

CMIP6 models overestimate sea ice melt, growth and conduction relative to ice mass balance buoy estimates

Alex E. West and Edward W. Blockley

Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK

Correspondence: Alex E. West (alex.west@metoffice.gov.uk)

Received: 28 June 2024 – Discussion started: 26 August 2024 Revised: 4 February 2025 – Accepted: 25 February 2025 – Published: 27 May 2025

Abstract. With the ongoing decline in Arctic sea ice extent, the accurate simulation of Arctic sea ice in coupled models remains an important problem in climate modelling. In this study, the substantial Coupled Model Intercomparison Project Phase 6 (CMIP6) model spread in Arctic sea ice extent and volume is investigated using a novel, process-based approach. An observational dataset derived from the Arctic ice mass balance buoy (IMB) network is used to evaluate fluxes of melt, growth and conduction produced by a subset of CMIP6 models, to better understand the model processes that underlie the large-scale sea ice states. Due to the sparse nature of the IMB observations, the evaluation is performed by comparing distributions of modelled and observed fluxes in the densely sampled regions of the North Pole and Beaufort Sea.

We find that all fluxes are routinely biased high in magnitude with respect to the IMB measurements by nearly all models, with too much melt in summer and too much conduction and growth in winter, even as a function of ice thickness. We also show that fluxes vary in ways which are physically consistent with the thermodynamic parameterisations used and that these effects likely modulate the large-scale relationship between ice thickness and ice growth and melt in the CMIP6 models.

1 Introduction

Arctic sea ice has declined substantially over the satellite record (since 1979), both in terms of extent (Stroeve et al., 2012; Stroeve and Notz, 2018; Cai et al., 2021) and thickness (Kwok, 2018). Model projections from the Coupled Model Intercomparison Project Phase 6 (CMIP6) suggest that an

ice-free Arctic in summer is a likely occurrence within the next 10-30 years: however, models tend to underestimate the sensitivity of summer sea ice to global temperature increase (e.g. Fig. 1d of Notz and SIMIP Community, 2020). In addition, there is substantial variation in the present-day sea ice area and volume simulated by models from the CMIP6 ensemble. The causes are not yet well-understood, although Long et al. (2021) found sea ice extent simulations to compare better to reference datasets in models with higher spatial resolution and greater physical complexity, particularly from December to June. Chen et al. (2023) found a similar, although very weak, association between model resolution and ice volume simulation accuracy relative to the PIOMAS forced ice-ocean model. The difficulty of finding clear associations between model complexity and either ice volume or summer ice extent simulation accuracy underlines the complexity of the processes driving sea ice evolution within the Arctic Ocean.

The mean state and future trend of Arctic sea ice are closely related, as annual mean ice thickness decreases more for thick ice than for thin ice for the same increase in atmospheric forcing (Holland et al., 2006; Chen et al., 2023). This is due to the thickness–growth feedback, by which thinner sea ice grows more quickly in winter (Massonnet et al., 2018). This negative feedback is nonlinear and operates more strongly as ice thickness approaches zero, opposing the direct effects of climate warming. Reducing, although not fully negating, its effects is the surface albedo feedback, a positive feedback operating over larger scales, by which areas of lower average sea ice thickness melt more quickly during summer, due to albedo falling to lower values sooner.

In fact, the sea ice volume is closely coupled to the seasonal ice growth and melt through these processes (West et al., 2022). Seasonal ice growth and melt drive the sea ice volume evolution in an obvious way, but sea ice volume, in turn, modulates how the ice growth and melt respond to thermodynamic forcing from above (atmospheric radiative fluxes and near-surface temperature and humidity) and from below (oceanic heat flux). A schematic view of this relationship is presented in Fig. 1 of West et al. (2020). To understand the drivers of sea ice melt in response to long-term climate warming, it is necessary to understand the evolution of this coupled system. Understanding the causes of variation in present-day Arctic sea ice area and volume is, therefore, an important step towards reducing uncertainty in future projections.

Ideally, then, a full evaluation of Arctic climate in CMIP6 would include not only ice area and volume but also ice growth and melt; the forcing variables of surface radiation, temperature and humidity, as well as oceanic heat flux; and internal ice thermodynamic quantities such as conduction. However, partly due to observational limitations, evaluation of Arctic climate variables apart from sea ice in CMIP6 has been sparse. For example, Henke et al. (2023) evaluated surface temperature in a subset of CMIP6 models relative to ERA5 and found a general cold bias. However, they noted that this could be caused by observational inaccuracy, as reanalysed temperatures are known to be too warm over Arctic sea ice due to lack of surface snow (Batrak and Müller, 2019).

Evaluation of the internal processes of the sea ice is, in principle, even more problematic, due to the extreme difficulties in measuring these quantities. However, in West et al. (2020), a dataset of conduction and mass balance fluxes was constructed from elevation and temperature data from the Arctic ice mass balance buoy (IMB) network (maintained by the Cold Regions Research and Engineering Lab, CRREL), and this dataset was used to evaluate the Coupled Model Intercomparison Project Phase 5 (CMIP5) model HadGEM2-ES. The evaluation produced results consistent with a previous surface radiation and sea ice study of the same model (West et al., 2019), confirming that ice growth, melt and conduction were all too strong in this model, causing the sea ice thickness seasonal cycle to be too amplified. It also elucidated the sea ice simulation further by showing that the model's lack of thermal inertia was likely causing winter ice growth to be too strong, a conclusion that would not have been possible without the IMB evaluation.

The purpose of this study is to apply the same method to the CMIP6 ensemble: to perform a detailed evaluation of CMIP6 internal sea ice thermodynamics – the energy fluxes associated with melt, growth and conduction – relative to fluxes derived from the IMB network using the methods described in West et al. (2020) and West (2021). This evaluation is restricted to a subset of 17 CMIP6 models that provide all of the relevant diagnostics and is combined with a full evaluation of sea ice extent and thickness, global and Arctic temperature, and surface radiative fluxes. Throughout this study, the mass fluxes associated with ice melt and growth are treated as synonymous with the energy fluxes driving these, related by the specific latent heat of fusion of ice $3.35 \times 10^5 \, \text{J kg}^{-1}$. Due to variations in ice salinity and density, the relationship between ice *volume* fluxes and energy fluxes is more complex for the IMBs (as discussed in West et al., 2020). It is also more complex for one particular group of models (Sect. 4.2 of this study).

The study is set out in the following way. In Sect. 2, the models, IMB data and other reference data are introduced. In Sect. 3, the climate states of the models are described by evaluating sea ice extent and other climate variables. In Sect. 4, fluxes of melt, growth and conduction are evaluated with respect to the IMB data. In Sect. 5, this evaluation is extended further to show how these fluxes vary with ice thickness and snow depth, and here we attempt to account for the sampling biases inherent in the IMB data. In Sect. 6, conclusions are presented.

2 Models and data

2.1 CMIP6 models

The CMIP6 data request gave scope for a much larger set of sea ice diagnostics compared with previous projects (Notz et al., 2016). In particular, diagnostics of the sea ice heat and mass budgets were requested, including full components of the sea ice mass balance and of the energy balance at the top and basal surfaces of the snow-ice column. However, not all models provided all or any of these diagnostics. In some cases, although by no means all, this was because sea ice components were sufficiently simple that they would not have been meaningful. In order to be included in this study, a model would provide the following diagnostics: sea ice area, thickness or volume, top and basal melting flux, and top and basal conduction flux. A total of 17 models were identified that provided all of these diagnostics (Table 1), representing contributions from 9 separate modelling centres. Hereafter this subset of CMIP6 models is referred to as the "CMIP6 subset".

The sea ice components of models in the CMIP6 subset share many common features. All use a sub-grid ice thickness distribution (ITD), in which ice in each grid cell is divided into distinct thickness categories. For each category, temperatures and energy fluxes are computed separately. The ITD is important because it allows for rapid ice growth at low thicknesses to be properly captured (Holland et al., 2006; Massonnet et al., 2019). All models in the CMIP6 subset allow sea ice to display thermal inertia, with multiple vertical layers of ice and at least one separate layer of snow simulated for each model. Thermal inertia is likely also important for achieving realistic basal conduction and, hence, winter ice growth (West et al., 2020). All models parameterise oceanto-ice heat flux in similar ways, using schemes derived from McPhee (1992).

set of models, their components and characteristics of thermodynamic parameterisations. In the melt pond column, H12 and H13 refer to Holland et al. (respectively. In the penetrating solar radiation column, "Y" and "N" refer to yes and no, respectively.
AIP6 subset of models, their	(2013), respectively. In the
Table 1. The CN	and Hunke et al.

Table 1. The and Hunke el	t al. (2013), res	of models, their compon pectively. In the penetrati	ents and characteris ng solar radiation c	stics of thern solumn, "Y" ;	nodynamic paran and "N" refer to	neterisations. In th	e melt pond col stively.	umn, H12 and H	113 refer to Hollar	ıd et al. (2012)
Grouping	Institute	Model	Atmosphere	Ocean	Sea ice	Thickness category bounds (m)	Vertical ther- modynamic layers (ice + snow)	Salinity treatment	Melt pond pa- rameterisation	Penetrating solar radiation
GSI8.1	CSIRO- ARCCSS	ACCESS-CM2	UM v10.6 (GA7.1)	GFDL MOM5	CICE5.1.2 (GSI8.1)	0.6, 1.4, 2.4, 3.6	4I + 1S	Prescribed	Topographic	Ν
	MOHC	HadGEM3-GC31-LL		NEMO3.6						
	MOHC	HadGEM3-GC31-MM		NEMO3.6						
	MOHC	UKESM1-0-LL		NEMO3.6						
	NIMS-KMA	UKESM1-0-LL		NEMO3.6						
Mushy- layer	NCAR	CESM2	CAM6	POP2	CICE5.1.2	0.645, 1.391, 2.470, 4.563, 9.338	8I + 3S	Fully prognostic (mushy-layer)	Level ice (H13)	Y
	NCAR	CESM2-WACCM	WACCM6	POP2		0.645, 1.391, 2.470, 4.563, 9.338				
	NCC	NorESM2-LM	CAM6-Nor	BLOM		0.6, 1.4, 2.4, 3.6				
	NCC	NorESM2-MM	CAM6-Nor	BLOM		0.6, 1.4, 2.4, 3.6				
CMCC	As grouping	CMCC-CM2	CAM5	NEMO3.6	CICE4	0.64, 1.39, 2.47, 4.57	4I + 1S	Prescribed	Level ice (H12)	Y
		CMCC-ESM2						I	Level ice (H12)	
IPSL	As grouping	IPSL-CM6A-LR	LMDZ6A-LR	NEMO3.6	LIM3.6	From formula	2I + 1S	Prognostic bulk salinity	Parameterised via albedo	Y
		IPSL-CM6A-LR- INCA								
MRI	As grouping	MRI-ESM2	MRI-AGCM3.5	MRI- COMv4	In-house (within ocean model)	0.6, 1.4, 2.4, 3.6	11 + 1S	Prescribed	Parameterised via albedo	Y
CNRM- CERFACS	As grouping	CNRM-CM6-1	ARPEGE-Climat v6.3	NEMO3.6	GELATO6	0.3, 0.7, 1.2, 2	9I + 1S	Prognostic bulk salinity	Parameterised via albedo	Y
		CNRM-CM6-1-HR								
		CNRM-ESM2-1								

Despite these common features, the models also differ considerably, and can be grouped, in terms of thermodynamic characteristics. Here, we describe these characteristic groups, which are referred to frequently in the paper. The first two groups contain models from multiple institutions; the remaining groups correspond to single institutions. These aforementioned groups are as follows:

- 1. *GSI8.1 group.* This group comprises the five CSIRO-ARCCSS, MOHC and NIMS-KMA models, which all use the GSI8.1 sea ice configuration (Ridley et al., 2018a). This configuration uses version 5.1.2 of the Los Alamos sea ice model CICE and features multilayer thermodynamics with a fixed salinity profile (Bitz and Lipscomb, 1999). It is notable due to its lack of solar radiation penetrating into ice; all incident shortwave radiation is either reflected or absorbed at the surface. Melt ponds are modelled using the topographic scheme of Flocco et al. (2012).
- 2. *Mushy-layer group*. This group comprises the four NCAR and NCC models, which use a different configuration of CICE5.1.2. In this configuration, penetrating solar radiation is modelled, and salinity is fully prognostic, with a "mushy" liquid-ice layer at the base of the ice (Turner and Hunke, 2015); as with the GSI8.1 models, melt ponds are explicitly modelled, using a level-ice rather than topographic scheme (Hunke et al., 2013).
- 3. *CMCC group.* The two CMCC models use a different configuration of CICE again, this time with CICE version 4; penetrating solar radiation is modelled, but salinity is prescribed. Melt ponds are simulated using the level-ice scheme of Holland et al. (2012), a simpler version of that used by the mushy-layer model group.
- 4. *IPSL group.* The two IPSL models use the LIM3 sea ice model; these models are distinguished by a salinity scheme of intermediate complexity, in which a linear profile is derived from a prognostic bulk salinity (Boucher et al., 2020). They use only two vertical ice layers, but penetrating solar radiation is permitted. They do not simulate melt ponds explicitly; rather, they model their effect on shortwave radiation through a parameterisation of albedo based on surface temperature.
- 5. MRI "group" (one model only). MRI-ESM2-0 uses a custom-built sea ice model with a single ice and snow layer, a fixed salinity profile, and penetrating solar radiation permitted, based on Mellor and Kantha (1989). Melt ponds are parameterised using a similar framework to that of the IPSL group.
- 6. *CNRM-CERFACS group*. The three CNRM-CERFACS models use GELATO6, a sea ice model with salinity, thermodynamics and melt pond treatment similar to LIM3 but with nine vertical ice layers instead of two (Voldoire et al., 2019).

It is important to note that a model's sea ice simulation is not entirely or even mostly controlled by the characteristics of its sea ice component. The forcings received from the ocean and, especially, the atmosphere component control the sea ice simulation to first order (e.g. Olonscheck et al., 2019). The atmosphere and ocean components of the CMIP6 subset are also shown in Table 1. More than half of the models feature the same ocean component (NEMO3.6). Models with identical sea ice components tend to use closely related atmosphere components: for example, the GSI8.1 group models all use UM version 10.6 (GA7.1 configuration; Walters et al., 2019), whereas the mushy-layer group models use atmospheric components derived from CAM6 (Danabasoglu et al., 2020).

2.2 The IMB data

An ice mass balance buoy (IMB) is a collection of instruments frozen into a sea ice floe (Richter-Menge et al., 2006). An IMB typically consists of acoustic sounders to measure ice surface and base elevation, a thermistor string to measure ice and snow temperatures at a 10 cm vertical resolution, and a data logger to record and transmit data. Some also include air temperature and sea level pressure sensors, but these variables will not be examined in this study. Since 1993, 110 IMBs have been deployed in the Arctic Ocean by the Cold Regions Research and Engineering Laboratory (CRREL), mainly in the North Pole and Beaufort Sea subregions (Fig. 1). Individual buoys have been analysed to provide useful process studies of variables such as ocean heat flux (Lei et al., 2014). We note that, since 2015, many IMBs have also been deployed by other institutions, notably in the course of the MOSAiC experiment (e.g. Koo et al., 2021). However, due to the complexity of data processing, we do not attempt to enlarge the dataset used in this study relative to that used in West et al. (2020).

In West et al. (2020), IMB data were systematically analysed to produce a dataset of monthly mean ice melt, growth and conduction fluxes for the North Pole and Beaufort Sea regions that was used to evaluate ice thermodynamics in a CMIP5 model, HadGEM2-ES, identifying a number of model biases (as described in Sect. 1). The processing of the raw IMB data is described fully in West et al. (2020), but it is briefly summarised again here. IMB raw time series of temperature or surface/interface/base elevation, measured at irregular times, are interpolated or binomially averaged in time to create regular time series which are then used to create time series of sea ice thickness and snow depth (West, 2020c). Fluxes of melt and growth are calculated from the elevation and temperature time series. Fluxes of conduction and heat storage are calculated from temperature gradients, using a reference layer 40-70 cm above the ice base for the basal conduction, as gradients are very weak at the ice base (West, 2020d). The IMB dataset contained around 500 data points for each analysed flux and displayed seasonal and spa-



Figure 1. A map of all IMB tracks in the Arctic between 1993 and 2015 used in this study. The Arctic Ocean region (blue shading) and the North Pole (dark-blue box) and Beaufort Sea (dark-red box) subregions used in the analysis are indicated.

tial variability consistent with observational and theoretical understanding of the Arctic Ocean climate. Due to the sparseness of the data, interannual variability could not be detected. Uncertainty in the derived fluxes due to ice salinity, conductivity and density was quantified, in addition to uncertainty due to the choice of the reference layer used to calculate basal conductive fluxes. While uncertainty due to measurement error was not quantified, the values identified by Lei et al. (2014) of 0.01 m and 0.1 K for elevation and temperature measurement, respectively, imply uncertainties over an order of magnitude lower than those identified for the factors listed above.

2.3 Other observational datasets used in this study

Other Arctic climate variables besides ice energy fluxes are evaluated in Sect. 3, and the datasets used are described here. For ice area, we use HadISST.2 (Titchner and Rayner, 2014). For surface temperature, we use GISTemp v4 (GISTemp Team, 2024; Lenssen et al., 2019), HadCRUT v5 (Morice et al., 2021), NOAAGlobalTemp v6.0.0 (Huang et al., 2024) and Berkeley Earth (Rohde and Hausfather, 2020), with these four datasets used to characterise the plausible range of observational uncertainty.

For ice thickness, we use the PIOMAS forced ice–ocean model reanalysis (Schweiger et al., 2011) and CryoSat-2 radar altimetry (Kurtz and Harbeck, 2017). To characterise observational uncertainty in ice thickness, we use a boot-

strapping method trained on the brief period of overlap of CryoSat-2 with our assessment period, 2011–2014, to generate, for each region and month, a plausible distribution of the discrepancy between PIOMAS and CryoSat-2. These were used to derive ranges of ice thickness for each month of the year as well as the annual mean ice thickness and seasonal cycle amplitude. It is hoped that the use of CryoSat-2 ameliorates any bias arising from the use of PIOMAS, which itself contains a sea ice model like many evaluated here, as a reference dataset.

In addition, surface radiative fluxes are evaluated in Sect. 3, with respect to the ERA5 analysis (Hersbach et al., 2023) and CERES-EBAF (Loeb et al., 2009). We do not attempt to explicitly characterise observational uncertainty in these variables.

3 Sea ice and Arctic climate state evaluation

In this section, the sea ice state (area and thickness) simulated by the CMIP6 subset models is evaluated. Throughout, we restrict evaluation to the Arctic Ocean region (Fig. 1). The evaluation period chosen is the last 30 years of the CMIP6 historical simulations, 1985–2014.

Because ice expansion in the Arctic Ocean region is largely limited by the Eurasian and North American continents, most inter-model variation in ice area occurs in the summer (Fig. 2a). Six models achieve minimum ice area in August (the GSI8.1 models and NorESM2-MM); all others, like HadISST.2, achieve minimum area in September. In September, the highest area occurs in MOHC UKESM1-0-LL and NIMS-KMA UKESM1-0-LL ($5.9 \times 10^6 \text{ km}^2$). The lowest ice areas occur in the two CMCC models ($0.0 \times 10^6 \text{ km}^2$), but these are outliers: the next lowest ice area, of $2.6 \times 10^6 \text{ km}^2$, occurs in NCAR CESM2. Most models simulate year-round ice area that is either like or much lower than that suggested by HadISST.2.

There is considerable spread in annual mean ice thickness (Fig. 2b), with the thickest ice in MOHC UKESM1-0-LL (annual mean of 3.3 m) and the thinnest ice in the CMCC models (annual mean of 0.5 m). The PIOMAS and CryoSat-2 observational references sit roughly in the middle of the model range, with PIOMAS displaying an annual mean thickness of 1.7 m. All models achieve maximum ice thickness in either April (the CMCC and CNRM-CERFACS models) or May; PIOMAS also achieves maximum ice thickness in May, but CryoSat-2 is much earlier, in March, although, due to problems with retrieving sea ice thickness measurements during the summer months (e.g. as described in Tilling et al., 2018), it is not clear whether or not this indicates a general model inaccuracy. Minimum ice thickness is achieved in September in both PIOMAS and CryoSat-2, as well as for all models except CMCC-ESM2 and MRI-ESM2-0, which achieve their minimum in October.



Figure 2. (a) Total Arctic Ocean ice area and **(b)** average Arctic Ocean ice thickness in the CMIP6 subset (1985–2014 average with bars denoting 2 times the interannual standard deviation). HadISST.2 for area and PIOMAS and CryoSat-2 for ice thickness are shown for comparison; for CryoSat-2, the period shown is 2011–2020; however, for the other datasets, 1985–2014 is used. **(c)** Scatter plot of annual mean ice thickness against September ice area (top panel) and ice growth/melt diagnosed from the October–September mean minus the April–May mean ice thickness (bottom panel).

We define annual ice growth and melt as the difference in mean Arctic Ocean sea ice thickness between the seasonal maximum in April–May and the seasonal minimum in September–October. This quantity is highest in IPSL-CM6A-LR at 1.46 m and lowest in NorESM2-MM at 1.00 m (Fig. 2c), whilst the value for PIOMAS is 1.15 m. While annual ice growth and melt is, in theory, strongly influenced by the annual mean ice thickness via the surface albedo and thickness–growth feedbacks, these quantities are only weakly negatively correlated across the ensemble, with a correlation coefficient of -0.27. The CMCC and CNRM-CERFACS models are instrumental in this lack of correlation, displaying both low annual mean sea ice thickness and low annual ice growth/melt; without these models, the correlation is -0.79. For the CMCC models, the lack of growth/melt is likely related to the complete loss of sea ice in many parts of the Arctic during July/August; a possible reason for the CNRM-CERFACS models is discussed in Sect. 4 below. Among the other models, the GSI8.1 and the IPSL models tend to lie on a higher curve than the mushylayer and MRI models (ice growth/melt is higher for a given annual mean ice thickness); the two smaller model groups display correlation of -0.93 and -0.90, respectively. Compared to the models, ice growth/melt from PIOMAS is relatively low: it estimates ice growth/melt similar to that of the CMCC/CNRM-CERFACS models, but annual mean ice thickness is most similar to that of NCAR CESM2-WACCM.

There is strong correlation (0.81) between annual mean ice thickness and September ice area amongst the CMIP6 subset (Fig. 2c). This correlation remains strong when the outlier CMCC models are removed (0.79).

We compare the annual mean ice thickness to the anomaly in global 2 m air temperature relative to the 1850-1899 average (Fig. 3a) and to the 2 m air temperature anomaly averaged over the Arctic Ocean region (Fig. 3b). There is a strong correlation between 2 m air temperature over the Arctic Ocean and ice thickness (-0.85). The correlation between global temperature and ice thickness is also quite high (-0.79). We evaluate the models' global and Arctic Ocean 2 m air temperature using the four datasets described in Sect. 2 to represent observational uncertainty. We evaluate the models' sea ice thickness with PIOMAS and CryoSat-2. Models tend to simulate greater ice thickness for a given Arctic Ocean and global temperature anomaly than is suggested by observations, which may be related to model tendency to underestimate sea ice response to a given rise in global temperature (Notz and SIMIP Community, 2020).

We finally evaluate surface radiative fluxes relative to the ERA5 reanalysis and to the CERES-EBAF satellite dataset (Fig. 4). Surface radiative fluxes are important for the Arctic sea ice seasonal cycle, as they are the principal driver of the surface flux variation over sea ice (and, hence, ice melt and growth). The CMCC models are notably distinct from all other models, displaying higher summer net shortwave (SW) fluxes (due to upwelling SW differences; Fig. 4a) and higher absolute longwave (LW) fluxes in both directions during autumn (Fig. 4b); both are close corollaries of these models' very low summer sea ice cover. The mushy-layer models tend to display lower downwelling SW fluxes in summer and higher downwelling LW fluxes in spring and early summer, suggesting more extensive cloud cover in these models. The GSI8.1 models (and, to a lesser extent, the IPSL models) model lower absolute LW fluxes in both directions during the cold season; for the MOHC models, this bias was noted in West (2021) and is likely related to a low liquid-water fraction in clouds.

Annual mean net SW and net LW are anticorrelated across the CMIP6 subset (Fig. 4d) such that the total net radiative flux varies little between models, with all but the CMCC models averaging between -0.5 and 6.4 W m^{-2} ; the average for ERA5 and CERES-EBAF is 0.5 and 1.5 W m^{-2} , respectively. The CMCC models show far more net SW (and, hence, net radiation) than is indicated by ERA5 and CERES-EBAF; the mushy-layer models show somewhat more net LW and less net SW. All other models lie quite close to ERA5 and CERES-EBAF.

It is likely that the total net surface flux, unlike the net radiative flux, is net upwards, with turbulent fluxes over openwater areas in winter providing much of the additional negative component. For example, Table 1 of Winkelbauer et al. (2024) shows the Arctic Ocean average net surface flux to be upwards in the vast majority of CMIP6 models.

4 Mass balance and thermodynamics evaluation

In this section, modelled fluxes of ice growth and melt and those of vertical conduction at the ice surface and base are evaluated with respect to the IMB values. All evaluated fluxes are available as direct model diagnostics; the only processing required is to divide melt and growth fluxes by ice area fraction. This is because these fluxes are produced as grid box means, averages over both sea-ice-covered and open-water areas; for greater comparability with the IMB values, they must be converted to ice-only means by dividing out ice area fraction. The conduction fluxes are produced in their raw form as means over ice and, thus, do not require this processing step.

Throughout this section, modelled fluxes are compared and evaluated for the comparatively well-sampled North Pole and Beaufort Sea subregions shown in Fig. 1 using the equivalent fluxes derived from IMB values. Fluxes from model points within these regions are collected into distributions, similar to the distributions derived from the IMB data but with many more data points. All model statistics are computed from these distributions using weighting by both grid cell area and by ice concentration. The similarity of distributions is assessed using a Welch *t* test, and differences are considered significant at the 5 % level.

4.1 Melt and growth fluxes

4.1.1 Top melting

The IMB measurements show that top melting fluxes are near zero outside the summer months and that they achieve their maximum in July, at 23 and 41 W m⁻² in the North Pole and Beaufort Sea subregions, respectively (Fig. 5a, b). Most models reproduce this shape, except for the CNRM-CERFACS models which have a greatly delayed seasonal cycle, displaying their seasonal maximum in September. Of the remaining models, the CMCC models display the highest top melting fluxes in both regions, reaching around $100 \,\mathrm{W}\,\mathrm{m}^{-2}$ in July; CSIRO-ARCCSS ACCESS-CM2 is next highest, again in both regions. At the lower end of the distribution, NorESM2-MM displays the lowest maximum at 15.8 and $42.2 \,\mathrm{W}\,\mathrm{m}^{-2}$ in the North Pole and Beaufort Sea regions, respectively; it is the only model whose maximum is below the IMB average, in the North Pole region only, although the difference is not significant.



Figure 3. The 1985–2014 Arctic Ocean region mean ice thickness compared to the global (top) and 2 m air temperature anomaly over the Arctic Ocean region relative to 1850–1899 (bottom). The black symbols and grey shaded regions represent the respective observational estimates and uncertainty intervals for the ice thickness and temperature anomaly (derived as described in Sect. 2.3).

Model distributions significantly different from those of the IMB values are indicated using bold lines and black triangles for means in the box plots in Fig. 5; those not significantly different have box plots with fainter lines and green triangles for means. All models except NCC NorESM2-MM are either biased high relative to the observations or are not significantly different from the observations. Apart from the CNRM-CERFACS models, which are phase-shifted, and the CMCC models, which are the highest, there is a partition between the GSI8.1 models and the mushy-layer models, with the former tending to display much higher top melting fluxes than the latter. Among the remaining models, the IPSL models tend to lie closer to the GSI8.1 group, whereas MRI- ESM2-0 is closer to the mushy-layer group. With a handful of exceptions, in most regions and months, the GSI8.1 and IPSL models are biased significantly high with respect to the IMB values, whereas the mushy-layer models and MRI-ESM2-0 are not.

4.1.2 Basal melting

Basal melting values are small in the IMB measurements outside the months of June–September (Fig. 5c, d), although, unlike for top melting, a small number of nonzero winter basal melting fluxes occur in the North Pole region which includes warmer waters at the Atlantic sea ice edge. They display maximum basal melt in August, contrasting with the



Figure 4. Evaluation of radiative fluxes in the CMIP6 subset relative to the ERA5 reanalysis and CERES-EBAF satellite dataset. Panel (a) shows, from top to bottom, downwelling, upwelling and net down SW radiation; panel (b) shows, from top to bottom, downwelling, net down and upwelling LW radiation; panel (c) presents the total net radiative flux; and panel (d) is a scatter plot of the annual mean net SW versus annual mean net LW, with isolines of the total net radiative flux overplotted. For panel (a), net down SW radiation is distinguished by triangle markers where the model spread overlaps with downwelling SW.

July maximum for top melt. From June to September, the average IMB basal melting fluxes are 7.8, 12.8, 18.6 and 6.5 W m^{-2} in the North Pole region and 10.9, 27.3, 32.3 and 20.6 W m⁻² in the Beaufort Sea region. Most models reproduce the shape of this seasonal cycle; the CNRM-CERFACS

models are phase-shifted earlier in the season in the North Pole region particularly, but the phase shift is not as severe as for the top melt. The CMCC models display much greater basal melt values than the other models, with CMCC-CM2-SR5 reaching 374 Wm^{-2} in the Beaufort Sea region in Au-



Figure 5. Seasonal cycles of (**a**, **b**) top melting flux, (**c**, **d**) basal melting flux, and (**e**, **f**) basal growth flux in the North Pole (**a**, **c**, **e**) and Beaufort Sea (**b**, **d**, **f**) regions, with the ice-area-weighted mean and standard deviation shown. Means and standard deviations are taken across all grid cells in the respective regions and across all years in the study period (1985–2014). For each flux and region, insets show box plots of ice-area-weighted statistics across all grid cells for 3 key months of the year: June–August for melting fluxes; November, January and March for growth fluxes. The IMB distribution mean and standard deviation, as well as box plots (black box and shaded area), are shown for each flux and region for comparison. Modelled distributions judged to be significantly different from the IMB values are distinguished by bold lines and black triangles for means.

gust. In most cases, the CMCC models do not report data for September, as there is essentially no ice left in this month.

Among the other models, IPSL-CM6A-LR-INCA displays the highest maximum basal melting in the North Pole region (51.6 W m⁻²), while NCAR CESM2 is highest in the Beaufort Sea region (70.8 W m⁻²). NIMS-KMA UKESM1-

0-LL and MOHC UKESM1.0 display the lowest maxima for the respective regions (12.5 and 17.2 W m⁻², respectively). In contrast to the top melt, the GSI8.1 models tend to display among the lowest basal melting fluxes, with ACCESS-CM2 being a notable exception. This is consistent with the finding of Keen et al. (2021) that the portion of melt attributable to top melt is much higher in models that do not allow penetrating shortwave radiation.

In the North Pole region, basal melting in the two UKESM1-0 models is biased moderately low relative to the IMB values, while the remaining MOHC models and the NCC models are not significantly different from the IMB values. All other models are biased high relative to the IMB data in this region. In the Beaufort Sea region, the two UKESM1-0 representatives remain biased low, but many of the models that are biased high tend to overlap more with the IMB values than is the case in the North Pole region. For example, the quartiles of ACCESS-CM2 are nearly identical to those of the IMB distribution, but a smaller number of very high fluxes cause the mean to be much higher and the distribution to be significantly different.

4.1.3 Basal growth

Basal growth fluxes are near zero during the summer in the IMB data but increase in magnitude during middle to late autumn and early winter to reach their greatest magnitude of -25.5 and -14.1 W m⁻² in February in the North Pole and Beaufort Sea regions, respectively (Fig. 5e, f). The North Pole time series is somewhat noisy in winter, with much lower values in February than the other cold-season months. The CMCC and CNRM-CERFACS models display greatly enhanced basal growth fluxes relative to the IMB values, with their highest magnitude exceeding $-40 \,\mathrm{W}\,\mathrm{m}^{-2}$ or, in some cases, even approaching -100 W m^{-2} ; the CNRM-CERFACS models, in addition, display a severe phase lead (5 months), with minima occurring in September. The remaining models produce minima in a similar range to the IMB data in the North Pole region, but this is largely due to the lower IMB values in February; in December and January, most models are biased low, with only UKESM1-0-LL, MRI-ESM2-0 and the two NCC models being similar to the IMB distribution. In the Beaufort Sea region, all models are biased low relative to the IMB values (MRI-ESM2-0 barely). Aside from the CMCC and CNRM-CERFACS models, the CSIRO-ARCCSS, IPSL and NCAR models tend to be biased most severely, with minima in the region of -30 W m^{-2} .

There is a consistent phase lead of 2–3 months among the non-CNRM-CERFACS models, with all of these attaining their minimum in November or December, rather than February. Associated with this is a severe model bias in the autumn: the CMCC and CNRM-CERFACS models aside, the mean and standard deviation of the modelled autumn basal growth flux is $-15.8 \pm 3.8 \text{ W m}^{-2}$ in the North Pole region; this compares to -3.0 W m^{-2} in the IMB data. This is likely to be caused, at least partially, by a sampling bias in the IMB measurement collection: in autumn, the strongest ice growth tends to occur in thin, newly forming ice. For two reasons, this ice is undersampled by the IMBs. Firstly, IMBs are preferentially placed in thicker ice floes, as these are easier to access and enhance the expected survival period of the IMBs. Secondly, while models report fluxes from an Eulerian perspective that automatically includes contributions from all sea ice within a grid cell, IMBs report from a Lagrangian perspective that inherently biases the sampled ice thickness distribution towards ice floes undergoing slower thickness changes. Ice floes that grow rapidly do not tend to do so for long, as the thickening ice weakens the vertical temperature gradient and, hence, heat loss. An IMB that happens to measure fast ice growth remains trapped in the same ice floe and will not subsequently measure growth in new areas. This sampling bias was investigated in West et al. (2020) and West (2021), and it was found to contribute partially to, although not completely explain, the model biases identified in these studies. In Sect. 5, it is assessed again in the context of the biases identified for the CMIP6 models.

4.2 Conduction and heat storage

4.2.1 Top conductive flux

The top conductive flux represents the downwards flux of energy from the surface of the snow-ice column into the snow/ice interior; the sign convention used throughout is that downwards fluxes are positive. The IMBs measure the top conductive flux as being strongly upwards in winter, representing heat lost to the atmosphere, and weakly downwards in summer (Fig. 6a, b). From October to March, the average flux is between -18 and -21 W m⁻² in the North Pole region (with a minimum in December) and between -12 and -19 W m⁻² in the Beaufort Sea region (with a minimum in November). The summer maximum, achieved in July, is 4.4 and 2.9 W m⁻² in the respective regions.

All models simulate somewhat stronger top conduction in winter than that measured by the IMBs, showing more negative values, representing greater heat loss to the atmosphere. This is more marked in the Beaufort Sea than in the North Pole region. For example, the January average top conductive flux in the North Pole region ranges from -21 W m^{-2} (MOHC UKESM1-0 LL) to -39 W m^{-2} (CMCC-CM2-0); in the Beaufort Sea region, the average ranges from -26 to -50 W m^{-2} (the same models at the extremes). Models with a lower annual mean ice thickness tend to display greater conduction in winter and vice versa. The relationship between thickness and conduction is evaluated explicitly in Sect. 5.

Another notable feature is the five distinct clusters of models formed by the differing seasonal cycle shapes and amplitudes. The CMCC models display a very amplified seasonal cycle with maxima in excess of -100 W m^{-2} for both models and regions, consistent with their very low annual mean ice thickness. The CNRM-CERFACS models are not particularly amplified but, again, display a phase shift, with minima occurring around April. The remaining models display seasonal cycles more similar to the IMB values, although with smaller differences between them. The GSI8.1 models form



Figure 6. Seasonal cycles of (a, b) top conduction flux, (c, d) basal conduction flux, and (e, f) heat storage flux for the North Pole (a, c, e) and Beaufort Sea (b, d, f) regions, with the ice-area-weighted mean and standard deviation shown. Means and standard deviations are taken across all grid cells in the respective regions and across all years in the study period (1985–2014). For the conduction fluxes in each region, insets show box plots of ice-area-weighted statistics across all grid cells for January and November. The IMB distribution mean and standard deviation, as well as box plots (black box and shaded area), are shown for each flux and region for comparison.

a distinct cluster with the highest summer conduction fluxes (biased high relative to the IMB values), possibly due to the lack of solar penetration: full SW absorption at the surface causes a warmer surface and more conduction into the interior. These models also display amongst the weakest winter conduction (most similar to the IMB values). The mushylayer models of NCAR and NCC are similar to the GSI8.1 models in winter, but they display negative conduction in summer (heat is conducted upwards from the interior to the top surface), indicating that these models warm the ice interior more rapidly than the GSI8.1 models, possibly partly because they allow solar radiation to penetrate the ice. The remaining models, MRI-ESM2-0 and the two IPSL models, are most similar to the IMB data in summer, with a mixture of weakly positive and weakly negative fluxes, but they dis-

play somewhat stronger conduction in winter than the IMB values and the GSI8.1 and mushy-layer models.

4.2.2 Basal conductive flux

The basal conductive flux represents the downwards flux from the ice interior into the ice base, driven by heat loss from the ice surface but modulated by the ice heat capacity, and is the principal driver of winter ice growth. As for top conduction, the basal conductive flux is shown as positive downwards throughout this section; during the winter, when conduction tends to occur upwards from the warm ocean to the cold atmosphere, conductive fluxes are usually negative. During the summer, basal conduction fluxes are usually small; inter-model variation in basal melting tends to be driven by the oceanic heat flux, rather than by the basal conductive flux; and variations in oceanic heat flux are, in turn, mainly driven by variations in direct solar heating of the ocean (Maykut and McPhee, 1995; Steele et al., 2010; Keen and Blockley, 2018).

The IMBs' basal conductive fluxes are similar to top conduction but tend to be smaller in magnitude and with the seasonal cycle shifted slightly later in the year (Fig. 6c, d). In both regions, the maximum upwards flux occurs in January, at -14 to -15 W m⁻². Fluxes rise sharply towards zero between April and June, becoming very weakly positive in July and August. Mean fluxes fall sharply to negative values again in November (North Pole) and October (Beaufort Sea).

For the basal conductive flux, as for the top conductive flux, the CMIP6 subset of models form five qualitatively distinct model clusters. The CMCC models display the characteristic exceptionally high amplitude, with values in excess of -100 W m^{-2} occurring in October for all region and model combinations; they also display high positive fluxes in summer, reaching 20 to 40 W m^{-2} in July. The CNRM-CERFACS models also have an elevated amplitude in winter, reaching -40 to -60 W m^{-2} in the North Pole region but again approaching or exceeding -100 W m^{-2} in the Beaufort Sea region; they also display a small phase lead, with the highest negative mean values occurring in September. However, fluxes do not turn positive in summer.

The mushy-layer models (from NCC and NCAR) display much smaller basal conduction fluxes in winter than all other models and the IMB data, in most cases not exceeding -5 W m^{-2} . This is due to other terms in the basal heat balance not reported in CMIP6 diagnostics (Elizabeth Hunke , personal communication, 2023). Notably, because ice is formed at a much lower density in the mushy-layer scheme, a given energy flux is able to produce a much greater increase in ice thickness, with the associated increased shallowing of the temperature gradient and reduced conduction. While this may, in part, represent a real-world effect, the counterpart to this "missing" energy is the energy released during internal freezing of brine pockets within sea ice, which is not obviously reported in the CMIP6 diagnostics. Plante et al. (2024)

discuss additional problems with the mushy-layer formulation as used in CICE, in particular noting that the congelation growth flux does not include contributions from the full energy loss at the base of the ice, resulting in the imbalance being transferred to frazil formation instead. These issues, as well as the IMB evaluation, suggest that these models' basal conductive fluxes may be biased low in magnitude during the freezing season.

The mushy-layer models also display higher positive fluxes in summer than all but the CMCC models (denoting strong conduction to the ice base from the ice interior), in many cases exceeding 10 W m^{-2} . Among the remaining models, the GSI8.1 models display behaviour distinct from that of the other non-GSI8.1 models (the MRI and IPSL models). Firstly, the GSI8.1 models have smaller (i.e. less negative) winter fluxes, closer to the IMB values. Secondly, the GSI8.1 models, like the CNRM-CERFACS models, continue to simulate weakly negative fluxes in summer, whereas the MRI and IPSL models simulate positive values similar to the IMB data.

4.2.3 Heat storage flux

The heat storage flux is calculated as the top conductive flux minus the basal conductive flux, and it represents the rate of change of the heat content of the snow-ice column: negative represents ice cooling, whereas positive represents ice warming. The IMB data show maximum ice warming rates in May and June (5–7 W m⁻²). In the North Pole region, maximum ice cooling occurs in October at -19 W m⁻²; however, in the Beaufort Sea region, the heat storage term has few points and a high standard deviation at this time of year. This may be related to the dataset being dominated by thinner ice for which the calculation of the basal conductive flux 40–70 cm above the ice base is less valid.

As for the conductive fluxes, the IMB models form five clusters with their own distinctive behaviour. The GSI8.1 and MRI/IPSL clusters are the most qualitatively similar to the IMB values, with cooling from September to March and warming from April to August; all models but MRI-ESM2-0 show more cooling than the IMB data during winter, while the GSI8.1 models show much stronger sensible heating in summer than MRI/IPSL and the IMB data. The mushy-layer models show a similar seasonal cycle shape, although it is translated downwards: stronger cooling in winter and no warming in summer. This is likely related to the additional terms in the basal energy balance mentioned in the basal conduction evaluation above.

The CMCC models show similar cooling to the IMB measurements in winter but are very different in summer, showing stronger cooling than in winter (basal conduction higher than top conduction). The CNRM-CERFACS models show an even more curious seasonal cycle, with the strongest ice cooling in April and May but exceptionally strong ice warming in September and October driven by the strongly negative basal conduction flux.

4.3 Aggregate metrics

We compute the seasonal total top and basal melt for each model by multiplying, for each model, month and region, the top plus basal melting by the ice area in that region and then summing over the months of the year. We compute a similar metric for the IMBs by multiplying the average top melt flux, plus the average basal melt flux, estimated for each month and region by the total ice area for that month and region, obtained from HadISST.2. For the rest of this subsection, the top plus basal melt is shortened to "total melt", despite the fact that, in both models and reality, there is another process causing melt not evaluated here (lateral melting). The total melt flux thus obtained is plotted against annual mean ice thickness (squares in Fig. 7a and c) and against the seasonal cycle amplitude (squares in Fig. 7b and d), with observed values derived from PIOMAS and CryoSat-2 as described in Sect. 2.3. The top melt flux is plotted separately in both figures (stars). In a similar way, we compute the seasonal total congelation growth flux, shortened to "total growth flux" despite not including the frazil growth term, and scatter these against the same variables (triangles).

When plotted against annual mean ice thickness (Fig. 7a, b), a rough inverse relationship is seen: models with higher thickness tend to see less growth/melt and vice versa. This is expected due to the ice thickness–growth and ice–albedo feedbacks, discussed in more detail in the following. The IMB values lie on the lower boundary of the scatter (in terms of magnitude) for both total melt and growth, although not for the top melting alone.

The relationship between ice thickness and top melt alone is weaker, partly because of the effect of the GSI8.1 models which have not only among the highest ice thicknesses but also proportionately greater top melting. Correlation between ice thickness and top melt is stronger within model groups; for example, in the North Pole region, within the GSI8.1 and mushy-layer clusters, the correlation is -0.69and -0.99, respectively, compared to -0.51 across the ensemble as a whole. The picture in the Beaufort Sea region is similar.

One effect of weighting monthly fluxes by ice area is that the "high flux" models of CMCC and CNRM-CERFACS become less extreme relative to the other models and to the IMB values. This is because the highest fluxes in these models occur in months when the ice area is relatively low and, hence, do not represent an exceptionally large ice volume loss or gain value (as a high flux is spread over a relatively small area). Nevertheless, these models still report the highest total melt and growth fluxes.

When plotted against the amplitude of the ice thickness seasonal cycle, a positive correlation is seen. The correlation is not perfect for several reasons: missing processes not evaluated here cause additional ice growth or melt (such as lateral melting and frazil ice growth), spatial correlation of the ice concentration with melt and growth fluxes causes discrepancies when averaging, and any ice growth and melt terms which occur in the same month are effectively invisible to the ice thickness seasonal cycle. This last effect is extreme for the CNRM-CERFACS models, as discussed further in Sect. 4.4. In the Beaufort Sea region, the observed relationship between ice growth and melt from the IMB measurements and ice thickness amplitude from PIOMAS/CryoSat-2 is notably inconsistent with the models. This is likely another symptom of the ice thickness sampling bias.

4.4 Model group discussion: the influence of thermodynamic parameterisation and mean climate state

We briefly discuss the simulations of each model group in turn, linking the vertical energy fluxes to the Arctic climate variables evaluated in Sect. 3. The CMCC models are distinguished by very high melt, growth and conduction fluxes; however, in each case, the seasonal cycles are of a similar shape and phase to the IMB data. The high fluxes arise naturally in response to the unusually low annual mean ice thickness of the CMCC models. Thinner sea ice supports a steeper temperature gradient between the (relatively) warm ocean and cold atmosphere in winter and, hence, also grows more quickly in winter. On a large scale, thinner ice also melts more quickly in summer due to the ice–albedo feedback. Hence, the large biases in the CMCC fluxes reflect the sea ice state bias, to which the sea ice thermodynamics are responding in a physically realistic way.

The CNRM-CERFACS models also have amplified maxima and minima in melt, growth and conduction variables, but many also display large phase offsets. Of particular note are the top melt and basal growth fluxes which both attain roughly equal and opposite maxima in autumn. The result is that, while they have a rather damped ice thickness seasonal cycle (Fig. 2b), the CNRM-CERFACS models display among the highest total ice growth and melt (Fig. 7), as much of this happens at the same time of year, rather than being mostly compartmentalised into different seasons. Unlike for the CMCC models, the unusual behaviour of the CNRM-CERFACS fluxes is likely reflective of either a diagnostic error or a problem with the underlying GELATO ice thermodynamics but is not yet fully understood (David Salas and Rym Msadek, personal communication, 2024).

The mushy-layer models (NCC and NCAR) are distinguished most by exceptionally low basal conductive fluxes during the freezing season, partly because of additional unreported energy flux terms in the basal energy balance. These models are also characterised by relatively low melt and growth fluxes, more similar to the IMB data than some other models, particularly in the case of the top melting. The lower growth flux may be partly explained by more of the ice



Figure 7. Seasonally averaged melt and growth fluxes (weighted by ice area for each month and region) plotted against the annual mean ice thickness (**a**, **b**) and ice thickness seasonal cycle amplitude (**c**, **d**) for the North Pole (**a**, **c**) and Beaufort Sea (**b**, **d**) regions. Shown are the total melt (squares), top melt (stars) and total growth (triangles), as defined in the text. Shaded regions represent observational uncertainty: for melt and growth fluxes, IMB-derived uncertainty (as diagnosed in Sect. 3.3 of West et al. 2020); for ice thickness, uncertainty derived from the differences between PIOMAS and CryoSat-2 (as described in Sect. 2.3).

growth being contained in the frazil ice growth term (which cannot be evaluated by IMBs), as described in Sect. 4.2 above. However, the amplitude of the ice thickness seasonal cycle in Fig. 2c is also lower than for other models at similar ice thicknesses, suggesting that at least part of this is a genuine model difference. As a model group, the damped seasonal cycles are not associated with unusually high annual mean ice thickness which would otherwise be an obvious cause (the converse of CMCC). In fact, ice thickness varies

greatly across these models, correlated with the net radiative flux and global and Arctic temperature.

The other distinguishing feature of these models noted in Sect. 3 is their relatively low annual mean net SW flux and correspondingly high net LW flux. While these could be partly related to cloud properties, they are also consistent with the relatively damped seasonal cycle of ice growth and melt in these models. It is plausible that the mushy-layer scheme, through its action on the low basal conductive flux, may be playing a role in at least the reduced sea ice growth of these models. In Fig. 4d, it is seen that these models display a much higher net downwards LW flux than others, indicating that these models simulate lower surface temperatures for a given downwelling LW flux. Lower conduction of heat through sea ice is a possible mechanism responsible for this.

The GSI8.1 models (CSIRO-ARCCSS, MOHC and NIMS-KMA) have among the highest annual mean ice thicknesses, the smallest LW fluxes, and the coldest global and Arctic temperatures. They display somewhat higher growth and top melt fluxes than the mushy-layer models. They are distinguished particularly by high summer top melt and top conduction fluxes, correspondingly strong sensible heat gain, and notably low ocean heat fluxes, likely associated with the lack of penetrating solar radiation. Similarly to the CMCC models, many of these may be responding in a largely physically realistic manner to a cold climate state, but the seasonal cycle of growth and melt may still be too amplified.

The remaining models (from IPSL and MRI) share no common model components or code but are strikingly similar in their simulation of basal melt and top conduction (as well as in annual mean ice thickness), both being amplified relative to the IMB data. MRI-ESM2-0, however, is colder both globally and in the Arctic. It simulates less top melting in summer and basal conduction in winter and has a less amplified seasonal cycle, all realistic effects of a colder climate. It also simulates less net SW in summer and more net LW in winter.

In summary, the fluxes of various model groups largely respond in a physically realistic manner to the differences between the base model climates and to the differences between the model thermodynamic choices, with the exception of the autumn maxima of the CNRM-CERFACS models. As a group, however, fluxes tend to be higher as a function of ice thickness than is estimated by the IMBs. This discrepancy is now evaluated in more detail.

5 Accounting for the IMB sampling bias

As discussed in Sect. 4, thin ice is likely under-sampled by the IMBs, and this may introduce a bias to evaluations of fluxes that systematically vary with ice thickness (notably basal growth and the conduction fluxes). To account for this, we evaluate fluxes at each model grid point as a function of ice thickness, applying the analysis first to the top conductive flux in the Beaufort Sea region, for the months of December, January and February. For each model, we select 100 points at random and produce scatter plots of top conductive flux against ice thickness, comparing these to monthly mean top conductive flux and ice thickness from the IMBs (Fig. 8a). The same comparison is shown as a set of histograms, sampling the full sets of points for each model (Fig. 8b). The sampling bias of the IMBs is demonstrated, with ice thinner than 1 m barely sampled – although ice thicker than 3 m is also sampled very little. The thin (thick) ice tendencies

of the CMCC (GSI8.1) model groups are also demonstrated. Modelled fluxes are shown to be biased towards higher magnitudes, even for the same range of ice thicknesses. For the 1–2 and 2–3 m histogram categories, each of which contains large numbers of IMB points, all model distributions are significantly different from the IMB values. At first sight, this suggests that the model bias towards high conductive fluxes is real. However, two other effects must be considered.

Firstly, ice thickness is not the only variable directly affecting conductive flux; snow depth also has a strong effect, as snow is a very powerful insulator. To account for snow depth and ice thickness simultaneously, we define the thermal insulance of the snow-ice column as follows:

$$R_{\rm ice} = \frac{h_{\rm I}}{k_{\rm I}} + \frac{h_{\rm S}}{k_{\rm S}},\tag{1}$$

where h_{I} , k_{I} , h_{S} and k_{S} represent ice thickness, ice conductivity, snow depth and snow conductivity, respectively.

Secondly, the model points plotted in Fig. 8 do not represent single floes (like the IMBs). Instead, they are averages over grid cells typically tens of kilometres in width, all of which include ice of multiple thicknesses due to the models' sub-grid-scale ice thickness distributions. This also biases the comparison, as a model grid cell with a 2 m average ice thickness will tend to allow more conduction than a single ice column with a 2 m thickness. This is due to the nonlinear relationship between conduction and ice thickness, which is most sensitive for the thinnest ice. A grid cell with a mean thickness of 2 m will likely contain, in its various separate categories, both thinner ice, which will allow much greater conduction, and thicker ice, whose conduction, while lower than that of a 2 m floe, is closer to this in magnitude than that of the thinner ice. This issue is illustrated by the long negative tails on the right of Fig. 8b; even grid cells with a mean ice thickness of 4-5 m can produce substantial conductive fluxes, if they contain sufficient thin ice in their thickness distribution.

It is impossible to exactly account for this effect without knowledge of conduction flux per ice category, which is not provided by any model in the CMIP6 subset. However, the effect can be roughly estimated using category ice concentration, snow depth and ice thickness data, which are provided by six models: MOHC UKESM1-0-LL, MRI-ESM2-0, IPSL-CM6A-LR, IPSL-CM6A-LR-INCA, NorESM2-LM and NorESM2-MM. For each grid cell in the comparison, the conductance of the ice in each category is estimated from the ice thickness and snow depth. The individual category conductances are then summed together, weighted by category ice concentration, to obtain the total grid cell conductance. An "effective" thermal insulance is then obtained from this, which represents the ability of the ice in the grid cell to conduct energy in a manner comparable to the IMB point measurements.

For the six models, top conductive flux is plotted both against approximate thermal insulance (derived from grid



Figure 8. (a) Scatter plot of top conductive flux against ice thickness for models and IMB data; for clarity, a random set of 100 points is selected for each model. (b) The same comparison, shown as a series of box plots, but with all model points sampled. Note that the 0-1 and 3-4 m classes contain only two and three IMB points, respectively.

box mean ice thickness and snow depth; Fig. 9a) and against corrected thermal insulance (derived from ice thickness distribution information as described above; Fig. 9b). Using the corrected thermal insulance has the effect of pushing the model points to the left, by reducing the insulance. Hence, each insulance class is composed of points with thicker ice (and therefore weaker conduction) than is the case with the uncorrected insulance. This reduces the model bias further but does not eliminate it: for all ranges of thermal insulance for which substantial numbers of IMB points exist, all six models are still significantly negatively biased relative to the IMB values. For the 1–1.5 K m² W⁻¹ insulance range, the IMBs measure an average top conductive flux of -18.8 ± 4.3 W m⁻², whereas the six models simulate an average conductive flux of -29.5 ± 5.4 W m⁻².

The analysis is repeated for the North Pole region (not shown). Although top conductive flux is biased less strongly with respect to the IMB values in this region, mapping the distributions to thermal insulance classes again has the effect of reducing but not eliminating the bias. In the $1.5-2 \text{ Km}^2 \text{ W}^{-1}$ range, for example, the IMBs measure an average top conductive flux of $-13.0 \pm 3.0 \text{ Wm}^{-2}$, whereas the six models simulate an average of $-22.5 \pm 4.2 \text{ Wm}^{-2}$. All models are significantly biased low with respect to the IMB values in this class, with NorESM2-MM and the IPSL models also significantly biased low in the $1.0-1.5 \text{ Km}^2 \text{ W}^{-1}$ class.

The top conductive flux expresses the thermal forcing of the atmosphere on the ice, but it is the basal conductive flux that modulates how this forcing drives ice growth. The basal conductive flux is less strongly negatively biased than the



Figure 9. (a) Top conductive flux box plots as a function of thermal insulance, calculated from model grid box mean ice thickness and snow depth. **(b)** Top conductive flux box plots as a function of thermal insulance calculated from sub-grid ice thickness distribution properties. For each model and insulance class, distributions significantly different from the IMB values are highlighted in bold.

top conductive flux, whereas it is positively biased for the mushy-layer models. The IPSL and MOHC models display small negative biases as a function of grid box mean thermal insulance that are mostly removed when they are evaluated as a function of corrected insulance. MRI-ESM2-0 displays a larger bias that is not removed. The NCC models remain strongly positively (less conduction) biased for each insulance class.

The above analysis is for December–February; we also compare conductive fluxes for October–November, when ice growth is typically strongest, as ice is thinner and insulance is lower. As for the winter months, the conductive flux bias is stronger at the top than at the base of the ice, and it is greater in the Beaufort Sea than in the North Pole region. However, a larger number of IMB measurements are available for the smallest insulance classes during autumn. In the 0.5–1 K m² W⁻¹ insulance class, both top and basal conductive fluxes are significantly biased low in both regions for all models, with the exception of NCC basal conductive fluxes which remain biased high.

We consider also the relationship between ice thickness and melting fluxes. As seen in Fig. 7, the relationship between seasonal average top melting and ice thickness was rather weak. To illustrate this further, we compare July top melting fluxes against ice thickness by grid point for the Beaufort Sea region (not shown). This shows the cross-model relationship between ice thickness and top melt to be much weaker than the winter relationship between ice thickness and conductive fluxes, with wide distributions of top melting fluxes simulated at all thicknesses, albeit with more extremely high values at low thicknesses. Inter-model variability dominates variability due to ice thickness, with, for example, the CMCC and CNRM-CERFACS models displaying two nearly distinct clusters at similar ice thicknesses. This could be partly because the relationship between top melting and ice thickness operates on somewhat larger scales than individual grid cells, via ice area and the ice–albedo feedback.

The IMB data also do not display a clear relationship between ice thickness and top melt flux. In fact, there is a notable lack of high top melt fluxes for low ice thicknesses in the IMB measurements. This may point again to the sampling problem inherent in the Lagrangian nature of the IMB data: thin ice floes subject to high melting fluxes will not survive long enough to report a flux to the IMB dataset. To address this problem, we compute daily top melting fluxes for the IMB dataset and compare these to the monthly mean modelled fluxes (Fig. 10). While model grid box mean fluxes account for the fast melting rates of thin ice irrespective of meaning period, it is necessary to refine the IMB fluxes to a finer temporal resolution to detect these.

The 0–1 m range is still very poorly sampled by the daily IMB fluxes; examination shows that this is because it is rare for IMBs to report data right up until the point of melting out. Moreover, on the occasions that this actually happens, the final destruction of the floe is due to basal rather than top melting, which provides additional evidence for the relationship between ice thickness and top melting being rather weak. However, the 1–2 m category provides plenty of data points. With the exception of the CNRM-CERFACS models, which do not report maximum top melt until September, all models have monthly grid box mean fluxes much higher than the daily fluxes shown by the IMB values in this category.

6 Conclusions

We have evaluated ice energy fluxes for a range of CMIP6 models using a dataset of equivalent fluxes derived from IMB measurements. In most cases, the fluxes vary between the models in ways consistent with the diverse model climates and the underlying ice thermodynamics schemes. Collectively, however, the models tend to simulate greater melt, growth and conduction fluxes than are measured by the IMBs for similar ice thicknesses.

In more detail, the model states in the evaluated period from 1985 to 2014 vary from warm with thin sea ice (the CMCC models) to cold with thick sea ice (particularly some of the GSI8.1 models). Models with thicker sea ice tend to simulate smaller melt, growth and conduction fluxes (and vice versa), which is consistent with ice thermodynamics theory. Model clusters are also influenced by choices in the ice thermodynamics parameterisation. For example, the GSI8.1 models tend to model higher top melting relative to ice thickness than many others, and the ice thickness seasonal cycles are relatively more amplified and further from the IMB values. Conversely, the mushy-layer models tend to simulate lower basal conductive fluxes than others, and their seasonal cycles are relatively more damped and closer to the IMB values. The flux simulation differences in these model clusters are likely to be linked to the lack of penetrating solar radiation and to the inclusion of mushy-layer thermodynamics, respectively. Hence, while the atmosphere exerts a firstorder control on the sea ice state, the sea ice parameterisation choice has a second-order effect on how growth and melt respond to this.

A large part of the discrepancy between IMB data and models is influenced by the ice thickness sampling bias (the tendency for IMBs to be placed in thicker, more robust ice floes) and by the Lagrangian–Eulerian sampling bias (the tendency for thin ice floes to measure faster growth and melt fluxes). Despite this, the ice thickness distribution analysis provides strong evidence that a portion of the conduction/growth bias is real. Additional circumstantial evidence has been provided that a portion of the melting bias is also real, by comparing modelled fluxes to IMB-measured daily fluxes.

We note that, as a purely thermodynamic analysis, this study does not attempt to give any account of the influence of differing ice dynamics on modelled ice volume, either via dynamic atmospheric forcing or ice dynamic scheme choices. While this factor is likely of secondary importance to the atmospheric thermal forcing, its influence on modelled ice volume may be similar to that of ice thermodynamic choices. For example, Keen et al. (2021) found that ice advection accounted for between 9 % and 30 % of the total annual sea ice loss in CMIP6 models.

The evaluation underlines the value of the IMB observations in aiding detailed analysis of sea ice process modelling. Future deployments of these instruments would be very useful to study changes in ice thermodynamics as sea ice continues to become thinner and less extensive.

The evaluation also suggests that there is value in continuing to improve the accuracy of sea ice thermodynamics simulation. Differences in ice thermodynamics choices are shown to have a clear, measurable effect on the ice melt, growth and conduction fluxes that are simulated for particular ice thicknesses that is almost certainly reflected in how the sea ice state responds to a given atmospheric forcing. This is particularly the case for the penetrating solar radiation inclusion and for the mushy-layer parameterisation, although the latter may suppress conduction too much owing to the issues identified by Plante et al. (2024).

The study also proves the importance of providing detailed energy budget information to the CMIP6 archive. Al-



Figure 10. (a) Scatter plot of monthly mean top melting fluxes compared to daily mean IMB fluxes, for a randomly selected subset of the Beaufort Sea region. (b) Bar plot for the same comparison, using the full model distributions.

though the CMIP6 subset contained reasonable model diversity, many additional models were of a complexity sufficient to justify submitting the energy flux diagnostics to the CMIP6 archive, and it would be of great value if a wider variety of models provided these for future CMIP iterations.

Code availability. The code used to analyse the IMB data is published in two repositories, corresponding to two stages of the analysis. The code used to read, quality control and process the data into consistent quantities on consistent time points can be downloaded from https://doi.org/10.5281/zenodo.3975692 (West, 2020a). The code used to produce datasets of monthly mean energy fluxes from these processed data and that used to produce the daily data used in Fig. 10 can be downloaded from https://doi.org/10.5281/zenodo.3971736 (West, 2020b).

The code used to analyse model data and calculate time series, multiannual means and ice-area-weighted statistics as well as the code used to produce Figs. 1–10 is published at https://doi.org/10.5281/zenodo.12518762 (West, 2024).

Data availability. The raw IMB data are publicly available and can be downloaded from http://imb-crrel-dartmouth.org/ results/ (Perovich et al., 2020). The processed IMB data and the derived dataset of monthly mean fluxes are published at https://doi.org/10.5281/zenodo.3773811 (West, 2020c) and https://doi.org/10.5281/zenodo.3773997 (West, 2020d), respectively. The derived dataset of daily mean fluxes used in Fig. 10 is also published at https://doi.org/10.5281/zenodo.3773997 (West, 2020d).

The diagnostics from CMIP6 used in this study are all publicly available and can be downloaded from the CMIP6

archive at https://esgf-index1.ceda.ac.uk/projects/cmip6-ceda/ (last access: 16 January 2024): CMCC-CM2-SR5 (https: //doi.org/10.22033/ESGF/CMIP6.1362, Lovato and Peano, 2020), CMCC-ESM2 (https://doi.org/10.22033/ESGF/CMIP6.13164, Lovato et al., 2021), CNRM-CM6-1 (https://doi.org/10.22033/ ESGF/CMIP6.1375, Voldoire, 2018), CNRM-CM6-1-HR (https://doi.org/10.22033/ESGF/CMIP6.1387, Voldoire, 2019), CNRM-ESM2-1 (https://doi.org/10.22033/ESGF/CMIP6.9564, Seferian, 2019), CSIRO-ARCCSS ACCESS-CM2 (https: //doi.org/10.22033/ESGF/CMIP6.2281, Dix et al., 2019), IPSL IPSL-CM6A-LR (https://doi.org/10.22033/ESGF/CMIP6.1534, Boucher et al., 2018), IPSL IPSL-CM6A-LR-INCA (https://doi.org/10.22033/ESGF/CMIP6.13582, Boucher al., 2021), MOHC HadGEM3-GC31-LL (https://doi.org/10.22033/ ESGF/CMIP6.419, Ridley et al., 2018b), MOHC HadGEM3-GC31-MM (https://doi.org/10.22033/ESGF/CMIP6.420, Ridley et al., 2019), MOHC UKESM1-0-LL (https: //doi.org/10.22033/ESGF/CMIP6.1569, Tang et al., 2019), MRI (https://doi.org/10.22033/ESGF/CMIP6.621, MRI-ESM2-0 Yukimoto et al., 2019), NCAR CESM2-WACCM (https: //doi.org/10.22033/ESGF/CMIP6.10024, Danabasoglu, 2019a), NCAR CESM2 (https://doi.org/10.22033/ESGF/CMIP6.2185, Danabasoglu, 2019b), NCC NorESM2-LM (https://doi.org/ 10.22033/ESGF/CMIP6.502, Seland et al., 2019), NCC NorESM2-MM (https://doi.org/10.22033/ESGF/CMIP6.506, Bentsen et al., 2019), and NIMS-KMA UKESM1-0-LL (https://doi.org/10.22033/ESGF/CMIP6.2245, Shim et al.. 2020)..

Data from the ERA5 reanalysis are available at https://doi.org/10.24381/cds.bd0915c6 (Hersbach et al., 2023). NOAAGlobalTemp data are available at https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ncdc%3AC01704/html (Huang et al., 2024). CryoSat-2 data are available at https://doi.org/10.5067/96JO0KIFDAS8 (Kurtz and Harbeck, 2017).

Author contributions. The analysis was devised, undertaken and documented by AEW. EWB gave feedback at each stage of the analysis and on the final paper drafts.

Competing interests. The contact author has declared that neither 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. This study was supported by the Met Office Hadley Centre Climate Programme, funded by the Department for Science, Innovation and Technology. The authors thank Emma Fiedler and Jeff Ridley for helpful suggestions concerning reference datasets. The authors also thank the reviewers, John Toole and Mathieu Plante, for many constructive suggestions which improved the clarity and content of this study.

Financial support. This research has been supported by the Met Office Hadley Centre Climate Programme, funded by the Department for Science, Innovation and Technology.

Review statement. This paper was edited by Christopher Horvat and reviewed by John Toole and Mathieu Plante.

References

- Batrak, Y. and Müller, M.: On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice, Nat. Commun., 10, 4170, https://doi.org/10.1038/s41467-019-11975-3, 2019.
- Bentsen, M., Oliviè, D. J. L., Seland, Ø., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O. A., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: NCC NorESM2-MM model output prepared for CMIP6 CMIP. Version 20191108, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.506, 2019.
- Bitz, C. M. and Lipscomb, W. H.: An energy-conserving thermodynamic model of sea ice, J. Geophys. Res., 104, 15669–15677, https://doi.org/10.1029/1999JC900100, 1999.
- Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M., Meurdesoif, Y., Cadule, P., Devilliers, M., Ghattas, J., Lebas, N., Lurton, T., Mellul, L., Musat, I., Mignot, J., and Cheruy, F.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP. Version 20180803, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.1534, 2018.
- 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., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J., Guenet, B., Guez, L. E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J., 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.
- Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M., Meurdesoif, Y., Balkanski, Y., Checa-Garcia, R.,

Hauglustaine, D., Bekki, S., and Marchand, M.: IPSL IPSL-CM6A-LR-INCA model output prepared for CMIP6 CMIP. Version 20210216, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.13582, 2021.

- Cai, Q., Wang, J., Beletsky, D., Overland, J., Ikeda, M., and Wan, L.: Accelerated decline of summer Arctic sea ice during 1850–2017 and the amplified Arctic warming during the recent decades, Environ. Res. Lett., 16, 034015, https://doi.org/10.1088/1748-9326/abdb5f, 2021.
- Chen, L., Wu, R., Shu, Q., Min, C., Yang, Q., and Han, B.: The Arctic Sea Ice Thickness Change in CMIP6's Historical Simulations, Adv. Atmos. Sci., 40, 2331–2343, https://doi.org/10.1007/s00376-022-1460-4, 2023.
- Danabasoglu, G.: NCAR CESM2-WACCM model output prepared for CMIP6 CMIP. Version 20190917 (sidmassmelttop, sidmassmeltbot and sidmassgrowthbot); Version 20190227 (all other diagnostics), Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.10024, 2019a.
- Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 CMIP. Version 20190919 (sidmassmelttop, sidmassmeltbot and sidmassgrowthbot); Version 20190308 (all other diagnostics), Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.2185, 2019b.
- Danabasoglu, G., Lamarque, J., Bacmeister, J., Bailey, D. A., Du-Vivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), J. Adv. Model. Earth Sy., 12, e2019MS001916, https://doi.org/10.1029/2019MS001916, 2020.
- Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., Mackallah, C., Marsland, S., O'Farrell, S., Rashid, H., Srbinovsky, J., Sullivan, A., Trenham, C., Vohralik, P., Watterson, I., Williams, G., Woodhouse, M., Bodman, R., Dias, F. B., Domingues, C. M., Hannah, N., Heerdegen, A., Savita, A., Wales, S., Allen, C., Druken, K., Evans, B., Richards, C., Ridzwan, S. M., Roberts, D., Smillie, J., Snow, K., Ward, M., and Yang, R.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP. Version 20200817, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.2281, 2019.
- Flocco, D., Schroeder, D., Feltham, D. L., and Hunke, E. C.: Impact of melt ponds on Arctic sea ice simulations from 1990 to 2007, J. Geophys. Res.-Oceans, 117, C09032, https://doi.org/10.1029/2012JC008195, 2012.
- GISTemp Team: GISS Surface Temperature Analysis (GISTEMP), version 4, NASA Goddard Institute for Space Studies, https://data.giss.nasa.gov/gistemp/ (last access: 26 February 2024), 2024.
- Henke, M., Cassalho, F., Miesse, T., Ferreira, C. M., Zhang J., and Ravens, T. M.: Assessment of Arctic sea ice and surface climate conditions in nine CMIP6 cli-

mate models, Arct. Antarct. Alp. Res., 55, 2271592, https://doi.org/10.1080/15230430.2023.2271592, 2023.

- 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 hourly data on pressure levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.bd0915c6, 2023.
- Holland, M. M., Bitz, C. M., Hunke, E. C., Lipscomb, W. H., and Schramm, J. L.: Influence of the Sea Ice Thickness Distribution on Polar Climate in CCSM3, J. Climate, 19, 2398–2414, https://doi.org/10.1175/JCLI3751.1, 2006.
- Holland, M. M., Bailey, D. A., Briegleb, B. P., Light, B., and Hunke, E.: Improved Sea Ice Shortwave Radiation Physics in CCSM4: The Impact of Melt Ponds and Aerosols on Arctic Sea Ice, J. Climate, 25, 1413–1430, https://doi.org/10.1175/JCLI-D-11-00078.1, 2012.
- Huang, B., Yin, X., Menne, M. J., Vose, R., and Zhang, H.: NOAA Global Surface Temperature Dataset (NOAAGlobalTemp), Version 6.0.0 (gridded), NOAA National Centers for Environmental Information [data set], https://www.ncei.noaa.gov/metadata/ geoportal/rest/metadata/item/gov.noaa.ncdc%3AC01704/html, last access: 26 February 2024.
- Hunke, E. C., Hebert, D. A., and Lecomte, O.: Level-ice melt ponds in the Los Alamos sea ice model, CICE, Ocean Model., 71, 26– 42, https://doi.org/10.1016/j.ocemod.2012.11.008, 2013.
- Keen, A. and Blockley, E.: Investigating future changes in the volume budget of the Arctic sea ice in a coupled climate model, The Cryosphere, 12, 2855–2868, https://doi.org/10.5194/tc-12-2855-2018, 2018.
- Keen, A., Blockley, E., Bailey, D. A., Boldingh Debernard, J., Bushuk, M., Delhaye, S., Docquier, D., Feltham, D., Massonnet, F., O'Farrell, S., Ponsoni, L., Rodriguez, J. M., Schroeder, D., Swart, N., Toyoda, T., Tsujino, H., Vancoppenolle, M., and Wyser, K.: An inter-comparison of the mass budget of the Arctic sea ice in CMIP6 models, The Cryosphere, 15, 951–982, https://doi.org/10.5194/tc-15-951-2021, 2021.
- Koo, Y., Lei, R., Cheng, Y., Cheng, B., Xie, H., Hoppmann, M., Kurtz, N. T., Ackley, S. F., and Mestas-Nunez, A. M.: Estimation of thermodynamic and dynamic contributions to sea ice growth in the Central Arctic using ICESat-2 and MO-SAiC SIMBA buoy data, Remote Sens. Environ., 267, 112730, https://doi.org/10.1016/j.rse.2021.112730, 2021.
- Kurtz, N. and Harbeck, J.: CryoSat-2 Level-4 Sea Ice Elevation, Freeboard, and Thickness, Version 1, NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA [data set], https://doi.org/10.5067/96JO0KIFDAS8, 2017.
- Kwok, R.: Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018), Environ. Res. Lett., 13, 105005, https://doi.org/10.1088/1748-9326/aae3ec, 2018.
- Lei, R., Li, N., Heil, P., Cheng, B., Zhang, Z., and Sun, B.: Multiyear sea-ice thermal regimes and oceanic heat flux derived from an ice mass balance buoy in the Arctic Ocean, J. Geophys. Res.-Oceans, 119, 537–547, https://doi.org/10.1002/2012JC008731, 2014.
- Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP un-

certainty model, J. Geophys. Res.-Atmos., 124, 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.

- Loeb, N. G., Wielicki, B. A., Doelling, D. R., Louis Smith, G., Keyes, D. F., Kato, S., Manalo-Smith, N., and Wong, T.: Toward Optimal Closure of the Earth's TopofAtmosphere Radiation Budget, J. Climate, 22, 748–766, https://doi.org/10.1175/2008JCLI2637.1, 2009.
- Long, M., Zhang, L., Hu, S., and Qian, S.: Multi-Aspect Assessment of CMIP6 Models for Arctic Sea Ice Simulation, J. Climate, 34, 1515–1529, https://doi.org/10.1175/JCLI-D-20-0522.1, 2021.
- Lovato, T. and Peano, D.: CMCC CMCC-CM2-SR5 model output prepared for CMIP6 CMIP. Version 20200616, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.1362, 2020.
- Lovato, T., Peano, D., and Butenshön, M.: CMCC CMCC-ESM2 model output prepared for CMIP6 CMIP. Version 20210114, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.13164, 2021.
- Massonnet, F., Vancoppenolle, M., Goosse, H., Docquier, D., Fichefet, T., and Blanchard-Wrigglesworth, E.: Arctic sea-ice change tied to its mean state through thermodynamic processes, Nat. Clim. Change, 8, 599–603, https://doi.org/10.1038/s41558-018-0204-z, 2018.
- Massonnet, F., Barthélemy, A., Worou, K., Fichefet, T., Vancoppenolle, M., Rousset, C., and Moreno-Chamarro, E.: On the discretization of the ice thickness distribution in the NEMO3.6-LIM3 global ocean-sea ice model, Geosci. Model Dev., 12, 3745–3758, https://doi.org/10.5194/gmd-12-3745-2019, 2019.
- Maykut, G. A. and McPhee, M. G., Solar heating of the Arctic mixed layer, J. Geophys. Res., 100, 24691–24703, https://doi.org/10.1029/95JC02554, 1995.
- McPhee, M. G.: Turbulent heat flux in the upper ocean under sea ice, J. Geophys. Res., 97, 5365–5379, https://doi.org/10.1029/92JC00239, 1992.
- Mellor, G. L. and Kantha, L., An ice-ocean coupled model, J. Geophys. Res., 94, 10937–10954, 1989.
- Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R., An updated assessment of near-surface temperature change from 1850: the Had-CRUT5 data set, J. Geophys. Res.-Atmos., 126, e2019JD032361, https://doi.org/10.1029/2019JD032361, 2021.
- Notz, D. and SIMIP Community: Arctic sea ice in CMIP6, Geophys. Res. Lett., 47, e2019GL086749, https://doi.org/10.1029/2019GL086749, 2020.
- Notz, D., Jahn, A., Holland, M., Hunke, E., Massonnet, F., Stroeve, J., Tremblay, B., and Vancoppenolle, M.: The CMIP6 Sea-Ice Model Intercomparison Project (SIMIP): understanding sea ice through climate-model simulations, Geosci. Model Dev., 9, 3427–3446, https://doi.org/10.5194/gmd-9-3427-2016, 2016.
- Olonscheck, D., Mauritsen, T., and Notz, D.: Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations, Nat. Geosci., 12, 430–434, https://doi.org/10.1038/s41561-019-0363-1, 2019.
- Perovich, D., Richter-Menge, J., and Polashenski, C.: Observing and understanding climate change: Monitoring the mass balance, motion, and thickness of Arctic sea ice, The CRREL-Dartmouth Mass Balance Buoy Program, Cold Regions Research and Engineering Laboratory, Thayer School of Engineering at Dart-

mouth [data set], http://imb-crrel-dartmouth.org/results/, last access: 20 April 2020.

- Plante, M., Lemieux, J.-F., Tremblay, L. B., Tivy, A., Angnatok, J., Roy, F., Smith, G., Dupont, F., and Turner, A. K.: Using Icepack to reproduce ice mass balance buoy observations in landfast ice: improvements from the mushy-layer thermodynamics, The Cryosphere, 18, 1685–1708, https://doi.org/10.5194/tc-18-1685-2024, 2024.
- Richter-Menge, J. A., Perovich, D. K., Elder, B. C., Claffey, K., Rigor, I., and Ortmeyer, M.: Ice mass balance buoys: A tool for measuring and attributing changes in the thickness of the Arctic sea ice cover, Ann. Glaciol., 44, 205–210, https://doi.org/10.3189/172756406781811727, 2006.
- Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., and Schroeder, D.: The sea ice model component of HadGEM3-GC3.1, Geosci. Model Dev., 11, 713–723, https://doi.org/10.5194/gmd-11-713-2018, 2018a.
- Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP. Version 20200330, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.419, 2018b.
- Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 CMIP. Version 20191207, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.420, 2019.
- Rohde, R. A. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, Earth Syst. Sci. Data, 12, 3469–3479, https://doi.org/10.5194/essd-12-3469-2020, 2020.
- Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, J. Geophys. Res., 116, C00D06, https://doi.org/10.1029/2011JC007084, 2011.
- Seferian, R.: CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 LS3MIP. Version 20181206, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.9564, 2019.
- Seland, Ø., Bentsen, M., Oliviè, D. J. L., Toniazzo, T.,; Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O. A., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: NCC NorESM2-LM model output prepared for CMIP6 CMIP. Version 20191108 (siflcondtop and siflcondbot); Version 20190815 (all other diagnostics), Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.502, 2019.
- Shim, S., Lim, Y., Byun, Y., Seo, J., Kwon, S., and Kim, B.: NIMS-KMA UKESM1.0-LL model output prepared for CMIP6 CMIP. Version 20200611, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.2245, 2020.
- Steele, M., Zhang, J., and Ermold, W.: Mechanisms of summer Arctic Ocean warming, J. Geophys. Res.-Oceans, 115, C11004, https://doi.org/10.1029/2009JC005849, 2010.
- Stroeve, J. C. and Notz, D.: Changing state of Arctic sea ice across all seasons, Environ. Res. Lett., 13, 103001, https://doi.org/10.1088/1748-9326/aade56, 2018.
- Stroeve, J. C., Serreze, M. C., Holland, M. M., Kay, J. E., Maslanik, J., and Barrett, A. P.: The Arctic's rapidly shrinking sea ice

cover: a research synthesis, Climatic Change, 110, 1005–1027, https://doi.org/10.1007/s10584-011-0101-1, 2012.

- Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., Walton, J., and Jones, C.: MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP. Version 20200309, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.1569, 2019.
- Tilling, R. L., Ridout, A., and Shepherd, A.: Estimating Arctic sea ice thickness and volume using CryoSat-2 radar altimeter data, Adv. Space Res., 62, 1203–1225, https://doi.org/10.1016/j.asr.2017.10.051, 2018.
- Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, J. Geophys. Res.-Atmos., 119, 2864–2889, https://doi.org/10.1002/2013JD020316, 2014.
- Turner, A. K. and Hunke, E. C.: Impacts of a mushy-layer thermodynamic approach in global sea-ice simulations using the CICE sea-ice model, J. Geophys. Res.-Oceans, 120, 1253–1275, https://doi.org/10.1002/2014JC010358, 2015.
- Voldoire, A.: CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 CMIP. Version 20180917, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.1375, 2018.
- Voldoire, A.: CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 HighResMIP. Version 20191021, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.1387, 2019.
- Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J., Michou, M., Moine, M., 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.
- Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geosci. Model Dev., 12, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019, 2019.

- West, A. E.: alex-west-met-office/IMBS_MO: Stage 1 IMB code – production version 1.1, Zenodo [code], https://doi.org/10.5281/zenodo.3975692, 2020a.
- West, A. E.: alex-west-met-office/IMBS_Stage2_modeleval_MO: Stage 2 IMB code – production version, Zenodo [code], https://doi.org/10.5281/zenodo.3971736, 2020b.
- West, A. E.: Arctic ice mass balance buoy data for use in calculating quantities to evaluate climate models, Version v1_april_2020, Zenodo [data set], https://doi.org/10.5281/zenodo.3773811, 2020c.
- West, A. E.: Monthly mean variables derived from Arctic Ice Mass Balance Buoy data, for evaluation of sea ice models, Version v1_april_2020, Zenodo [data set], https://doi.org/10.5281/zenodo.3773997, 2020d.
- West, A. E.: New methods of evaluating sea ice in climate models based on energy budgets, PhD thesis, University of Exeter, http: //hdl.handle.net/10871/127762 (last access: 20 May 2024), 2021.
- West, A. E.: Figure & analysis code for paper "CMIP6 models overestimate melt, growth & conduction relative to ice mass balance buoy estimates", Zenodo [code], https://doi.org/10.5281/zenodo.12518762, 2024.
- West, A., Collins, M., Blockley, E., Ridley, J., and Bodas-Salcedo, A.: Induced surface fluxes: a new framework for attributing Arctic sea ice volume balance biases to specific model errors, The Cryosphere, 13, 2001–2022, https://doi.org/10.5194/tc-13-2001-2019, 2019.
- West, A., Collins, M., and Blockley, E.: Using Arctic ice mass balance buoys for evaluation of modelled ice energy fluxes, Geosci. Model Dev., 13, 4845–4868, https://doi.org/10.5194/gmd-13-4845-2020, 2020.
- West, A., Blockley, E., and Collins, M.: Understanding model spread in sea ice volume by attribution of model differences in seasonal ice growth and melt, The Cryosphere, 16, 4013–4032, https://doi.org/10.5194/tc-16-4013-2022, 2022.
- 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.
- Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y.: MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP. Version 20210311, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.621, 2019.