 using satellite observations, various reanalyses and ocean reanalyses as well as oceanic transport derived from moorings. As here we assess longer time periods, i.e. to reduce sampling uncertainties, we additionally calculate the major budget components using observations and reanalyses directly: Net TOA fluxes are compared with the DEEP-C dataset (Liu et al. 2020; Allan et al. 2014,; publicly available at <https://doi.org/10.17864/1947.271>), a backward extension of the net TOA fluxes from the Clouds and the Earth's Radiant Energy System-Energy Balanced and Filled (CERES-EBAF) satellite product in version 4.1 (Loeb et al. 2018), where fluxes prior to the CERES period have been reconstructed using satellite observations, atmospheric reanalysis and model simulations (Liu et al. 2020). F_s is compared with inferred net surface energy fluxes derived from mass-consistent energy budgets using ERA5 data (Mayer et al. 2021b). The snowfall term in Eqs. 1 and 2 as well as atmospheric transports (via the *vertical integral of divergence of total energy flux*) and the atmospheric tendency term (using central differences on the *vertical integral of total energy*) are additionally calculated using data from the ERA5 reanalyses (Hersbach et al. 2020). Energy tendency components OHCT and MET, as well as latent heat transports associated with sea ice transports (IHT) are estimated using the GREP reanalysis ensemble and oceanic transports of heat are calculated using GREP and mooring-derived transports from ArcGate. Additionally, we use the merged data product from CryoSat2 and

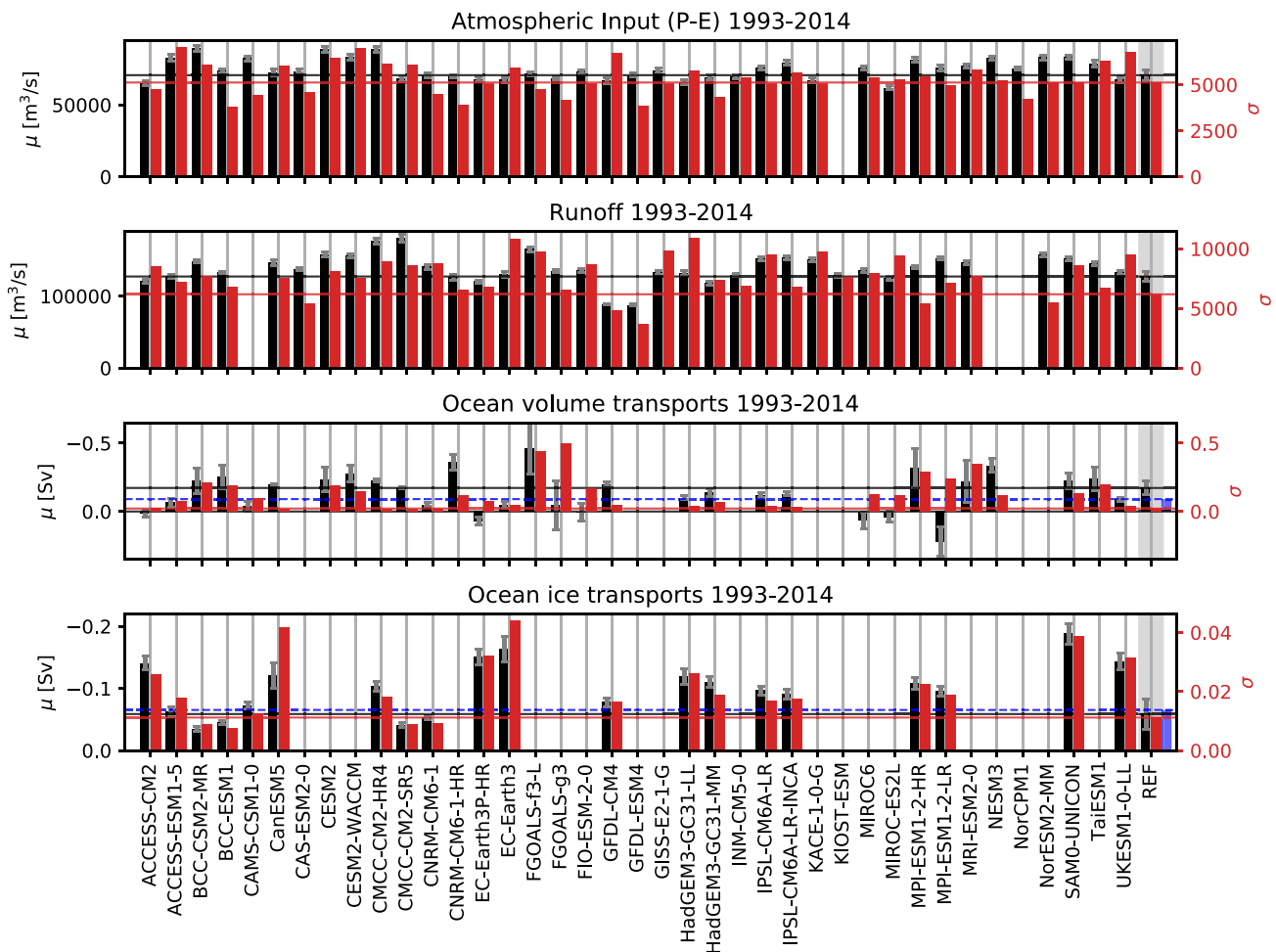


Fig. 2 Averages (black, left axis) and standard deviations of annual averages (red, right axis) for the major Arctic water budget components. Reference values (REF) for P-E and R are taken from Winkelbauer et al. (2022). They are indicated by horizontal lines and shown on the right hand side of the panels. For the oceanic transports REF

shows transports from the GREP reanalyses (1993–2014) and additionally also transports from the ArcGate project (2005–2010) are shown (blue bars and dashed lines). Error bars denote sampling errors for CMIP6 and errors calculated from the spread of used observational data for REF

the Soil Moisture and Ocean Salinity satellites (CS2SMOS Ricker et al. 2017), which has not been assimilated in the used ocean reanalyses, to validate sea ice thickness and MET data.

The datasets used to estimate the major energy and water budget components are summed up in Table 3. The calculation of reference uncertainties depends on the used data sources: for P-E and R we use uncertainties provided by Winkelbauer et al. (2022), for oceanic transports as well as OHCT and MET we use the spread of the GREP ensemble and the remaining uncertainties are based on the standard deviations of monthly mean values.

While Mayer et al. (2019) and Winkelbauer et al. (2022) provide closed budgets and therefore consistent estimates of the budget components, the various other, independent data products for some of the budget components (as described

above) are not expected to be fully consistent with each other and therefore budget closure for our observational reference estimates is not expected.

3.3 Study area

We consider the Arctic Ocean, which is bounded by hydrographic mooring lines in Fram Strait, Bering Strait, Davis Strait and the Barents Sea Opening (BSO). There are also two small passages, Fury and Hecla Straits, which connect the Arctic Ocean to Hudson Bay through the Canadian Arctic Archipelago (CAA). However, as Tsubouchi et al. (2012) and Bacon et al. (2022) pointed out, volume fluxes through these passages are very small and are not considered in this study. Figure 1 shows the study area, which was chosen to

Table 4 Long-term averages for the major water budget components

Units	P-E [10 ³ m ³ /s] ([frac/dec])	R	R _G	OVT [mSv = 10 ³ m ³ /s] ([frac/dec])	OIT
ACCESS-CM2	65.9 (0.04*)	122.9 (0.06*)	1.3 (0.12)	33 (− 0.80)	− 141 (0.13*)
ACCESS-ESM1-5	85.1 (0.01)	130.8 (0.01*)	7.6 (0.03)	− 66 (0.03)	− 63 (0.19*)
BCC-CSM2-MR	91.8 (0.03*)	149.4 (0.01)	9.4 (0.11*)	− 226 (0.08)	− 35 (0.28*)
BCC-ESM1	75.2 (0.01)	133.8 (0.03*)	9.5 (− 0.03)	− 255 (0.33)	− 45 (0.09)
CAMS-CSM1-0	82.5 (0.00)	—	—	− 37 (− 1.38*)	− 72 (0.0)
CanESM5	75.0 (0.01)	150.1 (0.03*)	8.2 (0.10)	− 194 (− 0.03)	− 120 (0.39*)
CAS-ESM2-0	74.7 (0.04*)	139.5 (0.00)	4.6 (0.07)	—	—
CESM2	89.8 (0.04*)	157.7 (0.02*)	4.7 (0.13)	− 234 (0.83*)	—
CESM2-WACCM	84.8 (0.03*)	159.4 (− 0.00)	4.5 (0.07)	− 277 (0.48*)	—
CMCC-CM2-HR4	91.6 (0.04*)	180.6 (0.01)	3.0 (0.13)	− 227 (0.02)	− 104 (0.01)
CMCC-CM2-SR5	69.3 (0.05*)	179.3 (0.01*)	4.4 (0.14)	− 174 (− 0.01)	− 41 (0.27*)
CNRM-CM6-1	70.9 (0.02)	145.2 (0.01)	9.7 (0.09)	− 50 (− 0.02)	− 53 (0.02)
CNRM-CM6-1-HR	71.2 (0.00)	129.4 (0.01*)	11.4 (0.00)	− 363 (0.21)	—
EC-Earth3P-HR	68.3 (0.00)	120.0 (0.03*)	1.5 (− 0.01)	70 (0.48)	− 151 (0.11*)
EC-Earth3	68.2 (− 0.00)	135.7 (0.07*)	1.8 (0.12)	− 48 (− 1.17)	− 163 (0.31*)
FGOALS-f3-L	72.8 (0.01)	170.0 (0.00)	2.5 (0.16)	− 465 (0.99*)	—
FGOALS-g3	70.2 (− 0.00)	136.7 (0.04*)	1.5 (0.09)	− 43 (1.31*)	—
FIO-ESM2-0	73.6 (0.02)	138 (0.02*)	1.9 (0.16)	6.6 (1.05)	—
GFDL-CM4	69.4 (0.08*)	89.0 (0.03*)	0.8 (0.03)	− 200 (− 0.15)	− 78 (0.18*)
GFDL-ESM4	72.2 (0.01*)	88.5 (− 0.01*)	0.4 (− 0.06)	—	—
GISS-E2-1-G	75.8 (− 0.02*)	136.0 (− 0.02*)	0.5 (0.01)	—	—
HadGEM3-GC31-LL	66.5 (0.04*)	134.6 (0.08*)	3.8 (0.27)	− 99 (− 0.44)	− 119 (0.31*)
HadGEM3-GC31-MM	69.5 (− 0.02*)	119.8 (0.06*)	2.4 (0.18)	− 139 (− 0.33)	− 111 (0.16*)
INM-CM5-0	71.1 (0.00*)	131.3 (0.00)	11.9 (0.11)	—	—
IPSL-CM6A-LR	75.9 (0.01*)	156.6 (0.05*)	12.9 (0.05)	− 120 (− 0.13)	− 97 (0.12*)
IPSL-CM6A-LR-INCA	80.9 (0.00)	154.4 (0.03*)	13.2 (− 0.00)	− 128 (− 0.09)	− 91 (0.07)
KACE-1-0-G	67.1 (0.00)	151.5 (0.01)	19.0 (0.03)	—	—
KIOST-ESM	—	132.6 (0.03*)	6.2 (0.03)	—	—
MIROC-ES2L	63.2 (0.07*)	127.5 (0.05*)	8.3 (0.04)	39 (1.76*)	—
MIROC6	76.7 (0.06*)	137.6 (0.03*)	9.9 (0.05)	64 (1.04*)	—
MPI-ESM1-2-HR	82.1 (− 0.02)	141.4 (0.02*)	0.3 (0.02)	− 316 (0.47*)	− 108 (0.22*)
MPI-ESM1-2-LR	76.4 (− 0.00)	153.8 (− 0.01*)	0.4 (0.04)	221 (− 0.53*)	− 96 (0.15*)
MRI-ESM2-0	78.0 (0.01)	149.2 (0.03*)	3.5 (0.30)	− 220 (− 1.26*)	—
NESM3	84.5 (0.02*)	—	—	− 336 (− 0.03)	—
NorCPM1	76.5 (0.04*)	—	—	—	—
NorESM2-MM	84.3 (0.07*)	158.8 (0.01*)	9.8 (0.05)	—	—
SAM0-UNICON	84.5 (0.04*)	154.8 (0.05*)	1.4 (− 0.06)	− 224 (− 0.14)	− 188 (0.15*)
TaiESM1	78.6 (− 0.00)	146.9 (0.03*)	8.0 (0.04)	− 238 (0.30*)	—
UKESM1-0-LL	68.4 (0.01)	135.5 (0.04*)	1.9 (0.37)	− 86 (− 0.33)	− 144 (0.27*)
MMM	75.9 (0.02*)	141.1 (0.02*)	5.6 (0.06)	− 145 (0.66)	− 101 (0.18*)
REF	69.2 ± 2.5 (0.02*)	127.0 ± 1.1 (0.02*)	11.9 ± 0.4 (0.13*)	− 151 ± 43 (− 0.00) / − 91 ^A	− 60 (0.19*) / − 65 ^A
Winkelbauer et al. (2022)	— " —	— " —	— " —	− 207	

The MMM is calculated using all available models and REF denotes the observation based reference values. Values in brackets show decadal relative trends (* are significant). Reference values for P-E, R and R_G are taken from Winkelbauer et al. (2022), while OVT and OIT are calculated using the GREP reanalyses (1993–2014) and ArcGate (2005–2010, denoted by ^A). OVT and OIT are defined as positive northward. Reference uncertainties (±) are taken from Winkelbauer et al. (2022) or calculated from the spread of the GREP ensemble

be the same as in Mayer et al. (2019) and Winkelbauer et al. (2022).

To analyse water entering the ocean from the surrounding land areas, we additionally introduce the terrestrial domain, which consists of all land areas draining into the Arctic Ocean, including the CAA as well as islands along the Eurasian coast. We use the catchments as defined by Winkelbauer et al. (2022) and use the same area for all models. The total oceanic and terrestrial areas are $11.3 \times 10^6 \text{ km}^2$ and $18.2 \times 10^6 \text{ km}^2$ respectively, and Greenland provides an additional terrestrial catchment area of $0.95 \times 10^6 \text{ km}^2$.

4 Results

4.1 Water budget

This section looks at the main components of the Arctic water budget. We assess their long-term averages, trends and seasonal cycles.

Figure 2 and Table 4 show long-term averages of surface fresh water flux (P-E) and runoff (R), as well as lateral oceanic fluxes of water volume and ice for the period 1993–2014 and compared with reference values from Winkelbauer et al. (2022). Figure 2 also shows standard deviations and values in brackets in Table 4 show decadal trends.

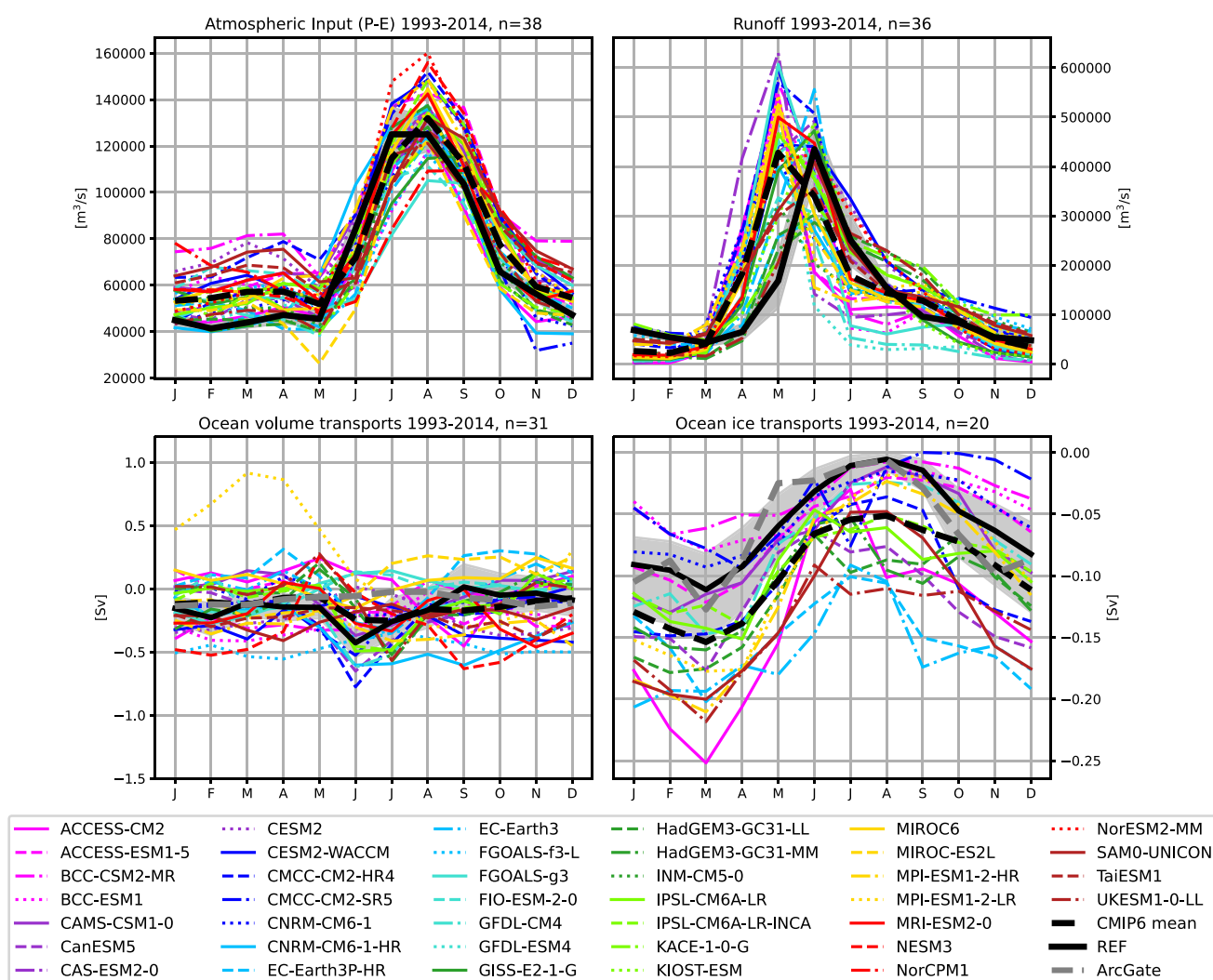


Fig. 3 Mean annual cycles of key terms of the oceanic Arctic water budget: **a** atmospheric freshwater input into the ocean (P-E), **b** runoff from Arctic lands (R), **c** oceanic volume transports and (d) oceanic ice transports across the main Arctic gateways. Ice exports from the ArcGate project are based on the PIOMAS reanalysis and not direct

measurements. Shading indicates the uncertainty range of the reference values and is either adopted from Winkelbauer et al. (2022) (top panels) or calculated from the spread (2σ) of the GREP ensemble (bottom panels)

Reference values indicate a long-term mean net freshwater input to the Arctic Ocean from surface fluxes of about $208 \times 10^3 \text{ m}^3 \text{ s}^{-1}$. About one-third comes from net precipitation ($69.2 \times 10^3 \text{ m}^3 \text{ s}^{-1}$), two-thirds from runoff from Arctic lands ($127.0 \times 10^3 \text{ m}^3 \text{ s}^{-1}$) and about 5% are melt water and ice discharge from the Greenlandic ice cap R_G ($11.9 \times 10^3 \text{ m}^3 \text{ s}^{-1}$). The MMMs for oceanic P-E (38 models) and R (36 models) are about 10 % higher than our observational references. Net precipitation ranges between $63.2 \times 10^3 \text{ m}^3 \text{ s}^{-1}$ and $91.8 \times 10^3 \text{ m}^3 \text{ s}^{-1}$, while runoff ranges between 88.5 and $180.6 \times 10^3 \text{ m}^3 \text{ s}^{-1}$, with the lowest values coming from GFDL-CM4 and GFDL-ESM4 and the highest values simulated by CMCC-CM2-SR5 and CMCC-CM2-HR4. Greenlandic runoff from CMIP6 models varies between 0.3 and $19.0 \times 10^3 \text{ m}^3 \text{ s}^{-1}$ and is underestimated by most models, with the MMM being about 50% smaller than the reference value. In contrast, the CMIP6 MMM of P-E over Greenland is about 10% higher than the reference estimate and there is a clear offset between runoff and P-E for most models. As soil moisture content and the surface snow amount do not change considerably (not shown) the mass balance over Greenland does not seem to be closed for the affected models. Possible reasons may include that the catchment area used, which is assumed to be the same for all models, might omit high runoff regions, that discharge coming directly from the ice sheet and/or solid discharge is underestimated or missing or, to a lesser degree, that the models feature conservation issues over Greenland. Further analyses would be needed to get to the origin of these discrepancies, which were not in the scope of this study.

Most models agree on an increase in freshwater input to the Arctic Ocean for the period 1993–2014: 25 models show a significant increase in R (only 3 show a significant decrease) and while all models agree on increasing precipitations and evaporations, trends in precipitation prevail in most models leading to significant positive trends in oceanic P-E for 19 models (only 2 show a significant decrease). The MMMs show an increase in oceanic P-E of 2% per decade and an increase in R of 2% per decade, which is in fairly good agreement with trends in the reference data. These increases in oceanic P-E and R contribute to an increase in liquid freshwater stored in the Arctic Ocean, which has been observed (Rabe et al. 2011; Proshutinsky et al. 2009; McPhee et al. 2009) and simulated by CMIP6 models (Zanowski et al. 2021; Wang et al. 2022), and further may lead to increased oceanic freshwater exports out of the Arctic system.

Reanalyses indicate a net outflow of liquid volume from the Arctic Ocean of $-151 \pm 43 \text{ mSv}$, while estimates derived from observation in the ArcGate project reach -91 mSv . Most CMIP6 models agree on an outflow of liquid volume out of the Arctic and the CMIP6 MMM stays within the reference estimates with -145 mSv , however

the inter-model variability is large. Some models significantly overestimate the net outflows (e.g., FGOALS-f3-L, MPI-ESM1-2-HR), while others indicate net inflows into the Arctic of up to 221 mSv (MPI-ESM1-2-LR). However, it has to be noted that diagnosed volume transports are very sensitive to the exact ocean bathymetry, where slight changes may lead to large deviations of multiple Sverdrups. As net Arctic volume transports are comparatively small values resulting from the sum of large in- and outflowing branches, small errors may lead to significant inconsistencies.

Ice volume transports have only been calculated for 20 models, with all models agreeing on an export of ice to the Atlantic. Ice transports vary between -188 and -35 mSv , with a MMM about 30% higher than our observational estimates ($-60 \pm 19 \text{ mSv}$ for GREP and -65 mSv for ArcGate).

Using those precisely calculated liquid and solid transports and taking into account all volume budget terms, we are still not able to close the simulated volume budgets for the individual models. Possible reasons for those shortcomings are discussed in Sect. 4.3.

While most models simulate an increase in liquid volume exports through the Fram Strait, an increase in imports through the Barents Sea opening and a decrease in exports through the Davis Strait (not shown, see e.g. Wang et al. 2022), the trends in net volume transports for the whole Arctic vary widely between models. For ice transports, the majority of models agree on a decrease in ice exports over the considered 22-year period, with significant trends between 11 and 39% per decade and a MMM trend of 18%. Long-term averages for volume transports and trends through the individual straits are shown in Table 6.

4.1.1 Long term mean seasonal cycles

Figure 3 shows the seasonal cycles of the main components of the Arctic water budget. Reference values (Winkelbauer et al. 2022) indicate a peak in net atmospheric freshwater input to the Arctic Ocean (P-E, Fig. 3a) from July to September and input minima during the cold season. The CMIP6 ensemble shows a large spread throughout the year. Most models are able to simulate the timing of P-E peaks and minima correctly, but tend to overestimate net P-E for most of the year (see also Table 1).

The annual cycles of terrestrial runoff are summarized in Fig. 3b). Observations show a strong runoff peak in June, mainly due to snowmelt and river ice break-up, and weak runoff during winter. CMIP6 models disagree on the timing of the runoff peak, with about two-thirds placing the runoff maximum in May. However, while observations are derived from gauge measurements at river mouths, the discharge estimates for CMIP6 are determined by calculating area integrals of runoff at each individual grid point over

the whole Arctic catchment. As we do not use any kind of river routing this may introduce an error in the runoff phase - especially for large catchments, routing can lead to delays of several months (Gosling and Arnell 2011). Hou et al. (2023) feed daily runoff outputs from 12 CMIP6 models into a state-of-the-art global river routing model to obtain discharge estimates at river gauges and compare the results with streamflow observations. In general, they find that models tend to perform better in non-cold regions than in cold environments. They find an early bias in the timing of the simulated maximum discharge for cold regions in most of the CMIP6 models evaluated. Therefore, in addition to differences in river routing, differences in runoff phase are most likely related to the ability of models to accurately simulate cryospheric hydrological processes such as snow and permafrost. Gosling and Arnell (2011) find that especially for catchments where the peak flow is strongly influenced by seasonal snowmelt (e.g. Ob and Mackenzie), models tend to overestimate the magnitude of the peak flow and show

an early bias of the seasonal peak flows. Kouki et al. (2022) analyse the seasonal snow cover for 33 CMIP6 models and find that the models generally overestimate the spring snowmelt rate, leading to early snowmelt. In addition, they found that the snow water equivalent is generally overestimated in winter, driven by precipitation biases and that, while temperature and precipitation can partly explain the biases in snowmelt, there may also be other contributing factors like inaccuracies in model parameterizations related to snow and the surface energy budget. The shift in the runoff phase also has implications for the seasonal cycles of terrestrial water storage. Wu et al. (2021) assess the annual cycles of terrestrial water storage for 25 CMIP6 models and find a shift in the phase of water storage for the four largest Arctic river basins compared to GRACE satellite data, with an earlier end of the recharge period and an earlier start of the discharge period, which is consistent with our results. Nevertheless, some models appear to get the timing of the runoff peak right (Fig. 3b), however whether this is caused by

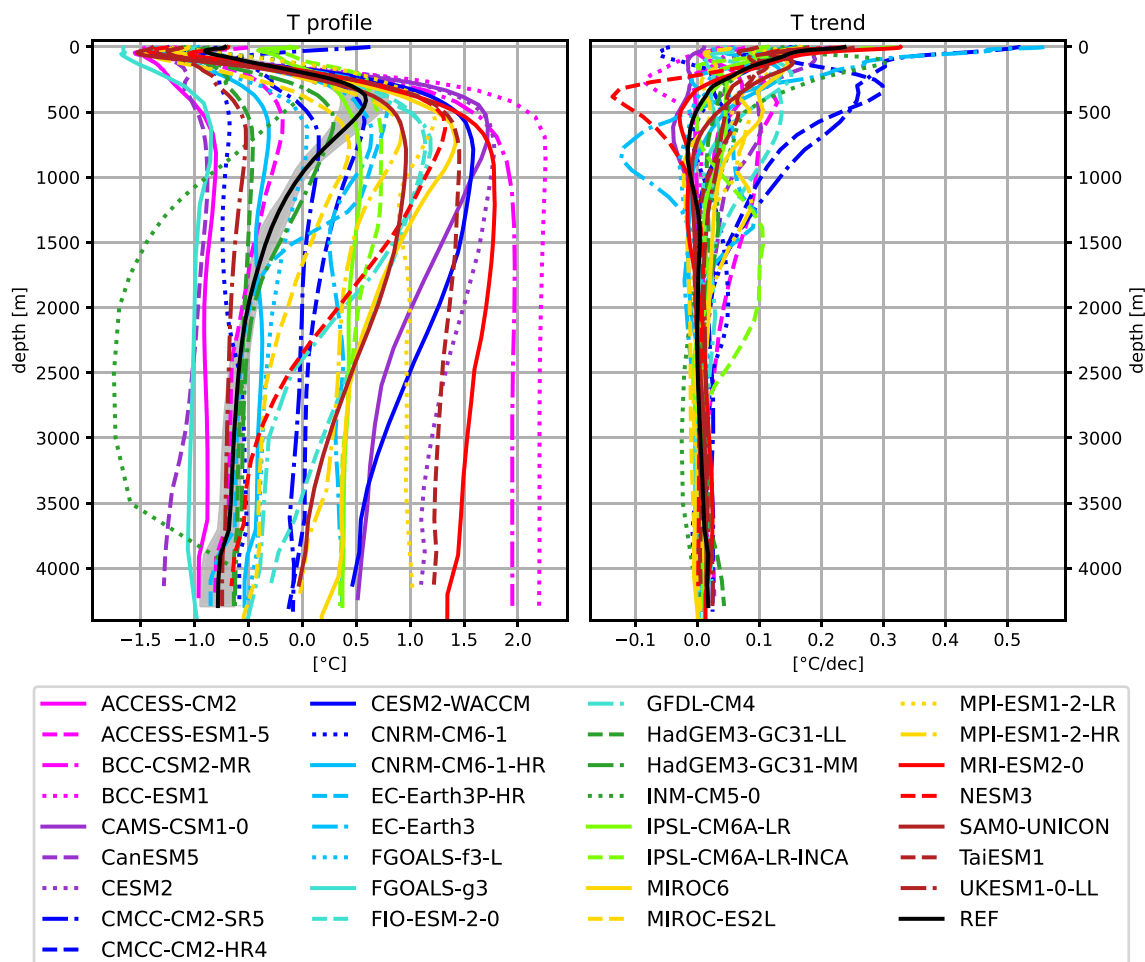


Fig. 4 Vertical profiles of area averaged Arctic temperatures and temperature trends over the 1993–2014 period. Reference values are taken from the GREP reanalyses. Shading indicates the spread (2σ) of the GREP ensemble

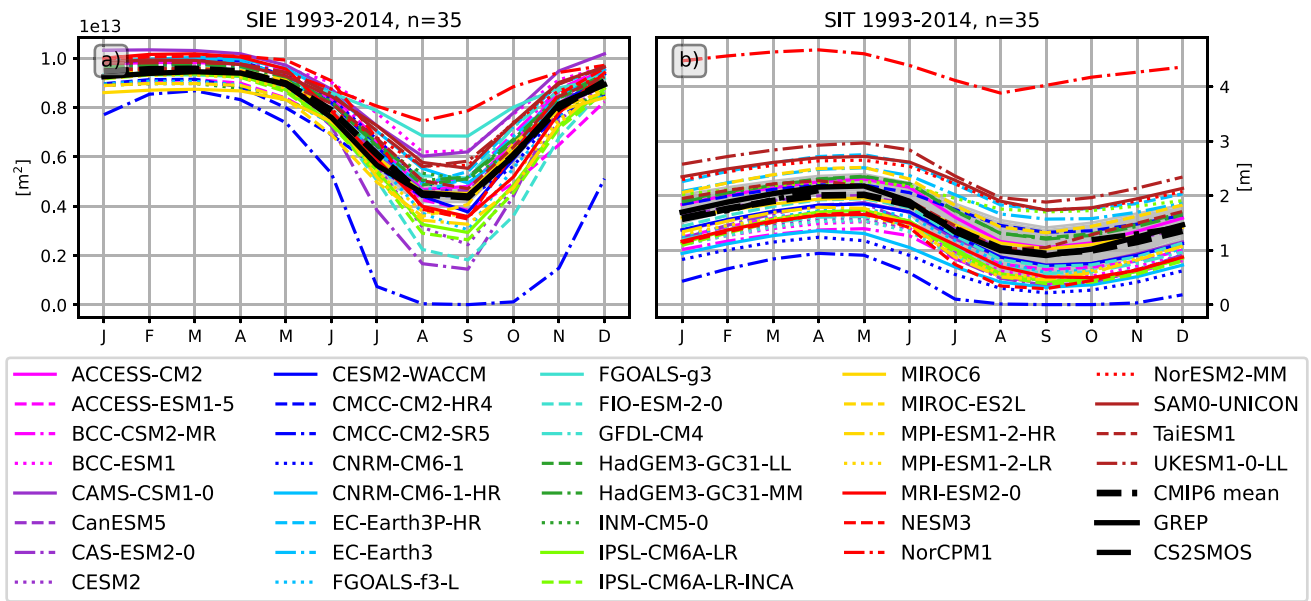


Fig. 5 Mean annual cycles of **a** the mean Arctic sea ice extent (SIE) and **b** the mean sea ice thickness for CMIP6 models and the GREP reanalysis. Additionally SIT estimates from CS2SMOS (10–2002–12–2014) are shown. Shading indicates the spread (2σ) of the GREP ensemble

an actual better representation of the cryospheric processes therein or whether they get the phase right for the wrong reason is not clear and would need further examination.

Figure 3c shows the seasonal cycles of oceanic volume transports through the main Arctic gateways. The GREP ocean reanalysis mean resembles the seasonal cycles of freshwater input to the ocean surface and shows an export maximum of 430 mSv in June, an almost instantaneous response of the ocean to surface freshwater input, as the ocean achieves mass adjustment within about a week through the generation of barotropic waves (Bacon et al. 2015). The observational ArcGate estimate does not show this peak in June, most likely because the mooring arrays are too sparse and the velocity field is not measured accurately enough (both in space and time) to resolve barotropic waves. Models using a non-linear free surface scheme (NLFS), where freshwater from sea ice melt is physically dumped into the ocean resulting in barotropic waves (Madec 2016; Roulet and Madec 2000), were corrected by subtracting the seasonal change in sea ice volume. Volume transports without the sea ice volume correction are shown in the Supplementary material (Fig. S1). The FOAMv2 reanalysis, which in contrast to the GREP ensemble also uses NLFS, as well as CMIP6 models with NLFS show much stronger amplitudes and summer peaks up to one order of magnitude larger than the other models and reanalyses. The effect of ice formation and growth on volume transport can be seen in the cold season, as freshwater is removed from the ocean, leading to a net import of water into the Arctic. However, this behaviour is physically not realistic, as in reality sea

ice melt and formation should not affect volume transports in and out of the Arctic. While the correction for net Arctic volume transports appears to be relatively straightforward, the correction for individual Arctic straits and heat transports is not as straightforward and is beyond the scope of this study. Unsurprisingly, the effect of the model mass adjustment of 2–3 Sv in summer is also visible to some extent in the heat transports and will be discussed in Sect. 4.2.1. Meanwhile, linear free surface models show smoother cycles with smaller amplitudes as salt is removed from the ocean during ice melt and added during ice growth to simulate brine rejection. After the correction (Fig. 3c) the seasonal cycles of the NLFS models appear smoother and the CMIP6 MMM stays within uncertainty bounds of our reference estimates during 9 out of 12 months. However, the inter-model spread is still large and there are also models showing some spurious patterns. For example, the MPI-ESM1-2-LR model features volume inflows of up to 900 mSv in spring. As mentioned above, due to the sensitivity of volume transports some errors may be a result of inaccuracies in the calculation process related to the ocean bathymetry. We discuss this further in Sect. 5 and Winkelbauer et al. (2023).

Annual cycles of oceanic ice transports are shown in Fig. 3d. Reanalyses (REF) and ArcGate estimates agree on the annual phase of ice export, with a maximum of ice export in March and a minimum from July to September. Of the 20 CMIP6 models used in this study, which provide all the necessary parameters to calculate ice transports, most agree with the observational estimates in terms of the timing of ice discharge, but differ widely in terms of magnitude. Ice

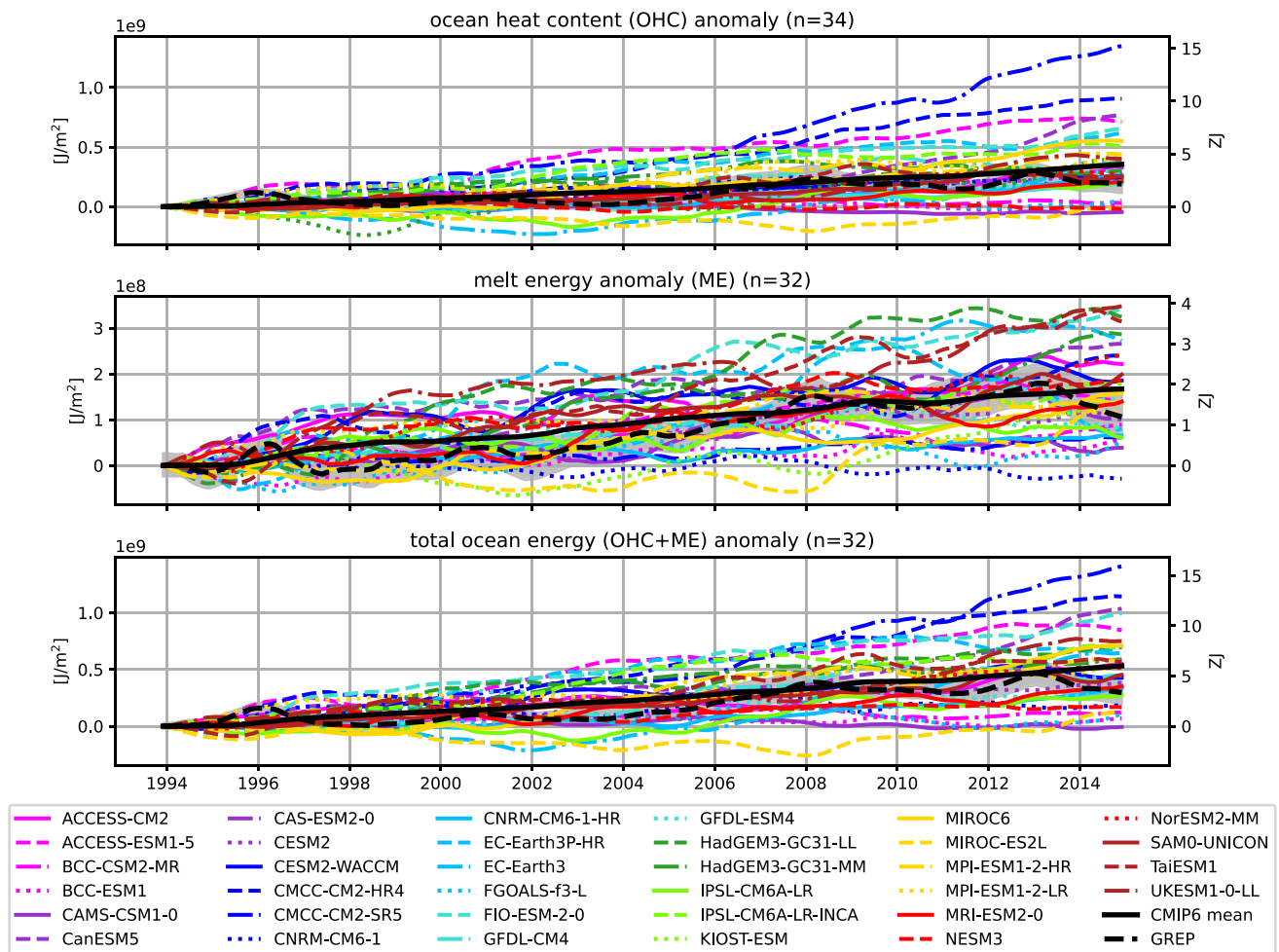


Fig. 6 Full-depth anomalous OHC (top) and ME (middle) accumulation as well as their sum (bottom) in the Arctic Ocean since 1993. Number of available CMIP6 models is given in the titles (n). Left axis

shows area-averaged changes in J/m² and the right axis shows area integrated changes in ZJ using a conversion factor of 11.3×10^{12} m². Shading indicates the spread (2σ) of the GREP ensemble

export maxima in March range from -250 mSv (ACCESS-CM2) to -70 mSv (BCC-CSM2-MR), with most models overestimating total sea ice export throughout the year.

4.2 Energy budget

In this section we will assess the key components of the coupled energy budget of the Arctic. We will start by looking at the mean state of the Arctic system and the accumulation of energy (enthalpy) in the Arctic ocean-sea ice system (heat content of the Arctic Ocean and the enthalpy due to sea ice melt).

Figure 4a shows depth profiles of average Arctic ocean temperatures integrated over our whole study area. The observed halocline lies in the uppermost 250 m, with the warmer and saltier Atlantic Water layer lying underneath. Vertical profiles from the GREP reanalyses (REF) are

quite consistent with observed profiles (see e.g. Khosravi et al. 2022) and show an Atlantic water core temperature of about 0.6–0.8 °C and a core depth of about 450 m. Temperature profiles from the CMIP6 models feature substantial biases, especially so in depths below 500 m. Consistent with Khosravi et al. (2022) and Heuzé et al. (2023), we find that CMIP6 models simulate the Atlantic layer too deep and too thick. Further, CMIP6 models feature a large inter-model spread of more than 3 °C for layers underneath the halocline.

Figure 5 shows the annual cycles of sea ice extent (SIE) and sea ice thickness (SIT). While the CMIP6 models again feature a large inter-model spread with obvious biases for several models, the MMM actually stays within the uncertainty bounds of our reference estimates for both SIE and SIT.

Figure 6 shows the heat accumulation in the Arctic since 1993. The starting year 1993 was chosen because of the

availability of our observational reference values. The top panel shows the increase in heat contained in the Arctic Ocean at full depth (OHC), as defined in Fig. 1. Ocean reanalyses show an increase of 0.2 GJ/m^2 (area integrated values may be calculated using the Arctic Ocean area of $11.3 \times 10^{12} \text{ m}^2$ and are provided on the right axes of Fig. 6). While most CMIP6 models agree on an increase in oceanic heat over the period 1993–2014, the amount of heat accumulation varies widely between models. Most models overestimate the heat accumulation, with CMCC-CM2-SR5 being the most extreme with an accumulation of 1.3 GJ/m^2 . The MMM of all 34 models is almost twice as high as the observational estimate, reaching 0.35 GJ/m^2 for the 22-year period. Three models (CAMS-CSM1-0, NESM3, MIROC-ES2L) show a slight decrease in oceanic heat storage over the 20-year period and another three models (BCC-ESM1, BCC-CSM2-MR, FGOALS-g3) show insignificantly small heat accumulations, about an order of magnitude smaller than the reference values.

The middle panel shows the accumulation of energy going into sea ice melt (ME). Ocean reanalyses show a heat accumulation of 0.1 GJ/m^2 over the 22-year period, about 57% less than the accumulated OHC change. All CMIP6 models (except one CNRM-CM6-1) agree on an increase of the ME over the last decades, but they again show a huge inter-model spread and range from a total accumulation of 0.0 to 0.4 GJ/m^2 . The MMM of all 32 models (0.2 GJ/m^2) is about 30% higher than indicated by the reanalyses, but remains within the uncertainty of the reanalysis ensemble. The total ocean energy accumulation (OHC+ME) is mostly dominated by ocean heat content and is shown in the bottom panel of 6.

For a deeper understanding of the OHC changes we assess the trends of Arctic Ocean temperatures with depth (Fig. 4b). Reanalyses reveal a strong increase of temperatures of about $0.25 \text{ }^\circ\text{C}$ per decade at the surface. Trends become weaker with depth and beneath about 500 m temperature changes become very small. CMIP6 models show quite diverse trends. While most models agree on an temperature increase at the surface, the strength of the trend ranges from close to zero up to an increase more than twice as high as shown by reanalyses. For the layers below the halocline models differ in terms of sign of the trend and trend strength. For the deep ocean all models agree on comparably small temperature changes. However, it has to be noted, that temperature trends are calculated over a 22-year period, a time-frame short enough that variabilities in inflowing Atlantic Waters may be of importance. Muilwijk et al. (2018) found that variabilities in northward ocean heat transports may impact temperature changes in the deeper Arctic Ocean, with prominent variability on perennial and decadal time scales as well as indicators of variability on multidecadal scales.

Nevertheless, for the 1993–2014 period the large OHC changes simulated by the CMCC-CM2-SR5 model are a result of strong temperature increases from the surface down to about 2000 m depth, while for instance the strong OHC changes for CMCC-CM2-HR4 are mainly driven by temperature changes in the depth of the Atlantic layer core. The NESM3 model, which simulates a slight decrease of heat accumulation, features plausible temperature trends at the surface, however those are compensated by a strong temperature decline around the Atlantic water core depth. For the other five models simulating either insignificantly small heat accumulations or even slight decreases,

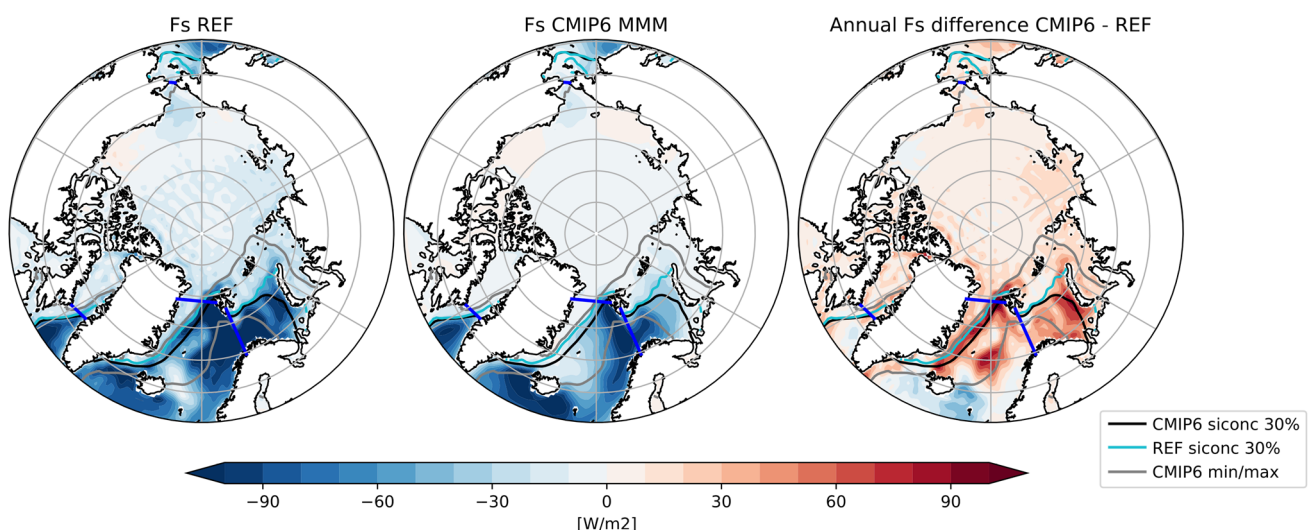


Fig. 7 1993–2014 average F_s for the observation based estimate (REF, left) and the CMIP6 MMM (middle) as well as their difference (right). 30% sea ice concentration lines are indicated in black, cyan and grey, borders of the study area are marked in blue

Table 5 Long-term averages for major energy budget components. The MMM is calculated using all available models and REF denotes the observation based reference values. Averaging periods are 1993–2014 for the MMM and REF and 2005–2009 for the Mayer et al. (2019) estimate. Reference values for OHT are taken from reanalyses and ArcGate (2005–2010, denoted by ^A). Oceanic transports are given in TW and may be converted to Wm^{-2} using an integration area of $11.3 \times 10^{12} \text{ m}^2$. Reference uncertainties (\pm) are based on the standard deviations of monthly mean values (F_S , F_{TOA} , AET) or calculated from the spread of the GREP ensemble (OHCT, MET, OHT)

Units	F_S [Wm^{-2}]	F_{TOA}	AET	OHCT	MET	OHT [TW]
ACCESS-CM2	−7.1	−109.3	0.1	0.4	0.3	87.2
ACCESS-ESM1-5	−9.7	−111.7	0.0	1.0	0.2	134.8
BCC-CSM2-MR	−2.8	−114.0	0.0	0.0	0.1	27.6
BCC-ESM1	−3.5	−112.5	−0.0	−0.0	0.1	23.6
CAMS-CSM1-0	−3.5	−110.0	0.0	−0.1	0.0	35.6
CanESM5	−8.0	−109.8	0.0	1.1	0.4	92.2
CAS-ESM2-0	0.1	−103.6	−0.0	-	0.2	-
CESM2	−6.3	−115.6	−0.1	0.4	0.2	68.0
CESM2-WACCM	−5.9	−118.4	0.0	0.3	0.3	60.4
CMCC-CM2-HR4	−14.6	−114.7	0.0	1.3	0.3	189.8
CMCC-CM2-SR5	−9.8	−98.0	0.0	1.9	0.1	144.6
CNRM-CM6-1	−8.1	−115.6	0.0	0.3	−0.0	100.7
CNRM-CM6-1-HR	−7.8	−117.1	−0.0	0.3	0.0	113.5
EC-Earth3P-HR	−9.5	−112.0	0.0	0.8	0.2	115.5
EC-Earth3	−9.8	−112.8	−0.0	0.6	0.3	103.2
FGOALS-f3-L	−5.3	−111.9	−0.0	0.1	0.1	64.4
FGOALS-g3	−5.4	−111.7	0.2	0.3	-	31.8
FIO-ESM2-0	−7.4	−104.7	−0.0	0.9	0.6	106.0
GFDL-CM4	−8.7	−112.9	−0.1	0.4	0.2	121.9
GFDL-ESM4	−7.7	−114.0	0.1	0.3	0.1	-
GISS-E2-1-G	−2.6	−113.3	−0.0	-	-	-
HadGEM3-GC31-LL	−9.9	−108.8	0.0	0.6	0.4	115.3
HadGEM3-GC31-MM	−9.8	−110.9	−0.0	0.4	0.4	122.5
INM-CM5-0	−10.1	−103.6	0.0	0.6	-	-
IPSL-CM6A-LR	−8.7	−114.9	0.0	0.4	0.2	110.6
IPSL-CM6A-LR-INCA	−8.0	−115.5	0.0	0.8	0.3	109.5
KACE-1-0-G	−12.8	−107.9	0.0	-	-	-
KIOST-ESM	-	−115.2	-	-	0.2	-
MIROC-ES2L	−6.7	−100.4	0.1	−0.0	0.2	94.1
MIROC6	−8.0	−109.6	0.0	0.8	0.3	119.3
MPI-ESM1-2-HR	−8.4	−115.7	0.1	0.6	0.2	102.4
MPI-ESM1-2-LR	−10.9	−115.3	0.0	0.6	0.2	124.3
MRI-ESM2-0	−6.1	−115.2	0.1	0.4	0.3	83.9
NESM3	−4.7	−112.6	0.0	−0.1	0.3	40.0
NorCPM1	−8.3	−111.9	−0.1	-	0.2	-
NorESM2-MM	−8.6	−116.3	0.0	0.4	0.4	-
SAM0-UNICON	−7.0	−111.8	−0.1	0.3	0.3	61.8
TaiESM1	−7.9	−113.0	0.1	0.3	0.5	78.8
UKESM1-0-LL	−8.3	−109.7	−0.0	0.5	0.6	93.0
MMM	−7.6	−111.6	0.0	0.5	0.2	92.9
REF	−18.0 ± 3.9	−116.7 ± 0.5	0.1 ± 0.9	0.3 ± 0.4	0.3 ± 0.1	126.7 ± 3.7 / 151.4 ^A
Mayer et al. (2019)	−16.2	−115.8	−0.1	0.3	0.4	175.2

temperature trends are rather small and partly negative already from the surface down. There are also some other spurious signals to be seen, for instance the EC-Earth3 model simulates net heat accumulations similar to the

observed values, however temperature trends at the surface are about twice as high as indicated by our observational reference and in return it features strongly negative temperature trends around 800 m depth. To assess the spurious

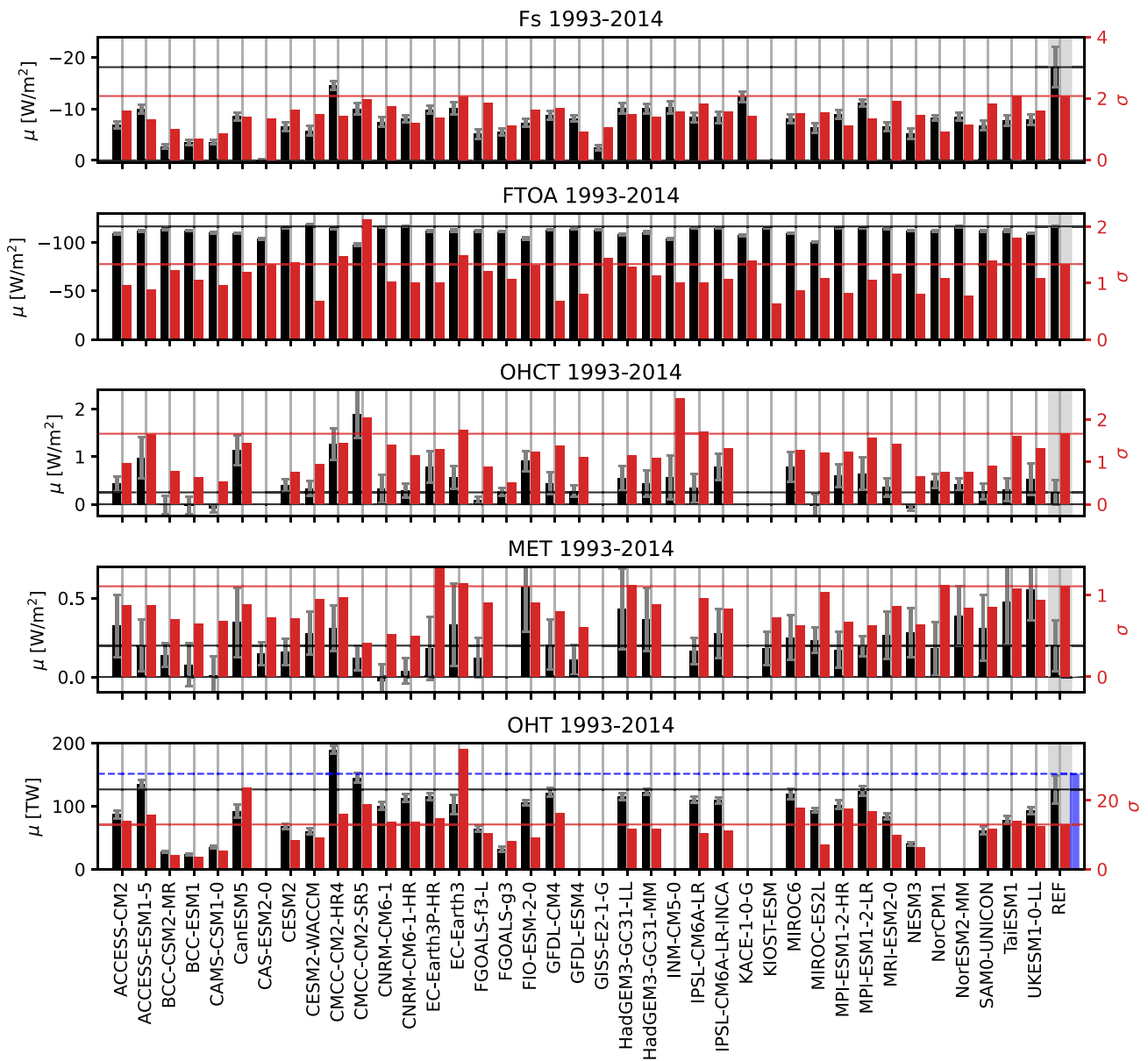


Fig. 8 Averages (black, left axis) and standard deviations of annual averages (red, right axis) for the major Arctic energy budget components. Reference values (REF) are indicated by horizontal lines and shown on the right hand side of the panels. They are taken from Mayer et al. (2021a), DEEPC, ERA5 and the GREP reanalyses

(1993–2014). Additionally also transports from the ArcGate project (2005–2010) are shown (blue bars and dashed lines). Error bars denote sampling errors for CMIP6 and errors calculated from the spread of used observational data for REF

trends more closely we looked at longer time periods and found some dubious jumps in the models' OHC time series and partly even changes in the sign of temperature trends when viewing other 22 year periods (not shown). This may indicate that equilibrium was not yet reached by the models and longer spin-up times may be required. We found no clear connection between temperature biases and the

strength of temperature trends. We additionally calculated some of the water and energy budget variables for a selection of intra-ensembles containing multiple members of the same models. While intra-model spreads are small for most variables, we found rather large ranges for OHC anomalies, with additionally strong variation from model to model, indicating model-dependence of simulated

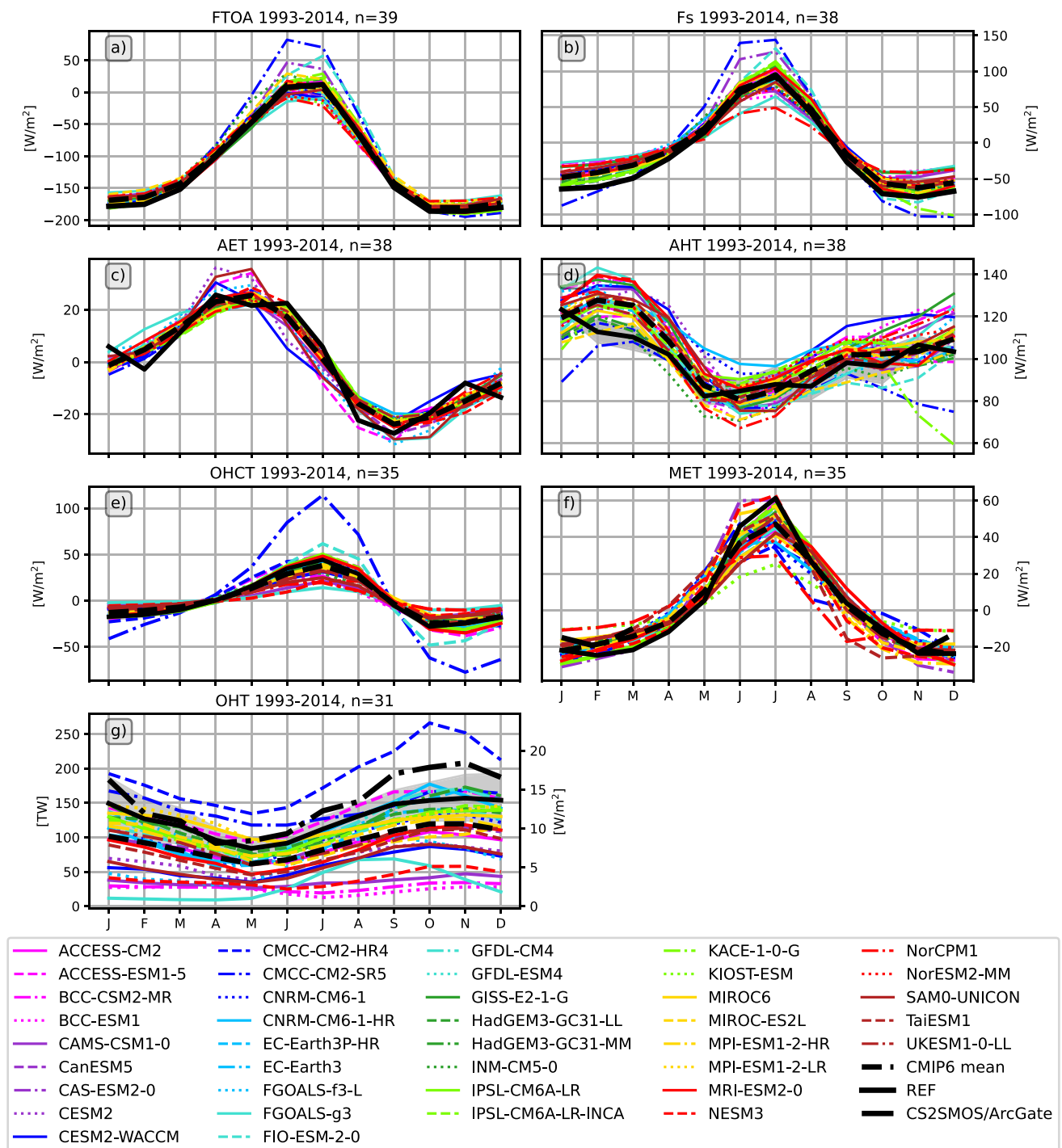


Fig. 9 Mean annual cycles of the key terms of the coupled Arctic energy budget: **a** net radiation at the top of the atmosphere F_{TOA} , **b** net vertical energy flux at the surface F_s , **c** atmospheric energy tendency AET, **d** atmospheric heat transport AHT, **e** full-depth ocean heat content tendency, **f** melt energy tendency (MET), and **g** the oceanic

heat transport across the main Arctic gateways. Shading indicates the uncertainty range of the reference values and is either based on the 2σ standard deviations of monthly mean values (F_s , F_{TOA} , AET, AHT) or calculated from the spread of the GREP ensemble (OHCT, MET, OHT)

internal variability. For instance, OHC anomalies at the end of 2014 for an ensemble of 11 CMCC-CM2-SR5 models range between 0.59 and 1.31 GJm^{-2} and for 11 CESM2 models only between 0.19 and 0.29 GJm^{-2} . The larger intra-model errors could again be a sign of internal variability or possible spin-up effects. Nevertheless, as errors estimated via our bootstrapping approach are of a similar or even higher value than those estimated from the model ensembles, we believe our uncertainty estimation to be valid.

The OHC and ME accumulations are converted into tendencies following Mayer et al. (2019) using the Theil-Sen trend estimator. Mean rates for 1993–2014 are given in Table 5. Reanalyses indicate a total ocean warming rate (OHCT+MET) of 0.4 Wm^{-2} for 1993–2014, of which about 40% is due to sea ice melting. The CMIP6 MMM shows a total warming rate of 0.7 Wm^{-2} , of which about one third is due to MET. The atmospheric warming rate (AET) is more than one order of magnitude smaller than OHCT and MET.

The CMIP6 models range between -0.1 and 0.1 Wm^{-2} (the exception being FGOALS-g3), with an MMM of 0.0 Wm^{-2} . Our reference estimate (ERA5) reaches $0.1 \pm 0.9 \text{ Wm}^{-2}$, while the estimate from Mayer et al. (2019) suggests -0.1 Wm^{-2} . As the latter was calculated only over the 2005–2009 period it may be affected by natural variability on various time scales, as AET is assumed to be positive but close to zero on longer time scales (von Schuckmann et al. 2020).

Table 5 also shows long-term averages of vertical and lateral energy fluxes into the Arctic and results suggest strong biases in several energy budget components. Satellite observations show a net radiation at TOA of $-116.7 \pm 1.2 \text{ Wm}^{-2}$ for the period 1993–2014 and for the area of interest. Most CMIP6 models show smaller fluxes, the whole ensemble ranging from -118.4 to -98.0 Wm^{-2} , with a MMM of -111.6 Wm^{-2} .

The net vertical energy flux at the ocean surface (F_s) from Mayer et al. (2021a) is $-18.0 \pm 2.1 \text{ Wm}^{-2}$ for the

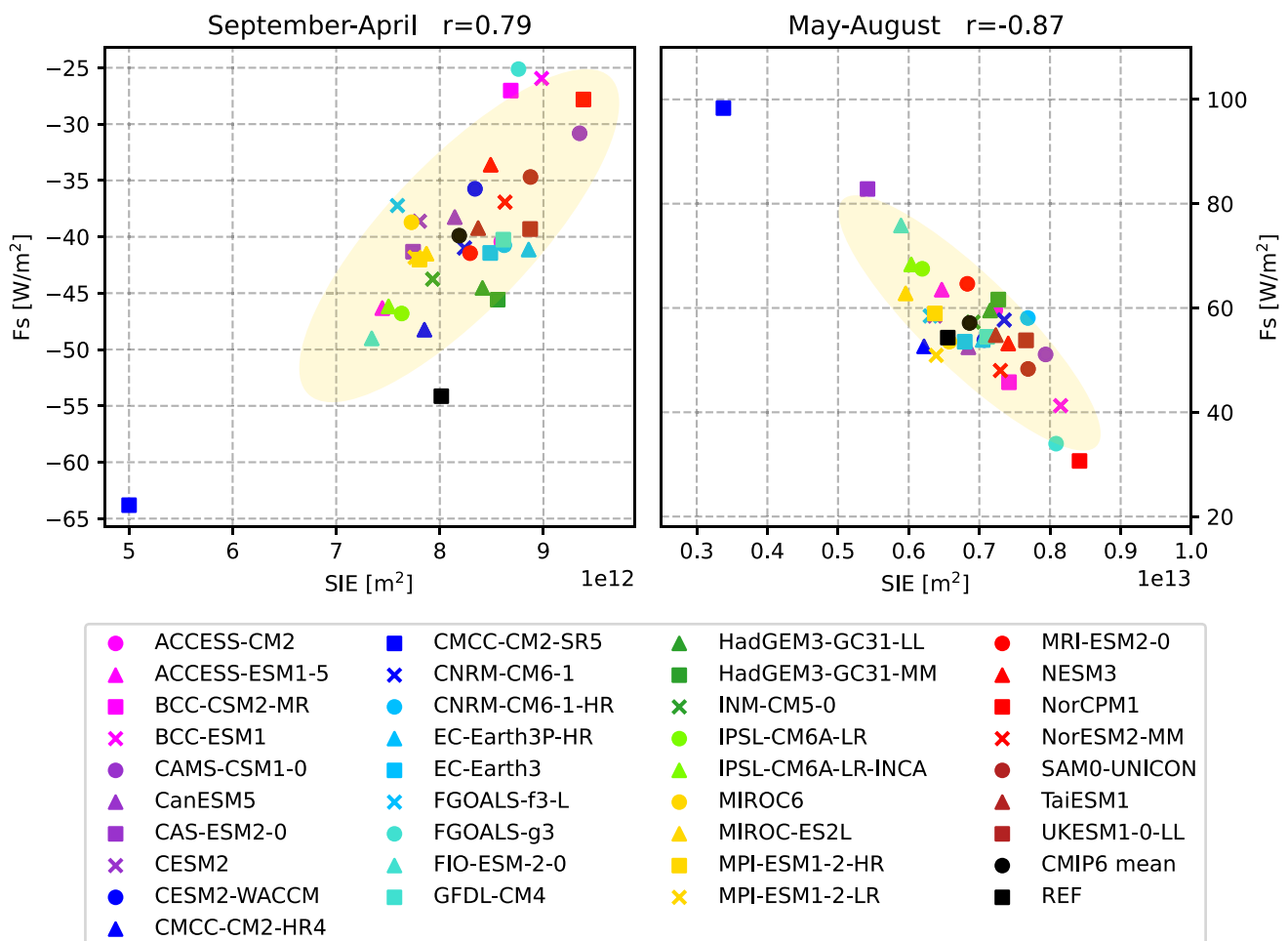


Fig. 10 Scatter plots between the net surface flux F_s and mean sea ice extent SIE averaged over 1993–2014. Left panel: September–April correlation, right panel: May–August correlation. Yellow ellipses show the 2-sigma confidence ellipses for the CMIP6 models

period 1993–2014, while Mayer et al. (2019) estimate a flux of -16.2 Wm^{-2} for 2005–2009. All CMIP6 models strongly underestimate the outgoing energy fluxes at the surface, ranging from -14.6 to -2.6 Wm^{-2} , with one model (CAS-ESM2-0) even showing a slightly positive annual F_S of 0.1 Wm^{-2} . Geographical maps of F_S for the individual models are not shown, but it should be noted that all models are able to simulate reasonable large-scale patterns, with low F_S values over sea ice and high values from the ocean in the Nordic Seas. However, net F_S in the Nordic Seas shows even larger biases with the CMIP6 MMM being about 30% lower than indicated by our reference (not shown). Figure 7 shows the long-term averaged F_S for our observational estimate (left panel), the CMIP6 MMM (middle panel) and their difference (right panel). Furthermore, 30 % sea ice concentration isolines are shown. Differences in F_S over the central, sea ice covered Arctic are small, with slightly higher values around the Kara, Laptev and Chukchi Seas. The largest differences in F_S occur near the sea ice edge between Greenland and Svalbard, in the Barents Sea and in the Norwegian Sea in proximity of the Lofoten Basin, with differences of up to 80 Wm^{-2} . The exact position of the sea ice edge in the Nordic Seas varies considerably between CMIP6 models (indicated by the grey lines in Fig. 7), with the MMM sea ice edge being positioned further south than the reference. Thus, most CMIP6 models simulate too little open water, resulting in smaller net outgoing energy fluxes. For the Labrador Sea and the Bering Sea, the sea ice concentration lines between our reference and CMIP6 are in good agreement and the differences in F_S are comparatively small. Apart from sea ice, the sea surface temperature has major effects on F_S . For example, F_S biases in Lofoten Basin are mainly caused by regional cold biases in the simulated sea surface temperatures (not shown).

The loss of energy to space over the Arctic is balanced by northward heat transports in atmosphere and ocean. The CMIP6 models show an enormous range of simulated oceanic heat transports, ranging from 20.30 to 189.82 TW (corresponding to a convergence of 1.80 to 18.80 Wm^{-2}), with a MMM of 93.3 TW (8.26 Wm^{-2}). For the same period, reanalyses indicate a long-term average heat flux of 126.7 TW for 1993–2014. Observational estimates (Tsubouchi et al. 2012) are only available for the period 10/2004–05/2010, but they show an even higher heat transport of 151.4 TW (13.40 Wm^{-2}). For the same period, the reanalysis is 136.2 TW (12.05 Wm^{-2}) and the CMIP6 MMM is 98.1 TW (8.68 Wm^{-2}), clearly underestimating the lateral energy input. Table 5 shows that while most models underestimate the reference value, there are 6 models in particular (BCC-CSM2-MR, BCC-ESM1, CAMS-CSM1-0, FGOALS-g3, FGOALS-f3-L and NESM3) that

have exceptionally low transports, with values more than 50% lower than our reference estimates. Some of these use the same ocean model component, BCC-CSM2-MR, BCC-ESM1 and CAMS-CSM1-0 use MOM4, while FGOALS-g3 and FGOALS-f3-L use LICOM3.0. Therefore, it would be a useful step to scale the models in terms of their independence through appropriate weighting algorithms to obtain reliable MMM.

Figure 8 shows long-term averages of the major energy fluxes and tendencies for all models and the reference-based estimates, whereby especially the models' biases in F_S and OHT stand out. Additionally, standard deviations of annual averages and sampling errors are shown. The large sampling errors for OHCT and MET highlight the high temporal variabilities in those variables, which, as discussed above, may indicate possible residual spin-up effects.

4.2.1 Long term mean seasonal cycles

Figure 9 shows the mean annual cycles of the main energy budget terms in Eqs. 1 and 2. Averaging periods depend on the availability of reference data and are indicated in the figure titles. In general, most models are able to simulate the general shape of the annual cycles accurately, but there are also some obvious biases and differences, which are discussed in more detail below.

The net radiation at TOA is shown in Fig. 9 a. It is strongly negative for most of the year and only slightly positive in June and July. This strong seasonal cycle is mainly driven by solar radiation. The spread (max-min) between CMIP6 models is relatively small in winter and the transition seasons, reaching a maximum of 95 Wm^{-2} in summer. A few models reach unrealistically high values during summer, in particular the CMCC-CM2-SR5 model shows a maximum of more than 80 Wm^{-2} (compared to 12 Wm^{-2} from observations), mainly due to strongly underestimated reflected shortwave radiation as a consequence of low sea ice biases (Fig. S2 in Supplementary material). About 20% of the models (8 out of 39) simulate negative F_{TOA} throughout the year. Inter-model spread is higher during summer, nevertheless the CMIP6 MMM is in quite good agreement with the observational estimate (DEEPC) and stays within the observational uncertainty bounds during those months. In winter, the inter-model spread is smaller, but most models underestimate the strong winter minima and the MMM is up to 10 Wm^{-2} lower than indicated by observations. The net radiation at TOA is an important driver of the annual cycle of the surface energy flux F_S , so F_S shows a similarly strong annual cycle. F_S remains negative (outgoing) during winter, and with the maximum of incoming shortwave radiation in May, F_S becomes positive and reaches its maximum in summer as sea ice melt progresses. Similar to F_{TOA} , some models

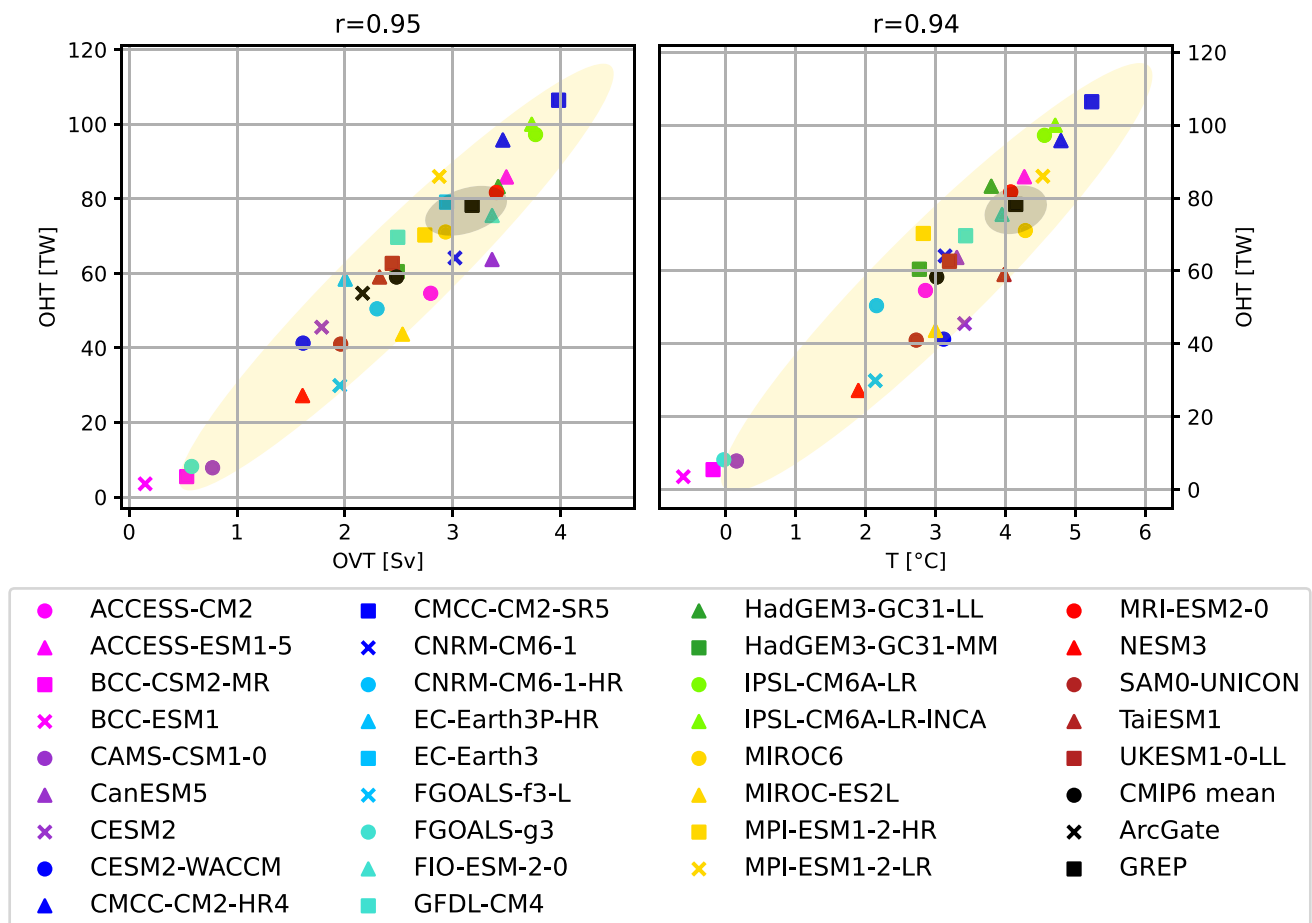


Fig. 11 Barents Sea Opening correlations of long-term annual averaged ocean heat transports (OHT) and ocean volume transports (OVT, left panel) as well as BSO average temperatures (right panel) for various CMIP6 models (1993–2014), the GREP reanalyses mean

(1993–2014) and ArcGate observations (2005–2010). Positive values denote transports into the Arctic. Yellow ellipses show the 2-sigma confidence ellipses for the CMIP6 models and grey ellipses for the GREP reanalyses

have unrealistically high summer maxima (CMCC-CM2-SR5, CAS-ESM2-0, FIO-ESM-2-0), caused by underestimated reflected shortwave radiation due to too little sea ice (not shown). The CMCC-CM2-SR5 model strongly underestimates both the total extent of sea ice in the area of interest and the thickness of the sea ice cover (Fig. 5), and in the summer months CMCC-CM2-SR5 even simulates an ice-free Arctic. CAS-ESM2-0 and FIO-ESM-2-0 are also at the lower end of the SIE ensemble, while models with high SIE during the summer months (e.g. NorCPM1, FGOALS-g3) also simulate low F_S during the summer months. In winter, most models simulate lower net upward F_S than our reference estimate (inferred F_S , Mayer et al. 2021a). However, CMCC-CM2-SR5 overestimates the winter minima because the SIE is quite small in winter, leading to unrealistically strong outgoing longwave radiation and latent heat fluxes (not shown). Figure 10 shows scatter plots between long-term average F_S and SIE. The correlations are divided into the season with negative net F_S (September - April) and the

season with positive net F_S (May - August). The correlations are high throughout the year. In summer, when incoming solar radiation is high, models with little sea ice simulate higher incoming net radiations, mostly caused by reduced reflected shortwave radiations (not shown). In autumn and winter, when the incoming solar radiation is low to non-existent, models with less sea ice simulate higher outgoing longwave radiations (not shown) and therefore lead to higher negative net radiations.

Figure 9c shows that the annual cycle of the atmospheric energy storage component AET is moderate compared to the other atmospheric components, and that CMIP6 models reproduce the observed cycles (ERA5) quite well. Atmospheric energy transport (AHT) for CMIP6 is estimated as residual using Eq. 1. Inter-model spread is relatively high throughout the year with most CMIP6 models simulating higher transports than indicated by our observational reference (ERA5). Biases are strongest from late autumn to early spring and are connected to biases in surface energy

fluxes and therefore biases in the position of the sea ice edge as well as sea surface temperatures. However, in summer, where biases in F_S and F_{TOA} are at their peaks in some models (e.g., CMCC-CM2-SR5, CAS-ESM2-0 and NorCPM1), compensating effects lead to smaller biases in AHT. Meanwhile, biases in AET play a less prominent role and adjust the total AHT biases with smaller reinforcing and compensating effects.

Figure 9e shows the annual cycles of the oceanic storage component OHCT. The models agree on ocean warming in the summer months and ocean cooling in the winter, but the amplitude of the cycles varies considerably. The model scatter is large for most of the year, with 95% of the models within $12\text{--}43\text{ Wm}^{-2}$ in summer and -24 and -4 Wm^{-2} in winter. The most obvious exception is CMCC-CM2-SR5, which has summer maxima about three times higher than the MMM and winter minima about three times lower. These large variations are again closely related to the underestimation of sea ice in CMCC-CM2-SR5. An amplification of the OHCT seasonal cycle with declining sea ice is expected,

and in fact has already been observed over recent decades (Mayer et al. 2016), but of course to a much lesser degree compared to CMCC-CM2-SR5. The annual cycles of the melt tendency (MET) are shown in figure 9f. The reference values are calculated from the GREP ensemble. The majority of models simulate the phase of the annual MET cycle correctly, but the inter-model spread is large throughout the year. Winter values range from -31 to -10 Wm^{-2} and summer peaks are between 16 and 52 Wm^{-2} . The MMM amplitude is generally lower than the reference estimate, with weaker freezing in late winter and early spring and weaker melting during the summer months.

The annual cycles of the net oceanic heat transports are shown in Fig. 9g. The large inter-model variability is evident throughout the year, but all models agree on an inflow of heat to the Arctic in all calendar months. Most models are able to simulate the timing of the inflow extremes correctly, with a minimum in May and a maximum in late autumn and early winter. Reference values (REF) are derived from the GREP ocean reanalysis ensemble, and the observational

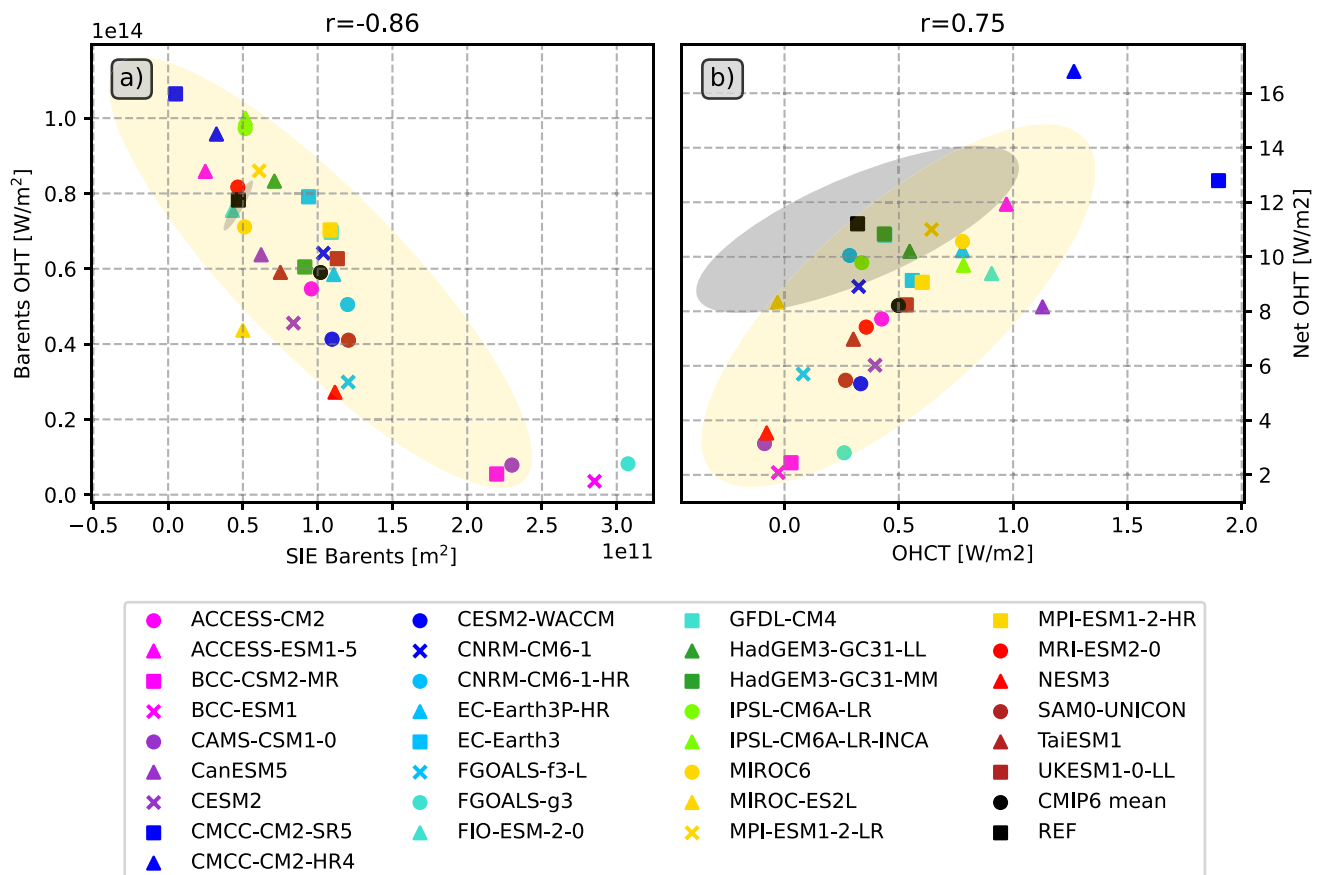


Fig. 12 Scatter plots of the effect of ocean heat transports on sea ice and the ocean warming rate. a) correlations between oceanic heat transports through the Barents Sea Opening and the mean sea ice extent in the Barents Sea, b) correlations between net Arctic oceanic

heat transports and the oceanic heat content tendency. All values are long-term annual averages over the 1993–2014 period. Yellow ellipses show the 2-sigma confidence ellipses for the CMIP6 models and grey ellipses for the GREP reanalyses

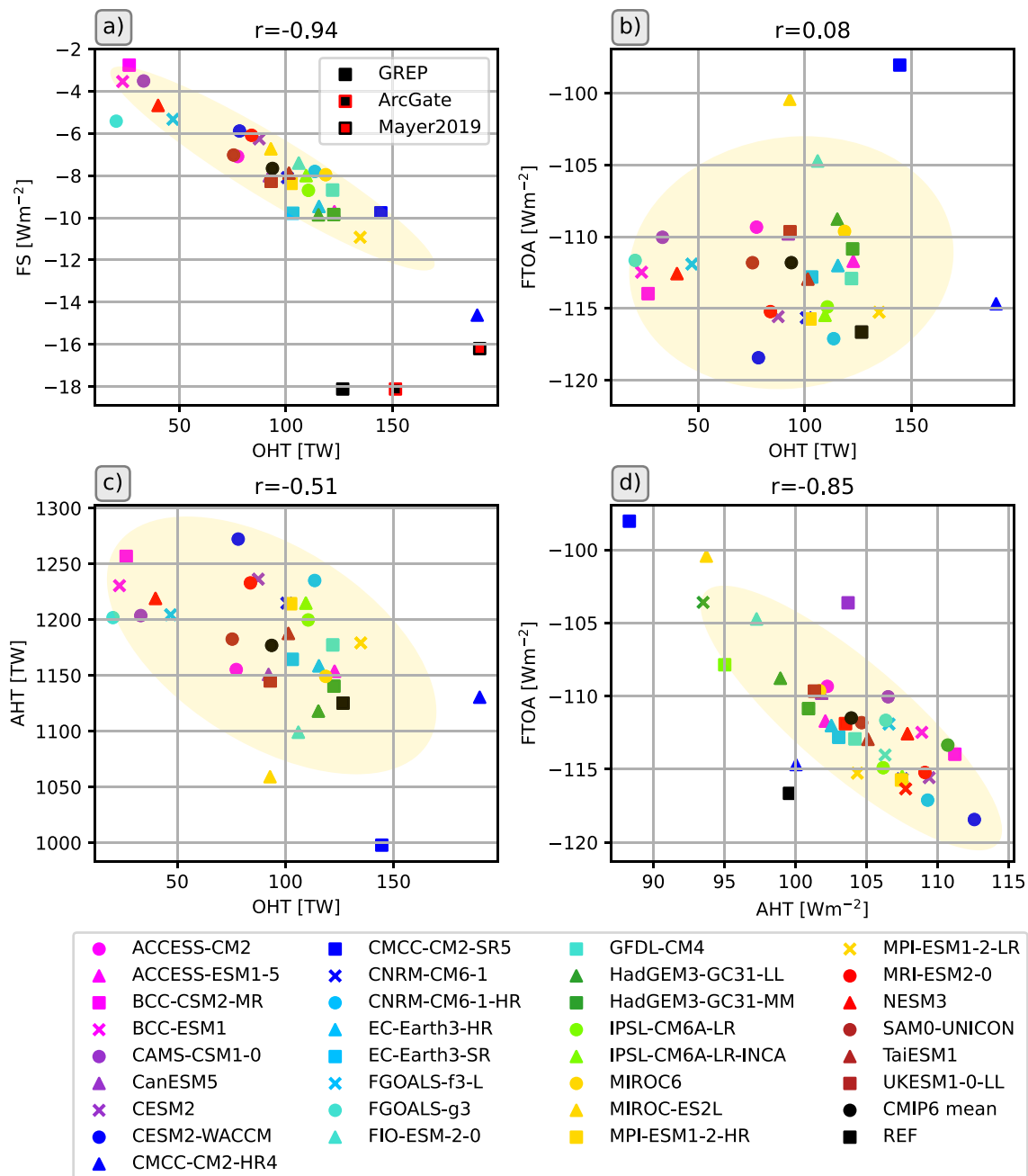


Fig. 13 Scatter plots of long-term annual averages of oceanic heat transports and atmospheric energy budget components. Correlations between **a** OHT and the net surface energy flux F_s , **b** OHT and the

net energy flux at the top of the atmosphere F_{TOA} , **c** OHT and atmospheric heat transports AHT, **d** AHT and F_{TOA} . Yellow ellipses show the 2-sigma confidence ellipses for the CMIP6 models

annual cycle from ArcGate is also shown. Almost all models underestimate the heat influx compared to REF and ArcGate. Only CMCC-CM2-HR4 simulates larger heat transports throughout the year than indicated by observations, with an October peak about 25% higher than the ArcGate estimate of about 200 TW. In addition, CMCC-CM2-SR5, MPI-ESM1-2-HR and MPI-ESM1-2-LR exceed observations in spring. The heat transports for BCC-CSM2-MR, BCC-ESM1,

CAMS-CSM1-0, FGOALS-g3, FGOALS-f3-L and NESM3 are too low in all calendar months and are mainly caused by biases in the inflow of Atlantic waters through the Barents Sea opening. Heat transports through the individual Arctic straits are shown in the Supplementary material (Fig. S3). In general, while most models are able to simulate the shape of the annual transport cycles to some extent, the inter-model spread is large for all Arctic straits. Seasonal cycles for the

BSO feature similar spreads and biases as the net Arctic heat inflow, reflecting the leading role of BSO in determining the amount of oceanic heat entering the central Arctic. Figure 11 shows correlations between BSO heat transports with BSO volume transports and BSO average ocean temperatures. Correlations are high both for volume transports and temperatures indicating biases in the simulated temperatures and currents. It is worth noting that volume transports and strait average temperatures are not independent of each other and feature moderate to high correlations (not shown). The models with exceptionally low OHT values show mean temperatures around 0 degrees Celsius or even slightly negative values. Additionally, they simulate Norwegian Coastal Currents (NCs) that are generally too weak, slowed down too far south, or even negative, while high OHT values are driven by high volume transports due to strong NCs and higher temperatures. Figure 4 revealed that some of the models (including BCC-CSM2-MR, BCC-ESM1 and CAMS-CSM1-0) feature large positive temperature biases in and underneath the Atlantic water layer. However, in the upper most layers some of those models show negative biases and underestimate the actual temperatures in the surface and halocline layers. Figure S4 shows temperature profiles averaged along the individual straits. The BSO profile reveals that while the largest temperature biases for BCC-CSM2-MR, BCC-ESM1, CAMS-CSM1-0 and FGOALS-g3 are found at the surface, where three of the models even simulate temperatures below 0°C, negative temperature biases are present in all layers of the rather shallow BSO. These low temperatures near the surface are tightly coupled to the overlying sea ice cover. However, the sea ice cover does not only affect heat transport, but the link between oceanic transports and sea ice goes both ways, as increased heat transports also lead to less sea ice (Årthun et al. 2019).

Heat transports through Fram Strait (Fig. S3a) are too small in the majority of models, but it is possible that models with the NLFS scheme are affected by the mass adjustment due to sea ice melt, as the differences in volume transports between models with non-linear and linear surfaces are largest for Fram Strait (not shown), leading to a simulated maximum volume outflow in summer for NLFS models and a minimum about 2 Sv smaller for those without NLFS. Volume transports through Fram Strait are generally biased low in CMIP6 (Heuzé et al. 2023). However, while net transports show a large spread similar to the BSO (Tab. 6), the MMM actually stays well within the uncertainty range of the reference values. Temperature profiles (Fig. S4) at Fram strait show that virtually all models feature a positive temperature bias below 500 m and above that the majority of models features a negative temperature biases. Therefore, the low biases in OHT are mainly caused by warm biased deep waters flowing out of the Arctic through the East

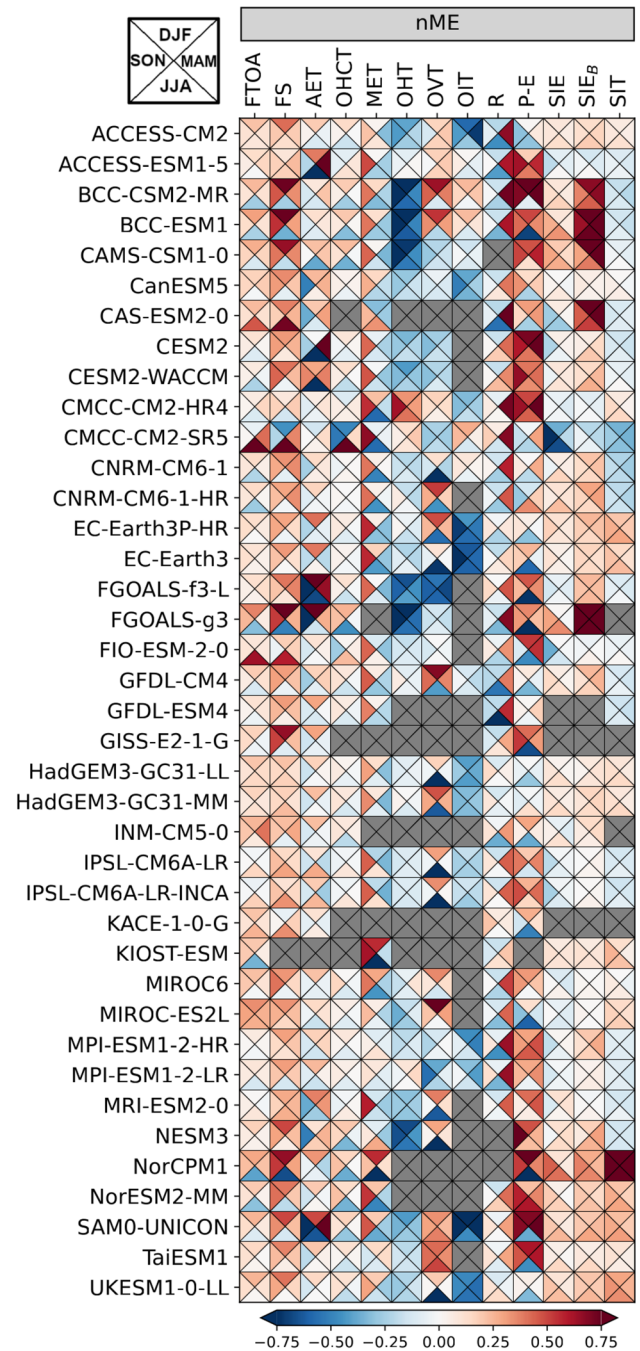


Fig. 14 “Portrait” diagram of seasonal normalized mean errors (nME) for various water and energy budget components for 1993–2014. Triangles indicate the respective seasons DJF (upper triangle), MAM (right triangle), JJA (bottom triangle), and SON (left triangle). SIE_B denotes the sea ice extent in the BSO

Greenlandic Current (EGC) and cold biased waters flowing into the Arctic via the more shallow West Spitsbergen Current (WSC).

The effect of model spatial resolution can be seen for heat transports through Davis Strait (Fig. S3c). All CMIP6 models with a horizontal resolution of 1/4 degree simulate an OHT peak in autumn, similar to the observational ArcGate estimate, while coarser resolution models do not show such a peak. Somewhat surprisingly, the reanalysis-based estimates, which also feature a horizontal resolution of 1/4 degree, do not simulate such a peak, however they are known to have a cold bias in the West Greenland Current (Pietschnig et al. 2017). The high resolution CMIP6 models however feature stronger and warmer West Greenlandic Currents and stronger, but similarly tempered, Baffin Island currents during autumn (see Fig. S5).

The strength of OHT has important implications for the state of the Arctic Ocean and sea ice. Figure 12 shows scatter plots of OHT, sea ice extent and the ocean warming rate OHCT. As mentioned above there is a tight coupling between sea ice and heat transports. The left panel shows this correlation for the BSO, as models with higher/lower heat transports simulate smaller/larger sea ice areas. This leaves two possibilities: either reduced OHTs allow more sea ice to form, or a larger sea ice cover slows down currents, cools the ocean and therefore leads to lower heat transports.

While the effect of OHT on Arctic sea ice has been discussed in various observational (e.g., Årthun et al. 2012; Onarheim and Årthun 2017) and modelling (e.g., Årthun et al. 2019; Dörr et al. 2021) studies, the influences of changes in Arctic sea ice on oceanic circulations, temperatures and therefore heat transports have been less investigated and still pose many unknowns (Docquier and Koenigk 2021). More thorough analysis and model experiments would be required to clarify this possible bidirectional effect, but, this is beyond the scope of this study.

Ocean heat transports also affect the change in oceanic temperature. Figure 12 b) shows the correlation of heat transports and the change in ocean heat content: models with larger/smaller OHT show a faster/slower warming of the Arctic Ocean.

The consequences of biases in the oceanic components may also pass over to the Arctic atmosphere, potential effects are shown in Fig. 13. There are strong correlations between simulated long-term averaged OHT and F_s (Fig. 13a), as OHT driven changes of sea ice and ocean temperature strongly affect the reflected shortwave radiation during summer and outgoing longwave radiations as well as turbulent energy fluxes. However, there are no significant correlations

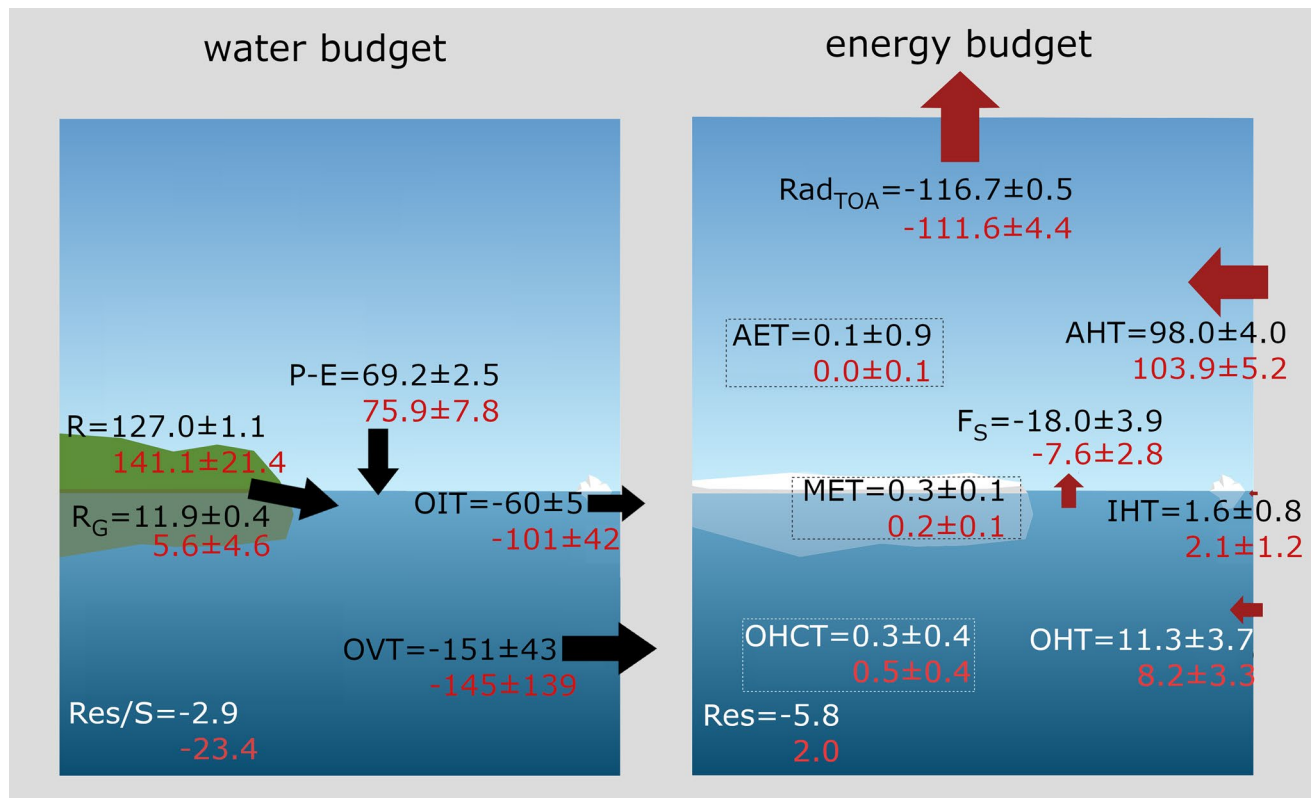


Fig. 15 Water (left, in 10³ m³/s) and energy (right, in Wm⁻²) fluxes and storage rates for the reference estimates (black) and the CMIP6 MMM (red) for 1993–2014. Additional estimates from ArcGate and

Winkelbauer et al. (2022) as well as Mayer et al. (2019) are given in Table 7. The graphic designs of the schematics are adapted from Winkelbauer et al. (2022) and Mayer et al. (2019)

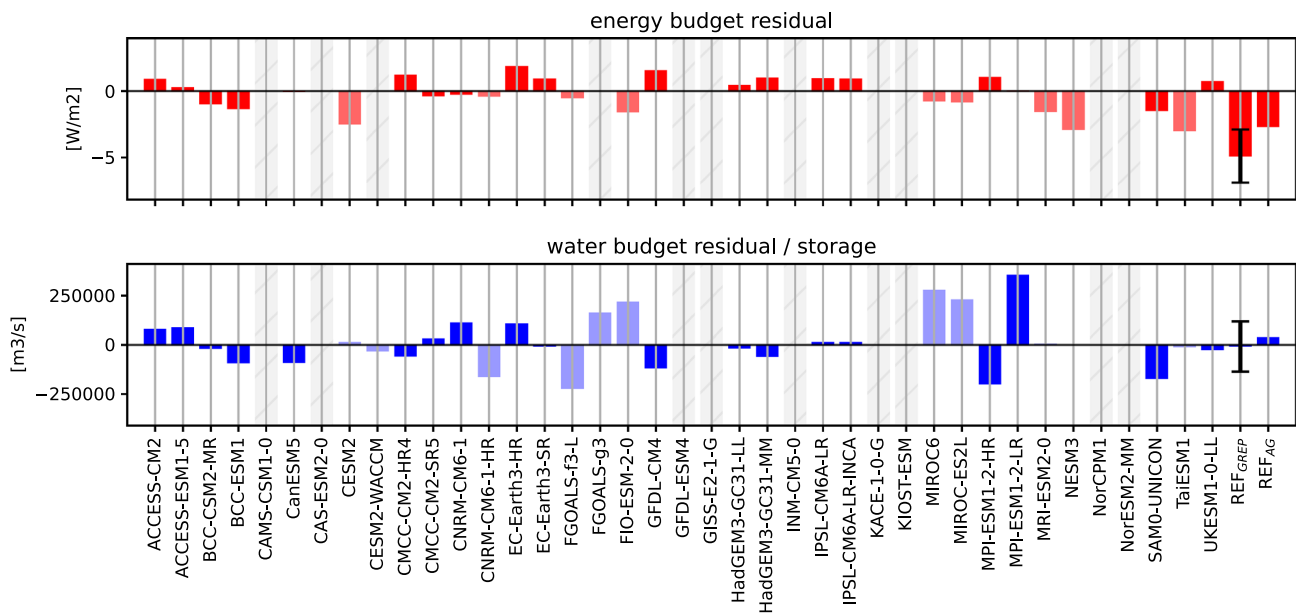


Fig. 16 Budget residuals for the energy (top) and water (bottom) budget of the Arctic Ocean

between OHT and the net radiation at the top of the atmosphere (Fig. 13b) and biases do not seem to reach up to the top of the atmosphere. In contrast, Fig. 13d shows high correlations between long-term averaged atmospheric heat transports AHT and F_{TOA} , as models with weaker outgoing F_{TOA} also feature weaker AHT. OHT and AHT feature moderate anti-correlation (Fig. 13c) and the atmosphere compensates for variances in OHT to the extent that its biases are not seen at the TOA. Models with stronger OHT feature stronger F_s and therefore weaken the atmospheric gradients and subsequently AHT. Note the deviation of the reference values from the model based reference ellipse in 13a. This is caused by inconsistency in our reference based estimates for F_s and OHT and will be discussed further in the next section.

The impact of OHT on other components of the Arctic system highlights the importance of detecting the exact source of any possible biases therein. To check whether the biases in OHT are also present further south, where less sea ice is present, we calculated transports through the Greenland-Scotland Ridge (GSR, dashed orange line in Fig. 1). Figure S6 shows heat transports at the GSR and figure S7a shows scatter plots between heat transports through the GSR and the sum of transports through Fram Strait and the BSO. They show a high correlation with biases of heat transports being also present further south in the Nordic Seas, tightly coupled to biases in the GSR across strait temperatures (Fig. S7b). Figure S7a shows a group of models slightly to the left of the reference estimates, simulating realistic GSR transports and lower Fram and BSO transports. This shift is actually caused by too much sea ice in the models, which forces the heat out of

the ocean in the Nordic Seas through higher outgoing surface energy fluxes.

Figure 14 summarises the seasonal performance of the models and shows normalised mean errors for all models and variables for the energy and water budgets. Seasons are subdivided by triangles, as indicated in the top left-hand corner of the figure. Mean errors for each variable have been normalised by the largest error of the concerning variable to allow for better inter-model comparisons. The closer the values are to 0, the smaller the model bias and the better the model performance. For instance, the net surface energy flux F_s is biased positive from autumn to spring and biased negative in summer for most models, meaning that there is less outgoing energy during the colder seasons and less net incoming energy during summer for those models (see Fig. 9). In contrast, the CMCC-CM2-SR5 model shows a positive bias during summer (more net incoming energy) and a negative bias during winter, caused mainly by its large negative sea ice bias. Further, the connection of biases in sea ice and oceanic heat transports is evident, as models with positive biases in sea ice extent have a negative OHT bias, while biases in sea ice thickness seem to be less relevant with regard to OHT. Seasonal biases in OVT for some models are caused by the models NFL scheme. The affected models are biased negative in summer (stronger outgoing flux due to the sea ice melt effect) and biased positive in winter (effect of ice formation and growth). However, also models without the NFL scheme tend to feature some spurious features. Biases in runoff are mostly due to the one-month shift in the simulated annual cycle, with summer runoff

being biased small and spring runoff being biased large. MET biases are largest in the transitional seasons (spring and autumn), while OHCT biases are largest in winter and summer, indicating a weakened amplitude of the annual cycle for most models and an enhanced amplitude for the CMCC-CM2-SR5 model.

4.3 Budget closure

Non-closure of global budgets may contribute to unforced long-term changes/trends in state variables, the so-called model “drift”. This may distort the estimate of forced changes in coupled climate simulations and lead to false interpretations. On a global scale Irving et al. (2021) find non-negligible drift trends in time-integrated ocean heat and freshwater fluxes, F_{TOA} and moisture flux into the atmosphere (evaporation minus precipitation), suggesting a considerable leakage of mass and energy in the simulated climate system. To our knowledge, budget closure on a more regional scale for the Arctic area has not been assessed yet for CMIP6. We use all terms from equations 2 and 3 and ignore changes in the oceanic volume storage, which are considered small (Winkelbauer et al. 2022), to assess the energy and water budget closure for the Arctic Ocean:

$$Res_{energy} = F_S - OHCT - MET - \nabla \cdot F_O - \nabla \cdot F_I + L_f(T_p)P_{snow} - L_f\rho_{snow}\frac{\partial d_{snow}}{\partial t} \quad (8)$$

$$Res_{water} = P + ET + R - \nabla \cdot F_{vol} \quad (9)$$

Annual mean fluxes and storage components for the water and energy budgets as simulated by the CMIP6 MMM (red values) and our reference estimates (black values) are shown in Fig. 15. Note that the reference estimates were taken from multiple independent data sources and are not consistent and therefore the observational budget estimates are not closed but rather feature budget residuals. Comparisons with estimates from Mayer et al. (2019) and Winkelbauer et al. (2022) are given in Table 7.

Figure 16 shows residuals for the energy (top) and water (bottom) budgets. With a snowfall term of 1 Wm^{-2} (ERA5) energy budget residuals for the reference estimates using oceanic transports from the GREP ensemble (REF_{GREP}) are at -4.8 Wm^{-2} . Residuals using oceanic transports from ArcGate (REF_{AG}) are smaller at -2.6 Wm^{-2} . As already seen in Fig. 13a, the largest inconsistencies are found between net surface energy fluxes derived from a combination of CERES-EBAF TOA fluxes and atmospheric energy budget quantities provided by Mayer et al. (2021a) and oceanic lateral heat transports from the GREP ensemble. Surface energy fluxes as seen by the ocean reanalyses are actually about 3 Wm^{-2} smaller (not shown),

explaining the observation based budget residuals. While Mayer et al. (2021a) use ERA5 data, the ocean reanalyses in GREP are not coupled and use atmospheric forcing from ERA-Interim, which already features significantly smaller surface energy fluxes than ERA5 (not shown). Further, the GREP reanalyses calculate their own upwelling fluxes influenced by their own ice thicknesses and skin temperatures, while ERA5 sees constant sea ice thickness of 1.5m and is known to have a warm temperature biases over sea ice (Wang et al. 2019). For the water budget reference based residuals are at $-2.9 \times 10^3 \text{ m}^3$ (REF_{GREP}) and $-52.1 \times 10^3 \text{ m}^3$ (REF_{AG}).

Figure 16 further shows budget residuals for the individual CMIP6 models. Residuals for the energy budget are comparatively small with values between -2.5 and 2 Wm^{-2} . Residuals for the water volume budget are mostly smaller than $\pm 100 \times 10^3 \text{ m}^3$. Some models feature larger residuals (e.g. MPI-ESM1-2-LR), however for those models, as discussed above, our volume transports calculations may not be accurate enough.

There are multiple potential reasons for non-closure, some of them are listed below:

- Even though we are confident in our methods of calculation, we still can not preclude problems with our technical analyses. Especially the calculation of Arctic ocean volume transports is very sensitive to the ocean bathymetry and many large fluxes of opposing sign sum up to a relatively small net transport. Therefore, small inaccuracies in the methods of calculation may lead to major errors in net integrated transports. It also has to be noted, that the needed information to calculate exact oceanic volume transports, like exact ocean depths, is not readily available for all models. This will be discussed further in the conclusions section.
- We consider the budget equations as complete as possible, however there is still the possibility that we are missing some smaller budget terms. While small themselves, they still could have effects when trying to close the budgets. For example, oceanic transports are calculated as the sum of the four major gateways, but we neglect transports through the smaller channels of Hecla and Fury Strait. Further, we ignore the small fluxes associated with the change in sensible heat content of ice (IHCT in Eq. 2) and also the temporal (sub-monthly) eddy component of oceanic transports. Also, it's possible that not all components are provided in the CMIP6 model output, e.g. small terms like numerical diffusion and mass leak increments.
- Imbalances may also arise from deficiencies in the models itself, including model coupling, numerical schemes and/or physical processes. While it is desirable for

regional budgets in climate models to be closed, achieving a perfect closure can be challenging. The closure of regional budgets depends on the accuracy and representation of processes within the model, the spatial and temporal resolution of the model, and the quality of parameterizations. However, due to the complexity of Earth's climate system, including interactions between different components (atmosphere, ocean, land, and ice), achieving complete closure at a regional scale is challenging (Lauritzen et al. 2022).

5 Summary and discussion

This study analyses the performance of 39 CMIP6 models in simulating the energy and water budgets of the Arctic. We find systematic biases in several energy and water budget components and large inter-model spreads in most evaluated parameters when compared to the uncertainty of the observationally constrained estimates.

We assessed model performance by comparing historical long-term averages and seasonal cycles of key energy and water cycle components with observational reference data. The main results of this study are summarised below.

Long-term averaged surface freshwater fluxes tend to be overestimated by most of the models analysed, and apart from large model spreads in their seasonal cycles, we also found an early timing bias of one month in the runoff cycle phase, most likely related to the models' disability to correctly simulate the timing of snow melt and permafrost degradation.

The introduction of the StraitFlux tools Winkelbauer et al. (2023) allowed the calculation of oceanic transports consistent with the discretization schemes of the respective models, allowing a fair comparison and avoiding spurious artifacts that would be caused by interpolation. However, the results of oceanic volume and ice transports show strong biases. Inter-model spread is large and the majority of models fail to simulate the annual cycles of the net Arctic volume transports correctly. The largest errors and some spurious peaks in summer are introduced by the use of the NLFS scheme for sea ice meltwater. Seasonal cycles of volume transports corrected for sea ice volume still show a large spread with some suspicious-looking models, but the MMM of the corrected

fluxes is in better agreement with the reference cycles and is within the uncertainty range of the reanalyses during 9 of 12 months. The calculation of volume transports is very sensitive to the exact ocean bathymetry. We found that for the individual Arctic straits biases due to inaccurate handling of the bathymetry are comparatively small and mostly amount to less than 10 %. However, as the individual fluxes sum up to a rather small net Arctic volume flux, those biases may cause some significant errors for the net transports. As discussed in Winkelbauer et al. (2023) caution is advised especially when calculating volume transports for shallow or bathymetrically more complicated straits, where currents are intensified in the proximity of the ocean ground or coast. While the calculation of heat and salinity transports is not as sensitive, it is still not neglectable in most cases. To improve the calculation of transports we would need the exact cell thicknesses either at the positions where the oceanic temperatures and velocities are defined or, if only thicknesses at the middle of the grid cell are supplied, provide the transformation equations to transform the thicknesses to the cell faces (Arakawa-C) or edges (Arakawa-B) for all models. Unfortunately, these data are not available for all CMIP6 models, and we hope the situation will improve in future CMIPs.

Surface energy fluxes in CMIP6 are generally strongly underestimated compared to the observationally constrained reference estimates, with the largest biases occurring in autumn, winter and spring. Radiative fluxes at TOA are closer to observations, but some models still show biases, especially in summer. Errors in F_s and F_{TOA} are closely related to the extent of the simulated sea ice area. Therefore, models with particularly large biases in their simulation of sea ice (CMCC-CM2-SR5, CAS-ESM2-0, FIO-ESM2-0, NorCPM1, FGOALS-g3) also have the largest errors in F_s and F_{TOA} . Also problems in the models' energy conservation may lead to errors in the net global energy budget at TOA and at the surface (Wild 2020), however those errors should be comparatively small.

As with the water budget, also for the energy budget the largest uncertainties and biases seem to be generated in the ocean. While most models are able to correctly simulate the timing of the oceanic lateral heat inflows, the inter-model spread is exceptionally large and most models show a systematic underestimation of the heat transports. Six models (BCC-CSM2-MR, BCC-ESM1, CAMS-CSM1-0, FGOALS-g3, NESM3) simulate

particularly small heat transports, mostly due to temperature biases but also because of too weak simulated Barents Sea volume transports. We find strong relationships between lateral oceanic heat transports and the mean state of the Arctic. Furthermore, oceanic transports have strong effects on sea ice cover and ocean warming rates, demonstrating the importance of the mean state on projected trends.

Biases in Arctic deep waters were shown to be caused by the lack of ventilation through shelf overflows and inaccurate oceanic transports (Heuzé et al. 2023). In addition, longer spin-up times may be required as deep waters may take longer to equilibrate to the initial conditions. A more detailed assessment of oceanic transports would be necessary to determine the exact source of these biases.

Despite the use of more accurate oceanic transport estimates and the assessment of more complete budgets, it was still not possible to close the energy and water budgets for the individual models completely. Nevertheless, energy budget residuals are smaller than 2 W m^{-2} for most models, which is still small when compared to the inter-model spreads in most energy budget components. Small residuals could be due to both technical issues on our side and deficiencies in the models, including model coupling, physical processes and numerical schemes. More extensive evaluations of these imbalances could help to further identify and address biases and limitations, leading to improved representations of regional processes and more balanced budgets.

Furthermore, it must be reiterated that all multi-model averages were computed using all available models without any kind of model weighting, which should be applied to mitigate biases, uncertainties and discrepancies between models and provide a more balanced representation of the overall model ensemble. The results of this study can nevertheless help us to understand typical model biases in the Arctic, and using these results it may be possible to generate physically based metrics to detect outliers from the model ensemble. These metrics may prove may be useful in reducing the spread of future projections of Arctic change.

Large model spreads can be exacerbated by several sources of error. First and foremost, we used only one realisation per model, which is known to introduce a sampling error as each different realisation simulates a different possible outcome of the chaotic climate system (Wang et al. 2022). However, past studies suggested intra-model biases to be quite small compared to inter-model biases (e.g., Zanowski et al. 2021; Khosravi

et al. 2022; Wang et al. 2022). We used a bootstrapping approach to estimate those sampling errors and found this to be true for for most variables in our study. Also, observations similarly account for only one realisation and therefore the sampling error should be of the same value for our observational estimates. So, in most cases, biases between models and observations when looking at long-term means are very likely to be true systematic biases inherent in the model. However, for variables with larger sampling errors, like e.g. OHCT and MET, and also when looking at trends of the relatively short period of 22-years, variabilities on different time scales may introduce sampling uncertainty. In those cases, the best solution would be to look at longer time scales, whereby this oftentimes is problematic due to the length of available observations and spinup effects during the earlier part of the model simulations.

In addition, errors may be introduced by missing processes or different treatment of processes in the models. For example, as we saw in Fig. 3, the inclusion of a non-linear free surface scheme leads to biased seasonal cycles of oceanic volume transports, at least in the current generation of climate models. Errors in the calculation of energy and water budget variables have been minimised by using the native grid files of all variables where interpolation can corrupt the result.

In conclusion, the biases we find in some of the Arctic's energy and water budgets of the evaluated models have substantial effects on the simulated mean state and changes within the system and therefore possibly also on projections of future warming of the Arctic. To obtain more realistic simulations of the Arctic and processes therein more observations would be needed to constrain the models, as well as higher resolution and improved parametrizations, as already discussed by e.g., Heuzé et al. (2023). Nevertheless, the diagnostics framework presented here can be applied to measure progress made with upcoming new versions of coupled model runs, performed, e.g., within CMIP7. The presented diagnostics may also be used to generate more process-based metrics compared to earlier studies (e.g., Brunner et al. 2020) that focused on state quantities to detect outliers from the model ensemble and therefore reduce the spread of future projections of Arctic change.

Appendix

See Tables 6, 7, 8.

Table 6 Averaged volume and heat transports through the main Arctic gateways for the 1993–2014 period

	Fram		Barents		Davis		Bering	
	Volume	Heat	Volume	Heat	Volume	Heat	Volume	Heat
ACCESS-CM2	− 3.3 (0.01)	27.6 (− 0.02)	2.8 (− 0.09*)	54.7 (− 0.04*)	− 0.2 (1.26*)	3.9 (0.21*)	0.7 (− 0.00)	1.1 (0.05)
ACCESS-ESM1-5	− 4.2 (− 0.09*)	35.6 (0.02)	3.5 (− 0.02*)	85.9 (− 0.01)	− 0.7 (0.64*)	6.3 (− 0.21*)	1.3 (0.02*)	7.0 (0.17*)
BCC-CSM2-MR	− 0.7 (− 0.14*)	21.8 (0.03*)	0.5 (0.21*)	5.5 (0.29*)	− 0.0 (4.65)	0.3 (0.04)	− 0.0 (0.04*)	0.1 (− 0.00)
BCC-ESM1	− 0.3 (− 0.02)	20.1 (0.08*)	0.1 (− 0.27*)	3.6 (− 0.02)	− 0.1 (1.24*)	− 0.2 (0.57)	− 0.0 (− 0.03*)	0.1 (− 0.14*)
CAMS-CSM1-0	− 0.9 (0.14*)	20.7 (− 0.02*)	0.8 (− 0.16*)	7.9 (− 0.29*)	− 0.5 (0.00)	7.1 (0.07*)	0.5 (− 0.08*)	− 0.1 (0.04)
CanESM5	− 3.7 (− 0.22*)	23.4 (0.26*)	3.4 (0.18*)	63.7 (0.35*)	− 0.6 (0.38*)	3.2 (− 0.16*)	0.8 (0.02)	3.0 (0.52*)
CAS-ESM2-0	−	−	−	−	−	−	−	−
CESM2	− 1.2 (− 0.12*)	10.2 (0.12*)	1.8 (0.07*)	45.6 (0.12*)	− 1.9 (0.09*)	6.2 (− 0.11*)	1.1 (0.02*)	6.0 (0.38*)
CESM2-WACCM	− 1.2 (− 0.17*)	9.0 (0.08*)	1.6 (0.12*)	41.3 (0.10*)	− 1.8 (0.07*)	6.4 (− 0.01)	1.1 (0.00)	3.7 (0.44*)
CMCC-CM2-HR4	− 2.5 (0.07*)	58.8 (0.15*)	3.5 (− 0.01)	95.8 (0.00)	− 2.7 (− 0.05*)	25.9 (0.09*)	1.5 (0.01)	8.7 (0.19*)
CMCC-CM2-SR5	− 3.3 (0.01)	17.1 (0.37*)	4.3 (0.01*)	106.4 (0.11*)	− 2.6 (− 0.05*)	8.2 (− 0.03)	1.3 (0.03*)	11.9 (0.34*)
CNRM-CM6-1	− 3.4 (− 0.23*)	24.0 (− 0.06*)	3.4 (− 0.14*)	64.1 (− 0.15*)	− 1.6 (0.76*)	8.8 (− 0.31*)	1.5 (0.01*)	7.0 (0.05)
CNRM-CM6-1-HR	− 1.0 (− 0.19*)	29.8 (0.30*)	1.9 (0.05*)	50.5 (0.18*)	− 2.2 (0.09*)	27.2 (− 0.04*)	0.9 (− 0.06*)	6.6 (0.02)
EC-Earth3P-HR	− 2.3 (− 0.14*)	34.9 (0.25*)	2.1 (0.08*)	58.4 (0.09*)	− 0.9 (0.15*)	16.2 (− 0.10*)	1.1 (0.01)	6.0 (− 0.08*)
EC-Earth3	− 2.6 (− 0.32*)	3.8 (1.80*)	3.2 (0.33*)	79.1 (0.45*)	− 2.4 (− 0.16*)	7.7 (− 0.13*)	1.8 (0.06*)	12.6 (0.37*)
FGOALS-f3-L	− 3.6 (0.11*)	6.1 (0.48*)	2.0 (− 0.04*)	29.9 (− 0.03*)	− 0.9 (0.16*)	14.4 (− 0.21*)	2.1 (− 0.02*)	14.0 (0.06*)
FGOALS-g3	− 1.1 (0.40*)	9.4 (0.29*)	0.6 (− 0.29*)	8.2 (0.50*)	− 1.4 (0.16*)	8.2 (− 0.17*)	1.9 (0.01*)	6.0 (0.43*)
FIO-ESM2-0	− 1.4 (− 0.03*)	11.1 (− 0.08*)	2.4 (− 0.02)	75.5 (− 0.04*)	− 1.9 (0.06*)	10.5 (− 0.10*)	0.9 (0.04*)	6.0 (0.43*)
GFDL-CM4	− 1.7 (− 0.29*)	26.1 (0.08*)	2.4 (0.14*)	69.7 (0.15*)	− 1.8 (0.07*)	24.8 (0.02)	0.9 (0.02*)	1.9 (0.37*)
GFDL-ESM4	−	−	−	−	−	−	−	−
GISS-E2-1-G	−	−	−	−	−	−	−	−
HadGEM3-GC31-LL	− 3.4 (− 0.04*)	23.5 (0.21*)	3.8 (0.00)	83.3 (0.05*)	− 1.8 (0.03*)	7.5 (− 0.11*)	1.3 (0.03*)	2.4 (0.62*)
HadGEM3-GC31-MM	− 2.3 (− 0.17*)	41.2 (0.03)	2.5 (0.06*)	60.5 (0.09*)	− 1.4 (0.07*)	17.9 (0.17*)	1.0 (0.09*)	2.9 (0.34*)
INM-CM5-0	−	−	−	−	−	−	−	−
IPSL-CM6A-LR	− 4.1 (− 0.13*)	2.3 (0.87*)	4.1 (− 0.02*)	97.2 (0.00)	− 1.3 (0.49*)	4.6 (− 0.08*)	1.1 (− 0.01)	2.9 (0.34*)
IPSL-CM6A-LR-INCA	− 3.6 (− 0.26*)	− 1.1 (− 1.59*)	4.0 (0.10*)	100.0 (0.08*)	− 1.7 (0.24*)	3.2 (0.14*)	1.1 (0.12*)	6.6 (0.20*)
KACE-1-0-G	−	−	−	−	−	−	−	−
KIOST-ESM	−	−	−	−	−	−	−	−
MIROC-ES2L	− 3.2 (0.02)	42.7 (0.10*)	2.5 (− 0.07*)	71.1 (0.13*)	0.2 (0.74*)	4.7 (0.03)	0.5 (− 0.03)	3.1 (0.26*)
MIROC6	− 2.8 (0.02*)	40.5 (0.16*)	2.9 (0.06*)	43.7 (− 0.07*)	− 1.1 (− 0.18*)	4.7 (0.04)	1.0 (− 0.00)	3.0 (0.36*)
MPI-ESM1-2-HR	− 2.8 (− 0.11*)	17.5 (− 0.06)	2.7 (0.18*)	70.3 (0.22*)	− 0.8 (− 0.10*)	8.6 (0.16*)	0.7 (0.07*)	3.3 (0.42*)
MPI-ESM1-2-LR	− 2.1 (− 0.32*)	26.0 (− 0.22*)	2.9 (0.12*)	86.0 (0.13*)	− 1.1 (0.23*)	10.7 (− 0.21*)	0.6 (0.04*)	1.6 (0.52*)
MRI-ESM2-0	− 2.6 (− 0.25*)	9.3 (− 0.29*)	2.5 (0.15*)	61.1 (0.22*)	− 1.7 (0.00)	6.8 (0.23*)	0.8 (− 0.05*)	6.6
NESM3	− 0.8 (− 0.29*)	10.2 (− 0.11*)	1.7 (0.11*)	27.2 (0.34*)	− 1.5 (0.01)	4.8 (− 0.16*)	0.3 (0.05*)	− 1.9 (0.21*)
NorCPM1	−	−	−	−	−	−	−	−
NorESM2-MM	−	−	−	−	−	−	−	−
SAM0-UNICON	− 1.2 (− 0.21*)	10.0 (− 0.08*)	2.0 (0.24*)	41.0 (0.28*)	− 2.0 (− 0.11*)	10.0 (0.15*)	1.0 (− 0.04*)	0.7 (1.37*)
TaiESM1	− 1.6 (− 0.12*)	8.0 (− 0.30*)	2.3 (0.08*)	59.0 (0.13*)	− 2.0 (0.01*)	9.8 (0.04*)	1.0 (0.04*)	2.0 (0.78*)
UKESM1-0-LL	− 2.0 (− 0.11*)	19.4 (0.07*)	2.8 (− 0.02*)	62.7 (0.06*)	− 2.2 (0.13*)	9.9 (− 0.30*)	1.3 (− 0.03*)	2.2 (0.59*)
MMM	− 2.3 (− 0.11)	20.2 (0.09)	2.4 (0.05)	57.4 (0.11)	− 1.3 (0.11)	9.1 (− 0.03)	0.8 (0.02)	3.6 (0.31)
REF	− 2.2 (− 0.10)	34.0 (0.03)	3.2 (0.07)	78.3 (0.12)	− 2.4 (− 0.00)	9.5 (− 0.16)	1.2 (− 0.02)	4.9 (0.12)
ArcGate	− 1.4	67.6	2.2	54.7	− 1.9	31.8	1.0	− 2.6

Volume Transports are given in SV and heat transports in TW. REF values are calculated using the GREP reanalyses ensemble. Values in brackets show Trends in [fraction/decade]

Table 7 Water and energy fluxes and storage rates for the reference estimates (Ref) and the CMIP6 MMM for 1993–2014

[10 ³ m ³ /s]	Ref	CMIP6	Winkelbauer et al. (2022)	ArcGate
P-E	69.2 ± 2.5	75.9 ± 7.8	Same as Ref	–
R	127.0 ± 1.1	141.1 ± 21.4	Same as Ref	–
R _G	11.9 ± 0.4	5.6 ± 4.6	Same as Ref	–
OVT	– 151 ± 43	– 145 ± 139	– 207	– 91
OIT	– 60 ± 5	– 101 ± 42		– 65
Res/S	– 2.9	– 23.4	– 1.1	– 52.1
[Wm ^{–2}]	Ref	CMIP6	Mayer et al. (2019)	ArcGate
Rad _{TOA}	– 116.7 ± 0.5	– 111.6 ± 4.4	– 115.8	–
F _S	– 18.0 ± 3.9	– 7.6 ± 2.8	– 16.2	–
AET	0.1 ± 0.9	0.0 ± 0.1	– 0.1	–
OHCT	0.3 ± 0.4	0.5 ± 0.4	0.3	–
MET	0.3 ± 0.1	0.2 ± 0.1	0.4	–
AHT	98.0 ± 4.0	103.9 ± 5.2	99.6	–
OHT	11.3 ± 3.7	8.2 ± 3.3	15.5	13.4
IHT	1.6 ± 0.8	2.1 ± 1.2	1.4	–
Res	– 5.8	2.0	0.0	– 3.6

Additionally ArcGate estimates and estimates from Winkelbauer et al. (2022) and Mayer et al. (2019) are given

Table 8 List of Acronyms

ArcGate	Mooring-derived data of oceanic fluxes through the Arctic gateways
BSO	Barents Sea opening
C2SMOS	Sea ice product merged from Cryosat-2 and Soil Moisture and Ocean Salinity satellites
CAA	Canadian Arctic Archipelago
CERES-EBAF	Clouds and the Earth's Radiant Energy System-Energy Balanced and Filled
CGLORS	Ocean reanalyses from the Euro-Mediterranean Center on Climate Change
CMEMS	Copernicus Marine Environment Monitoring Service
CMIP6	Coupled Model Intercomparison Project Phase 6
EGC	East Greenlandic Current
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF's fifth atmospheric reanalysis
ERA-Interim	ECMWF interim reanalysis
ESGF	Earth System Grid Federation
FOAM	Ocean reanalyses from the UK Met Office
GLORYS	Ocean reanalyses from Mercator Ocean
GRACE	Gravity Recovery and Climate Experiment
GREP	Global ocean Reanalysis Ensemble Product
MMM	Multi-model mean
NC	Norwegian Coastal Current
NEMO	Nucleus for European Modelling of the Ocean
NLFS	Non-linear free surface
NRMSE	Normalized root mean square error
ORAS5	ECMWF's Ocean Reanalysis System 5
OSR	Ocean Reanalysis
REF	Reference data
WSC	West Spitsbergen Current
AE(T)	Atmospheric energy (tendency)
AHT	Atmospheric energy/heat transport
c_p	Specific heat of seawater
d_{snow}	Snow thickness
$\nabla \cdot F_A$	Divergence of lateral atmospheric energy transports
$\nabla \cdot F_I$	Divergence of latent heat transport associated with sea ice transports
$\nabla \cdot F_O$	Divergence of ocean heat transports
$\nabla \cdot F_{vol}$	Divergence of lateral oceanic volume fluxes
ΔS_O	Change of ocean volume
ET	Evapotranspiration
F_s	Surface energy flux
F_{TOA}	Net radiation at the top of the atmosphere (TOA)
IHCT	Sea ice sensible heat content tendency
L_f	Latent heat of fusion
ME(T)	Sea ice melt energy (tendency)
\vec{n}	Vector normal to strait
nME	Normalised mean error
OHC(T)	Ocean heat content (tendency)
OHT	Oceanic transports of heat
OIT	Oceanic transports of ice
OVT	Oceanic transports of volume
P	Precipitation
P_{snow}	Snowfall rate
R	Runoff
ρ	Density of sea water
ρ_{snow}	Snow density

Table 8 (continued)

SIE	Sea ice extent
SIE _B	Sea ice extent in the Barents Sea
SIT	Sea ice thickness
θ	Potential temperature
θ_{REF}	Reference temperature
VIWVD	Vertical integral of divergence of moisture flux
\vec{v}_o	Ocean velocity vector
\vec{v}_i	Ice velocity vector
x	Width along the oceanic strait
x _s , x _e	Start and end point of oceanic strait
z	Ocean depth

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00382-024-07105-5>.

Acknowledgements Susanna Winkelbauer and Michael Mayer were supported by the Austrian Science Fund project P33177 and the Copernicus Marine Environment Service contract 21003-COP-GLORAN Lot 7. The authors sincerely thank Richard Renshaw from Met Office for the helpful discussions on GloRanV14 (FOAMv2).

Rights and permissions Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Author contributions SW, MM and LH conceptualized the study. SW performed the data analysis, including the production of the figures in the paper, and prepared the manuscript. MM and LH contributed to the interpretation of results and the writing of the manuscript. All authors have read and agreed to the publication of the present version of the manuscript.

Funding Open access funding provided by University of Vienna. This work was supported by the Austrian Science Fund project P33177 and the Copernicus Marine Environment Service contract 21003-COP-GLORAN Lot 7.

Data availability CMIP6 data is available through the Earth System Grid Federation (ESGF) website <https://esgf-node.llnl.gov/search/cmip6/>. ERA5 data (Hersbach et al. 2019) as well as mass-consistent energy fluxes derived from ERA5 (Mayer et al. 2021b) are available in the Copernicus Climate Change Service (C3S) Climate Data Store and DEEP-C fluxes are publicly available at <https://doi.org/10.17864/1947.271>.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allan RP, Liu C, Loeb NG et al (2014) Changes in global net radiative imbalance 1985–2012. *Geophys Res Lett* 41(15):5588–5597. <https://doi.org/10.1002/2014GL060962>
- Årthun M, Eldevik T, Smedsrud LH et al (2012) Quantifying the influence of Atlantic heat on Barents sea ice variability and retreat. *J Clim* 25(13):4736–4743. <https://doi.org/10.1175/JCLI-D-11-00466.1>
- Årthun M, Eldevik T, Smedsrud L (2019) The role of Atlantic heat transport in future arctic winter sea ice loss. *J Clim* 32(11):3327–3341. <https://doi.org/10.1175/JCLI-D-18-0750.1>
- Bacon S, Aksenov Y, Fawcett S et al (2015) Arctic mass, freshwater and heat fluxes: methods and modelled seasonal variability. *Phil Trans R Soc A*. <https://doi.org/10.1098/rsta.2014.0169>
- Bacon S, Garabato A, Aksenov Y et al (2022) Arctic ocean boundary exchanges: a review. *Oceanography*. <https://doi.org/10.5670/oceanog.2022.133>
- Bao Y, Song Z, Qiao F (2020) Fio-esm version 2.0: model description and evaluation. *J Geophys ResOceans*. <https://doi.org/10.1029/2019JC016036>
- Bethke I, Wang Y, Counillon F et al (2021) Norcpm1 and its contribution to cmip6 dcpp. *Geosci Model Dev* 14(11):7073–7116. <https://doi.org/10.5194/gmd-14-7073-2021>
- Bi D, Dix M, Marsland S et al (2020) Configuration and spin-up of access-cm2, the new generation Australian community climate and earth system simulator coupled model. *J South Hemisphere Earth Syst Sci*. <https://doi.org/10.1071/ES19040>
- Bintanja R, Selten FM (2014) Future increases in arctic precipitation linked to local evaporation and sea-ice retreat. *Nature* 509:479–482. <https://doi.org/10.1038/nature13259>

- Blackport R, Screen JA (2020) Weakened evidence for mid-latitude impacts of arctic warming. *Nat Clim Chang* 10(12):1065–1066. <https://doi.org/10.1038/s41558-020-00954-y>
- Bonan DB, Feldl N, Zelinka MD et al (2023) Contributions to regional precipitation change and its polar-amplified pattern under warming. *Environ Res Clim* 2(3):035010. <https://doi.org/10.1088/2752-5295/ace27a>
- Boucher O, Servonnat J, Albright AL et al (2020) Presentation and evaluation of the ipsl-cm6a-lr climate model. *J Adv Model Earth Syst*. <https://doi.org/10.1029/2019MS002010>
- Box J, Hubbard A, Bahr D et al (2022) Greenland ice sheet climate disequilibrium and committed sea-level rise. *Nat Clim Change*. <https://doi.org/10.1038/s41558-022-01441-2>
- Brunner L, Pendergrass AG, Lehner F et al (2020) Reduced global warming from cmip6 projections when weighting models by performance and independence. *Earth Syst Dyn* 11(4):995–1012. <https://doi.org/10.5194/esd-11-995-2020>
- Cai Z, You Q, Wu F et al (2021) Arctic warming revealed by multiple cmip6 models: evaluation of historical simulations and quantification of future projection uncertainties. *J Clim* 34:4871–4892. <https://doi.org/10.1175/JCLI-D-20-0791.1>
- Cao J, Wang B, Yang YM et al (2018) The nuist earth system model (nesm) version 3: description and preliminary evaluation. *Geosci Model Dev* 11(7):2975–2993. <https://doi.org/10.5194/gmd-11-2975-2018>
- Chen HM, Li J, Su JZ et al (2019) Introduction of cams-csm model and its participation in cmip6. *Adv Clim Chang Res* 15(5):540. <https://doi.org/10.12006/j.issn.1673-1719.2019.186>
- Cheng L, von Schuckmann K, Abraham JP et al (2022) Past and future ocean warming. *Nat Rev Earth Environ* 3(11):776–794. <https://doi.org/10.1038/s43017-022-00345-1>
- Coumou D, Di Capua G, Vavrus S et al (2018) The influence of arctic amplification on mid-latitude summer circulation. *Nat Commun*. <https://doi.org/10.1038/s41467-018-05256-8>
- Danabasoglu G, Lamarque JF, Bacmeister J et al (2020) The community earth system model version 2 (cesm2). *J Adv Model Earth Syst*. <https://doi.org/10.1029/2019MS001916>
- Dee DP, Uppala SM, Simmons AJ et al (2011) The era-interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137(656):553–597. <https://doi.org/10.1002/qj.828>
- Desportes C, Garric G, Régnier C et al (2017) CMEMS-GLO-QUID-001-026, E.U. Copernicus Marine Service Information
- Docquier D, Koenigk T (2021) A review of interactions between ocean heat transport and arctic sea ice. *Environ Res Lett* 16(12):123002. <https://doi.org/10.1088/1748-9326/ac30be>
- Dörr J, Arthun M, Eldevik T et al (2021) Mechanisms of regional winter sea-ice variability in a warming arctic. *J Clim* 34:1–56. <https://doi.org/10.1175/JCLI-D-21-0149.1>
- Döscher R, Acosta M, Alessandri A et al (2022) The ec-earth3 earth system model for the coupled model intercomparison project 6. *Geosci Model Dev* 15(7):2973–3020. <https://doi.org/10.5194/gmd-15-2973-2022>
- Dunne JP, Horowitz LW, Adcroft AJ et al (2020) The gfdl earth system model version 4.1 (gfdl-esm 4.1): overall coupled model description and simulation characteristics. *J Adv Model Earth Syst*. <https://doi.org/10.1029/2019MS002015>
- Eyring V, Bony S, Meehl GA et al (2016) Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geosci Model Dev* 9(5):1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Fasullo JT, Trenberth KE (2008) The annual cycle of the energy budget. Part I: global mean and land-ocean exchanges. *J Clim* 21(10):2297–2312. <https://doi.org/10.1175/2007JCLI1935.1>
- Fox-Kemper B, Hewitt H, Xiao C et al (2021) Ocean, cryosphere and sea level change. Cambridge University Press, Cambridge and New York, pp 1211–1362. <https://doi.org/10.1017/9781009157896.011>
- Francis JA, Vavrus SJ (2012) Evidence linking arctic amplification to extreme weather in mid-latitudes. *Geophys Res Lett*. <https://doi.org/10.1029/2012GL051000>
- Garric G, Parent L, Greiner E, et al (2017) Performance and quality assessment of the global ocean eddy-permitting physical reanalysis GLORYS2V4. In: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, p 18776
- Goosse H, Kay JE, Armour KC et al (2018) Quantifying climate feedbacks in polar regions. *Nat Commun*. <https://doi.org/10.1038/s41467-018-04173-0>
- Gosling SN, Arnell NW (2011) Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. *Hydrol Process* 25(7):1129–1145. <https://doi.org/10.1002/hyp.7727>
- Haarsma R, Acosta M, Bakhshi R et al (2020) Highresmp versions of ec-earth: Ec-earth3p and ec-earth3p-hr - description, model computational performance and basic validation. *Geosci Model Dev* 13(8):3507–3527. <https://doi.org/10.5194/gmd-13-3507-2020>
- Haine TW, Curry B, Gerdes R et al (2015) Arctic freshwater export: status, mechanisms, and prospects. *Global Planet Change* 125:13–35. <https://doi.org/10.1016/j.gloplacha.2014.11.013>
- Hajima T, Watanabe M, Yamamoto A et al (2020) Development of the miroc-es2l earth system model and the evaluation of biogeochemical processes and feedbacks. *Geosci Model Dev* 13(5):2197–2244. <https://doi.org/10.5194/gmd-13-2197-2020>
- He B, Bao Q, Wang X et al (2019) Cas fgoals-f3-l model datasets for cmip6 historical atmospheric model intercomparison project simulation. *Adv Atmos Sci* 36(8):771–778. <https://doi.org/10.1007/s00376-019-9027-8>
- Held IM, Guo H, Adcroft A et al (2019) Structure and performance of gfdl's cm4.0 climate model. *J Adv Model Earth Syst* 11(11):3691–3727. <https://doi.org/10.1029/2019MS001829>
- Hersbach H, Bell B, Berrisford P et al (2020) The era5 global reanalysis. *Q J R Meteorol Soc* 146(730):1999–2049. <https://doi.org/10.1002/qj.3803>
- Hersbach H, Bell B, Berrisford P, et al (2019) Era5 monthly averaged data on single levels from 1979 to present. <https://doi.org/10.24381/cds.f17050d7>
- Heuzé C, Zanowski H, Karam S et al (2023) The deep arctic ocean and fram strait in cmip6 models. *J Clim* 36(8):2551–2584. <https://doi.org/10.1175/JCLI-D-22-0194.1>
- Hou Y, Guo H, Yang Y et al (2023) Global evaluation of runoff simulation from climate, hydrological and land surface models. *Water Resour Res* 59(1):e2021WR031817. <https://doi.org/10.1029/2021WR031817>
- Irving D, Hobbs W, Church J et al (2021) A mass and energy conservation analysis of drift in the cmip6 ensemble. *J Clim* 34(8):3157–3170. <https://doi.org/10.1175/JCLI-D-20-0281.1>
- Kelley M, Schmidt GA, Nazarenko LS et al (2020) Giss-e2.1: configurations and climatology. *J Adv Model Earth Syst*. <https://doi.org/10.1029/2019MS002025>
- Khosravi N, Wang Q, Koldunov N et al (2022) The arctic ocean in cmip6 models: biases and projected changes in temperature and salinity. *Earth's Future* 10(2):e2021EF002282. <https://doi.org/10.1029/2021EF002282>
- Knutti R (2008) Should we believe model predictions of future climate change? *Philos Trans Royal Soc A Math Phys Eng Sci* 366(1885):4647–4664. <https://doi.org/10.1098/rsta.2008.0169>
- Kouki K, Räisänen P, Luojus K et al (2022) Evaluation of northern hemisphere snow water equivalent in cmip6 models during 1982–2014. *Cryosphere* 16(3):1007–1030. <https://doi.org/10.5194/tc-16-1007-2022>

- Kwok R (2018) Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018). *Environ Res Lett* 13(10):105005. <https://doi.org/10.1088/1748-9326/aae3ec>
- Lauritzen P, Kevlahan N, Toniazzo T et al (2022) Reconciling and improving formulations for thermodynamics and conservation principles in earth system models (esms). *J Adv Model Earth Syst*. <https://doi.org/10.1029/2022MS003117>
- Lee J, Kim J, Sun MA et al (2020) Evaluation of the Korea meteorological administration advanced community earth-system model (k-ace). *Asia Pac J Atmos Sci*. <https://doi.org/10.1007/s13143-019-00144-7>
- Li L, Yu Y, Tang Y et al (2020) The flexible global ocean-atmosphere-land system model grid-point version 3 (fgoals-g3): description and evaluation. *J Adv Model Earth Syst* 12(9):e2019MS002012. <https://doi.org/10.1029/2019MS002012>
- Liu C, Allan RP, Mayer M et al (2020) Variability in the global energy budget and transports 1985–2017. *Clim Dyn* 55:3381–3396. <https://doi.org/10.1007/s00382-020-05451-8>
- Loeb NG, Doelling DR, Wang H et al (2018) Clouds and the earth's radiant energy system (ceres) energy balanced and filled (ebaf) top-of-atmosphere (toa) edition-4.0 data product. *J Clim* 31(2):895–918. <https://doi.org/10.1175/JCLI-D-17-0208.1>
- MacLachlan C, Arribas A, Peterson KA et al (2015) Global seasonal forecast system version 5 (glosea5): a high-resolution seasonal forecast system. *Q J R Meteorol Soc* 141(689):1072–1084. <https://doi.org/10.1002/qj.2396>
- Madec G (2016) Nemo ocean engine—version 3.6. Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL) 27
- Mauritsen T, Bader J, Becker T et al (2019) Developments in the mpi-m earth system model version 1.2 (mpi-esm1.2) and its response to increasing co2. *J Adv Model Earth Syst* 11(4):998–1038. <https://doi.org/10.1029/2018MS001400>
- Mayer M, Haimberger L, Pietschnig M et al (2016) Facets of arctic energy accumulation based on observations and reanalyses 2000–2015. *Geophys Res Lett* 43(19):10,420–10,429. <https://doi.org/10.1002/2016GL070557>
- Mayer M, Haimberger L, Edwards JM et al (2017) Toward consistent diagnostics of the coupled atmosphere and ocean energy budgets. *J Clim* 30(22):9225–9246. <https://doi.org/10.1175/JCLI-D-17-0137.1>
- Mayer M, Tietsche S, Haimberger L et al (2019) An improved estimate of the coupled arctic energy budget. *J Clim* 32(22):7915–7934. <https://doi.org/10.1175/JCLI-D-19-0233.1>
- Mayer J, Mayer M, Haimberger L (2021) Consistency and homogeneity of atmospheric energy, moisture, and mass budgets in era5. *J Clim* 34(10):3955–3974. <https://doi.org/10.1175/JCLI-D-20-0676.1>
- Mayer M, Vidar SL, Kjell AM et al (2021) Ocean heat content in the highnorth. *J Oper Oceanogr* 14(Suppl. 1):S17–S23. <https://doi.org/10.1080/1755876X.2021.1946240>
- Mayer M, Tsubouchi T, von Schuckmann K et al (2022) Atmospheric and oceanic contributions to observed nordic seas and arctic ocean heat content variations 1993–2020. *J Oper Oceanogr* 15(Suppl. 1):S20–S28. <https://doi.org/10.1080/1755876X.2022.2095169>
- Mayer J, Mayer M, Haimberger L (2021b) Mass-consistent atmospheric energy and moisture budget monthly data from 1979 to present derived from era5 reanalysis. <https://doi.org/10.24381/cds.c2451f6b>
- McPhee MG, Proshutinsky A, Morison JH et al (2009) Rapid change in freshwater content of the arctic ocean. *Geophys Res Lett*. <https://doi.org/10.1029/2009GL013752>
- Moon T, Ahlstrøm A, Goelzer H et al (2018) Rising oceans guaranteed: arctic land ice loss and sea level rise. *Curr Clim Change Rep*. <https://doi.org/10.1007/s40641-018-0107-0>
- Mouginot J, Rignot E, Bjørk A et al (2019) Forty-six years of Greenland ice sheet mass balance from 1972 to 2018. *Earth Atmos Planet Sci* 116(19):9239–9244. <https://doi.org/10.1073/pnas.1904242116>
- Muilwijk M, Smedsrud LH, Ilicak M et al (2018) Atlantic water heat transport variability in the 20th century arctic ocean from a global ocean model and observations. *J Geophys Res Oceans* 123(11):8159–8179. <https://doi.org/10.1029/2018JC014327>
- Onarheim IH, Årthun M (2017) Toward an ice-free barents sea. *Geophys Res Lett* 44(16):8387–8395. <https://doi.org/10.1002/2017GL074304>
- Pak G, Noh Y, Lee MI et al (2021) Korea institute of ocean science and technology earth system model and its simulation characteristics. *Ocean Sci J*. <https://doi.org/10.1007/s12601-021-00001-7>
- Park S, Shin J, Kim S et al (2019) Global climate simulated by the Seoul national university atmosphere model version 0 with a unified convection scheme (sam0-unicon). *J Clim* 32(10):2917–2949. <https://doi.org/10.1175/JCLI-D-18-0796.1>
- Pietschnig M, Mayer M, Tsubouchi T et al (2017) Volume and temperature transports through the main arctic gateways: a comparative study between an ocean reanalysis and mooring-derived data. *Ocean Sci Discussions* 2017:1–32. <https://doi.org/10.5194/os-2017-98>
- Pithan F, Jung T (2021) Arctic amplification of precipitation changes—the energy hypothesis. *Geophys Res Lett* 48(21):e2021GL094977. <https://doi.org/10.1029/2021GL094977>
- Proshutinsky A, Krishfield R, Timmermans ML et al (2009) Beaufort gyre freshwater reservoir: state and variability from observations. *J Geophys Res Oceans*. <https://doi.org/10.1029/2008JC005104>
- Rabe B, Karcher M, Schauer U et al (2011) An assessment of arctic ocean freshwater content changes from the 1990s to the 2006–2008 period. *Deep Sea Res Part I Oceanogr Res Papers* 58(2):173–185. <https://doi.org/10.1016/j.dsr.2010.12.002>
- Rantanen M, Karpechko AY, Lipponen A et al (2022) The arctic has warmed nearly four times faster than the globe since 1979. *Commun Earth Environ*. <https://doi.org/10.1038/s43247-022-00498-3>
- Ricker R, Hendricks S, Kaleschke L et al (2017) A weekly arctic sea-ice thickness data record from merged cryosat-2 and smos satellite data. *Cryosphere* 11(4):1607–1623. <https://doi.org/10.5194/tc-11-1607-2017>
- Roulet G, Madec G (2000) Salt conservation, free surface, and varying levels: a new formulation for ocean general circulation models. *J Geophys Res Oceans* 105(C10):23927–23942. <https://doi.org/10.1029/2000JC900089>
- Rowland JC, Jones CE, Altmann G et al (2010) Arctic landscapes in transition: responses to thawing permafrost. *EOS Trans Am Geophys Union* 91(26):229–230. <https://doi.org/10.1029/2010EO260001>
- Schauer U, Beszczynska-Möller A (2009) Problems with estimation and interpretation of oceanic heat transport—conceptual remarks for the case of fram strait in the arctic ocean. *Ocean Sci* 5(4):487–494. <https://doi.org/10.5194/os-5-487-2009>
- Screen JA, Simmonds I (2013) Exploring links between arctic amplification and mid-latitude weather. *Geophys Res Lett* 40(5):959–964. <https://doi.org/10.1002/grl.50174>
- Seland Ø, Bentsen M, Olivié D et al (2020) Overview of the Norwegian earth system model (noresm2) and key climate response of cmip6 deck, historical, and scenario simulations. *Geosci Model Dev* 13(12):6165–6200. <https://doi.org/10.5194/gmd-13-6165-2020>
- Sellar AA, Jones CG, Mulcahy JP et al (2019) Ukesm1: description and evaluation of the u.k. earth system model. *J Adv Model Earth Syst* 11(12):4513–4558. <https://doi.org/10.1029/2019MS001739>
- Serreze MC, Barrett AP, Stroeve JC et al (2009) The emergence of surface-based arctic amplification. *Cryosphere* 3:11–19
- Shu Q, Wang Q, Song Z et al (2020) Assessment of sea ice extent in cmip6 with comparison to observations and cmip5. *Geophys*

- Res Lett 47(9):e2020GL087965. <https://doi.org/10.1029/2020GL087965>
- Shu Q, Wang Q, Årthun M et al (2022) Arctic ocean amplification in a warming climate in cmip6 models. *Sci Adv* 8(30):eabn9755. <https://doi.org/10.1126/sciadv.abn9755>
- Storto A, Masina S (2016) C-glorsv5: an improved multipurpose global ocean eddy-permitting physical reanalysis. *Earth Syst m Sci Data* 8(2):679–696. <https://doi.org/10.5194/essd-8-679-2016>
- Storto A, Alvera-Azcárate A, Balmaseda MA et al (2019) Ocean reanalyses: recent advances and unsolved challenges. *Front Marine Sci*. <https://doi.org/10.3389/fmars.2019.00418>
- Storto A, Masina S, Simoncelli S et al (2019) The added value of the multi-system spread information for ocean heat content and steric sea level investigations in the cmems grep ensemble reanalysis product. *Clim Dyn*. <https://doi.org/10.1007/s00382-018-4585-5>
- Stroeve J, Notz D (2018) Changing state of arctic sea ice across all seasons. *Environ Res Lett* 13(10):103001. <https://doi.org/10.1088/1748-9326/aade56>
- Swart NC, Cole JNS, Kharin VV et al (2019) The Canadian earth system model version 5 (canesm5.0.3). *Geosci Model Dev* 12(11):4823–4873. <https://doi.org/10.5194/gmd-12-4823-2019>
- Tatebe H, Ogura T, Nitta T et al (2019) Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in miroc6. *Geosci Model Dev* 12(7):2727–2765. <https://doi.org/10.5194/gmd-12-2727-2019>
- Trenberth KE, Fasullo JT, Mackaro J (2011) Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *J Clim* 24(18):4907–4924. <https://doi.org/10.1175/2011JCLI4171.1>
- Tsubouchi T, Bacon S, Naveira Garabato AC et al (2012) The arctic ocean in summer: a quasi-synoptic inverse estimate of boundary fluxes and water mass transformation. *J Geophys Res Oceans*. <https://doi.org/10.1029/2011JC007174>
- Tsubouchi T, Bacon S, Aksenov Y et al (2018) The arctic ocean seasonal cycles of heat and freshwater fluxes: observation-based inverse estimates. *J Phys Oceanogr* 48(9):2029–2055. <https://doi.org/10.1175/JPO-D-17-0239.1>
- Voldoire A, Saint-Martin D, Sénési S et al (2019) Evaluation of cmip6 deck experiments with cnrm-cm6-1. *J Adv Model Earth Syst* 11(7):2177–2213. <https://doi.org/10.1029/2019MS001683>
- von Schuckmann K, Cheng L, Palmer MD et al (2020) Heat stored in the earth system: where does the energy go? *Earth Syst Sci Data* 12(3):2013–2041. <https://doi.org/10.5194/essd-12-2013-2020>
- Walsh JE (2014) Intensified warming of the arctic: causes and impacts on middle latitudes. *Global Planet Change* 117:52–63. <https://doi.org/10.1016/j.gloplacha.2014.03.003>
- Wang C, Graham RM, Wang K et al (2019) Comparison of era5 and era-interim near-surface air temperature, snowfall and precipitation over arctic sea ice: effects on sea ice thermodynamics and evolution. *Cryosphere* 13(6):1661–1679. <https://doi.org/10.5194/tc-13-1661-2019>
- Wang YC, Hsu HH, Chen CA et al (2021) Performance of the Taiwan earth system model in simulating climate variability compared with observations and cmip6 model simulations. *J Adv Model Earth Syst* 13(7):e2020MS002353. <https://doi.org/10.1029/2020MS002353>
- Wang S, Wang Q, Wang M et al (2022) Arctic ocean freshwater in cmip6 coupled models. *Earth's Future* 10(9):e2022EF002878. <https://doi.org/10.1029/2022EF002878>
- Wild M (2020) The global energy balance as represented in cmip6 climate models. *Clim Dyn*. <https://doi.org/10.1007/s00382-020-05282-7>
- Williams KD, Copsey D, Blockley EW et al (2018) The met office global coupled model 3.0 and 3.1 (gc3.0 and gc3.1) configurations. *J Adv Model Earth Syst* 10(2):357–380. <https://doi.org/10.1002/2017MS001115>
- Winkelbauer S (2023) Straitflux. <https://doi.org/10.5281/zenodo.10053555>
- Winkelbauer S, Mayer M, Seitner V et al (2022) Diagnostic evaluation of river discharge into the arctic ocean and its impact on oceanic volume transports. *Hydrol Earth Syst Sci* 26(2):279–304. <https://doi.org/10.5194/hess-26-279-2022>
- Winkelbauer S, Mayer M, Haimberger L (2023) Straitflux—precise water strait fluxes on various modelling grids. [Manuscript in preparation]
- Wu T, Lu Y, Fang Y et al (2019) The Beijing climate center climate system model (bcc-csm): the main progress from cmip5 to cmip6. *Geosci Model Dev* 12(4):1573–1600. <https://doi.org/10.5194/gmd-12-1573-2019>
- Wu RJ, Lo MH, Scanlon BR (2021) The annual cycle of terrestrial water storage anomalies in cmip6 models evaluated against grace data. *J Clim* 34(20):8205–8217. <https://doi.org/10.1175/JCLI-D-21-0021.1>
- Yukimoto S, Kawai H, Koshiro T et al (2019) The meteorological research institute earth system model version 2.0, mri-esm2.0: description and basic evaluation of the physical component. *J Meteorol Soc Jpn Ser II* 97(5):931–965. <https://doi.org/10.2151/jmsj.2019-051>
- Zanowski H, Jahn A, Holland MM (2021) Arctic ocean freshwater in cmip6 ensembles: declining sea ice, increasing ocean storage and export. *J Geophys Res Oceans* 126(4):e2020JC016930. <https://doi.org/10.1029/2020JC016930>
- Zhang H, Zhang M, Jin J et al (2020) Description and climate simulation performance of cas-esm version 2. *J Adv Model Earth Syst* 12(12):e2020MS002210. <https://doi.org/10.1029/2020MS002210>
- Ziehn T, Chamberlain MA, Law RM et al (2020) The Australian earth system model: Access-esm1.5. *J South Hemisphere Earth Syst Sci*. <https://doi.org/10.1071/ES19035>
- Zuo H, Balmaseda M, Mogensén K (2015) The new eddy-permitting orap5 ocean reanalysis: description, evaluation and uncertainties in climate signals. *Clim Dyn* 49(3):791–811. <https://doi.org/10.1007/s00382-015-2675-1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Recent variations in oceanic transports across the Greenland–Scotland Ridge

Michael Mayer^{1,2,3}, Takamasa Tsubouchi⁴, Susanna Winkelbauer¹, Karin Margretha H. Larsen⁵,
Barbara Berx⁶, Andreas Macrander⁷, Doroteaciro Iovino⁸, Steingrímur Jónsson^{7,9}, and
Richard Renshaw¹⁰

¹Department of Meteorology and Geophysics, University of Vienna, 1090 Vienna, Austria

²Research Department, European Centre for Medium-Range Weather Forecasts (ECMWF),
53175 Bonn, Germany

³b.geos, 2100 Korneuburg, Austria

⁴Atmosphere and Ocean Department, Japan Meteorological Agency (JMA), Tokyo, 105-8431, Japan

⁵Faroe Marine Research Institute, Tórshavn, 100, Faroe Islands

⁶Marine Scotland, Aberdeen, AB11 9DB, United Kingdom

⁷Environmental division, Marine and Freshwater Research Institute, Hafnarfjörður, 220, Iceland

⁸Ocean Modeling and Data Assimilation Division, Centro Euro-Mediterraneo sui Cambiamenti
Climatici (CMCC), Bologna, 40127, Italy

⁹Faculty of Natural Resource Sciences, University of Akureyri, Akureyri, 600, Iceland

¹⁰Met Office, Exeter, EX1 3PB, United Kingdom

Correspondence: Michael Mayer (michael.mayer@univie.ac.at)

Received: 25 July 2022 – Discussion started: 7 October 2022

Revised: 20 January 2023 – Accepted: 20 February 2023 – Published: 27 September 2023

Abstract. Oceanic exchanges across the Greenland–Scotland Ridge (GSR) play a crucial role in shaping the Arctic climate and linking with the Atlantic meridional overturning circulation. Most considered ocean reanalyses underestimate the observed 1993–2020 mean net inflow of warm and saline Atlantic Water of 8.0 ± 0.5 Sv by up to 15 %, with reanalyses at 0.25° resolution additionally exhibiting larger biases in the single inflow branches compared to higher-resolution products. The underestimation of Atlantic Water inflow translates into a low bias in mean oceanic heat flux at the GSR of 5 %–22 % in reanalyses compared to the observed value of 280 ± 18 TW. Interannual variations in reanalysis transports correlate reasonably well with observed transports in most branches crossing the GSR. Observations and reanalyses with data assimilation show a marked reduction in oceanic heat flux across the GSR of 4 %–9 % (compared to 1993–2020 means) during a biennial (2-year-long) period centered on 2018, a record low for several products. The anomaly was associated with a temporary reduction in geostrophic Atlantic Water inflow through the Faroe–Shetland branch and was augmented by anomalously cool temperatures of Atlantic Water arriving at the GSR. The latter is linked to a recent strengthening of the North Atlantic subpolar gyre and illustrates the interplay of interannual and decadal changes in modulating transports at the GSR.

Table 1. CMEMS and non-CMEMS products used in this study, including information on data documentation.

Product ref. no.	Product ID and type	Data access	Documentation
1	GLOBAL_REANALYSIS_PHY_001_031 (GREPv2), numerical models	EU Copernicus Marine Service Product (2022a)	Quality Information Document (QUID): Desportes et al. (2022) Product User Manual (PUM): Gounou et al. (2022)
2	GLOBAL_MULTIYEAR_PHY_001_030 (GLORYS12V1), numerical models	EU Copernicus Marine Service Product (2022b)	Quality Information Document (QUID): Drévilion et al. (2022a) Product User Manual (PUM): Drévilion et al. (2022b)
3	Mooring-derived ocean heat transport into Arctic Mediterranean from 1993 updated to July 2021	Data available at http://metadata.mmc.no/metadata-api/landingpage/0a2ae0e42ef7a767a920811e83784b1 (last access: 21 March 2023); updated time series are available from the authors upon reasonable request.	Tsubouchi et al. (2020, 2021)
Observational input data for product no. 3			
	Iceland–Faroe branch: January 1993 to December 2020	Data for I–F branch, F–S branch, NIIC, and Faroe Bank Channel are available at OceansITES (http://www.oceansites.org/ima/gsr.html , last access: 22 March 2023). Data updates are available from the authors upon reasonable request.	Hansen et al. (2015)
	Faroe–Shetland branch: January 1993 to June 2021		Berx et al. (2013)
	North Icelandic Irminger Current: October 1994 to July 2021		Jónsson and Valdimarsson (2012)
	Faroe Bank Channel: December 1995 to April 2021		Hansen et al. (2016)
	Denmark Strait: May 1996 to 2021		Jochumsen et al. (2017)
	Bering Strait: August 1997 to August 2019	EU Copernicus Marine Service Product (2023)	Woodgate (2018)
	SEALEVEL_GLO_PHY_L4_MY_008_047 (DUACS), satellite observations		Quality Information Document (QUID): Pujol et al. (2023) Product User Manual (PUM): Pujol (2022)
5	Hindcast ocean simulations (no data assimilation) at 0.25° horizontal resolution and 50 vertical levels (GLOB4) and 0.0625° horizontal resolution and 98 vertical levels (GLOB16) provided by CMCC	Data available from the authors upon reasonable request.	Iovino et al. (2023)
6	Ocean reanalysis at 0.25° horizontal resolution and 75 vertical levels provided by UKMO (GloRan V14)	Data available from the authors upon reasonable request.	MacLachlan et al. (2015)

1 Introduction

The Greenland–Scotland Ridge (GSR), encompassing the Denmark Strait, Iceland–Faroe (I-F) branch, Faroe–Shetland (F-S) branch, and European shelf, represents the main oceanic gateway to the so-called Arctic Mediterranean (the ocean bounded by the GSR, Davis Strait, and Bering Strait). Oceanic transports across the GSR play an important role in the Arctic and global climate systems. In the surface layer, the warm and saline Atlantic Water moves northward across the GSR and the light Polar Water flows southward mainly through the Denmark Strait. In the lower layer, the cold and dense water is transported southward at depth into the North Atlantic, contributing to the lower limb of the Atlantic meridional overturning circulation (Hansen and Østerhus, 2000; Buckley and Marshall, 2016).

Transports across the GSR exhibit pronounced interannual variability and thereby play an important role in modulating the heat budget of the Arctic Mediterranean (e.g., Muilwijk et al., 2018; Mayer et al., 2016; Asbjørnsen et al., 2019). Specifically, the inflow of warm and saline Atlantic Water (AW) exhibits a strong co-variability with ocean heat content, especially in the AW layer of the Arctic Mediterranean (M. Mayer et al., 2022). Tsubouchi et al. (2021), using observation-based oceanic transport data (1993–2016), revealed a step change towards stronger oceanic heat transports (OHTs) across the GSR around 2002, suggesting an enhanced contribution of OHT to the observed warming of the Arctic Ocean. M. Mayer et al. (2022) temporally extended the monitoring of OHT at the GSR using ocean reanalyses and found a pronounced reduction in OHT around 2018, which could not be verified with observational data at that time and the causes of which were not explored in detail.

Here, we use observational oceanic transport data at the boundaries of the Arctic Mediterranean updated to 2021 and an extended set of ocean reanalyses to explore the pronounced reduction in OHT in more detail, track it to the main contributing oceanic branch, and relate these changes to larger-scale climate variations on interannual and decadal timescales. An additional aspect of this study is a more detailed validation of reanalysis-based oceanic transports at the GSR at the scale of single branches to further build trust in the usefulness of these products for monitoring Arctic climate and its oceanic drivers.

2 Data and methods

We use monthly data from a comprehensive set of ocean reanalyses to compute oceanic transports across the GSR, Davis Strait, and Bering Strait. The latter two straits are calculated to close the volume budget and obtain unambiguous net heat transport into the Arctic Mediterranean (Schauer and Beszczynska-Möller, 2009). Inflow (positive) has been defined as positive towards the Arctic Mediterranean. The employed products are an updated ensemble based on the

Copernicus Marine Environment Monitoring Service Global Reanalysis Ensemble Product (CMEMS GREPv2, product ref 1), consisting of ORAS5, CGLORS, GLORYS2V4, and GloRanV14 (an improvement of GloSea5, also known as the FOAM product, product ref 6; MacLachlan et al., 2015). These are all run at 0.25° horizontal resolution with 75 vertical levels and use atmospheric forcing from ERA-Interim (Dee et al., 2011). The ensemble is complemented with GLORYS12 version 1 (product ref 2), a reanalysis at 0.083° horizontal resolution with 50 vertical levels driven by ERA-Interim atmospheric forcing. Furthermore, two hind-cast ocean simulations (i.e., with no data assimilation) at 0.25° horizontal resolution with 50 vertical levels (GLOB4) and 0.0625° horizontal resolution and 98 vertical levels (GLOB16; Iovino et al., 2023) driven by JRA55-do (Tsujino et al., 2018) are employed to investigate the impact of resolution on oceanic volume fluxes. Heat transports from GLOB4 and GLOB16 are not assessed as these are biased due to their setup as forced runs without data assimilation. Transports are computed on the native grid through line integrals similar to Pietschnig et al. (2018).

Observational mass-consistent estimates of oceanic transports (product ref 3) are updated to July 2021 following Tsubouchi et al. (2021; i.e., using the same strategy to infill data gaps, estimate uncertainty, and create a box inverse model to close the volume budget). Temporal coverage and references for the single observational estimates used as input are provided in the data table (Table 1). Surface freshwater inputs by river discharge and precipitation minus evaporation for 1993–2021 used as input to the box inverse model are based on Winkelbauer et al. (2022). The used reanalyses assimilate temperature and salinity profiles available from databases such as Hadley EN4 (Good et al., 2013), which according to our investigations include only a small subset of the mooring data used for our observational transport estimates. Currents are generally not assimilated in ocean reanalyses. Hence, the observation-based volume fluxes represent fully independent data, while temperature fluxes represent largely independent validation data.

We note that quantification methods of oceanic transports in reanalyses and observations are fundamentally different, which needs to be kept in mind when intercomparing. The reanalysis-based estimate is based on surface-to-bottom, coast-to-coast temperature and velocity sections across the Arctic Mediterranean. This ensures conservation of volume and avoids projection of potentially biased positioning of currents in the reanalyses onto the transport estimates. The observational estimate is based on the sum of 11 major ocean current transport estimates categorized into three major water masses – Atlantic Water (AW), Polar Water (PW), and Overflow Water (OW) (Tsubouchi et al., 2021). An assumption is that the 11 major ocean currents represent the major water mass exchanges well across the Arctic Mediterranean. This means it is important that no recirculation, e.g., of AW waters, remains unobserved, as this would introduce

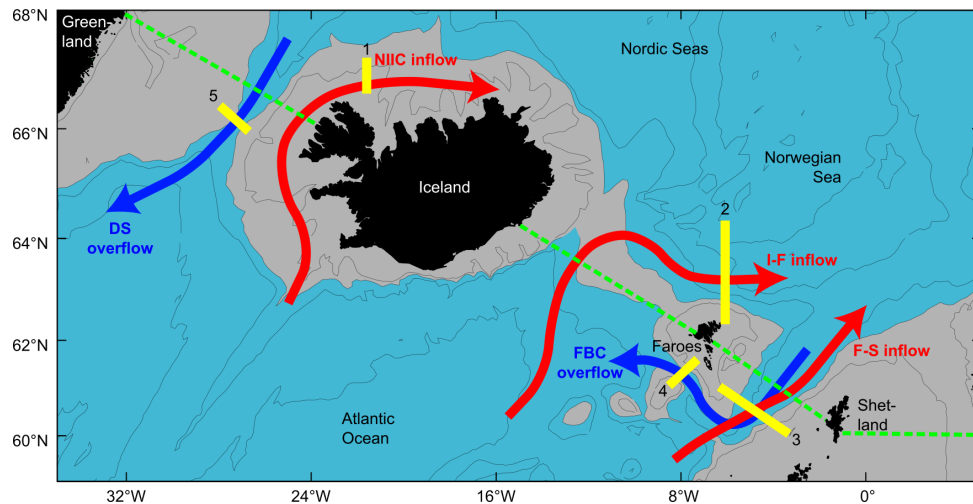


Figure 1. Map of the Greenland–Scotland Ridge region, schematically depicting shallow shelf water and deep water, major AW inflow (red arrows) and OW outflow (blue arrows) branches (DS stands for Denmark Strait, and FBC stands for Faroe Bank Channel), the location of oceanic moorings used for deriving the observation-based transport estimates (yellow bars), and the section used for the computation of transports from ocean reanalyses (dashed green line). Adapted from OceanSITES (<http://www.oceansites.org/tma/gsr.html>, last access: 20 March 2023, Fig. 1).

biases into the observational estimate. This assumption has been assessed and confirmed many times over the last 2 decades from establishment of sustained hydrographic sections in the GSR in the 1990s (e.g., Dickson et al., 2008) to recent oceanographic surveys to capture ocean circulation in the GSR for AW (e.g., Berx et al., 2013; Hansen et al., 2017; Rossby et al., 2018; Jónsson and Valdimarsson, 2012) and OW (Hansen et al., 2018). We also note that remaining uncertainties arising from potential undersampling are taken into account in the observational estimate obtained through the inverse model. For reference, Fig. 1 shows pathways of major flows across the GSR and the locations of considered oceanic moorings and the sections used for evaluation of reanalyses.

As in M. Mayer et al. (2022), we assume total uncertainties in monthly mean observations (provided in Tsubouchi et al., 2021) to consist of roughly half systematic and half random errors; i.e., the two contributions are the total uncertainty reduced by a factor of $1/\sqrt{2}$. Consequently, the contribution of random errors to uncertainties in long-term mean observational estimates is further reduced by a factor of $1/\sqrt{N}$, where N is the number of years, and deseasonalized anomalies only include the random errors.

Transported water masses at the GSR are decomposed into AW, PW, and OW, largely following Eldevik et al. (2009). PW is defined as $T < 4^{\circ}\text{C}$ and $\sigma_{\theta} < 27.7\text{ kg m}^{-3}$. OW is defined as $\sigma_{\theta} > 27.8\text{ kg m}^{-3}$. The rest (i.e., waters with $\sigma_{\theta} < 27.8\text{ kg m}^{-3}$ with PW taken out) is considered AW. Note that these definitions have been revised from M. Mayer et al. (2022). Water mass decomposition is performed each month based on the monthly temperature and salinity fields in the reanalyses. These definitions are similar to those used

for observational products (see references for more details, e.g., Eldevik et al 2009).

We additionally use sea level anomaly (SLA) data provided through CMEMS (product ref 4) for investigating drivers of observed OHT anomalies. The global mean SLA trend is removed before computation of the presented diagnostics.

Deseasonalized anomalies are based on the 1993–2019 climatologies, i.e., the period for which all data are available. Anomaly time series have a 12-monthly temporal smoother applied to emphasize interannual variations. Confidence levels (95 % is set as threshold for significance testing) for temporal correlations take auto-correlation of the involved time series into account (see Oort and Yienger, 1996).

3 Results

3.1 General evaluation of transports of water masses across the GSR

Figure 2 presents mean annual cycles and anomaly time series of relevant oceanic transport quantities at the GSR. It is complemented with long-term averages shown in Table 2 (volume fluxes) and Table 3 (heat fluxes). Observations show seasonally varying AW inflow across the GSR ($8.0 \pm 0.5\text{ Sv}$; mean \pm standard deviation of the mean reported throughout, unless explicitly stated) with a maximum in December and minimum in June–July (Fig. 2a). The AW inflow is largely balanced by PW (Fig. 2c; $-1.8 \pm 0.7\text{ Sv}$) and OW (Fig. 2b; $-5.6 \pm 0.3\text{ Sv}$) outflow, yielding a relatively small net volume flux across the GSR of 0.7 Sv (balanced by flows through Bering and Davis straits). The PW outflow exhibits

an annual cycle balancing the AW inflow (i.e., maximum outflow in boreal winter), while the OW exchange is more stable throughout the year (i.e., small annual cycle).

All reanalyses but one (CGLORS) underestimate AW inflow across the GSR when compared to observations, but the shape of the annual cycle of all estimates is in good agreement with observations. The AW net flow from high-resolution products does not stand out compared to the 0.25° products and is close to the GREP mean. Agreement is also good for the PW outflow, where all products show the observed seasonal maximum flow in boreal winter. The range of reanalysis-based estimates is large in a relative sense, with the observations lying in the middle of the range. There is less coherence across products concerning the OW transports. GLORYS12, GLORYS2V4, and CGLORS are close to observations, with rather persistent overflow on the order of -5.4 Sv and seasonal variations that agree with the observations. Other products tend to have overflows that are too weak (most notably ORAS5), and some also exhibit biases in the representation of the annual cycle (e.g., GloRanV14).

Table 2 additionally includes long-term average volume fluxes in the main AW inflow branches (North Icelandic Irminger Current (NIIC), I-F branch, F-S branch, and European shelf). The reanalysis-based estimates generally agree well with observations. The main discrepancy is the underestimation of I-F inflow and overestimation of F-S inflow by all GREP reanalyses, while the high-resolution products GLORYS12 and GLOB16 are in much better agreement with observations. Direct comparison of GLOB16 to GLOB4 confirms that overestimation of F-S volume flux is reduced when going from 0.25 to 0.0625° resolution. This suggests that increased resolution, along with more realistic bathymetry, improves representation of inflow pathways in the reanalyses. We also note that temporal anomaly correlations with observed I-F volume fluxes are low (Pearson correlation coefficients r range from -0.03 to 0.26 and are statistically insignificant) for all reanalyses but are substantially higher for F-S volume fluxes (r ranges from 0.30 to 0.71 and is statistically significant for all products with data assimilation).

Figure 2d shows the mean annual cycle of heat flux across the GSR, i.e., the sum of sensible heat transported by all waters crossing the GSR. The mean annual cycle of GSR heat fluxes generally follows that of AW volume fluxes, with a minimum between boreal spring and early summer and a maximum in fall and early winter. Seasonal minima and maxima in GSR heat flux co-occur with those of AW volume flux; i.e., seasonal variations in heat flux are largely volume flux driven, and the seasonal cycle in volume-weighted temperatures is in phase.

Since net volume flux across the GSR is small, the ambiguity arising from the choice of reference temperature can be considered small as well. However, for the long-term averages we focus on net heat transport into the Arctic Mediterranean, i.e., the sum of heat fluxes across the GSR, plus those through Bering and Davis straits. Values in Table 3 show that

all reanalyses exhibit lower net heat transport (by $\sim 16\%$ for the GREP mean of 256 ± 19 TW) than that observed (306 ± 19 TW; 311 ± 20 TW when including sea ice). GLORYS12 exhibits a mean net heat flux similar to the GREP mean, but we note that all GREP reanalyses have AW inflow that is too strong in the F-S branch (Table 2), where the climatologically warmest waters cross the GSR, and thus enhance the heat flux in those products for the wrong reason.

The net heat transport is clearly related to the strength of AW inflow, but even CGLORS, which has a higher AW mean volume flux (8.3 Sv) than observations, has a negative net heat flux bias. The reason is that all reanalyses exhibit a warm bias in outflowing OW (not shown) and a cold bias in Davis Strait inflow (see Pietschnig et al., 2018). Based on M. Mayer et al. (2022) and taking oceanic storage into account, the energy-budget-based estimate of the net heat boundary transport suggests even higher values (~ 348 TW) than observations. This inferred value appears high, but we note that this indirect approach has been applied successfully to infer observation-based oceanic transports in the North Atlantic (Trenberth and Fasullo, 2017; Liu et al., 2020; J. Mayer et al., 2022) and the central Arctic (Mayer et al., 2019), and hence it is deemed credible. This estimate at least adds confidence to the conclusion that the observational estimate of oceanic heat transport is not biased high. Table 3 also confirms that long-term averages for heat flux across the GSR are qualitatively very similar to the net heat transport; i.e., heat fluxes across the GSR are the dominant contributor to oceanic heat transport into the Arctic Mediterranean.

Figure 2e shows deseasonalized anomalies of AW volume flux, with a 12-monthly smoother applied to emphasize interannual variability. Typical variability is similar across observations (temporal standard deviation $\sigma = 0.35$ Sv) and reanalyses (σ ranges between 0.22 and 0.47 Sv). Temporal correlations between reanalyzed and observed AW inflow anomalies are moderately high (r ranges between 0.39 and 0.60 and is statistically significant for all products; see the legend of the plot for values).

Figure 2f shows anomalies of total oceanic heat flux across the GSR, which show similar variability to AW volume flux; i.e., the strength of AW inflow modulates not only the seasonal cycle of the total GSR heat flux but also its interannual variations (r ranges from 0.86 to 0.91). GSR total heat flux from reanalyses is in slightly better agreement with observations than AW volume fluxes (r ranges between 0.51 and 0.63 and is statistically significant for all products; see the legend of the plot for values). Figure 2f also shows a prominent negative heat flux anomaly centered around the year 2018, which has already been noted by M. Mayer et al. (2022) for net heat transport into the Arctic Mediterranean.

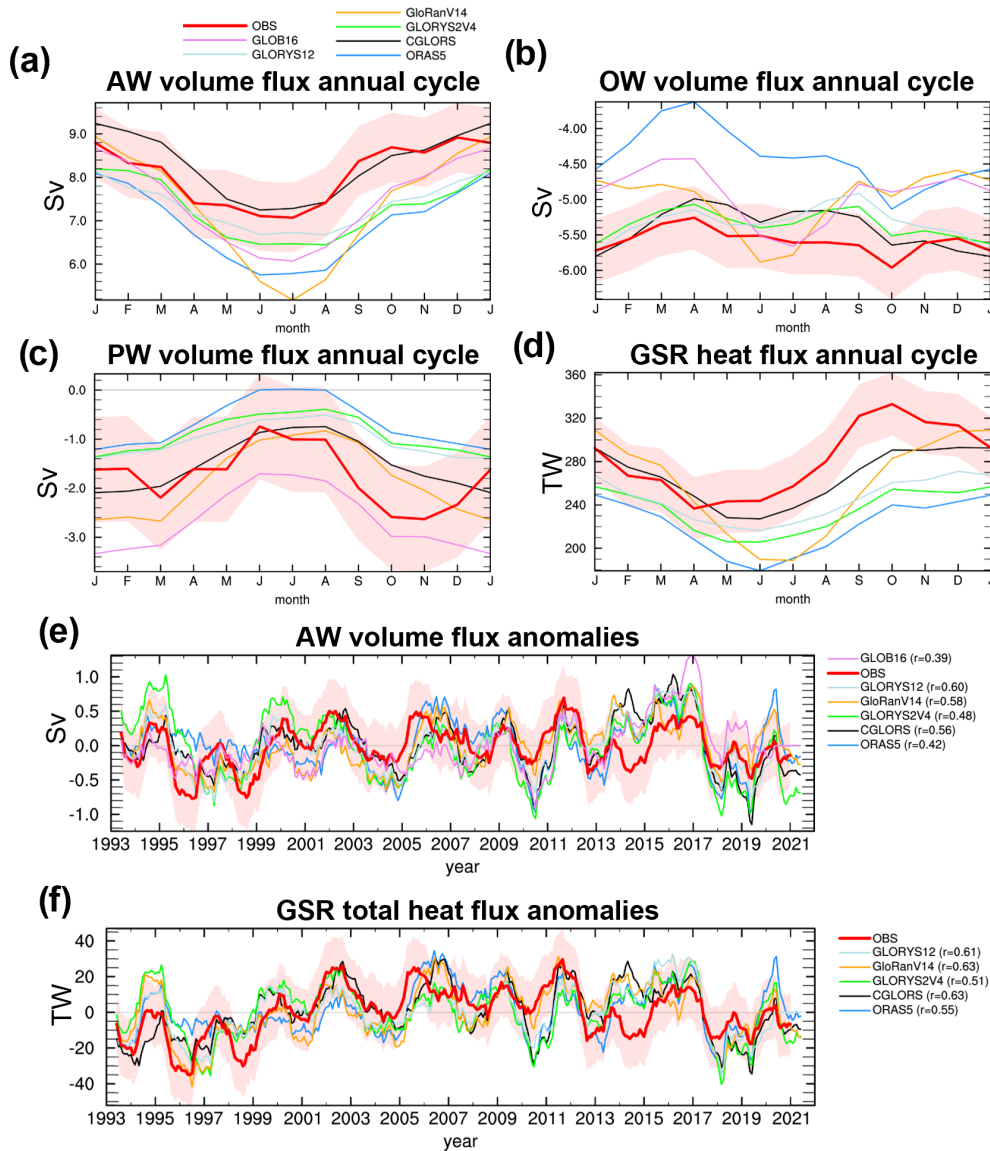


Figure 2. Mean annual cycle of (a) AW volume flux, (b) OW volume flux, (c) PW volume flux, and (d) GSR total (AW + PW + OW) heat flux. Temporal anomalies of (e) AW volume flux and (f) GSR total heat flux. The red shading indicates ± 1 standard error in the observational data. Temporal correlations of reanalyses with observations are provided in the legends.

3.2 Spatial structure of the July 2017–June 2019 transport anomaly

To set the scene for further investigation, we present climatological temperatures and currents at the GSR based on GLOBYS12 in Fig. 3a and b, respectively. Comparison with analogous figures based on the GREP (shown in M. Mayer et al., 2022) allows us to appreciate the benefits of increased resolution (0.08 vs. 0.25° resolution), including a more distinct representation of inflow and outflow branches and a more spatially variable bathymetry, especially in the I-F branch.

Next, we investigate the recent reduction in AW volume and GSR heat fluxes in more detail. This is most prominent

in the biennial (2-year-long) signal of average anomalies in July 2017–June 2019. Figure 3c shows that during this period strong warm anomalies were present over 0 – 400 m depth in the eastern Denmark Strait, and warm anomalies are also seen in the F-S branch. The latter suggests a temporary deepening of the AW layer. Velocity anomalies for the July 2017–June 2019 period (Fig. 3d) suggest that the positive temperature anomalies in the eastern Denmark Strait are driven by enhanced NIIC transports. The strongest and deepest velocity anomaly in July 2017–June 2019 is located in the eastern part of the F-S branch, where reduced inflow is present from the surface down to the interface at ~ 600 m, and hence we focus on this branch next. Negative anomalies are also seen in

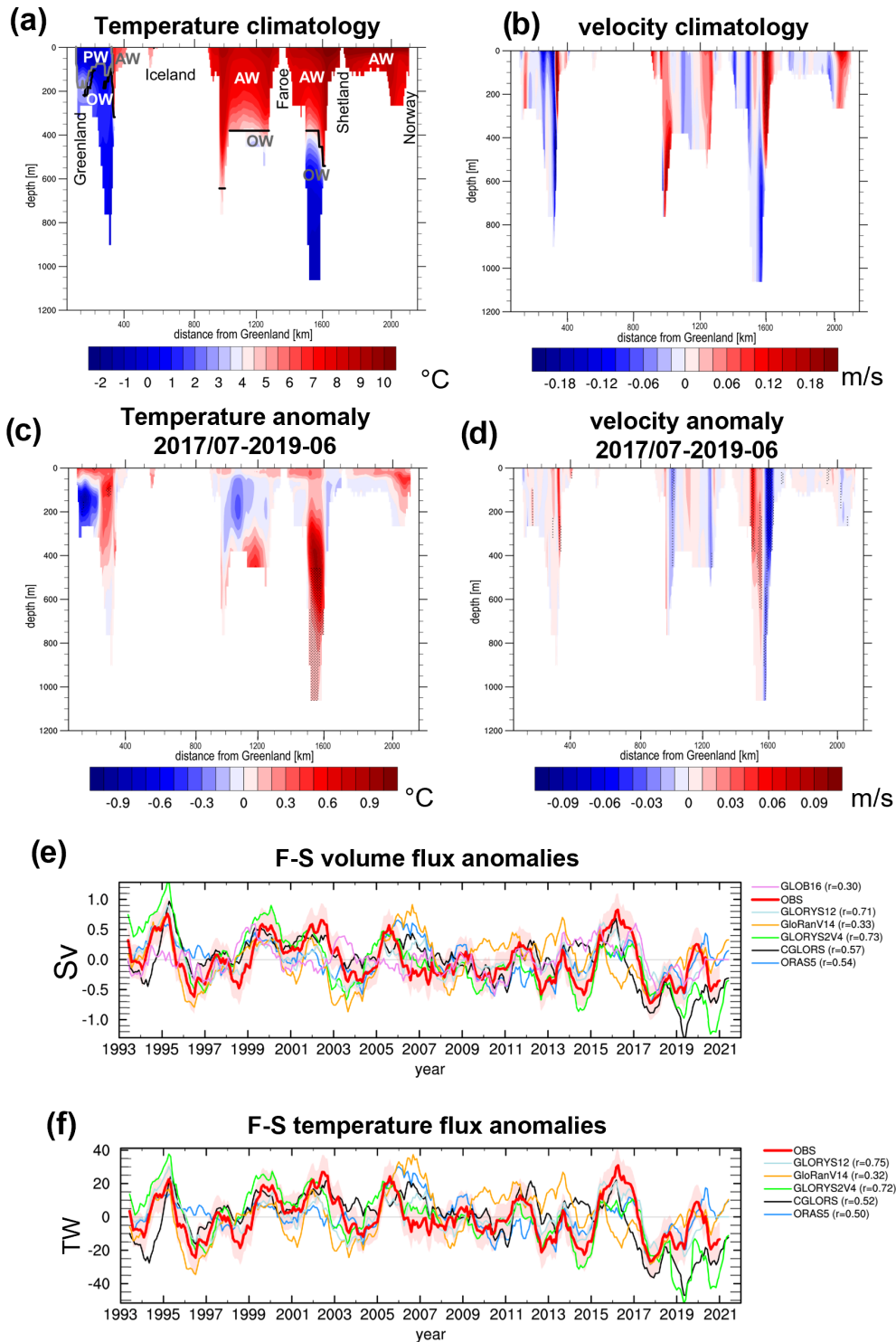


Figure 3. Longitude–depth sections of (a) mean temperature (with water mass boundaries indicated), (b) mean velocity, (c) July 2017–June 2019 anomalous temperature, and (d) July 2017–June 2019 anomalous velocity across the GSR based on GLORYS12 (stippling denotes grid cells where anomalies are $> 2\sigma$ of biennial anomalies). Note that the section does not go along the shallowest part of the Denmark Strait and I-F Ridge everywhere, leading to deeper trenches in some places (see also Fig. 1). Time series of (e) anomalous volume and (f) anomalous temperature flux through the Faroe–Shetland branch, where the red shading indicates ± 1 standard error in the observational data. Temporal correlations of reanalyses with observations are provided in the legends.

Table 2. Long-term mean of volume flux in different water masses and branches across the GSR. Observational European shelf volume fluxes are based on Østerhus et al. (2019). All values are based on 1993–2020 (since observational data do not completely cover 2021) data, except for GLOB16 and GLOB4 (based on 1993–2019 data). Values are given in Sv.

	GSR total	PW	OW	AW	AW NIIC	AW I-F	AW F-S	Shelf
Observations	0.7 ± 0.9	-1.8 ± 0.7	-5.6 ± 0.3	8.0 ± 0.5	0.9 ± 0.1	3.8 ± 0.3	2.7 ± 0.3	0.6 ± 0.3
GREP mean	1.2 ± 0.6	-1.2 ± 0.5	-5.0 ± 0.5	7.4 ± 0.6	0.9 ± 0.2	2.5 ± 0.4	3.5 ± 0.6	0.5 ± 0.2
ORAS5	1.8	-0.6	-4.4	6.8	1.2	2.6	2.8	0.2
CGLORS	1.4	-1.5	-5.4	8.3	0.8	2.8	3.9	0.8
GLORYS2V4	1.0	-0.9	-5.4	7.3	0.8	2.6	3.3	0.6
GloRanV14	0.4	-1.8	-5.0	7.2	0.9	1.9	4.1	0.3
GLORYS12	1.0	-1.0	-5.3	7.3	1.0	3.3	2.4	0.6
GLOB16	-0.2	-2.6	-4.9	7.4	0.6	3.2	3.0	0.6
GLOB4	1.2	-2.7	-4.4	8.3	0.8	3.2	3.7	0.6

Table 3. Long-term mean of total heat flux across the GSR and into the Arctic Mediterranean (total GSR transports plus Bering Strait and Davis Strait transports). The energy-budget-based transport estimate is taken from M. Mayer et al. (2022) and by definition can only be provided for a closed area like the Arctic Mediterranean. All values are based on 1993–2020 data. Values are given in TW.

	GSR total	Arctic Mediterranean
Observations	280 ± 18	306 ± 19
GREP mean	243 ± 21	256 ± 19
ORAS5	219	239
CGLORS	265	276
GLORYS2V4	234	242
GloRanV14	255	269
GLORYS12	243	252
Based on the energy budget	–	348

the overflow from 600 to 1000 m depth. There is some compensation by positive velocity anomalies in the western F-S (meaning reduced southward flow there), but the effect of the eastern F-S anomaly dominates, and the net F-S volume flux anomaly was clearly reduced during this period (see below). We note that these main features are also similar in anomaly sections based on the GREP ensemble mean (not shown).

Observations and all reanalyses show large negative F-S volume inflow anomalies of -0.61 , -0.56 , -0.34 , and -0.24 Sv in the observations, GREP_mean, GLORYS12, and GLOB16, respectively, during July 2017–June 2019 (Fig. 3e). In four out of seven datasets, this is the overall biennial minimum of the 1993–2021 record (not shown). The total AW volume flux anomaly for that period was -0.24 , -0.53 , -0.37 , and $+0.08$ Sv in the observations, GREP_mean, GLORYS12, and GLOB16, respectively. Thus, F-S volume flux anomalies were partly compensated by other AW branches, with no very clear signal in any of them (not shown). Only GLOB16 exhibits a positive AW volume flux anomaly during that period, which appears

to be related to a shift towards generally higher AW volume flux around 2016 (see Fig. 2e). This is not seen in any of the other products.

Temperature transport anomalies in the F-S branch are strongly correlated with volume flux anomalies (compare Fig. 3e and f), and there is a clear reduction in F-S branch temperature flux in July 2017–June 2019 of -21.9 , -24.1 , and -16.1 TW in the observations, GREP_mean, and GLORYS12, respectively. The contribution of temperature anomalies in the F-S branch during that time was small, with biennial anomalies of observed volume-weighted temperatures between -0.2 and -0.1 K (similar for all products). The total AW heat flux anomaly for July 2017–June 2019 was -11.0 , -22.5 , and -19.4 TW in observations, GREP_mean, and GLORYS12, respectively. Thus, the reduction in GSR heat transports during that period was mainly driven by a reduction in volume inflow through the F-S branch. It was only partly offset by compensating transport anomalies in other branches. Very similar biennial heat transport anomalies for the total GSR (-11.1 , -20.9 , and -19.3 TW for observations, GREP_mean, and GLORYS12, respectively) confirm AW as the main driver of heat flux variability across the GSR.

3.3 Relationships between sea level and AW inflow

For a better understanding of mechanisms contributing to the GSR transport anomaly around 2018, we first consider basic statistical relationships between SLA and oceanic quantities. We find a statistically significant temporal correlation of the zonal SLA gradient at GSR with observed AW volume flux anomalies (Fig. 4a), which is plausible in terms of geostrophic balance. The correlation pattern looks very similar when performed with the F-S branch volume flux alone (not shown). The pattern of temporal correlation between the SLA field and observed anomalies of volume-weighted temperature of AW transports (Fig. 4b) is distinct from the relationship with AW volume flux (compare Fig. 4a). It emphasizes SLA in the North Atlantic subpolar gyre (SPG),

with higher SLA in the SPG (i.e., a weaker gyre) associated with higher volume-weighted temperature and vice versa. Although correlations are high (r up to 0.69) in the SPG region, they are not statistically significant. The cause may be the low number of degrees of freedom, as the SLA in the SPG exhibits high temporal auto-correlation (see Fig. 4d discussed below).

Actual SLA anomalies averaged over July 2017–June 2019 (Fig. 4c) indicate a weakened zonal SLA gradient at the GSR, albeit not very strongly pronounced, and anomalously low SLA in the SPG region. According to the correlation patterns discussed above, these two features suggest reduced AW volume flux (as suggested by patterns in Fig. 4a) and anomalously low AW volume-weighted temperatures (as suggested by patterns in Fig. 4b). We also note the positive SLA anomalies north of the GSR with a maximum in the central Nordic Seas.

To put these results in context, we define two SLA-based indices (shown in Fig. 4d) from the correlation patterns found in Fig. 3a and b. The similarity of the correlations in Fig. 4a to the correlations between total OHT at the GSR and SLA shown in M. Mayer et al. (2022) reinforces use of their gradient-based index (i.e., standardized SLA difference between $58\text{--}60^\circ\text{N}$, $2\text{--}0^\circ\text{W}$ and $63\text{--}67^\circ\text{N}$, $20\text{--}15^\circ\text{W}$). This index is correlated with AW volume flux anomalies ($r = 0.62$ for observed transports, ranging from 0.47 to 0.70 for reanalyses, which is statistically significant in all cases). The second index uses spatial SLA averages in the North Atlantic region ($55\text{--}60^\circ\text{N}$ and $40\text{--}15^\circ\text{W}$) as an inverse proxy for SPG strength. This index is correlated with anomalies of volume-weighted temperatures of AW transports. The two indices in Fig. 4d show different characteristics, with the SPG index varying on decadal timescales, while the gradient index shows stronger interannual variations. The SPG index has been negative since ~ 2014 , which suggests a strong SPG and lower inflow temperatures in recent years, in agreement with results by Hátún and Chafik (2018). This SPG index exhibits two minima during the July 2017–June 2019 period, although these are not extreme relative to the entire time series.

Figure 4e shows volume-weighted temperatures of AW waters from different products and confirms their overall decrease in recent years. Comparison with the SPG index in Fig. 4d suggests a generally delayed response of volume-weighted temperature in Atlantic inflow water at the GSR to SPG strength. Lagged correlation analysis indeed suggests positive correlations peaking (at values around 0.3 to 0.5, depending on the dataset) when temperatures are lagging the SPG index $\sim 2\text{--}3$ years (not shown), but the correlations are not significant due to the high auto-correlation of time series.

4 Conclusions

Reanalysis-based oceanic transports show generally good agreement with observations on the scale of single branches of the GSR, in terms of both mean and variability in volume and heat fluxes. There is some indication that the higher-resolution products have a better representation of AW inflow in the I-F and F-S branches. All considered products underestimate net heat flux into the Arctic Mediterranean. The magnitude of the low bias is correlated with the strength of AW volume flux, but a warm bias in OW and cold bias in Davis Strait inflow further add to the found net heat flux bias. We reiterate that reanalysis-based and observational transport estimates are obtained in different ways (closed line integrations versus measurements from 11 branches with an inverse model applied), but, as elaborated in Sect. 2, we deem this a fair and robust approach for an intercomparison. The energy-budget-based estimate from M. Mayer et al. (2022) suggests even higher net heat flux than oceanic observations, which confirms the underestimation of heat transports by the ocean reanalyses. A much smaller discrepancy was found in an analogous comparison of observed oceanic transports into the central Arctic and an energy-budget-based estimate (Mayer et al., 2019), potentially reflecting the different observational strategies in Fram Strait and the Barents Sea Opening compared to the GSR (see, e.g., Dickson et al., 2008) or potential biases in the employed energy budget fields over the Nordic Seas.

All reanalyses with data assimilation and observations show a pronounced reduction in OHT during the 2-year period July 2017–June 2019, with some recovery after that. Comparison of observed SLA patterns during this period with statistical relationships between SLA and oceanic transports suggests that this reduction arose from a combination of interannual- (i.e., reduced zonal SLA gradient at the GSR) and decadal-scale changes (i.e., strong SPG in recent years). Another potential factor contributing to the OHT reduction during July 2017–June 2019 may have been the positive SLA anomalies centered in the Nordic Seas (Fig. 4c), which Chatterjee et al. (2018) have related to a weakened gyre circulation in the Nordic Seas and may have contributed to the weakened AW inflow as well.

Our results also reveal a delayed response of AW inflow temperatures to SPG strength. This is consistent with earlier studies finding anti-correlation between SPG strength and GSR heat transport (Häkkinen et al., 2011; Hátún et al., 2005). Specifically, the generally weaker SPG during ~ 1997 and ~ 2014 (with more pronounced minima in 1997–1998, 2010–2011, and 2003–2006; see Hátún and Chafik, 2018) was associated with warm inflow temperatures and stronger OHT after 2001 (Tsubouchi et al., 2021). After that point, the SPG strengthened and the inflow temperatures declined, which is also consistent with generally reduced oceanic heat transports in recent years.

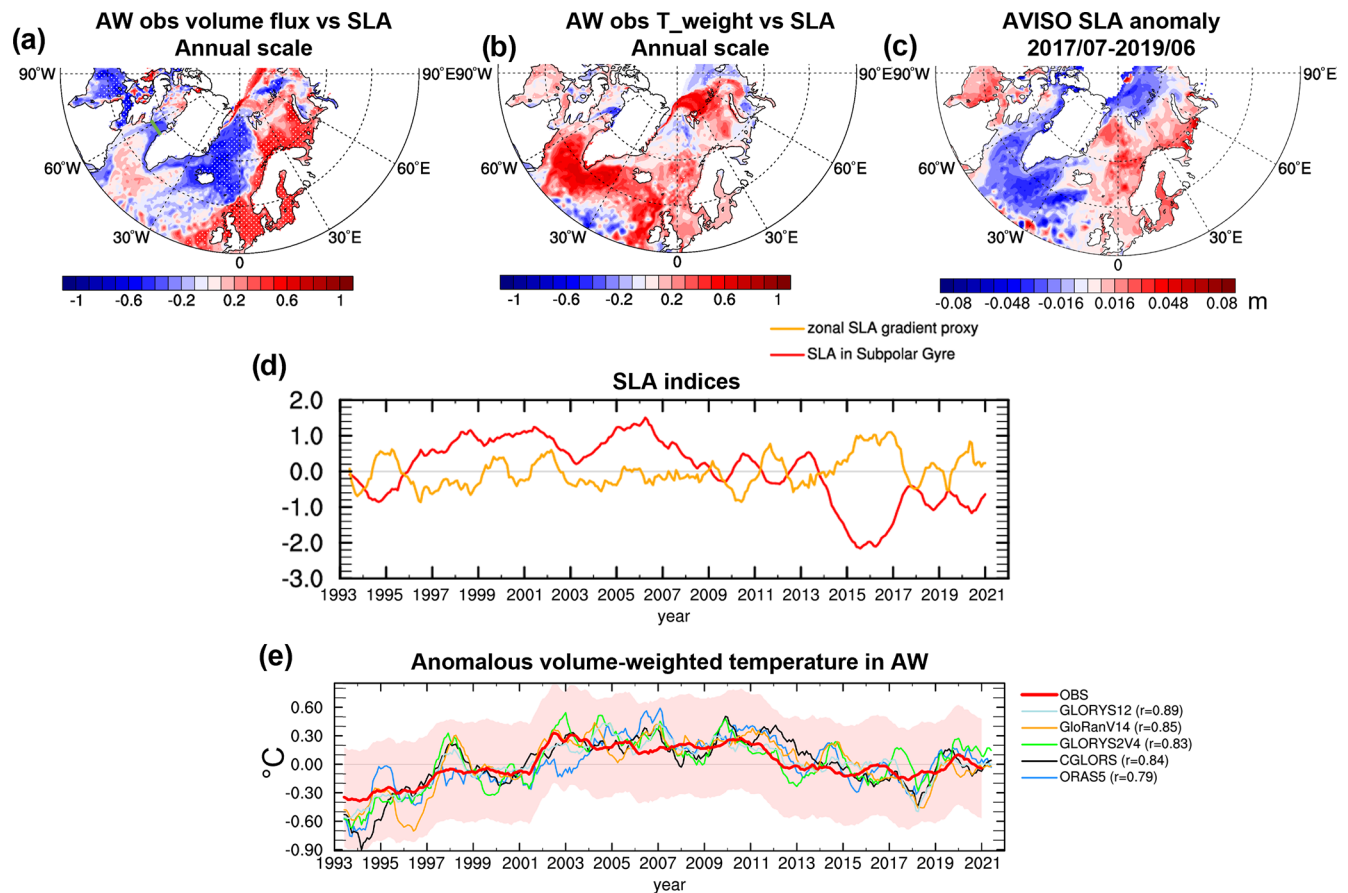


Figure 4. Correlation of observation-based (a) AW volume flux and (b) AW flux-weighted temperature with SLA with a 12-monthly smoother applied. (c) SLA anomaly in July 2017–June 2019. (d) Temporal evolution (standardized) of two SLA-based indices based on the zonal SLA gradient at GSR and the SLA in the SPG region. (e) Temporal evolution of volume-weighted temperature (T_{weight}) anomalies of Atlantic waters at the GSR (red shading indicates ± 1 standard error in the observational data, and temporal correlations of reanalyses with observations are provided in the legend). Stippling in (a) and (b) denotes statistically significant correlations on the 95 % confidence level.

The presented results indicate that decadal predictions of the SPG strength, which have been shown to exhibit skill (e.g., Robson et al., 2018; Borchert et al., 2021), may also allow us to infer near-term trends in OHT across the GSR. Another implication is that the strong interannual-to-decadal variability in OHT across the GSR hampers detection of longer-term (forced and unforced) trends in observed OHT, an aspect in which climate simulations show large spread (Burgard and Notz, 2017). Continued in situ monitoring of OHT, complemented with reanalysis efforts, is thus needed to provide observationally constrained time series of sufficient length for climate model validation.

Data availability. The data products used in this article, as well as their names, availability, and documentation, are summarized in Table 1.

Author contributions. MM and TT conceptualized the study. All co-authors were involved with data preparation and analysis, as well as interpretation of results. MM prepared the manuscript with contributions from all co-authors.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The Iceland–Faroe branch and Faroe Bank Channel overflow data collections received funding from the Danish Ministry of Climate, Energy and Utilities through its climate support program to the Arctic. The Denmark Strait overflow time series was generated by the Institute of Oceanography, Hamburg, and the Marine and Freshwater Research Institute (Iceland). Observational

data collection was further supported through funding from Nordic WOCE, VEINS, MOEN, ASOF-W, NACLIM, RACE II, RACE-Synthese, THOR, AtlantOS, and Blue-Action (EU Horizon 2020 grant agreement no. 727852).

Financial support. This research has been supported by the Austrian Science Fund (grant no. P33177) and Copernicus Marine Service (grant no. 21003-COP-GLORAN Lot 7).

Review statement. This paper was edited by Gilles Garric and reviewed by two anonymous referees.

References

- Asbjørnsen, H., Årthun, M., Skagseth, Ø., and Eldevik, T.: Mechanisms of ocean heat anomalies in the Norwegian Sea, *J. Geophys. Res.-Oceans*, 124, 2908–2923, 2019.
- Berx, B., Hansen, B., Østerhus, S., Larsen, K. M., Sherwin, T., and Jochumsen, K.: Combining in situ measurements and altimetry to estimate volume, heat and salt transport variability through the Faroe–Shetland Channel, *Ocean Sci.*, 9, 639–654, <https://doi.org/10.5194/os-9-639-2013>, 2013.
- Borchert, L. F., Menary, M. B., Swingedouw, D., Sgubin, G., Hermanson, L., and Mignot, J.: Improved decadal predictions of North Atlantic subpolar gyre SST in CMIP6, *Geophys. Res. Lett.*, 48, e2020GL091307, <https://doi.org/10.1029/2020GL091307>, 2021.
- Buckley, M. W. and Marshall, J.: Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review, *Rev. Geophys.*, 54, 5–63, 2016.
- Burgard, C. and Notz, D.: Drivers of Arctic Ocean warming in CMIP5 models, *Geophys. Res. Lett.*, 44, 4263–4271, 2017.
- Chatterjee, S., Raj, R. P., Bertino, L., Skagseth, Ø., Ravichandran, M., and Johannessen, O. M.: Role of Greenland Sea gyre circulation on Atlantic water temperature variability in the Fram Strait, *Geophys. Res. Lett.*, 45, 8399–8406, 2018.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Desportes, C., Garric, G., Régnier, C., Drévilion, M., Parent, L., Drillet, Y., Masina, S., Storto, A., Mirouze, I., Cipollone, A., Zuo, H., Balmaseda, M., Peterson, D., Wood, R., Jackson, L., Mulet, S., Grenier, E., and Gounou, A.: EU Copernicus Marine Service Quality Information Document for the Global Ocean Physics Reanalysis, GLOBAL_REANALYSIS_PHY_001_031, Issue 1.1, Mercator Ocean International, <https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-GLO-QUID-001-031.pdf> (last access: 31 March 2023), 2022.
- Dickson, R., Meincke, J., and Rhines, P.: Arctic-Subarctic Ocean Fluxes: Defining the Role of the Northern Seas in Climate, Springer, Dordrecht, the Netherlands, ISBN 978-1-4020-6774-7, 2008.
- Drévilion, M., Lellouche, J.-M., Régnier, C., Garric, G., Bricaud, C., Hernandez, O., and Bourdallé-Badie, R.: EU Copernicus Marine Service Quality Information Document for the Global Ocean Physics Reanalysis, GLOBAL_REANALYSIS_PHY_001_030, Issue 1.6, Mercator Ocean International, <https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-GLO-QUID-001-030.pdf> (last access: 31 March 2023), 2022a.
- Drévilion, M., Fernandez, E., and Lellouche, J.-M.: EU Copernicus Marine Service Product User Manual for the Global Ocean Physics Reanalysis, GLOBAL_REANALYSIS_PHY_001_030, Issue 1.4, Mercator Ocean International, <https://catalogue.marine.copernicus.eu/documents/PUM/CMEMS-GLO-PUM-001-030.pdf> (last access: 31 March 2023), 2022b.
- Eldevik, T., Nilsen, J. E. Ø., Iovino, D., Anders Olsson, K., Sandø, A. B., and Drange, H.: Observed sources and variability of Nordic seas overflow, *Nat. Geosci.*, 2, 406–410, 2009.
- EU Copernicus Marine Service Product: Global Ocean Ensemble Physics Reanalysis v2, Mercator Ocean International [data set], <https://doi.org/10.48670/moi-00024>, 2022a.
- EU Copernicus Marine Service Product: Global Ocean Physics Reanalysis GLORYS12V1, Mercator Ocean International [data set], <https://doi.org/10.48670/moi-00021>, 2022b.
- EU Copernicus Marine Service Product: Global Ocean Gridded L 4 Sea Surface Heights And Derived Variables Reprocessed 1993 Ongoing, Mercator Ocean International [data set], <https://doi.org/10.48670/moi-00148>, 2023.
- Gounou, A., Drévilion, M., and Clavier, M.: EU Copernicus Marine Service Product User Manual for the Global Ocean Ensemble Physics Reanalysis, GLOBAL_REANALYSIS_PHY_001_031, Issue 1.1, Mercator Ocean International, <https://catalogue.marine.copernicus.eu/documents/PUM/CMEMS-GLO-PUM-001-031.pdf> (last access: 31 March 2023), 2022.
- Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *J. Geophys. Res.-Oceans*, 118, 6704–6716, 2013.
- Häkkinen, S., Rhines, P. B., and Worthen, D. L.: Warm and saline events embedded in the meridional circulation of the northern North Atlantic, *J. Geophys. Res.*, 116, C03006, <https://doi.org/10.1029/2010JC006275>, 2011.
- Hansen, B. and Østerhus, S.: North atlantic–nordic seas exchanges, *Prog. Oceanogr.*, 45.2, 109–208, 2000.
- Hansen, B., Larsen, K. M. H., Hátún, H., Kristiansen, R., Mortensen, E., and Østerhus, S.: Transport of volume, heat, and salt towards the Arctic in the Faroe Current 1993–2013, *Ocean Sci.*, 11, 743–757, <https://doi.org/10.5194/os-11-743-2015>, 2015.
- Hansen, B., Húsgarð Larsen, K. M., Hátún, H., and Østerhus, S.: A stable Faroe Bank Channel overflow 1995–2015, *Ocean Sci.*, 12, 1205–1220, <https://doi.org/10.5194/os-12-1205-2016>, 2016.

- Hansen, B., Poulsen, T., Húsgarð Larsen, K. M., Hátún, H., Østerhus, S., Darelius, E., Berx, B., Quadfasel, D., and Jochumsen, K.: Atlantic water flow through the Faroe Channels, *Ocean Sci.*, 13, 873–888, <https://doi.org/10.5194/os-13-873-2017>, 2017.
- Hansen, B., Larsen, K. M. H., Olsen, S. M., Quadfasel, D., Jochumsen, K., and Østerhus, S.: Overflow of cold water across the Iceland–Faroe Ridge through the Western Valley, *Ocean Sci.*, 14, 871–885, <https://doi.org/10.5194/os-14-871-2018>, 2018.
- Hátún, H. and Chafik, L.: On the recent ambiguity of the North Atlantic subpolar gyre index, *J. Geophys. Res.-Oceans*, 123, 5072–5076, 2018.
- Hátún, H., Sandø, A. B., Drange, H., Hansen, B., and Valdimarsson, H.: Influence of the Atlantic subpolar gyre on the thermohaline circulation, *Science*, 309, 1841–1844, <https://doi.org/10.1126/science.1114777>, 2005.
- Iovino, D., Fogli, P. G., and Masina, S.: Evaluation of the CMCC global eddying ocean model for the Ocean Model Intercomparison Project (OMIP2), *EGUsphere* [preprint], <https://doi.org/10.5194/egusphere-2023-469>, 2023.
- Jochumsen, K., Moritz, M., Nunes, N., Quadfasel, D., Larsen, K. M. H., Hansen, B., Valdimarsson, H., and Jónsson, S.: Revised transport estimates of the Denmark Strait overflow, *J. Geophys. Res.-Oceans*, 122, 3434–3450, <https://doi.org/10.1002/2017JC012803>, 2017 (data available at: <https://www.cen.uni-hamburg.de/en/icdc/data/ocean/denmark-strait-overflow.htm>, last access: 22 March 2023).
- Jónsson, S. and Valdimarsson, H.: Water mass transport variability to the North Icelandic shelf, 1994–2010, *ICES J. Mar. Sci.*, 69, 809–815, <https://doi.org/10.1093/icesjms/fss024>, 2012.
- Liu, C., Allan, R. P., Mayer, M., Hyder, P., Desbruyères, D., Cheng, L., Xu, J., Xu, F., and Zhang, Y.: Variability in the global energy budget and transports 1985–2017, *Clim. Dynam.*, 55, 3381–3396, 2020.
- MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system, *Q. J. Roy. Meteor. Soc.*, 141, 1072–1084, <https://doi.org/10.1002/qj.2396>, 2015.
- Mayer, J., Mayer, M., Haimberger, L., and Liu, C.: Comparison of Surface Energy Fluxes from Global to Local Scale, *J. Climate*, 35, 4551–4569, 2022.
- Mayer, M., Haimberger, L., Pietschnig, M., and Storto, A.: Facets of Arctic energy accumulation based on observations and reanalyses 2000–2015, *Geophys. Res. Lett.*, 43, 10420–10429, <https://doi.org/10.1002/2016GL070557>, 2016.
- Mayer, M., Tietsche, S., Haimberger, L., Tsubouchi, T., Mayer, J., and Zuo, H.: An improved estimate of the coupled Arctic energy budget, *J. Climate*, 32, 7915–7934, 2019.
- Mayer, M., Tsubouchi, T., von Schuckmann, K., Seitner, V., Winkelbauer, S., Haimberger, L.: Atmospheric and oceanic contributions to observed Nordic Seas and Arctic Ocean Heat Content variations 1993–2020, in: *Copernicus Ocean State Report*, Issue 6, *J. Oper. Oceanogr.*, 15, s119–s126, <https://doi.org/10.1080/1755876X.2022.2095169>, 2022.
- Muilwijk, M., Smedsrud, L. H., Ilicak, M., and Drange, H.: Atlantic Water heat transport variability in the 20th century Arctic Ocean from a global ocean model and observations, *J. Geophys. Res.-Oceans*, 123, 8159–8179, 2018.
- Oort, A. H. and Yienger, J. J.: Observed interannual variability in the Hadley circulation and its connection to ENSO, *J. Climate*, 9, 2751–2767, 1996.
- Østerhus, S., Woodgate, R., Valdimarsson, H., Turrell, B., de Steur, L., Quadfasel, D., Olsen, S. M., Moritz, M., Lee, C. M., Larsen, K. M. H., Jónsson, S., Johnson, C., Jochumsen, K., Hansen, B., Curry, B., Cunningham, S., and Berx, B.: Arctic Mediterranean exchanges: a consistent volume budget and trends in transports from two decades of observations, *Ocean Sci.*, 15, 379–399, <https://doi.org/10.5194/os-15-379-2019>, 2019.
- Pietschnig, M., Mayer, M., Tsubouchi, T., Storto, A., Stichelberger, S., and Haimberger, L.: Volume transports and temperature distributions in the main Arctic Gateways: A comparative study between an ocean reanalysis and mooring-derived data, *EarthArXiv* [preprint], <https://doi.org/10.31223/osf.io/5hg3z>, 2018.
- Pujol, M.-I.: EU Copernicus Marine Service Product User Manual for the Global Ocean Gridded L4 Sea Surface Heights And Derived Variables Reprocessed 1993 Ongoing, SEALEVEL_GLO_PHY_L4_MY_008_047, Issue 7.0, Mercator Ocean International, <https://catalogue.marine.copernicus.eu/documents/PUM/CMEMS-SL-PUM-008-032-068.pdf>, last access: 31 March 2023, 2022.
- Pujol, M.-I., Taburel, G., and SL-TAC Team: EU Copernicus Marine Service Quality Information Document for the Global Ocean Gridded L4 Sea Surface Heights And Derived Variables Reprocessed 1993 Ongoing, SEALEVEL_GLO_PHY_L4_MY_008_047, Issue 8.2, Mercator Ocean International, <https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-SL-QUID-008-032-068.pdf>, last access: 31 March 2023.
- Robson, J., Polo, I., Hodson, D. L., Stevens, D. P., and Shaffrey, L. C.: Decadal prediction of the North Atlantic subpolar gyre in the HiGEM high-resolution climate model, *Clim. Dynam.*, 50, 921–937, 2018.
- Rosby, T., Flagg, C., Chafik, L., Harden, B., and Sørland, H.: A direct estimate of volume, heat, and freshwater exchange across the Greenland–Iceland–Faroe–Scotland Ridge, *J. Geophys. Res.-Oceans*, 123, 7139–7153, <https://doi.org/10.1029/2018JC014250>, 2018.
- Schauer, U. and Beszczynska-Möller, A.: Problems with estimation and interpretation of oceanic heat transport – conceptual remarks for the case of Fram Strait in the Arctic Ocean, *Ocean Sci.*, 5, 487–494, <https://doi.org/10.5194/os-5-487-2009>, 2009.
- Trenberth, K. E. and Fasullo, J. T.: Atlantic meridional heat transports computed from balancing Earth’s energy locally, *Geophys. Res. Lett.*, 44, 1919–1927, 2017.
- Tsubouchi, T., Våge, K., Hansen, B., Larsen, K. M. H., Østerhus, S., Johnson, C., Jónsson, S., and Valdimarsson, H.: Increased ocean heat transport into the Arctic Mediterranean over the period 1993–2016, NMDC [data set], <http://metadata.nmdc.no/metadata-api/landingpage/0a2ae0e42ef7af767a920811e83784b1> (last access: 21 March 2023), 2020.
- Tsubouchi, T., Våge, K., Hansen, B., Larsen, K. M. H., Østerhus, S., Johnson, C., Jónsson, S., and Valdimarsson, H.: Increased ocean heat transport into the Nordic Seas and Arctic Ocean over the period 1993–2016, *Nat. Clim. Change*, 11, 21–26, 2021.
- Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L.,

- Bentsen, M., Böning, C. W., Bozec, A., Chassignet, E. P., Curchitser, E., Boeira Dias, F., Durack, P. J., Griffies, S. M., Harada, Y., Ilicak, M., Josey, S. A., Kobayashi, C., Kobayashi, S., Komuro, Y., Large, W. G., Le Sommer, J., Marsland, S. J., Masina, S., Scheinert, M., Tomita, H., Valdivieso, M., and Yamazaki, D.: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do), *Ocean Model.*, 130, 79–139, 2018.
- Winkelbauer, S., Mayer, M., Seitner, V., Zsoter, E., Zuo, H., and Haimberger, L.: Diagnostic evaluation of river discharge into the Arctic Ocean and its impact on oceanic volume transports, *Hydrol. Earth Syst. Sci.*, 26, 279–304, <https://doi.org/10.5194/hess-26-279-2022>, 2022.
- Woodgate, R. A.: Increases in the Pacific inflow to the Arctic from 1990 to 2015, and insights into seasonal trends and driving mechanisms from year-round Bering Strait mooring data, *Prog. Oceanogr.*, 160, 124–154, <https://doi.org/10.1016/j.pocean.2017.12.007>, 2018 (data available at: <http://psc.apl.washington.edu/HLD/Bstrait/bstrait.html>, last access: 22 March 2023).



StraitFlux – precise computations of water strait fluxes on various modeling grids

Susanna Winkelbauer^{1,2}, Michael Mayer^{1,2,3}, and Leopold Haimberger¹

¹Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

²b.geos, Korneuburg, Austria

³European Centre for Medium-Range Weather Forecasts, Bonn, Germany

Correspondence: Susanna Winkelbauer (susanna.winkelbauer@univie.ac.at)

Received: 1 December 2023 – Discussion started: 2 January 2024

Revised: 10 April 2024 – Accepted: 15 April 2024 – Published: 12 June 2024

Abstract. Oceanic transports shape the global climate, but the evaluation and validation of this key quantity based on reanalysis and model data are complicated by the distortion of the used curvilinear ocean model grids towards their displaced north poles. Combined with the large number of different grid types, this has made the exact calculation of oceanic transports a challenging and time-consuming task. Use of data interpolated to standard latitude/longitude grids is not an option, since transports computed from interpolated velocity fields are not mass-consistent. We present two methods for transport calculations on grids with variously shifted north poles, different orientations, and different Arakawa partitions. The first method calculates net transports through arbitrary sections using line integrals, while the second method generates cross sections of the vertical–horizontal planes of these sections using vector projection algorithms. Apart from the input data on the original model grids, the user only needs to specify the start and endpoints of the required section to get the net transports (for the first method) and their cross sections (for the second method). Integration of the cross sections along their depth and horizontal extent yields net transports in very good quantitative agreement with the line integration method. This allows us to calculate oceanic fluxes through almost arbitrary sections to compare them with observed oceanic volume and energy transports at available sections, such as the RAPID array or at Fram Strait and other Arctic gateways, or to compare them amongst reanalyses and to model integrations from the Coupled Model Intercomparison Projects (CMIPs).

We implemented our methods in a Python package called StraitFlux. This paper represents its scientific documenta-

tion and demonstrates its application on outputs of multiple CMIP6 models and several ocean reanalyses. We also analyze the robustness and computational performance of the tools, as well as the uncertainties in the results. The package is available on GitHub and Zenodo and can be installed using pypi.

1 Introduction

Oceanic transports of heat, volume, and salt are integral components of the Earth's energy and mass budgets, playing a key role in regulating the Earth's climate. For instance, the Atlantic Meridional Overturning Circulation (AMOC) plays a crucial role in heating the North Atlantic by transporting warm surface waters from the tropics to the North Atlantic via the Gulf Stream and cold, dense waters southward at depth. It influences the weather and climate of eastern North America and western Europe (Jackson et al., 2015) and subsequently also affects the Arctic climate and sea ice (Liu and Fedorov, 2022; Mahajan et al., 2011). A weakening of the AMOC has been reported (Caesar et al., 2018; Mayer et al., 2023a; Rahmstorf et al., 2015) over recent decades, and a potential future collapse of the AMOC (Ditlevsen and Ditlevsen, 2023) would have major effects on the North Atlantic region and beyond. Monitoring oceanic currents is therefore particularly important in today's rapidly changing climate.

Several mooring arrays and other measurement devices exist that are capable of recording deep-water velocities and other sea state variables in the oceans. For example, there

are mooring lines in the Arctic gateways (Tsubouchi et al., 2012, 2018) or the Rapid Climate Change–Meridional Overturning Circulation and Heatflux Array (RAPID-MOCHA; e.g., Rayner et al., 2011) and the Overturning in the Subpolar North Atlantic Program (OSNAP; Lozier et al., 2017) for measuring the AMOC. It is desirable to compare transports calculated from those instruments with ocean reanalyses and climate models. This is challenging because the moorings are not aligned with the model grids, and the grids of ocean models, particularly in the Arctic, are complicated. A tool that facilitates a consistent comparison of flux estimates from this growing set of sources is therefore needed.

The convergence of meridians towards the North Pole poses challenges in ocean modeling. Murray (1996) proposed several global orthogonal curvilinear grids in which the North Pole is placed over land areas in order to avoid singularities over the ocean. Those ideas were picked up by many modeling centers and are now commonly used in the world of ocean modeling. Figure 1 shows examples of the two most common grid types. Many ocean models use tripolar grids, where two mesh north poles are placed over North America and Eurasia and at which the exact location of those two north poles varies between models. Ellipses around those poles and their normals create the new grid lines (Madec and Imbard, 1996) which are strongly displaced in the northern latitudes when compared to a regular dipolar grid. The second main grid type is the displaced dipolar grid, where the North Pole is displaced to somewhere over land areas, usually Greenland. Hereby, especially the grid lines in the proximity of the artificial pole feature a strong distortion. While solving the numerical problem of a singularity over the ocean, those curvilinear grids complicate the calculation of oceanic transports, especially in the proximity of the poles, as velocities in the direction of the artificial poles do not point in the direction of the true north and artificial zonal velocities do not point to the true east. The exact position of the poles, the angle between the native grid lines and regular longitude–latitude lines, as well as the horizontal and vertical resolution, varies between different models, forming a vast number of different grid types that complicate intercomparison between different models and with observations.

For optimal accuracy and consistency, the transports have to be calculated on the native grids. Horizontal interpolation of vector quantities (u , v) onto more convenient regular grids prior to the transport calculation compromises the conservation properties of the respective models, potentially leading to spurious effects and misleading results (an example is provided in Sect. 3). Finding the nearest points of the model grid has so far been done mostly manually for selected straits (e.g., Heuzé et al., 2023). This is time-consuming and becomes increasingly impractical when dealing with multiple models and multiple straits and increasing model resolution.

We have developed two methods for calculating oceanic transports on arbitrary oceanic sections that are independent of grid pole placement, orientation, and Arakawa partition.

The first method, using line integration, yields net transports of volume, heat, salt, and ice across defined straits. The second method employs vector projection algorithms to generate cross sections of currents, temperature, and salinity in the vertical plane. We will refer to the methods as the line integration method (LM) and vector projection method (VPM). For both methods, the tedious point selection process is fully automatized. We tested our methods on various tri- and dipolar grids from the Climate Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016) and show some exemplary results in Sect. 3.1.

This paper is structured as follows: in Sect. 2, we describe the fundamental concepts of both calculation methods and their mathematical foundation. Furthermore, in Sect. 2.2, we describe the implementation of the methods in the open-source Python package StraitFlux. Section 3 assesses the robustness of the tools, examines their accuracy, provides application examples, and analyzes their computational performance. The final section outlines the strengths and weaknesses of StraitFlux and draws conclusions regarding its utility.

2 Methods

2.1 Mathematics/general idea

The general idea is to calculate oceanic volume, heat, salinity, and ice transports across any chosen vertical section, typically a straight section between two land masses. We define the oceanic transports of volume (OVT), heat (OHT), salinity (OST), and ice (OIT) through a given strait as follows:

$$\text{OVT} = \int_{x_1}^{x_2} \int_0^{z_b(x)} \mathbf{v}_o(x, z) \cdot \mathbf{n} \, dz \, dx, \quad (1)$$

$$\text{OST} = \int_{x_1}^{x_2} \int_0^{z_b(x)} S \mathbf{v}_o(x, z) \cdot \mathbf{n} \, dz \, dx, \quad (2)$$

$$\text{OHT} = c_p \rho \int_{x_1}^{x_2} \int_0^{z_b(x)} (\theta(x, z) - \theta_{\text{ref}}) \mathbf{v}_o(x, z) \cdot \mathbf{n} \, dz \, dx, \quad (3)$$

$$\text{OIT} = \int_{x_1}^{x_2} d(x) \mathbf{v}_i(x) \cdot \mathbf{n} \, dx, \quad (4)$$

where \mathbf{v}_o (\mathbf{v}_i) represents the velocity vector of liquid water (sea ice), and \mathbf{n} is the vector normal for the strait – therefore their product gives the velocities normal for the considered coast-to-coast section. Furthermore, x defines the along-strait extent, and z is its depth. The boundaries z_b , x_1 , x_2 have to be chosen such that no water can “escape” the desired coast-to-coast section. This can be ensured if x_e and x_w are land points and the auxiliary fields describing model

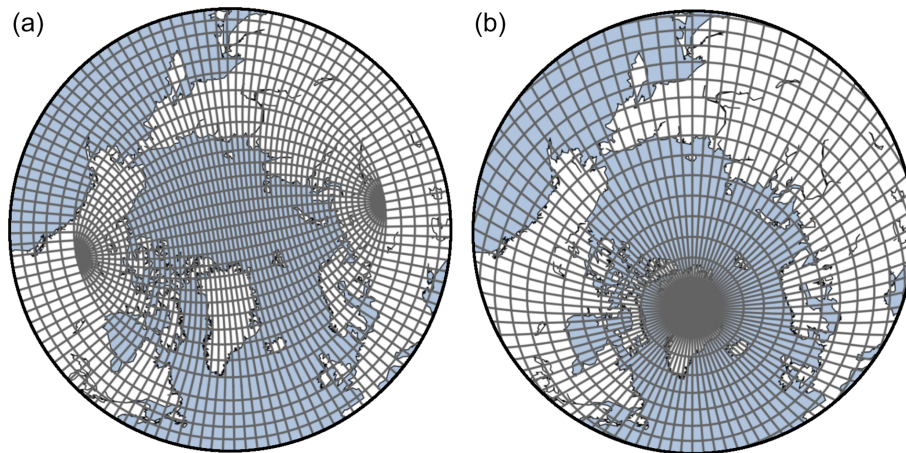


Figure 1. Examples of two curvilinear grids typically used for ocean modeling. (a) A tripolar grid with two northern poles (one over Eurasia and one over North America). (b) A displaced dipolar grid with one northern pole displaced over Greenland.

ocean depths are used appropriately. It is also possible to calculate transports between two water points; however, results should be viewed with caution, and their meaningfulness is left to the discretion of the user. S is the seawater salinity, c_p the specific heat of seawater, ρ the density of seawater, and θ the potential temperature. Throughout this study, we will use both θ and T synonymously for the potential temperature of seawater. For the validation in Sect. 3, c_p and ρ are set to $3996 \text{ J kg}^{-1} \text{ K}^{-1}$ and 1026 kg m^{-3} , respectively, as default values. However, it is easy to adapt those values to individual model needs. Previous studies (Schauer and Losch, 2019; Schauer and Beszczynska-Möller, 2009) correctly point out that true heat transports would actually demand closed-volume transports through the examined straits. This is generally not the case for partial sections, as transports may be compensated by flows through other passages, and unbalanced volume transports would generally introduce the dependence of heat transports on the chosen temperature scale via θ_{ref} . However, the “heat fluxes” as defined above are commonly used to ensure comparability with transports derived from observations. Therefore, we will further refer to heat transports when calculating “heat fluxes” as defined here. Additionally each model’s heat flux should be computed relative to a reference temperature θ_{ref} , representing the mean temperature of the assessed flow. The validation part of this paper focuses on Arctic straits; therefore, we follow, e.g., Heuzé et al. (2023) and Mulwijk et al. (2018) and choose a universal reference temperature of 0°C . Generally, in StraitFlux, reference temperatures are set per default to 0°C but may be changed to individual needs. Furthermore, following the approach of Schauer and Losch (2019) and Heuzé et al. (2023), salinity fluxes are calculated instead of freshwater fluxes to avoid the need for a reference salinity, which would vary for each model. This simplification facilitates model comparisons.

2.1.1 Line integration method (LM)

The basic principle of LM is outlined in Fig. 2a. A closed line is generated following the strait (red line) as closely as possible by connecting the faces of the individual grid cells (blue line). For an Arakawa C grid (see the inset in Fig. 2a for a definition), the u and v components are positioned exactly at the faces of the grid cells. This positioning allows a straightforward integration of the u/v components along the meridional/zonal width of the grid cell and its depth. The sum of all grid cells vertically and horizontally then provides the net transport through the strait. In the case of an Arakawa B grid, however, the u and v components are defined at the edges of the grid cells and must be transformed to the faces of the tracer grid cells. To obtain u and v velocities at the faces of the tracer cells, we adapt Eqs. (6.46) and (6.50) from Griffies et al. (2004) (in accordance, we also use the same multi-letter variable names). The u velocities at the eastern faces of the tracer cells (uet) are calculated as

$$uet_{i,j,k} = \frac{dyu_j dhu_j u_j + dyu_{j-1} dhu_{j-1} u_{j-1}}{2dyet_{i,j}}, \quad (5)$$

with dyu_j as the meridional width of the u/v cell, dhu_j as the depth of the u/v cell, and $dyet_{i,j}$ as the meridional width of the tracer cell’s east side. The v velocities at the northern faces of the tracer cells (vnt) are calculated as

$$vnt_{i,j,k} = \frac{dxu_i dhu_i v_i + dxu_{i-1} dhu_{i-1} v_{i-1}}{2dxnt_{i,j}}, \quad (6)$$

with dxu_j as the zonal width of the u/v cell, dhu_j as the depth of the u/v cell, and $dxnt_{i,j}$ as the zonal width of the tracer cell’s north side. After transformation, transports are calculated identically (as done for the Arakawa C grid). For the case of an Arakawa A cell, where T , u , and v are placed in the middle of the grid cell, we implement a similar method to

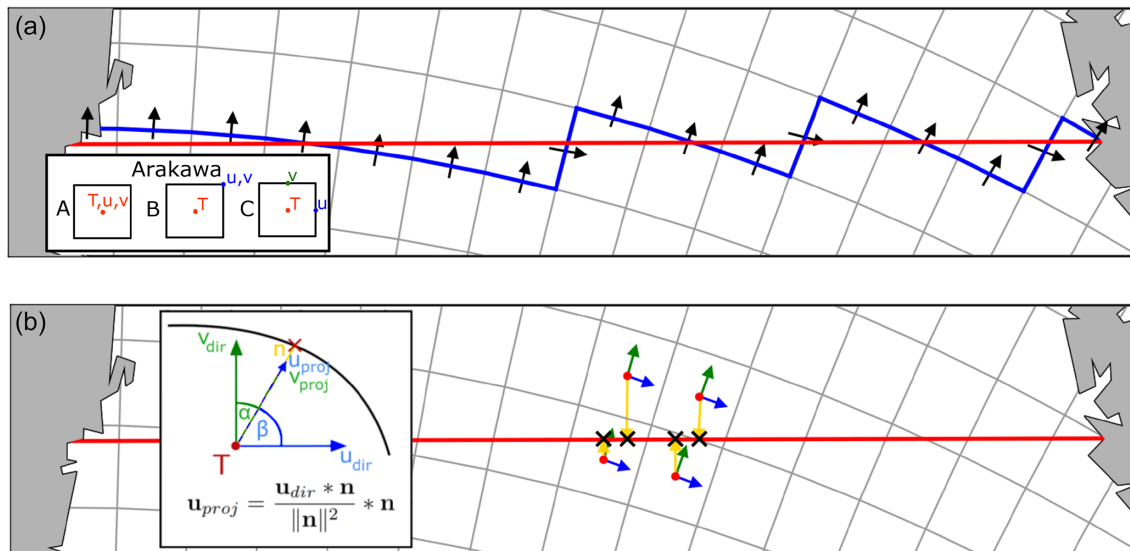


Figure 2. (a) Schematic of the LM, with the reference line (bold red line), the closed line generated on the native model grid (blue line), and used u and v components (black arrows). (b) Schematic of the VPM; the direct u and v vectors (blue and green) are projected onto the normal vectors (yellow) to find the portions of u/v that pass orthogonally through the strait (red line).

that for Arakawa B and move the u and v components onto the cell faces. Note that while volume transports are calculated without any further use of interpolation, for heat and salinity transports, the scalar quantities of T and S have to be interpolated to the faces of the tracer grid cells. This is done using linear interpolation (similar to Madec et al., 2017, Sect. 12.3.1).

2.1.2 Vector projection method (VPM)

The second method uses simple vector projection algorithms to obtain the share of the u and v components that passes orthogonally through the strait. Figure 2b shows a schematic of the VPM. For every grid cell touching the strait and their neighboring cells (needed for the interpolation onto the strait), we calculate direction vectors of the u and v components (blue and green arrows) and normal vectors pointing from the tracer grid cell in the direction of the strait (yellow arrows). Then, using Eq. (7), the u and v vectors are projected onto the normal vector:

$$\mathbf{u}_{\text{proj}} = \frac{\mathbf{u}_{\text{dir}} \cdot \mathbf{n}}{\|\mathbf{n}\|^2} \cdot \mathbf{n}; \quad \mathbf{v}_{\text{proj}} = \frac{\mathbf{v}_{\text{dir}} \cdot \mathbf{n}}{\|\mathbf{n}\|^2} \cdot \mathbf{n}, \quad (7)$$

where \mathbf{u}_{dir} and \mathbf{v}_{dir} represent the direction vectors of the u and v components, \mathbf{n} are the normal vector between tracer grid cell and strait, and \mathbf{u}_{proj} and \mathbf{v}_{proj} are the projection vectors of \mathbf{u}_{dir} and \mathbf{v}_{dir} passing orthogonally through the strait. The closer the angle between the direction vector and normal vector (α and β in Fig. 2b) to 0 (90) or 180 (270), the larger (smaller) the amount that actually passes through the strait. Using the magnitudes of the projection vectors, we calculate the u and v components pointing orthogonally onto the

strait at all grid cells touching the strait and their neighboring cells (needed for the bilinear interpolation). Then, we multiply them with the respective vertical cell thicknesses at the u/v points and interpolate those orthogonal “transports” bilinearly onto the closest points on the reference line (black crosses in Fig. 2b; henceforth called T_{proj}). In a final step, we divide the interpolated transports by the vertical thickness of the cells on the reference line to obtain velocities again. This results in velocity cross sections of the vertical plane which are spaced irregularly along the along-strait distance (x) in accordance with the distribution of T_{proj} points. The interpolation onto evenly distributed points on the section to, e.g., enable the calculation of differences with other models/reanalyses is initially left to the user and eventually will be included in a future version of StraitFlux. By calculating transports (scalar quantities) prior to the interpolation onto the strait, we ensure that the conservation properties of the models are maintained. This ensures that the integration of the cross sections along the along-strait distance (x) and depth (z) provides net transports which agree very well with the values obtained by the LM (see Sect. 3). To obtain heat and salinity transports, the velocity cross sections simply have to be multiplied with the T and S cross sections, which are obtained using bilinear interpolation, before integration.

2.2 Software implementation

In this section, we describe the implementation of the transport calculation tools into the open-source Python package StraitFlux. We provide an overview of the code structure and its usage.

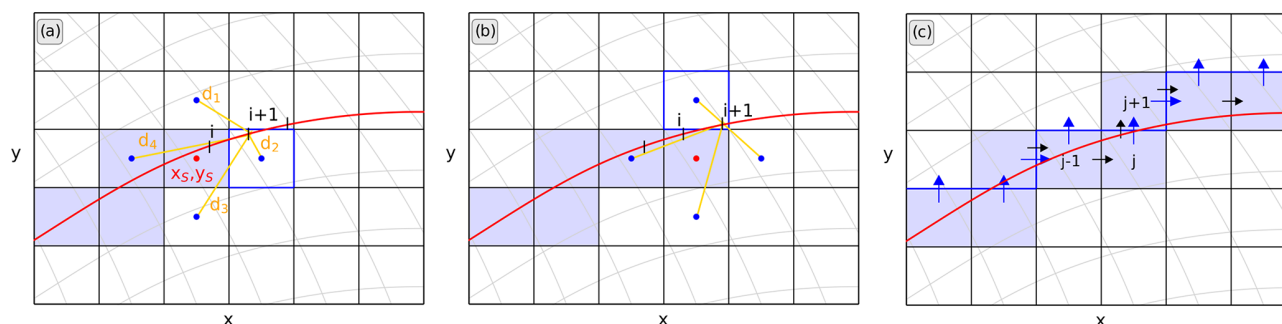


Figure 3. Illustration of the indices selection process using `select_points`. Lines of constant latitude/longitude are shown in grey. (a, b) Determination of consecutive grid cells on the native grid. Distances (orange lines) of the four grid cells (blue dots) surrounding the current grid cell (red dot) are compared for all equally spaced points i along the reference line. (c) Specification whether a u or v component should be taken. Black arrows show movement from one grid cell to the next, and blue arrows show the chosen u or v component.

2.2.1 Preprocessing

The script is designed to work with ocean reanalyses and various CMIP6 models which may differ in terms of grid orientation, partitioning, coordinate names, units, and dimension names. Therefore, prior to the transport calculations, data preprocessing is conducted to standardize these attributes. This includes ensuring consistency in dimension names (x , y , lev , and $bnds$), coordinate names (long and lat), units (SI), and the shape of longitude and latitude coordinates (2D). For CMIP6 models, preprocessing is carried out using selected tools from the open-source Python package `xMIP` (<https://cmip6-preprocessing.readthedocs.io/en/latest/>, last access: 3 June 2024). Ocean reanalyses are treated in a similar fashion by adapting some of the `xMIP` tools. There is no regridding involved.

To integrate transports along cell faces and depths accurately, precise information about the cell extents is required. As horizontal grid metrics are not always provided by CMIP6 models, the script automatically determines the zonal and meridional extents using the `calc_dxdy` function. The vertical extent has to be specified prior to the calculation process. For instance, for CMIP6 the variable “cell thickness” (“`thkcello`”), which is available through the Earth System Grid Federation (ESGF) website (<https://esgf-node.llnl.gov/search/cmip6/>, last access: 3 June 2024) for most models, may be used. Since the cell thickness is not available for all models, the function `calc_dz` enables the calculation of cell thicknesses by supplying the variable “total ocean depth” (“`deptho`”; also available via ESGF) and the vertical level boundaries (“`lev_bnds`” for CMIP6 that are contained in every three-dimensional ocean variable). If `deptho` is not available, then it is also possible to supply it from another model, preferably one with a similar grid, and the variable will automatically be interpolated to the needed grid. However, especially volume transport calculations are very sensitive to the exact ocean bathymetry used by the model. Therefore, if

possible, it is advised to supply the exact fields (e.g., `thkcello` for CMIP6) and not necessarily to rely on `calc_dz`.

2.2.2 Index extraction

The determination of section positions for transport calculations is accomplished in the `def_indices` function. Users can specify the start and end points of a section using the `coords` parameter in the `transports` function. The section will then follow the shortest distance along the sphere. Alternatively, users may pass specific coordinates by setting the `set_latlon=True` parameter and providing a list of latitude (`lat_p`) and longitude (`lon_p`) points. The latter option also allows the calculation of “kinked” sections. Using the “`coords`” option, the function generates a reference line (`ref_line`) consisting of equally spaced latitude–longitude pairs at which point the interval between points on the reference line is set automatically to be suitable for the resolution of the model. When passing coordinates via the `set_latlon=True` option, we advise the user to use intervals not larger than 0.4 times the resolution of the model (e.g., intervals of 0.1° for models with a resolution of about 0.25°) as coarser intervals might lead to the skipping of grid points and generate broken lines. Providing coordinates at high resolution might create duplicates in the indices found; however, those will be removed automatically.

In order to calculate the net transports via line integration, a polyline along the edges of grid boxes has to be generated following the `ref_line` as closely as possible (red line in Fig. 2). The function `check_availability_indices` determines the indices of the u and v components along the closed-integration line.

The first point of the integration line is found by selecting the nearest point on the native grid (center of a cell) within a selection window with the size of 2° around the first point on the reference curve; the size of the selection window was chosen to work properly on grids with a resolution of 1° and higher (if needed, the window size can be adapted to coarser resolutions). The subsequent points on the reference

closed boundaries. For the cyclic boundaries, the values of the last 1–2 columns are set to the values of the first 1–2 columns; therefore, whatever flows out of the western end of the basin enters the eastern end, and vice versa (Madec et al., 2017). For the closed-boundary conditions, solid walls are enforced at all model boundaries, and the first and last columns are set to zero. The same is true for the north–south boundary conditions. For our application, the displaced dipolar model grids are uncomplicated concerning the north–south conditions. As both poles are placed over land, the southern and northern most grid cells consist of land areas only, and no oceanic transports pass the northern boundary – the grids are topologically equivalent to a cylinder (periodic in x but not in y ; Smith et al., 2010). The tripolar grids, however, require additional consideration along the northern boundary, as the ocean is divided by the line between the two northern grid poles. For instance, the three-polar ORCA grid (used in multiple CMIP6 models) uses a north fold boundary with a T point pivot (see Sect. 8.2.2 in Madec et al., 2017), where the upper three grid cells are duplicated and pivoted around the line connecting the two north poles. For more information on the model-specific boundary conditions see, e.g., Madec et al. (2017), Griffies et al. (2004), and Smith et al. (2010).

These conditions have to be handled with care, as especially the volume transport calculation is very sensitive and can yield useless results when there is a gap in the integration line or if any grid cells are counted twice. StraitFlux automatically checks for overlapping cyclic boundary points and drops any duplicates. This should ensure correct transport calculations across the zonal boundaries independent of how the models deal with periodicity. Similarly, concerning the north boundary conditions, StraitFlux automatically selects the correct indices and avoids gaps and/or duplicates. We tested this successfully for an arbitrary line going over the top boundary of the model grids for various CMIP6 models with different boundary conditions (CMCC-CM2-SR5, EC-Earth3, CanESM5, ACCESS-CM2, CAMS-CSM1-0, and IPSL-CM6A-LR). Figure A1 in the appendix shows an example for the CMCC-CM2-SR5 model. StraitFlux correctly chooses the indices so that a continuous line without overlaps is formed. While the indices selection worked for the tested models, the generated indices should also still be checked to ensure a continuous line for more complicated boundary conditions. Therefore, the code automatically outputs the warning “Attention: Strait crossing the northern boundary – make sure correct indices are chosen!” when moving across the boundary of the grid.

2.2.4 Line integration method (LM)

The actual oceanic transports are calculated through the `transports` function by integrating the velocities at the chosen u and v indices over the cells’ zonal or meridional

extents and their actual vertical extent. See Sect. 2.2.1 for the computation of the needed mesh files.

For heat transports, the indexed cells are additionally multiplied by the ocean’s potential temperature prior to integration and for salinity transports with the cells salinity. These are defined at grid cell midpoints for Arakawa B and C grids. Therefore, the fields of T and S have to be interpolated to the faces of the grid cells first, which is done using linear interpolation (similar to Madec et al., 2017, Sect. 12.3.1). Furthermore, heat/salinity transports have to be multiplied by specific heat and density (see Eq. 3). Those are set as constant to $3996 \text{ J kg}^{-1} \text{ K}^{-1}$ and 1026 kg m^{-3} per default; however, it is possible to adapt them to individual needs.

Transports may be calculated for longer time periods at once, where the time period may be set by the “time_start” and “time_end” arguments in `transports`. The final result of `transports` gives net integrated transports through the strait with the coordinate time which are saved per default as a netCDF file. The used indices and values through the cell faces can be saved on request.

2.2.5 Velocity projections

The LM has good conservation properties, but since the faces of the polyline can point into very different directions, it is difficult to plot cross sections. Therefore a second method for calculating the cross-strait transports at points on the reference line has been developed. The direction of the normal vectors thus changes smoothly and allows the calculation of horizontally and vertically resolved contributions to the total transport through a respective strait.

As for the LM, the first step is to find the closest points on the native grids to the reference line. The selection of the indices proceeds similar to the selection for LM. However, in addition to the closest points to the reference line, the four immediate neighboring cells of the closest points are also used. Those are needed for the interpolation of the transports onto the reference line (as described in Sect. 2.1.2). Again, any duplicate indices are removed automatically.

For the projection of the u and v velocities onto the strait, direction vectors and normal vectors for every grid cell are determined using the functions `calc_dir_vec` and `calc_normvec`. Direction vectors are assumed to point from one grid cell to the neighboring ones and are simply calculated by taking the difference between the Cartesian coordinates of $u_{x,y}$ and $u_{x+1,y}$ for \mathbf{u}_{dir} and the difference between the Cartesian coordinates of $v_{x,y}$ and $v_{x,y+1}$ for \mathbf{v}_{dir} .

Normal vectors are calculated using three consecutive cross products. The basic principle is illustrated in Fig. 6. For each point, T , u , and v , we find the two closest points $r1$ and $r2$ on the reference line. Transforming them into Cartesian coordinates, we can take their cross-product and get the vector \mathbf{rn} standing orthogonally on the surface spanned by $r1$ and $r2$ (blue surface). The cross-product of \mathbf{rn} with T yields the vector \mathbf{Tr} (orthogonal to the green surface). Finally, tak-

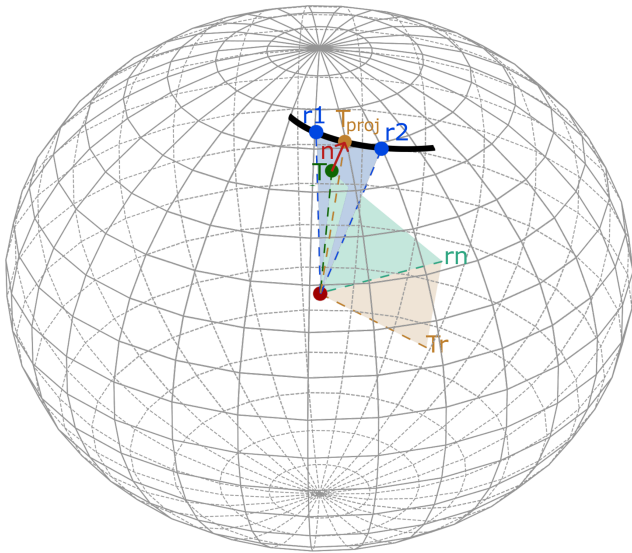


Figure 6. Illustration of the normal vector computation process using three consecutive cross-products; $r1$ and $r2$ feature the two closest points to T on the reference line, rn is the normal vector on the surface spanned by $r1$ and $r2$, Tr is the normal vector of rn with T , and T_{proj} is the projection point on the reference line and results from normalizing the cross-product of rn with Tr with the Earth's radius. T and T_{proj} then provide the normal vector n .

ing the cross-product of rn with Tr and normalizing the resulting vector with the Earth's radius yields the point T_{proj} on the reference line. The normal vector n for every grid cell is given as the vector pointing from point T to T_{proj} . The projection points T_{proj} are later used as interpolation points on the reference line.

The function `proj_vec` uses the direct and normal vectors and Eq. (7) to calculate the projection vectors at every grid cell. The actual calculation is done using `vel_projection`. Using the norm of the projection vectors, the u and v components of every vertical layer are projected orthogonally onto the strait. To enhance conservation properties, we additionally multiply by the actual cell thicknesses of the cells before interpolation. The signs of the u and v components are determined by comparing the angles between direct and normal vector – the general idea is outlined in Fig. 7. Then, in a final step, the scaled fluxes for every vertical layer are interpolated onto the reference line and divided by the respective layer thickness at the reference line.

The final result of `vel_projection` gives cross sections of the velocities passing through the strait, with coordinates time, depth, and x – the along-strait distance. Exemplary cross sections are shown in Sect. 3.1. To simplify the integration of the cross sections to net transports, the horizontal (dx_{int}) and vertical extents (dz_{int}) of every point on the reference line are output as well. Therefore, the net volume transport – similar to the end-product of the LM – can be

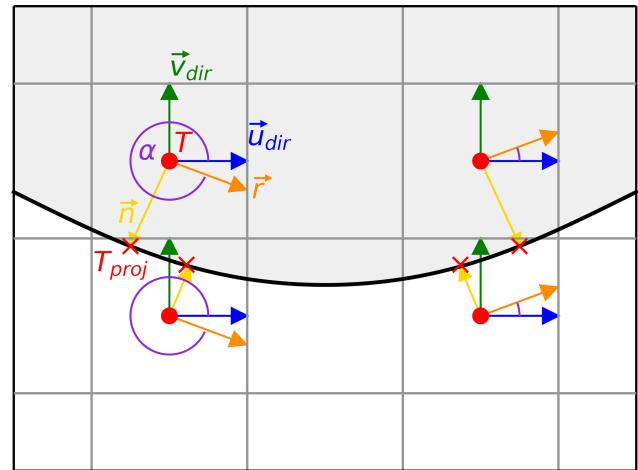


Figure 7. Determination of the u component signs using the angle α between the direction vector of u and the direction vector of the reference line r at the corresponding projection point. When $\alpha < 180$, the u component is counted negative, and when $\alpha > 180$, the u component is counted positive. The v signs are calculated similarly using v_{dir} .

calculated by multiplying the section with dx_{int} and dz_{int} and summing up over x and depth.

Cross sections of temperature and salinity are simply calculated by interpolating the scalar quantities of T and S onto the strait defined by the T_{proj} points. To obtain the vertical profiles of heat and salinity transports, the T and S cross sections have to be multiplied with the velocity cross sections.

The velocity projection method has originally been developed for visualization purposes; however, as shown in the next section, almost everywhere it provides nearly as accurate estimates of total fluxes through a strait as in the LM.

3 Validation

In this section, we will assess the robustness of the tools and their accuracy with multiple approaches. First, the results of our computations are compared with naive calculation of the fluxes from interpolated velocity fields. Second, we specify simple u , v , and T fields where the transports can be calculated analytically. These fields are then transformed exactly to the respective ocean model grids using the analytical mapping functions. The transports are then calculated using our LM and compared to the analytic solutions. Third, we show the consistency between VPM and LM and then we check the correspondence between area-integrated divergence fields and the transports through the array boundary. Last, we compare our results to results taken from an independent study where transports through Fram Strait were calculated by picking indices and signs for all grid points by hand. The exact definitions (start and endpoints) of all straits

used throughout this paper are given in the Appendix (Table A1).

For verification, we use harmonic functions to specify simple two-dimensional u , v , and T fields where the transports can be calculated analytically:

$$\begin{aligned} u(\lambda, \varphi) &= 0; & v(\lambda, \varphi) &= v_r(\varphi) + v_0 \cos \varphi \sin k\lambda; \\ T &= T_r + T_0 \cos \varphi \sin(k(\lambda + \psi)), \end{aligned} \quad (8)$$

with longitude λ , the latitude φ , wavenumber k , and phase shift ψ given in radians. With the Earth radius a , we get the following for the transport:

$$\begin{aligned} F(\varphi) &= a \cos \varphi \int_0^{2\pi} T v d\lambda \\ &= v_r(\varphi) T_r \cos \varphi + \frac{v_0 T_0 \cos^3 \varphi}{2} \cos \psi. \end{aligned} \quad (9)$$

The second term on the right-hand side of Eq. (9) vanishes for phase shift $\pi/2$. The defined v and T fields are transformed to different ocean model grids (CMIP6); the four used modeling grids are shown in Fig. 8 (top) and were chosen to be as different as possible in terms of horizontal resolution, number and location of poles, strength and extent of the distortion, and used Arakawa partition. Transports are then calculated for full parallels at different latitudes ($\varphi = -70^\circ$ to 85° N) and for different wavenumbers ($k = 1$ to 100) using our LM, which leads to the generation of nontrivial polylines, depending on the curvature of the respective grid. Solutions are then compared to the analytic solutions of the transport integral (Eq. 9). Differences remain low (mostly below $\pm 1\%$) for all four grid types over all assessed latitudes and wavenumbers, as shown in Fig. 8 (bottom). The biggest errors occur for the lower-resolution grids at higher latitudes and higher wavenumbers. This is most likely caused by the coarse resolution and discretization of the models which are not able to resolve the smaller generated waves and less likely caused by the curvature of the grid alone, as the higher-resolution model features very low errors up to $k = 100$. At latitudes with regular grid lines, errors due to discretization are deemed to be small. We calculated transports using the defined spherical harmonics for the tripolar CanESM5 grid at 20° N for various horizontal resolutions and found a difference of just about 0.015% between a 1° and a 10° resolution. Other small differences are mainly caused due to inaccuracies in the latitude selection. While analytic solutions are calculated at full latitudes, the position of the polylines may be shifted north or south due to the grids' resolution. For instance, the latitudinal shift in the 20° N line in the CanESM5 model leads to an error of 0.2% , which explains the total recorded error practically. Furthermore, differences in the Earth radius (we assume $a = 6371$ km) may lead to minor discrepancies.

The grey areas in the lower-left panels of Fig. 8 indicate missing values and are due to the nature of the two dipolar

grids as they skip areas over the Greenlandic ice cap around the artificial north poles. This is problematic for this application as we define complete parallels for our analytic solutions. However, it does not affect the actual transport calculations as this complication does not occur over oceanic areas.

In summary, results from the LM correspond very well with analytical results.

Interpolation of the vector components u and v onto regular grids is quite complex and may lead to significant errors in the calculated transports. The complexity arises from the rotation of the u and v components in comparison to the directions on a geographic latitude–longitude grid. Regridding would involve rotating the ocean velocity components to the new flow direction (eastward/northward) prior to the interpolation as done, e.g., by He et al. (2019). However, for the rotation, the exact grid angle at each grid cell is needed, which is not standard output for most CMIP6 models and reanalyses. Outten et al. (2018) found that small inaccuracies in the used angles, e.g., the exact position of the angles in the grid cell (center vs. cell edges or corners) may lead to differences in the calculated transports. Even if the model configurations and grid angles were archived correctly, it is still hard to guarantee the conservation properties of the interpolated fluxes. Figure 9 compares transports calculated from interpolated u and v values on a regular grid with those derived from u and v on the native grid for the CanESM5 model. Transports are calculated through the Drake Passage and the RAPID array, two straits where the native grid of the CanESM5 model is not distorted, and therefore, any errors connected to the rotation of the velocity components are avoided. Even here, interpolation (both bilinear and conservative as defined in xESMF) leads to significant deviations from the actual transports obtained through both StraitFlux methods. This is especially so at the RAPID array, which is a very long strait with a relatively small net volume transport. It is thus clear that the use of interpolated vector components is inappropriate for all kinds of transport calculations. An alternative approach would be to write each vector in terms of scalar vorticity and streamfunction using Helmholtz decomposition (e.g., Watterson, 2001), remap those scalar quantities to a regular grid, and then recover eastward and northward velocity components using gradients.

While we have not compared the cross section method with the analytical solution as we did for the LM, we show the credibility of the VPM by comparing volume, heat, and ice transports obtained through the LM and the VPM (Fig. 10). Ideally, both methods should provide the same results; however, due to differences in the calculation process, small differences are expected. We choose a strait in the Arctic region – Fram Strait – in order to come close to the strongest distortion of the curvilinear grids and show volume transports for three models that use different grid types and Arakawa partitions. The CMCC-CM2-SR5 model uses a tripolar grid with an Arakawa C partition, IPSL-CM6A-LR (Boucher et al., 2020) uses a tripolar grid with an

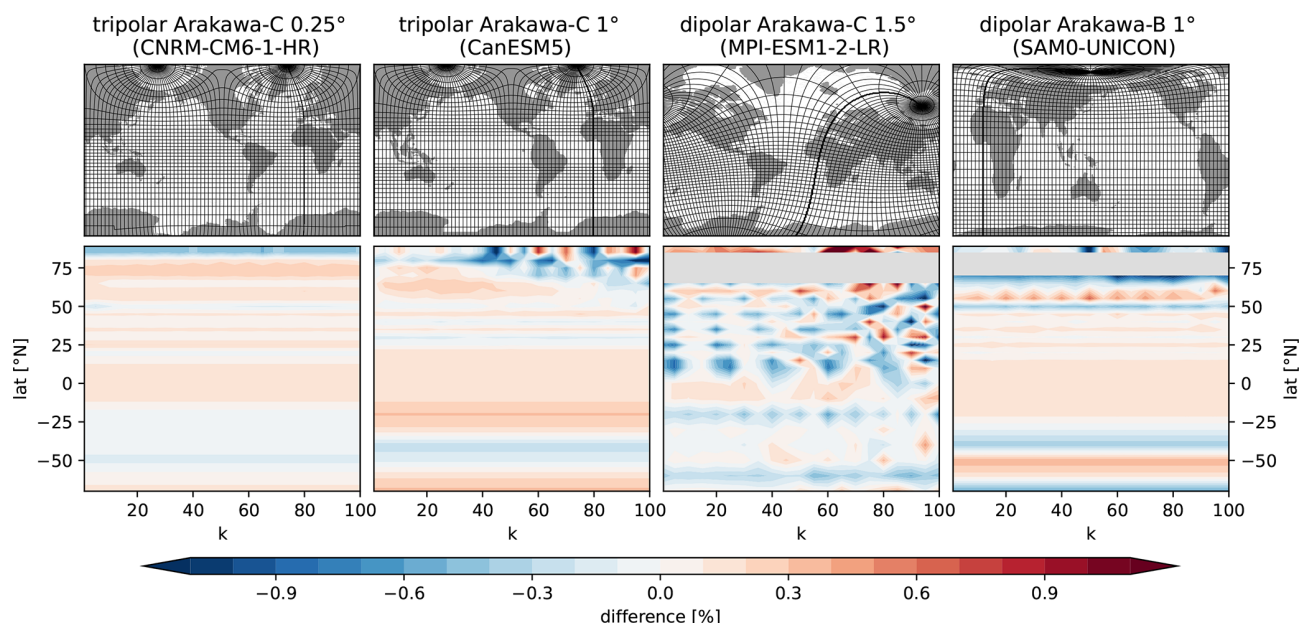


Figure 8. Top: grids used for the generation of polylines for the transport calculation via LM. The grid lines show the position of the regular grid lines on the distorted modeling grids (resolution of 5°). Bottom: differences between the LM and analytic solutions of transports using spherical harmonics as fields. Grey areas in the bottom plots indicate areas where the calculation was not performed due to the absence of grid points in the dipolar grids over parts of Greenland.

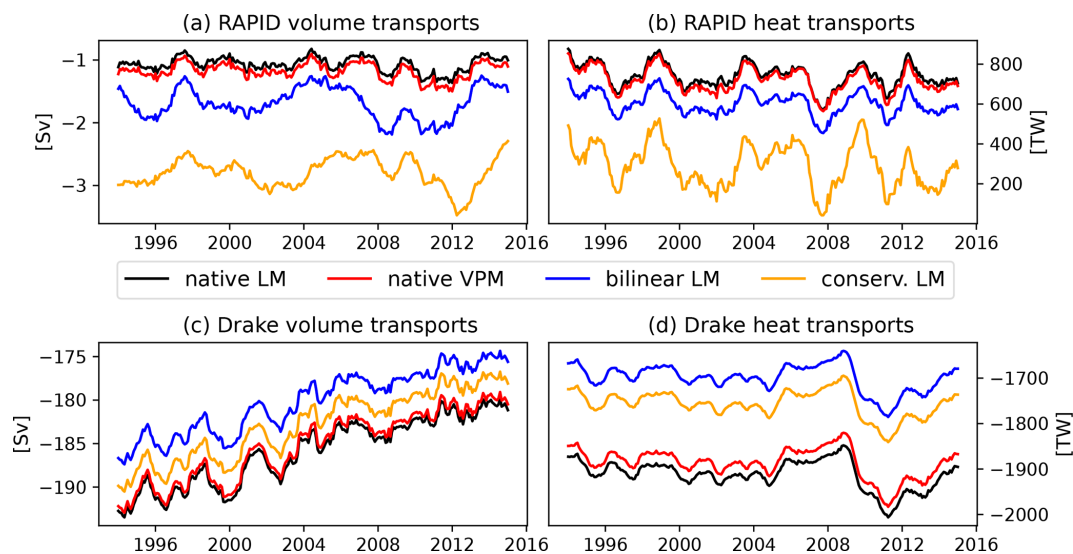


Figure 9. Display of the interpolation error for volume (a, c) and heat (b, d) transports at the RAPID array and at Drake Passage from the CanESM5 model (1° resolution).

Arakawa B partition, and SAM0-UNICON uses a displaced dipolar grid with an Arakawa B partition. Depending on the model, transports obtained through the different calculation methods match within a few percent of their total value. We consider this a very good result, given that this method was more designed for plotting purposes than for maximum accuracy of the integrated result.

Another test to validate our transport calculation tools is the comparison of the transports across a whole latitudinal circle to the divergence of transports north of that latitude (i.e., validation of the sentence of Gauss). This is done for the ORAS5 ocean reanalysis (Zuo et al., 2019) and shown in Fig. 11. While the VPM differs from the values obtained by the LM and the divergence integral, the differences are still very small compared to those found in Fig. 9. Those differ-

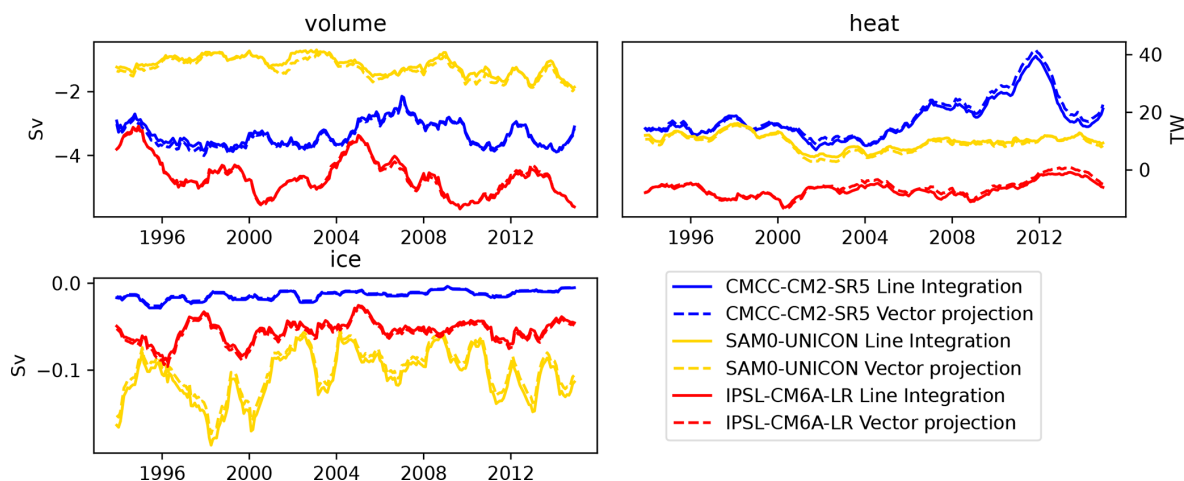


Figure 10. Comparison of volume, heat, and ice transports obtained through the LM (solid) and through integration of cross sections obtained through the VPM (dashed). The selected models use different grid types (see text) all with a horizontal resolution of about 1° .

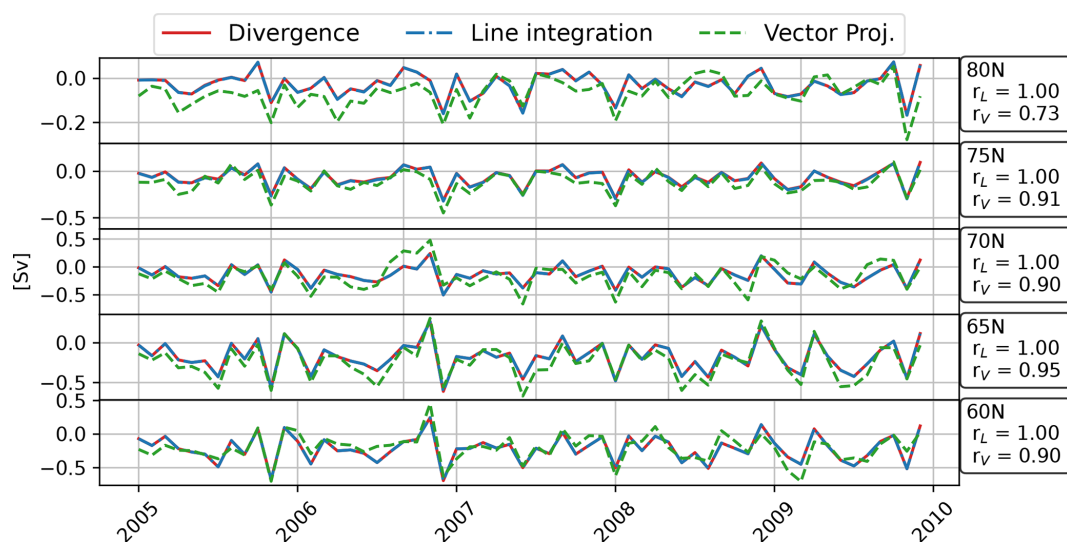


Figure 11. Integrated volume fluxes across different circles of latitude derived from ORAS5. Transports computed using the LM and VPM, as well as through the integration of the divergence of transports north of the section in question, are shown.

ences may be caused by the increasing difference in integration area between the two methods with stronger grid curvature further north or also by an inaccuracy in the treatment of the north fold boundary points. This needs to be further investigated and may be resolved in a later version of the software.

Last, we compared our methods to transports obtained by Heuzé et al. (2023), available via PANGAEA (Zanowski et al., 2023), who calculate transports of salinity, heat, and volume through Fram Strait for various CMIP6 models by choosing the coordinates for each model by hand. Figure 12 shows the comparison of our transports to those obtained by Heuzé et al. (2023) for 10 selected models. For most models, the results match within an expected range of uncertainty; differences may arise from differences in the exact position-

ing of the straits and differences in the definitions of ρ and c_p .

3.1 Application examples

To illustrate the abilities of StraitFlux, we present some sample results and refer to studies in which the tools have already been successfully used.

Results from the LM and VPM have already been shown in Sect. 3. Additionally to the net integrated transports the VPM also provides cross sections of the vertical plane. Figure 13 shows exemplary cross sections of currents, temperature, and salinity for the Greenland–Scotland Ridge (GSR) for two CMIP6 models with different horizontal resolutions.

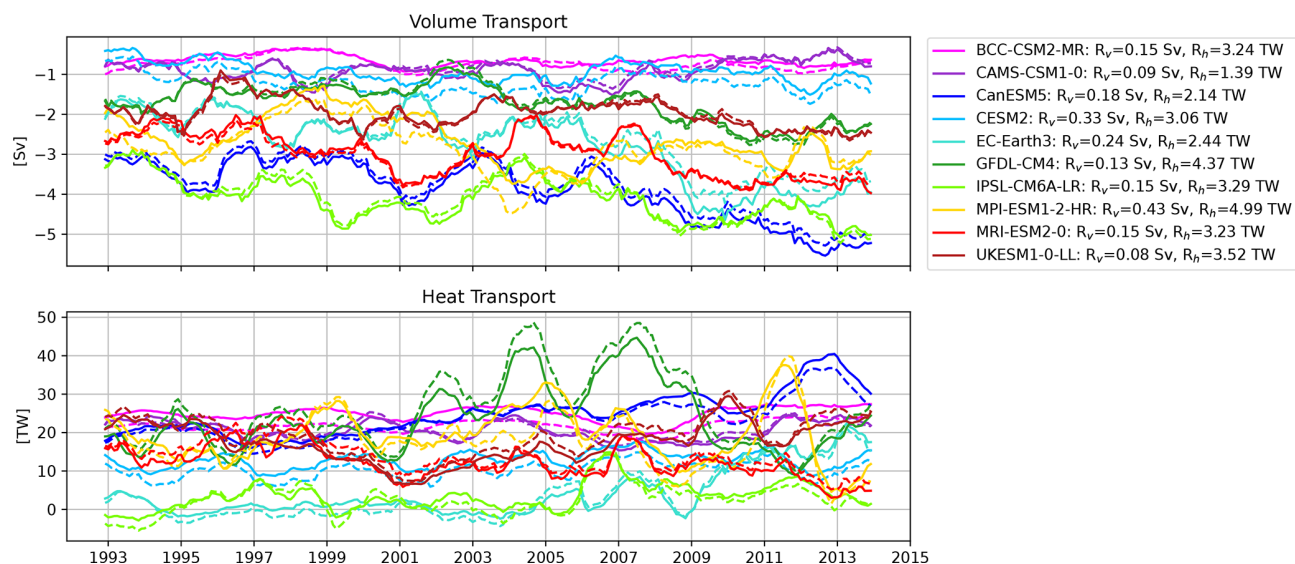


Figure 12. Volume and heat transport time series at Fram Strait for 10 selected CMIP6 models from Heuzé et al. (2023) (solid) and our LM estimates (dashed). Root mean square differences between the Heuzé et al. (2023) estimates and ours for volume (R_v) and heat (R_h) transports over the given time range are added in the line labels.

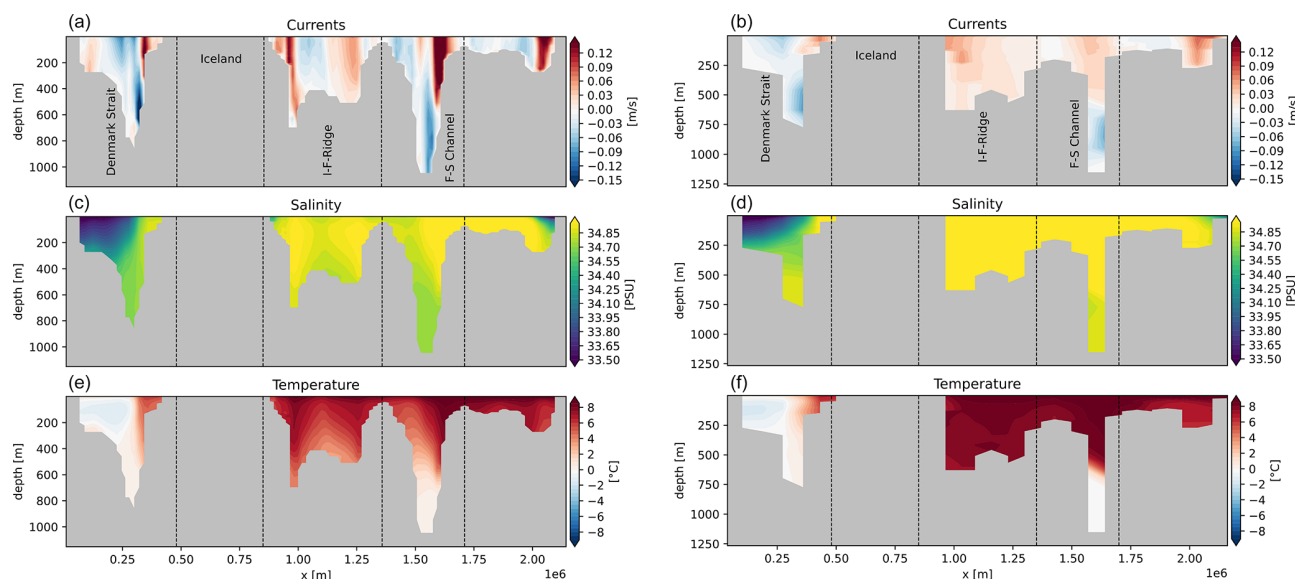


Figure 13. Sample cross sections of currents (a, b), salinity (c, d), and temperature (e, f) for the Greenland–Scotland Ridge for the CNRM-CM6-1-HR (0.25° horizontal resolution; Voldoire et al., 2019) and CNRM-CM6-1 (1° horizontal resolution; Voldoire et al., 2019) CMIP6 models.

Note the big difference in bottom topography and also the depiction of individual currents between the models.

Mayer et al. (2023b) use StraitFlux to compare oceanic transports across the GSR from ocean reanalyses against largely independent observations. They use the results from StraitFlux to partition the water masses into Atlantic, overflow and polar water, enabling a more in-depth analysis. They find that ocean reanalyses underestimate the observed Atlantic water inflow by up to 15 %, causing a low bias in

oceanic heat transports (OHT) of 5 %–22 %. Furthermore, they attribute a pronounced anomaly in OHT during the 2-year period around 2018 to a reduction in Atlantic water inflow through the Faroe–Shetland branch in combination with anomalously cool temperatures of Atlantic water arriving at the GSR due to a recent strengthening of the North Atlantic subpolar gyre. Winkelbauer et al. (2024) use StraitFlux to calculate net transports of volume and heat passing into and out of the Arctic through Fram Strait, Davis Strait, Bering

Table 1. Approximate dimensions and sizes of variables for typical 1/4° models needed for the transport calculations of one time step.

Field	Approx. dimensions	Approx. size [MB]
Seawater <i>X</i> velocity <i>u</i>	1400 × 1000 × 75	300
Seawater <i>Y</i> velocity <i>v</i>	1400 × 1000 × 75	300
Seawater potential temperature <i>T</i>	1400 × 1000 × 75	300
Ocean model cell thickness (thkcello)	1400 × 1000 × 75	300

Table 2. Approximate calculation times for 12 months (m) of monthly data for two exemplary models with 1° and 1/4° resolution and 75 vertical layers (l) at Fram Strait. Functions are divided into “used by both” (top), LM (middle), and VPM (bottom). The single asterisk * indicates functions that only have to be calculated once per model, and the double asterisks ** indicate functions that need to be calculated once per model and strait. Total times are given for the first calculation performed (left values) and every consecutive calculation (right values) when the functions’ parameter “saving” is set to “True” (default).

Horizontal resolution [°]	1° × 1°, 75 l, 12 m	0.25° × 0.25°, 75 l, 12 m
Read files for mesh and index calculation**	1.5 s	3.5 s
Calculate indices**	0.9 s	1.4 s
Determine Arakawa partition*	3 ms	3 ms
Read files (sub-selected)	2.2 s	3.5 s
Calculate dz at cell faces**	25 ms	0.15 s
Line integration method (LM)		
Calculate mesh files*	10 s	44 s
Calculate transports	50 ms	80 ms
Total	15 s/2.3 s	55 s/3.8 s
Vector projection method (VPM)		
Calculate projection vectors and constants**	8.2 s	31 s
Calculate transports	0.2 s	3.3 s
Regrid to section	0.1 s	0.2 s
Total	12.5 s/2.5 s	43 s/7.5 s

Strait, and the Barents Sea Opening (BSO). They assess the transports’ seasonal cycles and find clear correlations between oceanic transports and the Arctic’s mean state. Fritz et al. (2023) use StraitFlux to assess transports in the Indonesian Throughflow (ITF) region and find reasonable agreements between reanalysis-based transports and observations in terms of means, seasonal cycles, and variability. Furthermore, transports have been calculated at the RAPID and OS-NAP sections.

3.2 Computational performance

The transport calculations usually need to involve only a small fraction of the 3D field values stored in the Coupled Model Intercomparison Project (CMIP) or reanalysis archives. As the current archives do not support extraction of subareas, the global fields need to be downloaded, and consequently, a fair amount of the total computational time is spent in reading and preprocessing the files.

For instance, in order to calculate the temperature flux for a typical 1/4° CMIP6 model, the fields listed in Table 1 are needed.

To detect the indices of the section, calculate horizontal meshes, and determine the Arakawa partition, the software reads one vertical layer of the data files. Once the section is known, the software chooses the subregion so that not the whole file needs to be read. While reading the files takes up the majority of the calculation time, the calculation itself is performed relatively fast. Approximate times for the major calculation steps for a Xeon Gold 6148 CPU for a 1° model (CanESM5) and a 1/4° model (EC-Earth3P-HR) at Fram Strait are given in Table 2. The calculation of the mesh files and detection of the Arakawa partition has to be performed only once per model, and the calculation of indices and parameters for the VPM (normal and direct vectors, signs of velocity components, etc.) has to be performed once per model and strait when the functions’ parameter `saving` is set to `True` (default). This speeds up subsequent calculations, e.g., for different months or straits, considerably.

With a monthly time resolution, for $1/4^\circ$ models it is possible to calculate transports directly for multiple years (e.g., the calculation of the 65-year period takes about 60 s); for higher-resolution models; we advise looping the calculation (e.g., over 12 months) to avoid high memory consumption. For faster performance, calculations may of course be done in parallel. Also, the flux calculation for other ensemble members can be done in parallel as well.

3.3 Availability

StraitFlux is available as an open-source Python package at GitHub and Zenodo and can be installed from pypi. The GitHub repository also contains an example script and some example datasets, as well as a requirements file, to simplify the installation and usage of StraitFlux.

StraitFlux is a free software and can be redistributed and/or modified under the terms of the GNU General Public License version 3, as published by the Free Software Foundation.

4 Conclusions

In this study, we have introduced StraitFlux, an open-source Python package designed to facilitate the calculation and analysis of oceanic transports through arbitrary oceanic straits and sections. We give a comprehensive overview of StraitFlux, including its underlying principles, software implementation, validation, and application examples. StraitFlux facilitates scientific studies to validate models and gain valuable insights into ocean circulation, heat transports, and water mass exchanges, making it a useful tool for climate scientists, oceanographers, and modelers.

StraitFlux works on various curvilinear ocean modeling grids and is written so flexibly that it is expected to work for future versions (e.g., CMIP7) as well. Unstructured grids are not included in this release. However, the methods have already been successfully adapted and tested for the FESOM2 ocean model (Danilov et al., 2017), the successor of FESOM which is, for instance, used in the Alfred Wegener Institute Climate Model (AWI-CM) in CMIP6 and is planned to be included in future versions of StraitFlux as well. The tools include two methods for calculating oceanic transports: the line integration method (LM) and the vector projection method (VPM). The LM creates a closed polyline along grid cell faces to compute net integrated transports, while the second method employs vector projection algorithms to estimate the share of u and v components passing orthogonally through the strait and generates cross sections of velocities, temperatures, and salinities in the vertical plane.

Both methods have been thoroughly validated and produce reliable results across various ocean models and grids. Our validation efforts have demonstrated that StraitFlux consistently matches analytical solutions, even in complex grid configurations and regions with strong distortion. Both meth-

ods deliver net transports that match within a few percent of their total value, even at the most distorted sections. The tool's accuracy is further affirmed by comparisons with the divergence of transports and independent transport calculations. One problem remains at the northernmost latitudes for the VPM, which we hope to resolve soon.

The applications of StraitFlux extend to a wide range of research areas. Researchers can use the package to analyze seasonal cycles, mean states, and variability in oceanic transports. Furthermore, the ability to generate cross sections of currents, temperature, and salinity provides a detailed view of the ocean's vertical structure and flow patterns.

In summary, the user-friendly implementation and broad applicability make it a valuable tool for studying the Earth's climate system and its dynamics. The simplified comparison to observational data highlights its suitability for model validation and assessment. We hope that StraitFlux empowers researchers to explore and understand oceanic transports more thoroughly, given their importance in the climate system and changes therein.

Appendix A

Table A1. Start and endpoints for the straits used. The net Arctic transports are calculated using the sum of the Fram, Davis, and Bering straits and the Barents Sea Opening. Additionally, coordinates of the RAPID array and Drake Passage are given, which are used to assess the interpolation error (see Fig. 9). The Greenland–Scotland Ridge is defined as kinked line with four intermediate points.

Strait	Start point (lat [° N], long [° E])	Intermediate points (lat [° N], long [° E])	Endpoint (lat [° N], long [° E])
Fram	(78.82, −20.7)	–	(78.83, 12.00)
Davis	(66.65, −61.80)	–	(67.31, −52.50)
Barents Sea Opening	(78.00, 18.00)	–	(69.20, 19.80)
Bering	(65.99, −170.50)	–	(66.75, −166.00)
Greenland–Scotland Ridge	(68.53, −30.82)	(66.01, −23.24), (64.41, −15.07) (62.07, −6.87), (60.28, −1.17)	(59.47, 6.11)
RAPID	(26.00, −80.50)	–	(26.00, −13.50)
Drake Passage	(−55.70, −66.92)	–	(−64.10, −59.20)

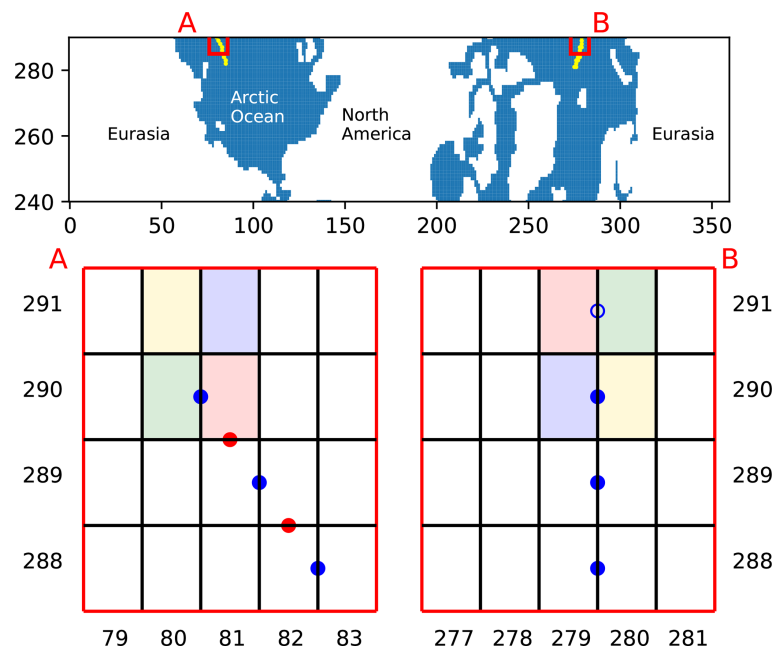


Figure A1. Indices selection across the northern boundary for the CMCC-CM2-SR5 model. The top two rows of grid cells are rotated along the northern boundary, and colorful cells show duplicate cells which are pivoted at the top boundary (= same cells but upside down). Filled blue dots show selected u indices, and filled red dots show selected v indices. The empty blue dot shows the overlapping point which is not selected.

Code and data availability. The Python implementation of StraitFlux is available at <https://github.com/susannawinkelbauer/StraitFlux> (last access: 19 March 2024) and long-term archived on Zenodo (<https://doi.org/10.5281/zenodo.10053554>, Winkelbauer, 2024). It can be installed using pypi. The GitHub repository additionally contains the notebook “Examples.ipynb” with some easy examples to get started with the transport calculations. Data files used in the notebook as well as CMIP6 data used in the validation section of this paper may be downloaded via the Earth System Grid Federation (ESGF) using for instance the Lawrence Livermore National Laboratory (LLNL) node: <https://esgf-node.llnl.gov/search/cmip6/> (ESGF, 2024) or the German Climate Computing (DKRZ) node: <https://esgf-data.dkrz.de/search/cmip6-dkrz/> (ESGF, 2024). Reanalyses data are available via the Copernicus Marine Service website (<https://doi.org/10.48670/moi-00024>, CMEMS, 2022).

Author contributions. SW, MM, and LH conceptualized the study. SW developed the Python implementation of StraitFlux with the help of MM and LH. SW performed the data analysis, including the production of the figures in the paper, and prepared the article. All authors wrote, edited, and reviewed the paper and agreed to the publication of the present version of the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. Susanna Winkelbauer and Michael Mayer have been supported by the Austrian Science Fund project (grant no. P33177) and the Copernicus Marine Service contract (grant no. 21003-COP-GLORAN, Lot 7). The authors sincerely thank Céline Heuzé from the University of Gothenburg for the provision of manually calculated transport data to use as validation and helpful discussions on the transport calculation methodology. We acknowledge work by Vanessa Seitner, who developed an earlier version of StraitFlux. We thank Shizhu Wang and an anonymous referee for their constructive comments that improved this study.

Financial support. This research has been supported by the Austrian Science Fund (grant no. P33177) and the Copernicus Marine Service (grant no. 21003-COP-GLORAN, Lot7).

Review statement. This paper was edited by Qiang Wang and reviewed by Shizhu Wang and one anonymous referee.

References

- Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D’Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Èthè, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levassasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR Climate Model, *J. Adv. Model. Earth Sy.*, 12, e2019MS002010, <https://doi.org/10.1029/2019MS002010>, 2020.
- Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V.: Observed fingerprint of a weakening Atlantic Ocean overturning circulation, *Nature*, 556, 191–196, <https://doi.org/10.1038/s41586-018-0006-5>, 2018.
- Danilov, S., Sidorenko, D., Wang, Q., and Jung, T.: The Finite-volume Sea ice–Ocean Model (FESOM2), *Geosci. Model Dev.*, 10, 765–789, <https://doi.org/10.5194/gmd-10-765-2017>, 2017.
- Ditlevsen, P. and Ditlevsen, S.: Warning of a forthcoming collapse of the Atlantic meridional overturning circulation, *Nat. Commun.*, 14, 4254, <https://doi.org/10.1038/s41467-023-39810-w>, 2023.
- Earth System Grid Federation (ESGF): WCRP CMIP6 search interface, Deutsche Klimarechenzentrum node [data set], <https://esgf-data.dkrz.de/search/cmip6-dkrz/>, last access: 3 June 2024a.
- Earth System Grid Federation (ESGF): WCRP CMIP6 search interface, Lawrence Livermore National Laboratory (LLNL) node [data set], <https://esgf-node.llnl.gov/search/cmip6/>, last access: 3 June 2024b.
- E.U. Copernicus Marine Service Information (CMEMS): Global Ocean Ensemble Physics Reanalysis, Marine Data Store (MDS) [data set], <https://doi.org/10.48670/moi-00024>, 2022.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Fritz, M., Mayer, M., Haimberger, L., and Winkelbauer, S.: Assessment of Indonesian Throughflow transports from ocean reanalyses with mooring-based observations, *Ocean Sci.*, 19, 1203–1223, <https://doi.org/10.5194/os-19-1203-2023>, 2023.
- Griffies, S., Harrison, M., Pacanowski, R., Rosati, A., Liang, Z., Schmidt, M., Simmons, H., and Slater, R.: A Technical Guide to MOM4, Ocean Group Technical Report No. 5, NOAA/Geophysical Fluid Dynamics Laboratory, 2004.
- He, C., Liu, Z., and Hu, A.: The transient response of atmospheric and oceanic heat transports to anthropogenic warming, *Nat. Clim. Change*, 9, 222–226, <https://doi.org/10.1038/s41558-018-0387-3>, 2019.

- Heuzé, C., Zanowski, H., Karam, S., and Muilwijk, M.: The Deep Arctic Ocean and Fram Strait in CMIP6 Models, *J. Climate*, 36, 2551–2584, <https://doi.org/10.1175/JCLI-D-22-0194.1>, 2023.
- Jackson, L. C., Kahana, R., Graham, T., Ringer, M. A., Woollings, T., Mecking, J. V., and Wood, R. A.: Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM, *Clim. Dynam.*, 45, 3299–3316, <https://doi.org/10.1007/s00382-015-2540-2>, 2015.
- Liu, W. and Fedorov, A.: Interaction between Arctic sea ice and the Atlantic meridional overturning circulation in a warming climate, *Clim. Dynam.*, 58, <https://doi.org/10.1007/s00382-021-05993-5>, 2022.
- Lozier, M. S., Bacon, S., Bower, A. S., Cunningham, S. A., de Jong, M. F., de Steur, L., deYoung, B., Fischer, J., Gary, S. F., Greenan, B. J. W., Heimbach, P., Holliday, N. P., Houpert, L., Inall, M. E., Johns, W. E., Johnson, H. L., Karstensen, J., Li, F., Lin, X., Mackay, N., Marshall, D. P., Mercier, H., Myers, P. G., Pickart, R. S., Pillar, H. R., Straneo, F., Thierry, V., Weller, R. A., Williams, R. G., Wilson, C., Yang, J., Zhao, J., and Zika, J. D.: Overturning in the Subpolar North Atlantic Program: A New International Ocean Observing System, *B. Am. Meteorol. Soc.*, 98, 737–752, <https://doi.org/10.1175/BAMS-D-16-0057.1>, 2017.
- Madec, G. and Imbard, M.: A global ocean mesh to overcome the North Pole singularity, *Clim. Dynam.*, 12, 381–388, <https://doi.org/10.1007/BF00211684>, 1996.
- Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D., Ethé, C., Flavoni, S., Graham, T., Harle, J., Iovino, D., Lea, D., Lévy, C., Lovato, T., Martin, N., Masson, S., Mocavero, S., Paul, J., Rousset, C., Storkey, D., Storto, A., and Vancoppenolle, M.: NEMO ocean engine. In Notes du Pôle de modélisation de l'Institut Pierre-Simon Laplace (IPSL) (v3.6-patch, Number 27), Zenodo, <https://doi.org/10.5281/zenodo.3248739>, 2017.
- Mahajan, S., Zhang, R., and Delworth, T. L.: Impact of the Atlantic Meridional Overturning Circulation (AMOC) on Arctic Surface Air Temperature and Sea Ice Variability, *J. Climate*, 24, 6573–6581, <https://doi.org/10.1175/2011JCLI4002.1>, 2011.
- Mayer, J., Haimberger, L., and Mayer, M.: A quantitative assessment of air–sea heat flux trends from ERA5 since 1950 in the North Atlantic basin, *Earth Syst. Dynam.*, 14, 1085–1105, <https://doi.org/10.5194/esd-14-1085-2023>, 2023a.
- Mayer, M., Tsubouchi, T., Winkelbauer, S., Larsen, K. M. H., Berx, B., Macrander, A., Iovino, D., Jónsson, S., and Renshaw, R.: Recent variations in oceanic transports across the Greenland–Scotland Ridge, *State of the Planet*, 1–osr7, 14, <https://doi.org/10.5194/sp-1-osr7-14-2023>, 2023b.
- Muilwijk, M., Smedsrud, L. H., Ilicak, M., and Drange, H.: Atlantic Water Heat Transport Variability in the 20th Century Arctic Ocean From a Global Ocean Model and Observations, *J. Geophys. Res.-Oceans*, 123, 8159–8179, <https://doi.org/10.1029/2018JC014327>, 2018.
- Murray, R. J.: Explicit Generation of Orthogonal Grids for Ocean Models, *J. Comput. Phys.*, 126, 251–273, <https://doi.org/10.1006/jcph.1996.0136>, 1996.
- Outten, S., Esau, I., and Otterå, O. H.: Bjerknes Compensation in the CMIP5 Climate Models, *J. Climate*, 31, 8745–8760, <https://doi.org/10.1175/JCLI-D-18-0058.1>, 2018.
- Park, S., Shin, J., Kim, S., Oh, E., and Kim, Y.: Global Climate Simulated by the Seoul National University Atmosphere Model Version 0 with a Unified Convection Scheme (SAM0-UNICON), *J. Climate*, 32, 2917–2949, <https://doi.org/10.1175/JCLI-D-18-0796.1>, 2019.
- Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S., and Schaffernicht, E. J.: Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation, *Nat. Clim. Change*, 5, 475–480, <https://doi.org/10.1038/nclimate2554>, 2015.
- Rayner, D., Hirschi, J. J.-M., Kanzow, T., Johns, W. E., Wright, P. G., Frajka-Williams, E., Bryden, H. L., Meinen, C. S., Baringer, M. O., Marotzke, J., Beal, L. M., and Cunningham, S. A.: Monitoring the Atlantic meridional overturning circulation, *Deep-Sea Res. Pt II*, 58, 1744–1753, <https://doi.org/10.1016/j.dsr2.2010.10.056>, 2011.
- Schauer, U. and Beszczynska-Möller, A.: Problems with estimation and interpretation of oceanic heat transport – conceptual remarks for the case of Fram Strait in the Arctic Ocean, *Ocean Sci.*, 5, 487–494, <https://doi.org/10.5194/os-5-487-2009>, 2009.
- Schauer, U. and Losch, M.: “Freshwater” in the Ocean is Not a Useful Parameter in Climate Research, *J. Phys. Oceanogr.*, 49, 2309–2321, <https://doi.org/10.1175/JPO-D-19-0102.1>, 2019.
- Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., Dukowicz, J., Eden, C., Fox-Kemper, B., Gent, P., Hecht, M., Jayne, S., Jochum, M., Large, W., Lindsay, K., Maltrud, M., Norton, N., Peacock, S., Vertenstein, M., and Yeager, S.: The Parallel Ocean Program (POP) Reference Manual, <https://opensky.ucar.edu/islandora/object/manuscripts:825> (last access: 3 June 2024), 2010.
- Tsubouchi, T., Bacon, S., Naveira Garabato, A. C., Aksenov, Y., Laxon, S. W., Fahrbach, E., Beszczynska-Möller, A., Hansen, E., Lee, C. M., and Ingvaldsen, R. B.: The Arctic Ocean in summer: A quasi-synoptic inverse estimate of boundary fluxes and water mass transformation, *J. Geophys. Res.-Oceans*, 117, C01024, <https://doi.org/10.1029/2011JC007174>, 2012.
- Tsubouchi, T., Bacon, S., Aksenov, Y., Garabato, A. C. N., Beszczynska-Möller, A., Hansen, E., de Steur, L., Curry, B., and Lee, C. M.: The Arctic Ocean Seasonal Cycles of Heat and Freshwater Fluxes: Observation-Based Inverse Estimates, *J. Phys. Oceanogr.*, 48, 2029–2055, <https://doi.org/10.1175/JPO-D-17-0239.1>, 2018.
- Voltaire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, *J. Adv. Model. Earth Sy.*, 11, 2177–2213, <https://doi.org/10.1029/2019MS001683>, 2019.
- Watterson, I. G.: Decomposition of Global Ocean Currents Using a Simple Iterative Method, *J. Atmos. Ocean. Tech.*, 18, 691–703, [https://doi.org/10.1175/1520-0426\(2001\)018<0691:DOGOCU>2.0.CO;2](https://doi.org/10.1175/1520-0426(2001)018<0691:DOGOCU>2.0.CO;2), 2001.
- Winkelbauer, S.: StraitFlux, Zenodo [code], <https://doi.org/10.5281/zenodo.10053554>, 2024.
- Winkelbauer, S., Mayer, M., and Haimberger, L.: Validation of key Arctic energy and water budget components in CMIP6,

- Clim. Dynam., 62, 3891–3926, <https://doi.org/10.1007/s00382-024-07105-5>, 2024.
- Zanowski, H., Heuzé, C., Muilwijk, M., and Karam, S.: Historical (1985–2015) Fram Strait volume, heat, and salt transports in CMIP6 models, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.949182>, 2023.
- Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, Ocean Sci., 15, 779–808, <https://doi.org/10.5194/os-15-779-2019>, 2019.