Sea ice cover in the Copernicus Arctic Regional Reanalysis

Yurii Batrak[1], Bin Cheng[2], and Viivi Kallio-Myers[2]

[1]Development Centre for Weather Forecasting, Norwegian Meteorological Institute, Oslo, Norway
[2]Finnish Meteorological Institute, Helsinki, Finland

Correspondence: Yurii Batrak (yurii.batrak@met.no)

Abstract. The Copernicus Arctic Regional Reanalysis (CARRA) is a novel regional high-resolution atmospheric reanalysis product that covers a considerable part of the European Arctic including substantial amounts of ice-covered areas. Sea ice in CARRA is modelled by means of a one-dimensional thermodynamic sea ice parameterisation scheme, which also explicitly resolves the evolution of the snow layer over sea ice. In the present study we assess the representation of sea ice cover in the

5   CARRA product and validate it against a wide set of satellite products and observations from ice mass balance buoys. We show that sea ice cover in CARRA adequately represents general interannual trends towards thinner and warmer ice in the Arctic. Compared to ERA5, sea ice in CARRA shows a reduced warm bias in the ice surface temperature. The strongest improvement was observed for winter months over the Central Arctic, and the Greenland and Barents seas where a 4.91°C median ice surface temperature error of ERA5 is reduced to 1.88°C in CARRA on average. Over the Baffin Bay, intercomparisons suggest the

10   presence of a cold winter-time ice surface temperature bias in CARRA. No improvement over ERA5 was found in the ice surface albedo with spring-time errors in CARRA being up to 8% higher on average than those in ERA5 when computed against the CLARA-A2 satellite retrieval product. Summer-time ice surface albedos are comparable in CARRA and ERA5. Sea ice thickness and snow depth in CARRA adequately resolve the annual cycle of sea ice cover in the Arctic and bring added value compared to ERA5. However, limitations of CARRA indicate potential benefits of utilising more advanced approaches

15   for representing sea ice cover in next generation reanalyses.

1 Introduction

Many scientific and engineering applications require, or can benefit from, information about the past and present states of the Earth's atmosphere provided by atmospheric reanalysis products (see, e.g., Frank et al., 2020; Chung et al., 2013; Serreze et al.,

20   2003). These products offer a pragmatic solution to the problem of obtaining a consistent multiyear gridded data set of atmospheric and surface variables by utilising a numerical weather prediction (NWP) system. Operational NWP systems applied in routine weather forecasting are under constant development, and as a result, archives of weather forecasts comprise subsets produced with different versions of atmospheric models, each having its own biases and limitations, resulting in inconsistent

data sets. Therefore, to have a consistent gridded data set of past atmospheric states, a series of objective analyses is repeated
25 using the same version of an NWP system, which results in an atmospheric *reanalysis* data set (Bengtsson and Shukla, 1988).

Reanalysis systems are usually based on a short range operational NWP system (using an up-to-date model version at the
time of the start of reanalysis production), which is kept unmodified. In the same way as the underlying NWP systems, atmo-
spheric reanalysis data sets can be split into two categories: global and regional. Global reanalyses such as ERA5 (Hersbach
et al., 2020), NCEP-DOE Reanalysis version 2 (Kanamitsu et al., 2002) or MERRA-2 (Gelaro et al., 2017) provide consistent
30 gridded series of atmospheric and surface variables spanning multiple decades and covering the whole Earth. However, these
reanalysis products usually have relatively coarse spatial resolution, ranging from hundreds of kilometres (in older products) to
a few tens of kilometres in latest generation products (see, e.g., Fujiwara et al., 2017). Regional atmospheric reanalysis systems,
unlike their global counterparts, are based on limited area NWP systems. Thus, they are less computationally expensive and
allow higher spatial resolution and more advanced model formulations. Contemporary regional atmospheric reanalysis systems
35 provide gridded data sets with spatial resolution close to 10 km and below (see, e.g., Kaiser-Weiss et al., 2019).

Ongoing climate change is leading to unprecedented modern-time warming of the Arctic, which is stronger than in any
other region on Earth (Cohen et al., 2014; Serreze and Barry, 2011). Retreating sea ice and growing economic activity result
in increasing scientific attention to the region and demand on accurate and reliable atmospheric data sets. However, the Arctic
region is challenging to accurately model in NWP systems due to several factors. Firstly, the remote location of the Arctic
40 limits the availability of in situ observations that can be used for constraining the models, although this lack of conventional
observations is partly compensated by higher availability of the satellite observations from polar orbiting satellites (Lawrence
et al., 2019). Secondly, operational short range NWP systems that are used as the modelling component in atmospheric reanal-
ysis systems are usually tuned to perform best in the mid-latitude regions. Additionally, accurate representation of the surface
processes in NWP systems is crucial for modelling interactions between the surface and the model atmosphere, however, tradi-
45 tionally applied parameterisation schemes employ various simplifications to reduce their computational costs, which can lead
to modelling errors and biases. One example, more specific to the Arctic region, is the representation of sea ice cover. Sea ice,
still abundant in the present-day Arctic, moderates the heat exchange between the ocean and the atmosphere, meaning that
its accurate representation is vital to proper modelling of the surface energy balance, and, as a consequence, of near-surface
atmospheric variables in the Arctic. Despite that fact, short-range NWP systems, and in turn, reanalysis systems that are based
50 upon them, both global and regional, traditionally apply simplified one-dimensional parameterisation schemes for representing
sea ice. So called fully-coupled short-range NWP systems, which represent the atmosphere and the ocean three-dimensionally,
as well as the dynamics and the evolution of sea ice cover, even though they are in the active development, to our knowledge are
yet to be applied in contemporary atmospheric reanalysis systems (it must be noted, however, that fully coupled atmospheric
reanalyses exist, for example, the Climate Forecast System Reanalysis (CFSR, Saha et al., 2010), albeit not based on a short
55 range NWP system).

Contemporary atmospheric reanalysis systems utilise sea ice parameterisation schemes of various complexity ranging from
those representing ice by means of computing the thermal balance of a thin ice layer (MERRA2) to thermodynamic sea ice
models, often with prescribed ice thickness (ASRv2, ERA5), or the snow layer omitted (ERA5). These schemes are developed

with a focus on representing the surface energy balance of the ice layer since sea ice specific variables such as ice thickness or
60 ice salinity are of secondary interest in an atmospheric reanalysis. However, errors and biases found in the reanalysis products
over the Arctic ocean (Graham et al., 2019; Wang et al., 2019) suggest potential benefits to implementing a more detailed
representation of the evolution of sea ice in the current and next generation reanalysis systems.

In the present study we assess the performance of the Copernicus Arctic Regional Reanalysis (CARRA), a novel regional
atmospheric reanalysis product for Greenland and the European Arctic based on the HAMRONIE-AROME NWP system
65 (Bengtsson et al., 2017), in representing the evolution of sea ice cover. Additionally, we compare the representation of sea ice
cover in CARRA, which employs a considerably more advanced sea ice parameterisation scheme, against the ERA5 reanalysis
product used for obtaining lateral boundary conditions in the CARRA system. We focus on the following sea ice specific
variables: ice surface temperature, ice albedo, ice thickness and snow depth over sea ice. However, sea ice concentration,
which is prescribed over the Arctic ocean in both CARRA and ERA5 from well-established satellite-based products is not
70 validated in the present study. Near surface atmospheric variables over sea ice, such as two metre air temperature or ten metre
wind speed, are not discussed in the present study mainly due to a limited number of available observations of these variables
over sea ice within the area represented in CARRA. When performing comparisons a wide set of remote sensing products is
employed for assessing the performance of the reanalysis on the large scale. Additionally, observations reported from a set of
ice mass balance buoys are used to complement comparisons against the remote sensing products.

75 The paper is organised as follows. Section 2 provides an overview of the studied atmospheric reanalysis products and
underlying modelling systems with special attention to the applied parameterisations of the sea ice cover. Section 3 describes
observational data sets utilised and the analysis methods applied in the present paper. Section 4 evaluates sea ice cover in
CARRA and ERA5 by comparison against observational products. The final section provides a short summary of the obtained
results and discusses their implications as well as opportunities for further improvements in representing sea ice cover in next
80 generation reanalysis systems.

2 Representation of sea ice in CARRA and ERA5

2.1 CARRA

CARRA is a regional atmospheric reanalysis product which covers a sector of Arctic between 56°N and 86°N spanning from
the Baffin Bay in the west to the Kara Sea in the east. The data set covers the time period from 1990 to the present (2023, at
85 the moment of writing this manuscript) with an analysis provided every 3 hours. Additionally, the CARRA data set includes
the output from model integration, which is provided with hourly temporal resolution for the lead times under 6 hours and
with three-hourly resolution for lead times over 6 hours (lead times longer than 3 hours are available only for the forecasts
initialised at 00 and 12 UTC). The CARRA system is based on the HARMONIE-AROME NWP system (Bengtsson et al.,
2017) and is forced by ERA5 on the boundaries. The reanalysis product is provided for two overlapping model domains: a
90 larger western domain centred on Greenland, and a smaller eastern domain covering the European Arctic (see Fig. 1). For both
model domains a Lambert conformal conic grid with a horizontal resolution of 2.5 km is used.

Sea ice in the CARRA system is represented by a one-dimensional thermodynamic sea ice scheme (SICE, Batrak et al., 2018; Batrak and Müller, 2019) which resolves the processes of thermodynamic ice growth and melting. Snow cover on top of the sea ice is explicitly modelled by an adapted version of a multilayer parameterisation scheme originally developed for snow cover over land (Boone, 2000; Boone and Etchevers, 2001). The ice scheme of CARRA does not resolve the processes of snow-ice formation and internal melting of the ice, and sea ice salinity in the scheme is prescribed and constant. Surface albedo of the sea-ice covered grid cells in CARRA is computed by applying simple parameterisation schemes. For snow-free ice cover, a temperature-dependent broadband albedo scheme is applied (defined as HIRHAM in Liu et al., 2007), and when ice is covered by snow an adapted version of the broadband snow albedo scheme by Douville et al. (1995) is used. When computing albedo of cold dry snow covering sea ice in the CARRA system, the albedo scheme of Douville et al. (1995) is modified to increase the value of the lowest possible albedo in the dry albedo degradation term from the original 0.5 to 0.75. Sea ice albedo schemes applied in CARRA do not distinguish between direct and diffuse components of surface albedo. The HARMONIE-AROME NWP system does not produce grid-cell averaged albedo as an output variable, therefore in the CARRA product the surface albedo field is computed from the hourly accumulated downwelling and upwelling shortwave radiation fluxes and available only from the model integration output.

The CARRA system is based on a classic non-coupled regional short-range NWP system that does not include a prognostic ocean model nor a slab mixed layer parameterisation scheme. The sea surface temperature and ice concentration fields in CARRA are prescribed from observational data sets and, as a result, the sea ice scheme can not freeze new ice during the model integration. Therefore, sea ice extent in the CARRA system is updated only at the analysis time by means of updating the ice concentration field and using a simple extrapolation-like procedure for initialisation of the prognostic variables of the sea ice scheme. This new ice is always snow-free and snow cover over sea ice in CARRA is accumulated during the model forecast from the model precipitation. The following satellite sea ice concentration products are used in CARRA: ESA CCI SICCI (Toudal Pedersen et al., 2017), which is applied whenever available, and OSISAF OSI-450 (Tonboe et al., 2016) as a fallback data set (Yang et al., 2020).

Over the ice-covered grid cells the CARRA system does not apply any surface data assimilation or procedure, thus the sea ice model is not constrained by observations (except for prescribing the sea ice concentration from an external data set). At the initial cold start of a reanalysis production stream the system is initialised with snow-free ice cover with a uniform thickness of 0.75 m and the temperature set to the freezing point of the sea water. Then, a one year spin-up period is used for preparing the initial model state in that CARRA production stream.

Coupling between the ice surface and the model atmosphere follows the original implementation of Batrak et al. (2018), however, form drag over sea ice is not taken into account in the CARRA system.

2.2 ERA5

ERA5 (Hersbach et al., 2020) is a fifth-generation global atmospheric reanalysis product developed by the ECMWF, which covers the time period from 1950 to the present (2023, at the moment of writing this manuscript). The ERA5 reanalysis system

125 is based on the IFS-HRES NWP system by ECMWF and provides data on a reduced Gaussian grid with a nominal horizontal resolution of 31 km.

Sea ice cover in the ERA5 system is modelled in a way similar to CARRA by using a one-dimensional sea ice parameterisation scheme, although the sea ice model of ERA5 is considerably simplified compared to that of CARRA. In ERA5, sea ice has a constant and uniform thickness of 1.5 m without explicit prognostic parameterisation of the snow cover. Sea ice

130 concentration in ERA5 is also provided from an external source (several data sets are used throughout the time period covered by the ERA5 product, see Hersbach et al. (2020) for additional details) and not modified by the modelling system during the model integration. Surface albedo of the sea ice cover in the ERA5 system is represented by time-interpolated monthly values of Ebert and Curry (1993). For winter months the dry snow albedo is used to simulate the effects of snow cover in the snow-free parameterisation scheme (ECMWF, 2016).

3 Observational data sets and methods

135

To assess the performance of CARRA in representing Arctic sea ice cover, we validate the model output against a wide set of remote-sensing and observational products as well as in situ observational data sets. In this section we provide a summary of the applied processing methods and an overview of the utilised data sets.

3.1 General design of the validation procedure

140 Where applicable, we use verification scores computed for ERA5 as a baseline against which to show the added value of the new regional reanalysis product. However, ERA5 has much lower spatial resolution than CARRA and in the present study we use ERA5 data interpolated from the native reduced Gaussian grid to a 0.25° regular latitude-longitude grid. Thus, when computing scores based on high-resolution observational data sets (fore example, for the ice surface temperature) for the ERA5 reanalysis we resample the data from the regular 0.25° grid onto the 2.5 km CARRA grid using nearest neighbour interpolation.

145 After applying such a procedure the ERA5 fields on the CARRA model grid still represent the variability of the original ERA5 data set thus potentially degrading some of the verification scores due to oversampling when using observational data sets with high spatial resolution. However, in the present study we do not penalise the regional reanalysis field by discarding extra information available due to a denser model grid. When comparing reanalyses against coarse-resolution products, both CARRA and ERA5 are aggregated on the product grid as a first step before computing scores.

150 The model domains of CARRA include considerable parts of the European and Canadian Arctic, although the characteristics of sea ice cover vary in different parts of the area represented in CARRA. For example, Baffin Bay, which is locked between Baffin Island and Greenland, and connected to the Central Arctic by a few straits, is primarily covered by first-year ice and has a low amount of multiyear ice transported from the Central Arctic (Tang et al., 2004; Dunbar, 1973). In contrast, sea ice cover of the Greenland Sea includes a considerable amount of old multiyear ice exported through the Fram Strait by the East

155 Greenland Current (Aagaard and Coachman, 1968; Schmith and Hansen, 2003). The Barents Sea, unlike Baffin Bay, is not locked between land masses and is better connected with the central Arctic Ocean, thus it has a very dynamic ice cover (Vinje

and Kvambekk, 1991). Therefore, in the present study we assess performance of CARRA in representing the ice conditions for a selected set of regions in addition to verifying the performance of the system over the whole ice-covered part of the model domain. The following four areas of interest are introduced (see Fig. 1): zone A – Baffin Bay (including the Nares Strait) and Davis Strait; zone B – Greenland Sea and the part of the North Atlantic Ocean adjacent to the Greenland coast; zone C – Barents Sea, Kara Sea, White Sea; zone D – central part of the Arctic Ocean within the CARRA domains defined by the northern borders of zones A, B and C. Borders of the zones A–D are set following the definitions of sea boundaries by IHO (International Hydrographic Organization) (1953) complemented by the proposed by IHO boundaries of the Iceland sea for the sake of convenience.

To study the long-term evolution of sea ice in the CARRA product, in addition to validation against observational products, we assess the series of mean monthly anomalies of sea ice surface temperature, ice thickness and snow depth. The anomalies are computed against reference multiyear mean fields constructed using the CARRA data over a 20 year time period from 2000 to 2020. The reference period of 20 years was selected to allow comparisons of the the sea ice surface temperature anomaly trends in CARRA to these derived from an observational product, which is not available prior to 2000.

The large data volumes of the CARRA product often do not allow for the direct computation of quantiles of a parameter of interest due to limitations of the processing hardware. Thus, when direct computation is not feasible, we use the algorithm suggested by Greenwald and Khanna (2001) to compute quantiles which, while not being mathematically precise, are accurate enough for the purposes of the present study. In the text we explicitly distinguish between approximate and precise quantiles by using the term 'estimated qunatiles' for the former case.

3.2 MODIS ice surface temperature products

The majority of the atmospheric reanalyses are based on adapted versions of operational NWP systems, and sea ice in these products is often represented by simplified one-dimensional sea ice parameterisation schemes. Ice surface temperature is one of the most important parameters in such schemes since ice is treated as a lowest boundary condition for an atmospheric model of an NWP system and not as one of the main prognostic components of the system. Other parameters such as ice thickness, snow depth or snow-ice interface temperature, while undoubtedly important to accurately represent the evolution of ice cover and valuable for end users of a reanalysis product, do not *directly* affect the energy exchange between the ice surface and the model atmosphere, thus their quality (as long as the produced surface temperature is realistic) is less critical to the reanalysis system itself.

In situ observations are local and sparse in the Arctic, thus, to obtain a general overview of the quality of ice surface temperature in CARRA, remote-sensing products are employed as the main source of observational data. In the present study we use near real time (NRT) level-2 (Parkinson et al., 2006) ice surface temperature products based on data from the MODIS instrument onboard the Terra and Aqua satellites (Hall and Riggs, 2015a, b). The MODIS sea ice surface temperature product is provided in 5 minute swathes, which have a nominal resolution of 1 km. Since the product is based on data from infrared-sensitive channels of the instrument it provides estimates of the ice surface temperature only in cloud-free conditions. Therefore, MODIS retrievals of ice surface temperature tend to have a cold bias when compared to in situ observations. More-

over, errors in cloud-detection, which can be challenging over sea ice, would result in spurious 'cold' pixels in the product. Nevertheless, the MODIS product has been shown to provide ice surface temperature fields of high enough accuracy (Hall et al., 2004) for the needs of the present study. Additionally, the MODIS data record covers a considerable time period allowing for assessing the multiyear performance of a reanalysis product without employing different products based on data from
195   different satellite instruments which simplifies intercomparisons and analysis.

For the intercomparisons we use the MODIS product data sets from both Terra and Aqua satellites covering the period from 2000 (Terra, the product from the Aqua satellite is available from 2002) to 2020. When processing, the two observational products are treated as a single merged data set and referred to as the MODIS ice surface temperature product in the following text. To reduce the impact of the misrepresented pixels of the MODIS product, we select only the pixels marked as 'good
200   quality' by the quality assessment procedure of the ice surface temperature retrieval algorithm. To compare gridded reanalysis fields against the non-projected satellite product, MODIS data are aggregated on the CARRA model grid (separately for the two CARRA model domains). When aggregating for a selected valid time of a reanalysis product all the MODIS swathes within the $[-30; 30)$ min interval are used without any time interpolation or adjustment. When comparing model data against the MODIS product, only the cloud-free sea ice grid cells of a reanalysis product (total cloud cover is less than 0.125) with ice
205   concentration over 15% are used.

3.3   Satellite albedo products

Similarly to the ice surface temperature, sea ice albedo is an important parameter which has a strong effect on the surface energy budget through the albedo feedback mechanism: a decreased albedo leads to more absorbed radiation, which again leads to higher surface temperatures and loss of sea ice (Riihelä et al., 2021; Pistone et al., 2014; Curry et al., 1995). Therefore,
210   to assess the modelled albedo fields in CARRA we compare them to a satellite-based surface albedo product. Additionally we perform a similar comparison using ERA5 to study the potential added value of the regional reanalysis product.

In the present study we use the surface albedo product (SAL) of the CLARA-A2 data record, a 34-year time series of black-sky surface albedo (covering the time period from 1982 to 2019), which is based on Advanced Very High Resolution Radiometer (AVHRR) data from the polar orbiting NOAA and METOP satellites (Karlsson et al., 2017). This product has been
215   validated against in situ observation in earlier studies (see, e.g., Karlsson et al., 2017; Anttila et al., 2016) and it is known to perform reasonably well over sea ice. Technically, SAL is provided on a 25×25 km equal-area grid (over the polar regions) and it is available as monthly or 5-day means.

For the albedo comparison we selected a 15-year time period, 2000-2015, which reflects modern sea ice conditions of the Arctic well. To match the monthly means of the SAL product, we perform similar averaging of the hourly output data from
220   both CARRA and ERA5 over each month from April to September (for other months the SAL product does not provide enough observations over the study area due to insufficient light conditions). Only the product grid cells with SAL monthly means derived from over 100 valid clear-sky AVHRR observations at global-area coverage resolution (4 km) are included in the analysis. Both CARRA domains are included in the intercomparison, and for the overlapping region an average of the albedo fields from both domains is used. To compare the albedo from the coarser-gridded SAL to the reanalyses, we aggregate

225 CARRA and ERA5 albedo fields in the product grid. Additionally, for CARRA we consider only the aggregates with at least 40 CARRA grid cells within a SAL product pixel, and for ERA5 with at least 2 grid cells. For both reanalyses this extra check ensures that no less than a half of the SAL grid cell area is represented, without extending to the adjacent grid cells. Finally, the monthly mean error is computed for each grid cell of the aggregated reanalysis fields.

3.4 Satellite sea ice thickness retrievals

230 The evolution of sea ice thickness is not the main target process of the one-dimensional sea ice scheme of HARMONIE-AROME applied in CARRA, and, in absence of ice dynamics, it can not be reproduced with all its complexity. Nevertheless, since ice thickness is provided as one of the model parameters in the CARRA product, we compare it against an observational product to highlight the limitations of the produced data set.

For this task, similarly to ice surface temperature and ice albedo, we use a satellite product to obtain a considerable spatial 235 and temporal coverage of sea ice within the area covered by CARRA. Specifically, we utilise a weekly combined CryoSat-2 and SMOS product (Ricker et al., 2017), which uses satellite altimetry data from the CryoSat-2 satellite for estimating the thickness of thick ice while taking estimates based on passive-microwave observations taken by the SMOS satellite over thin ice. The product covers the time period from 2010 to 2021 (at the moment of writing this paper) and it is provided on a 25 km grid. Data gaps in the estimated weekly ice thickness fields are filled by means of an optimal interpolation procedure where 240 the background field is produced by merging past and future (relative to the valid time of the produced analysis field) estimates derived from CryoSat-2 and SMOS (Ricker et al., 2017). Due to the limitations of the retrieval algorithms, the product does not provide ice thickness estimates between May and October.

When comparing the ice thickness reported by CARRA against the observational product, the three-hourly CARRA fields are aggregated on the 25 km grid of the product and then weekly-average values are computed. Over the overlap area of the 245 western and eastern CARRA model domains, the final ice thickness within a grid cell is computed as a mean of the values obtained from the two domains. Additionally, to assess the potential added value of applying a thermodynamic sea ice model in the CARRA system we use the uniform value of 1.5 m of ERA5 as a baseline.

3.5 Satellite snow depth over sea ice retrievals

The quality of snow cover in the CARRA product is of higher interest than that of ice thickness since misrepresented snow 250 cover can result in larger errors in the modelled surface energy balance compared to the effects induced by errors of a similar scale in a misrepresented ice layer beneath the snow cover. Despite that, satellite-based snow depth retrievals are much more uncertain and less reliable compared to ice thickness retrievals. Moreover, drifting ice mass balance buoys, which can be a valuable source of in situ observations tend to enter the area covered by the CARRA model domains in the spring time when snow cover starts actively melting thus provide little insight into the snow accumulation period.

255 Thus, in the present study we use a satellite-based snow depth product (Lee et al., 2021) for assessing the snow depth in CARRA and complement it by the Operation IceBridge (Sec. 3.6) flight campaign retrievals. The utilised product is based on the algorithm of Shi et al. (2020) where monthly estimates of the snow depth over sea ice are computed using sea ice freeboard

derived from passive-microwave data. The applied algorithm uses monthly surface and snow ice interface temperature fields for estimating the snow depth to ice thickness ratio, which is, in turn, used to derived the snow depth from the estimated freeboard.

260 The product covers the time period from 2003 to 2020 and provides pan-Arctic fields of the estimated snow depth on a 25 km grid for January, February, and March.

Similar to the ice thickness fields, snow depth over sea ice in the CARRA product is aggregated on the 25 km grid of the product and monthly means are computed. The region where western and eastern CARRA domains overlap is treated in the same way as when assessing the ice thickness.

265 Additionally, when comparing the CARRA data against the observational product, monthly CARRA snow depth 'estimates' are computed by applying the algorithm of Shi et al. (2020) to the model freeboard obtained by using the actual snow and sea ice parameters (i.e. the model snow water equivalent, model snow density and ice thickness). Applying the same algorithm as was used in the product to the model data reduces the discrepancy between the model snow depth in CARRA and the product arising from the differences in the snow and ice parameters. The snow depth in CARRA retrieved using the Shi et al. (2020)

270 algorithm is referred to as 'corrected snow depth' later in the text.

Since the ERA5 reanalysis system does not resolve the evolution of the snow cover over sea ice, ERA5 was excluded from the snow depth intercomparisons.

3.6 Operation IceBridge snow depth data

Since the satellite snow depth retrievals tend to have high uncertainty, we use an additional independent data set to complement

275 the comparisons against the satellite product. In the present study we use snow depths obtained from the radar altimetry observations taken during the Operation IceBridge (OIB) flights (Kurtz et al., 2015, 2016). This data set spans over the time period from 2009 to 2019 and has uneven spatial coverage with most of the flights within the CARRA domains conducted over the north of Greenland and only few tracks entering the areas south of 80°N.

To compare the CARRA snow depth against the OIB data, the snow depth estimates along the OIB flight tracks, which have

280 a spatial resolution of 40 m, are aggregated on the 2.5 km grid of the CARRA product. For intercomparisons, snow depths from the CARRA analysis with the closest valid time are considered for each data point of the aggregated OIB track.

3.7 In situ data and ice mass balance buoys

Satellite retrievals discussed so far in the previous sections provide estimates of the sea ice properties in the Eulerian frame, or in other words, over a prescribed grid. Thus, for these products changes in sea ice state within each grid cell arise due to

285 contributions from both thermodynamic and dynamic processes. However, the CARRA system uses a greatly simplified sea ice parameterisation scheme which represents only thermodynamic processes in the ice column. Therefore, to better assess the performance of the CARRA system in representing these processes we compare the CARRA product against a set of in situ observations reported by drifting ice mass balance buoys.

The unmanned ice mass balance buoys (IMB) usually measure snow depth, ice thickness and temperature and can vary

290 in design and complexity. In the present study we use data from two types of IMBs: acoustic sounder-based buoys (Richter-

Menge et al., 2006), referred to as CRREL buoys in the text, and simpler thermistor string-based buoys (Jackson et al., 2013), referred to as SIMBA buoys in the text. The CRREL IMBs measure the distances between the downward-looking sounder and the snow surface, and the upward-looking sounder and the ice bottom. Based on initial sea ice conditions at the time of buoy deployment, these distances can be converted to snow depth and ice thickness. Additionally, CRREL buoys employ a

295 separate thermistor string that measures the vertical temperature profile through air-snow-sea ice-ocean. The thermistor string of CRREL buoys has individual sensors located at a distance of 10 cm between each one. The SIMBA IMBs measure only the series of vertical temperature profiles by means of a thermistor string with sensors located every 2 cm. However, two types of temperatures are measured by the SIMBA buoys. Firstly, they report the environment temperature of air, snow, ice and water where SIMBA thermistor sensors are located, which is consistent with the temperature profiles reported by CRREL buoys.

300 Secondly, SIMBA buoys measure the temperature change after each thermistor is applied with an identical amount of heat by means of heating elements adjacent to the sensors. The changes in the temperature reading after a heating cycle depend on the thermal properties of air, snow, ice and water, and therefore can be used to identify the type of medium surrounding the sensors. Thus, temperatures reported by SIMBA buoys can be used to derive snow depth and ice thickness manually (Lei et al., 2018) or automatically (Liao et al., 2018; Cheng et al., 2021). Both types of IMBs are normally deployed on undeformed ice

305 floes at a selected location and then drift along with the ice floe. The standard observations are made every 6 hours, and the buoy's GPS location is recorded every hour. Both types of IMB have been deployed in the Arctic Ocean for many years. Their data are representative for regional, seasonal and interannual sea ice mass balance (Perovich and Richter-Menge, 2015; Lei et al., 2018) and air-sea ice-ocean interactions along IMB drift trajectories (Provost et al., 2017; Koo et al., 2021; Cheng et al., 2021; Lei et al., 2022). In this study, we use data from 19 individual IMBs (see Table S5) collected from various field programs

310 and compare them against the CARRA product (using ERA5 as a baseline, where applicable). Other IMB buoys, which also entered the CARRA domains throughout the time period covered by the product, were excluded from the intercomparisons due to issues with the reported parameters. We target four parameters: snow depth, ice thickness, surface temperature and snow-ice interface temperature. The surface temperature was obtained by linear interpolation based on snow depth (or ice thickness) and readings from the thermistor sensors closest to the snow-air interface (or ice-air interface in case of missing snow cover). The

315 snow-ice interface is assumed to remain unchanged from its initial position when an IMB was deployed. Although, dynamic and thermodynamic interactions between snow and ice may result in a moving snow-ice interface because of snow-ice and superimposed ice formation, especially during the early melting season and early winter when ice is still thin (Cheng et al., 2003, 2008, 2021). However, the IMBs used in this study were deployed in late autumn on thick ice floes when the ice was about to freeze up and the snow was thin, thus reducing the chances of snow-ice formation processes affecting the IMB reading.

320 Therefore, the assumption of a static snow-ice interface is adequate for the purposes of the present study.

When processing IMB data we first identify parts of an IMB trajectory that are located within the two CARRA model domains, and a corresponding subset of observed parameters is extracted. Then, for each GPS position reported by an IMB, the same parameters from the nearest CARRA model grid cell are retrieved. Note that in the CARRA product there is no dedicated snow-ice interface temperature field, therefore we used the temperature of the top-most ice layer (which can be up to 5 cm

325   thick, see Batrak et al. (2018) for the details) as an analogue. To facilitate intercomparisons both IMB and CARRA data were
resampled to hourly temporal resolution.

4   Results and discussion

4.1   Ice surface temperature

When assessing the quality of ice surface temperature in CARRA we first study the multiyear performance of the product
330   in order to evaluate whether it reasonably represents temperature trends linked to ongoing climate change. As a second step
we evaluate the annual cycle of modelling errors computed against the MODIS satellite product. Figure 2 shows the obtained
anomalies as well as the computed ice surface temperature anomaly trends for both CARRA and the MODIS satellite product.
As can be seen from the figure both CARRA domains show a positive anomaly trend with a value of $0.08 \,^{\circ}\mathrm{Cy}^{-1}$ (95% CI
$[0.06; 0.11] \,^{\circ}\mathrm{Cy}^{-1}$) and of $0.20 \,^{\circ}\mathrm{Cy}^{-1}$ (95% CI $[0.16; 0.25] \,^{\circ}\mathrm{Cy}^{-1}$) for the western and eastern CARRA model domains,
335   respectively. Monthly sea ice surface temperature anomaly trends found in the MODIS product show comparable values for
both CARRA model domains; $0.07 \,^{\circ}\mathrm{Cy}^{-1}$ (95% CI $[0.05; 0.10] \,^{\circ}\mathrm{Cy}^{-1}$) for the western domain and $0.17 \,^{\circ}\mathrm{Cy}^{-1}$ (95%
CI $[0.13; 0.21] \,^{\circ}\mathrm{Cy}^{-1}$) for the eastern domain. These values are in line with the findings of previous studies (see, e.g., Rantanen
et al., 2022; Nielsen-Englyst et al., 2023), but differences in the lengths of the anomaly series and covered areas, as well
as the shorter period used for computing multiyear means in the present study (20 versus 30 years), do not allow the direct
340   comparisons. The eastern CARRA model domain showing a considerably larger anomaly trend than that found for the western
domain, is also in agreement with earlier works, which suggest the Barents Sea region has higher warming rates than the
Greenland Sea and the Central Arctic region (see, e.g., Screen and Simmonds, 2010; Comiso and Hall, 2014; Isaksen et al.,
2022; Nielsen-Englyst et al., 2023).

    After assessing the multiyear trends in the CARRA product we focus on annual variability found in the CARRA data
345   and on the performance of the regional reanalysis compared to the ERA5 data set. First, we assess the general performance
of CARRA in terms of ice surface temperature throughout the year. Figures 3 and 4 show monthly estimated quantiles of ice
surface temperature in CARRA and ERA5 compared to the estimated quantiles of the MODIS product computed for the period
from January 2000 to January 2020. The figures show that for both model domains CARRA tends to have lower ice surface
temperature than ERA5 for all months, except September for the western CARRA domain where ERA5 is slightly colder. The
350   lower temperatures of the CARRA product better agree with MODIS, especially during the winter and spring seasons over the
eastern CARRA domain. During the active melting season in the summer, both CARRA and ERA5, compared to MODIS, tend
to have higher ice surface temperatures than in the retrieval and show comparable performance. With the onset of the freezing
season (starting from September) and until December the two reanalysis products show a considerable warm bias similar to the
summer months. Moreover, during this period CARRA does not show any noticeable improvement over ERA5 for the western
355   domain. For the eastern domain, CARRA is slightly colder than ERA5 in October and November, however it still has a higher
ice surface temperature than that reported by the MODIS product.

Differences in the ice surface temperature quantiles between the two CARRA model domains suggest that sea ice cover is represented with a varying degree of accuracy over the different parts of the joint area covered by the CARRA product. Thus, Fig. 5 shows the annual evolution of the ice surface temperature bias over the four selected areas of interest computed

360   against the MODIS product for the period from January 2000 to January 2020. As can be seen from the figure, evolution of the ice surface temperature bias differs considerably over the selected areas, although CARRA still shows a lower mean error compared to ERA5 for all zones and months except for January and September in zone A, and September in zone B. While ERA5 almost universally has a warm bias when compared to the MODIS product (except for September in zone A) CARRA shows periods of distinct negative median bias within zone A from December to April.

365   Over zone A, which includes the Baffin Bay and the Davis Strait, ERA5 shows relatively low variation of the ice surface temperature bias, which has a value close to 2 °C with the only major drop to a value of approximately 0 °C observed in September. Contrary to ERA5, CARRA has the highest positive bias in September with the value reaching 2.05 °C which is reduced to zero by December and then becomes negative. A negative median bias in CARRA over zone A is found throughout the period from December to April with the lower-most values of -1.66 °C observed in January–February. Over the summer

370   season CARRA has a near-constant median bias, with values within the range of 0.87–1.45 °C, which starts growing in August.

For zone B, covering the Greenland Sea and the North Atlantic Ocean, both CARRA and ERA5 show positive median bias throughout the year. ERA5 has the highest bias in December with a value of 5.75 °C which then gradually decreases over the following months and reaches a minimum value of 1.41 °C in August before starting to grow again. For this zone CARRA shows a similar annual cycle of the median bias, although it is shifted with a maximum value of 3.63 °C observed in November

375   and a minimum value of 0.79 °C found in July. For the period from mid-August to the beginning of October CARRA tends to have a higher positive bias than ERA5 due to a shift in the annual cycle of modelling errors.

For ERA5, zones C and D show a qualitatively similar evolution of the median ice surface temperature bias with the annual maximum observed in the autumn months and lowest bias found in July, although over zone D bias is higher on average. On the other hand, the CARRA product features noticeable differences in the annual cycle of the median ice surface temperature

380   bias for these two zones. For zone C, CARRA, while having the highest median bias of 3.83 °C in September–October (similar to ERA5, which has a value of 4.62 °C), shows a period of the median bias reduced to near-zero from January to March (unlike ERA5 with a winter-time median bias close to 3 °C). This bias starts growing again in April to reach the summer value of approximately 1.5 °C, which is close to that of ERA5. Over zone D, CARRA does not show a winter-time near-zero median ice surface temperature bias as was found over zone C and exhibits a similar to ERA5 annual cycle with a maximum of 5.54 °C in

385   September–October and a minimum of 0.97 °C in July. However, the autumn-time maximum of the bias, which is not present in ERA5, is well-pronounced in CARRA, similar to zone C. For zone D, the median ice surface temperature bias in ERA5 has a maximum value of 7.17 °C (November) and a minimum of 2.37 °C (July).

In general, CARRA shows the most noticeable decrease of the median bias during the winter months, when this difference can reach values of up to 4 °C, and during the melting season the difference between CARRA and ERA5 is reduced. These

390   results, arising from including an explicit representation of the snow cover over sea ice in the CARRA systems are in line with the result of Arduini et al. (2022) assessing the effects of resolving the snow layer over sea ice in IFS-HRES. Additionally, the

year-to-year variability of the ice surface temperature bias is noticeably different between CARRA and ERA5 with CARRA
tending to have more variability than the global reanalysis product. This variability in CARRA is considerably higher in zones B
and D than in zone A. Notably, scores over zone C in CARRA show increased variability mainly during the autumn freeze-up
395 season, similarly to ERA5.

4.2 Ice albedo

To assess the sea ice albedo fields in CARRA and ERA5 reanalysis products, we examine the monthly mean error maps for
CARRA and ERA5 SALs.

Qualitatively, as can be seen from Fig. 6, sea ice albedo in CARRA and ERA5 show similar features throughout the studied
400 months, and in both reanalyses it is higher in April, May and September compared to the satellite-based product, while in July
the sea ice albedo is underestimated. For June and August CARRA and ERA5 show noticeable difference in the sea ice albedo
errors. Specifically, in June CARRA shows a positive bias in the albedo field over the northern Barents Sea, where bias is
weakly negative in ERA5. In August, ERA5 shows good agreement with the CLARA-A2 product, while albedo in CARRA is
overestimated in the northernmost areas of the model domain (zone D in Fig. 1).

405 Quantitatively, sea ice albedo in CARRA is consistently higher than values reported by the SAL product for all studied
months except June and July. These high albedo values lead to larger errors than in ERA5 on average, which can be traced in
the error probability density functions (PDFs, see Fig. S1). For example, in the April PDF of the albedo error, CARRA has a
mode of 14%, which is considerably higher than the 6% found in ERA5, although these values are reduced for other months.
In June, sea ice albedo errors in CARRA are distributed around zero with a mode of 0.1%, but in July both CARRA and ERA5
410 show a clear underestimation of the albedo with the mode of the error PDF close to -5%. In August, the error PDF of the sea ice
albedo in CARRA shows a bimodal distribution which is attributed to the fresh snow accumulation and temperature drop over
the ice-covered central Arctic. As a result, the first mode of 8% indicates the previously observed characteristic overestimation
of the surface albedo while the second mode of -1% is related to the snow-free ice cover within the model domain. This value
of the second mode of the albedo error PDF is close to that found in the ERA5 data, namely 0.1%.

415 In general, the observed errors in the CARRA sea ice albedo can be attributed to several factors. Firstly, sea ice albedo
parameterisation schemes applied in CARRA and ERA5 systems differ in their philosophy: CARRA uses modelled albedos,
while ERA5 relies on time-interpolated observation-based albedos. Therefore, in CARRA surface albedo over sea ice has more
degrees of freedom and depends on the surface temperature and model precipitation. Applying an unconstrained parameterisa-
tion can result in considerable errors, even though such an approach gives a more consistent model state. Secondly, snow cover
420 over sea ice is represented as a flat and uniform layer covering a whole grid cell with the surface albedo computed by the snow
scheme. This idealised approach results in a more reflective ice surface compared to real sea ice cover. Thirdly, the negative
summer-time sea ice albedo bias found in CARRA (and also observed in ERA5) indicates shortcomings in the representation
of the melting regime of the sea ice. Specifically, processes of melt pond formation and evolution are not explicitly represented
in the CARRA system and a simple temperature-based sea ice albedo scheme can not accurately reproduce all the effects of
425 melt ponds. Finally, it is also possible that the SAL of the CLARA-A2 product underestimates sea ice albedo in the spring

months. However, characteristic biases found in CARRA and ERA5 (for example, spring-time overestimation of the sea ice albedo) are not unique to the CLARA-A2 SAL product, and similar performance of ERA5 was observed when using other satellite based albedo retrievals as a reference (Pohl et al., 2020). Nevertheless, uncertainties in the intercalibration method of the AVHRR data record can influence the average level of the albedo, and it is expected that the upcoming next edition of the

430   albedo product, CLARA-A3 SAL, will have somewhat higher sea ice albedo values in spring (personal communication, Aku Riihelä).

4.3   Ice thickness

The CARRA system is based on a non-coupled atmospheric NWP system, therefore it uses a simplified one-dimensional parameterisation scheme for representing sea ice cover in the model. However, CARRA uses a more advanced sea ice scheme

435   compared to ERA5 and the CARRA data set includes such fields as ice thickness and snow depth. Therefore, in the present study, we use available remote-sensing and in situ observations for assessing the performance of the sea ice scheme in the CARRA system with respect to these additional parameters.

Sea ice thickness, specifically in the grid cells with perennial ice ice cover, is the prognostic model variable with longest memory in the CARRA system since it is not constrained by observations and does not disappear during summer melts, unlike

440   snow over sea ice, for example. Thus, consistent long-term performance of the sea ice scheme becomes more important to avoid unrealistic features in the produced data set. The long memory of sea ice becomes especially important when considering the initial production of a reanalysis data set, which represents multiple decades of data and is therefore usually generated by means of a number of separate production streams to reduce the integration time. In these streams sea ice cover is initialised independently and it can be challenging to achieve a seamless transition from one stream to another if the scheme is not

445   constrained. Therefore, we assess the long-term performance of the CARRA system in representing sea ice cover by computing monthly mean ice thickness anomalies over the period covered by the product. Figure 7 shows the computed anomalies as well as the fitted anomaly trend for both model domains of the CARRA system. As can be seen from the figure, the CARRA product shows a general trend towards decreasing ice thickness for both model domains, namely -1.2 $\mathrm{cm\,y^{-1}}$ (95% CI [-1.4; -1.1] $\mathrm{cm\,y^{-1}}$) for the western domain, and -1.4 $\mathrm{cm\,y^{-1}}$ (95% CI [-1.5; -1.2] $\mathrm{cm\,y^{-1}}$) for the eastern domain. These values are

450   in line with the general trend towards thinner ice in the Arctic observed and reported by multiple studies (see, e.g., Renner et al., 2014; Hansen et al., 2013; Lindsay et al., 2009), albeit with weaker thinning rates. However, Fig. 7 reveals an inconsistency in the computed anomaly series caused by separating the CARRA production into a set of parallel production streams. This inconsistency, which can be illustrated as a sudden anomaly drop between streams BE and S1 of the western CARRA domain as shown in Fig 7a, affects the long-term ice thickness statistics. Similar feature can be traced for the eastern domain as well,

455   although much less apparent.

The sea ice scheme applied in the CARRA system (Batrak et al., 2018) does not resolve ice dynamics and represents only thermodynamic processes in the ice column. Thus, comparing CARRA against a gridded satellite product can highlight the limitations of the reanalysis. On the other hand, comparisons against measurements taken by drifting platforms (for example,

ice mass buoys), which essentially observe only the thermodynamic processes in a single ice floe, can highlight the performance
460   of the parameterisation scheme applied.

An initial intercomparison against the composite CryoSat-2/SMOS satellite product for all available dates shown in Fig. 9a indicates high spatial non-uniformity of the modelling errors. In general, sea ice thickness in the CARRA data set tends to be underestimated over the coast of Greenland within zone B and in the central Arctic (zone D). For other zones and areas sea ice in CARRA is thicker than reported by the satellite product. Similarly, non-systematic errors in the modelled ice thickness
465   (expressed as the standard deviation of errors (ESTD), see Fig. 9b) are very non-uniform within the model domain. The highest ice thickness ESTD values are found over the Greenland Sea while over Baffin Bay and the Kara Sea CARRA shows mainly systematic errors. This distinction can be attributed to the characteristic sea ice regime of the Greenland Sea where transport of old ice from the central Arctic makes ice cover more variable and challenging to reproduce.

The annual evolution (limited to the period of availability of the satellite ice thickness retrieval, namely from October to
470   April) of the average mean error of the sea ice thickness modelled by the CARRA system is presented in Fig. 8. As can be seen from the figure, and supported by the features found in Fig. 9, CARRA shows distinct differences in the ice thickness modelling errors and their temporal evolution between the four zones of interest. The figure shows persistent negative bias in zones B and D where the multiyear average ice thickness mean error (ME) reaches -0.88 m over zone B in January, and -1.05 m over zone D in April. For zones A and C, where ice is thinner on average, CARRA shows better performance, although the
475   modelled ice thickness shows a positive average ME, which tends to grow throughout the winter. Thus the ice thickness ME for zone A at the beginning of the freeze-up period is 0.07 m on average, but by mid-April it reaches values of 0.57 m. For zone C the situation is similar with the average ME ranging from 0.17 m in October to 0.52 m in April. Therefore, thermodynamic ice growth rates in CARRA over zones A and C tend to overestimate the values observed in reality. On the other hand, the zones B and D annual variability of the multiyear average ME is less pronounced.

480   The thermodynamic sea ice model applied in CARRA is more advanced than a scheme with a prescribed and uniform ice thickness, such as that used in ERA5. To assess the improvement in the modelled ice thickness (if any) resulting from applying a thermodynamic sea ice model in a reanalysis system we compare the ice thickness ME of the CARRA product to the simulated ME of a CARRA system version with a prescribed and uniform ice thickness of 1.5 m. These additionally computed scores are presented in Fig. 8. The figure suggests that having a prescribed ice thickness in the CARRA system would show reduced
485   on average ME compared to the actual CARRA system for zones B and D but considerably increased ME for zones A and C. In general, when ice thickness is prescribed, annual series of the ME show a negative slope and the difference between the CARRA system (where ice thickness ME grows throughout the winter season or remains relatively constant) and persistent 1.5 m ice is greatest in October and reduced by April. For example, for zone D, having prescribed ice thickness would result in an October average ME of -0.05 m which is a considerably lower ME than the value of -0.78 found in CARRA, however,
490   by the end of the winter season in April, this difference in ME is greatly reduced and ME takes values of -1.05 m and -1.03 m, respectively. For zones A, B and C the ME difference between modelled and persistent ice thickness evolves in a similar way. The reduced growth rate of the ME in CARRA compared to the persistent ice thickness indicates the benefits of applying a

thermodynamic sea ice model in the reanalysis system. However, the offset errors found over zones B and D suggest potential advantages of constraining the ice thickness by means of observational data sets.

4.4 Snow depth

Similar to the ice surface temperature and ice thickness we first assess the long term performance of the CARRA system by studying the monthly anomalies of the snow depth in the reanalysis product. As can be seen from Fig. 10, snow cover over sea ice shows a similar trend to that found for ice thickness towards more frequent negative anomalies over the last decade compared to the first years of the anomaly series. However, compared to the ice thickness, snow depth anomalies are smaller, and anomaly trends are less pronounced. Specifically, for the western CARRA model domain, the product shows a very weak negative trend of -0.09 $\mathrm{cm\,y^{-1}}$ (95% CI [-0.06; -0.12] $\mathrm{cm\,y^{-1}}$), and for the eastern model domain the anomaly trend is stronger with a value of -0.28 $\mathrm{cm\,y^{-1}}$ (95% CI [-0.24; -0.34] $\mathrm{cm\,y^{-1}}$). The general trend towards diminishing snow depth over the Arctic sea ice in both observations and modelling data sets is noted in multiple studies (see, for example, Webster et al., 2014; Chen et al., 2021; Zhou et al., 2021). The more pronounced decrease in the snow depth for the eastern CARRA model domain is in line with the modelling results of Chen et al. (2021) and Zhou et al. (2021), which show stronger negative trends in the snow depth series over the Barents Sea region.

When assessing the snow layer over sea ice in the CARRA system we use both satellite retrievals and direct observations from OIB flights. However, since the satellite retrievals of snow depth over sea ice are highly uncertain we only use them for a general qualitative assessment because they cover a much larger area compared to OIB. Figure 11 shows the mean error of the modelled snow depth in CARRA compared to the satellite retrieval product and OIB data. For both observation types CARRA shows a similar distribution of the modelling errors with generally overestimated snow depth within the model domain on average. The largest errors are found in the Greenland Sea along the eastern coast of Greenland. For this area OIB reports a snow depth of 0.24 m on average while the modelled snow depth in CARRA is 0.71 m on average. However, it must be noted that most of the OIB data in that region originate from a very limited number of flight tracks, thus the drawn conclusions should be taken with care and not generalised. For the satellite snow retrieval product a similar pattern can be traced in the Greenland Sea, which supports the aforementioned findings and suggests that snow depth is overestimated in general in the CARRA product for this region. Over zone D, or in other words, in the part of the central Arctic present in the CARRA domains, snow depth in CARRA is considerably lower than over the Greenland Sea and, when compared to both OIB and the satellite product, shows close to zero and negative ME. A similar distribution of the modelling errors is observed when comparing CARRA against OIB around the Svalbard archipelago where the reanalysis product shows clear underestimation of the snow depth. This behaviour can be partially attributed to the misrepresentation errors of the sea ice cover in CARRA combined with the crude initialisation procedure for the newly ice-covered grid cells which always start from the snow-free state. However, in reality, sea ice is a drift medium and areas within and close to the marginal ice zone may contain ice floes that originate from the remote parts of the Arctic ocean and have accumulated snow cover throughout their drift. The OIB data set does not provide snow depths over zones A (Baffin Bay) and C (Barents and Kara seas), although comparisons against

the satellite retrieval product suggest that CARRA has lower snow depth modelling errors for these regions compared to the Greenland Sea.

Estimated PDFs of the snow depth in the observational products and in the reanalysis data set provided in Fig. 12 complement the findings made from assessing modelling errors. Although the differences in the PDFs computed from the satellite product

530   and OIB data suggest that the satellite product underestimates the snow depth, as can be seen from the figure, thicker snow layers occur more often in the CARRA data than in both the satellite product and OIB observations. Modelled snow cover in CARRA has a median depth which grows throughout the winter from 0.28 m in January to 0.33 m in February and finally reaches a value of 0.37 m by March. In contrast, the satellite product reports much thinner snow cover on average with a median snow depth close to 0.13 m for all three months (see Fig. 12a). The PDFs of the corrected snow depth in CARRA show less

535   frequent occurrence of extreme snow depths (both low and high) which indicates that the processing algorithm applied in the satellite product of Lee et al. (2021) underrepresents cases of thick snow cover within the CARRA model domains. Along the OIB tracks, median values of the snow depth in CARRA and OIB observations are 0.42 m and 0.27 m, respectively. Notably, the CARRA snow depth PDF in Fig. 12b shows a peak at zero snow depth not present in the OIB data, which suggests that there are instances of ice cover in CARRA misrepresenting the state of the actual ice cover near the ice edge, which is again

540   attributed mainly to the effects of the applied initialisation procedure for the newly ice-covered grid cells.

4.5 Additional validation against in situ observations from buoys

So far in the present study we have used remote sensing data from satellite- and airborne instruments for assessing the performance of the CARRA system in reproducing the evolution of sea ice cover. Although providing valuable information about the sea ice state on large scales, these products rely on multiple assumptions and tend to have their own biases and limitations.

545   Therefore we additionally assess the CARRA product using observations reported by a set of ice mass balance buoys.

Figure 13 summarises the intercomparisons between the CARRA and ERA5 reanalysis products, and ice mass balance buoy data. As can be seen from Fig. 13a, for most of the IMB trajectories CARRA report ice thickness close to or below the observed values, although there is a singe case where ice thickness is considerably overestimated in CARRA, being 0.35 m thicker, on average, than the reported ice thickness. ERA5, which has ice cover of uniform thickness, while having smaller median ME

550   compared to CARRA (-0.08 m and -0.20 m, respectively) shows larger spread of the modelling errors throughout the set of selected buoys. Specifically, in CARRA, ME values for individual buoys range from -0.67 m to 0.35 m, and in ERA5 they are within the interval from -0.78 m to 0.61 m. Such a discrepancy in the performance of the two reanalysis products suggests that the thermodynamic sea ice scheme applied in CARRA, while having obvious deficiencies, is more suitable for representing the evolution of a drifting ice floe. The standard deviation of errors in the modelled ice thickness in CARRA is higher of that

555   in ERA5, which is in line with the increased complexity of the sea ice model in the regional reanalysis system (in ERA5 ice thickness is prescribed and constant, thus computed ESTDs simply represent variability within the observational series).

The snow depth over sea ice in CARRA, when compared to buoy data, also shows similar variability between individual buoys, although the median value of the per-buoy ME is positive and has a value of 0.10 m, while errors from individual buoys range from -0.34 m to 0.34 m. This result is in line with the findings of earlier intercomparisons against the satellite snow depth

560   product and OIB observations where CARRA shows consistent overestimation of the snow depth in the area. However, there
      is a substantial number of buoys (5 of 21 in total) that report a considerable underestimation of snow depth in CARRA, with
      along-track ME below -0.10 m. For four of these five buoys, underrepresented ice cover in CARRA can be partially attributed
      to the offset error, as these buoys report a snow layer that is considerably thicker than in CARRA already at the beginning of
      the buoy's trajectory section within the CARRA model domains (this initial offset error takes values from 0.14 to 0.79 m).
565   The last buoy in this group enters the western CARRA domain during the melting season 2012 and shows much faster snow
      accumulation compared to the reanalysis product during the following autumn months.

      In the presence of offset errors, correlation can be used as an additional indicator of the level of agreement between the ob-
      servational and modelled values. However, for the discussed set of ice mass balance buoys, computed values of the correlation
      coefficient show considerable variation ranging from strong correlation to strong anticorrelation for both snow depth and ice
570   thickness (see Table S5 for the actual values). Such a discrepancy between the modelled and observed parameters can be partly
      attributed to boundary effects near the edge of model domain and to the crude procedure of the initialisation of new ice.

      Ice and snow temperature series reported by buoys are a valuable source of in situ observations, which can be used to assess
      the performance of a reanalysis system, especially for the parameters that are not available from the satellite products, such as
      the snow-ice interface temperature. Figure 13d shows estimated PDFs of the modelling errors of the ice surface temperature
575   in ERA5 and CARRA reanalysis products, as well as the PDF of the snow-ice interface temperature errors in CARRA. As
      can be seen from the figure, CARRA shows a lower probability of positive errors in the modelled ice surface temperature
      than ERA5 and a higher probability of the modelled ice surface being colder than reported by buoys. Additionally, the mode
      of the CARRA ice surface temperature error PDF is -0.19 °C, which is closer to zero than in ERA5, where the mode has a
      value of 0.62 °C. These error distributions support the conclusions drawn from comparing the reanalysis products against the
580   MODIS ice surface temperature product. Similar to those results, CARRA, when compared to buoys, tends to show lower than
      ERA5 surface temperatures with ice surface temperature MEs of 0.04 °C and 1.48 °C, respectively. Unlike the ice surface
      temperature, the snow-ice interface temperature, when modelled by the CARRA system, shows a notable warm bias with the
      error PDF having a mode of 0.77 °C. The positive bias in the snow-ice interface temperature found in the CARRA data (ME
      is 0.61 °C) is attributed to the commonly occurring overestimation of the snow depth found when comparing CARRA against
585   satellite and OIB snow depth, which is also identifiable when comparing reanalysis data against IMBs (see Fig. 13a). In such
      cases, the insulating effect of the snow layer is too strong, which leads to a higher snow-ice interface temperature, especially
      if the ice thickness is also underestimated. When assessing the modelled and observed temperature series for individual buoys,
      CARRA shows consistently high correlation with the buoy data for both surface and snow-ice interface temperatures. For
      surface temperature, the correlation coefficient ranges from 80% to 97%, while for the snow-ice interface temperature the
590   correlation is slightly lower on average with the lowest value of 63% (see Table S5).

      Since the main purpose of the sea ice parameterisation scheme in the CARRA system is representing the heat exchange
      between the ice and the model atmosphere, we additionally assess the heat fluxes throughout the snow and ice layers as they
      govern the heat transport from the ocean to the atmosphere. Figures 13b and 13c show the temperature gradient within the
      ice and snow layers, respectively, as reported by the CARRA product compared against the values computed from the buoy

data. The figures show good agreement between the modelled and observed gradient in most cases, although CARRA tends
to underestimate the highest values of the ice temperature gradient compared to buoy data, which can be attributed to the
warm bias of the snow-ice interface temperature. One notable exception is a single CRREL buoy deployed in 2010 for which
CARRA reports much higher temperature gradients within the ice layer than observed as the buoy drifts within the CARRA
model domains. For this buoy, CARRA shows a considerable underestimation of the ice thickness at the point where the buoy
entered the model domain. This initial error of $0.75\,\mathrm{m}$ in the modelled ice thickness resulted in an overestimated temperature
gradient in the ice layer with a maximum value of $0.19\,^{\circ}\mathrm{C}\,\mathrm{cm}^{-1}$ compared to $0.06\,^{\circ}\mathrm{C}\,\mathrm{cm}^{-1}$ as computed from buoy data.
Snow temperature gradients, as shown in Fig. 13c, exhibit similar features to the ice temperature gradient in the CARRA
data, although the snow temperature gradients have more spread in the computed values due to higher variability of the snow
layer (primarily in terms of surface temperature) compared to the ice layer. Similarly to the ice temperature gradients, Fig. 13c
suggests some underestimation of the gradient in CARRA for the strongest gradients (with a value over $0.50\,^{\circ}\mathrm{C}\,\mathrm{cm}^{-1}$ when
computed from the buoy data). Additionally, CARRA seldomly shows negative values of the snow temperature gradient (or
cases when snow ice interface is colder than the snow surface). However, some of the negative temperature gradients in the
buoy data may arise from the buoys reporting positive snow surface temperatures during the melting season, thus these results
should be taken with care.

5 Conclusions

Numerous research and engineering studies benefit from using atmospheric reanalysis products, which are sometimes treated
as providing information about the *true* atmospheric state. However, since atmospheric reanalyses are generated by employing
NWP systems, they are not devoid of modelling errors and biases. Moreover, a reanalysis product is produced by the same
unmodified version of an NWP system that quickly becomes outdated after a few years of production. All of these factors
necessitate the production of new reanalysis products that incorporate the latest developments in NWP and reflect the advances
in high performance computing.

The Copernicus Arctic Regional Reanalysis (CARRA) is a novel regional atmospheric reanalysis product that focuses on the
Canadian and European Arctic. This product has a considerably higher spatial resolution compared to the global ERA5 product
($2.5\,\mathrm{km}$ versus $30\,\mathrm{km}$) and is based on a non-hydrostatic regional NWP system, HARMONIE-AROME. CARRA covers the
time period from 1990 to present (2023 at the moment of publication) and represents a region defined by two overlapping
model domains. Compared to ERA5, CARRA uses a more advanced sea ice parameterisation scheme, which includes explicit
representation of thermodynamic ice growth and evolution of the snow cover.

In the present study we assessed the sea ice surface temperature, surface albedo, ice thickness and snow depth fields provided
by the CARRA product and validated them against an extensive set of remote-sensing and in situ observations, with focus on
the recent decades. Additional comparisons against a selected set of IMBs complement and support initial validation against
the satellite products. Sea ice extent in CARRA was not discussed in the present study since the CARRA system uses satellite-

based sea ice concentration products, which are well-covered by earlier studies (Lavergne et al., 2019), to define ice-covered regions.

The sea ice cover in CARRA adequately represents general multiyear trends towards thinner and warmer ice cover, con-
630   nected to the ongoing climate change in the Arctic. Comparisons against the satellite-based and in situ sea ice observations show generally improved representation of sea ice in CARRA (using ERA5 as a baseline), although this improvement is not universal. The main difference between the sea ice schemes in ERA5 and CARRA is the presence of an explicitly resolved snow layer, which allows for much lower ice surface temperature in the CARRA system, therefore reducing the warm ice surface temperature bias found in ERA5. However, for the area covering Baffin Bay and the Davis Strait the verification scores
635   suggest that a warm winter-time bias of ERA5 is replaced with a cold bias in CARRA. For sea ice albedo, the CARRA product does not show improvement compared to ERA5, which uses observation-based albedo estimates over sea ice, and displays considerable overestimation of the spring-time surface albedo, when compared to the CLARA-A2 satellite product. This result suggests that future applications could benefit from a more detailed representation of the sea ice albedo in the sea ice model of Batrak et al. (2018), or from constraining model albedo against an observational product. The general limitation of the sea
640   ice scheme applied in CARRA, namely the absence of ice dynamics and external constraints, strongly manifests itself in the verification scores for ice thickness and snow depth. Even though, unlike ERA5, CARRA has spatially non-uniform ice thick-
ness, it can not accurately resolve thick multiyear ice in the central Arctic leading to a negative ice thickness bias. Additionally, due to a simplified initialisation procedure for new ice, thin first-year ice is thicker in CARRA than suggested by the reference satellite-based product. Snow cover over sea ice in CARRA exhibits similar features, such a extensive snow accumulation
645   along the eastern coast of Greenland where missing ice transport can not aid at removing snow-covered ice. Thus, prognostic snow cover over sea ice and ice thickness fields computed by the thermodynamic sea ice model of the CARRA system, which are available within the reanalysis product and showing a reasonable annual cycle, should be used with great care.

An additional intercomparison performed against ice mass balance buoys shows good agreement between the modelled in CARRA and observed temperatures, although the snow-ice interface temperature in CARRA shows a consistent warm bias,
650   with a mode comparable to the that of the ice surface temperature bias found in ERA5. Due to the location of the CARRA model domains most of the buoys enter them at the final stage of the drift. As a consequence, ice thickness and snow depth over sea ice show less agreement to observational series and instances of considerable offset errors were noted.

Sea ice cover in CARRA reflects current approaches applied in operational short-range applications utilising the HARMONIE-
AROME NWP system. The shortcomings and limitations of representing sea ice with non-constrained one-dimensional sea
655   ice parameterisation schemes, found in the present study, suggest that future generation Arctic reanalysis projects can strongly benefit from applying more advanced approaches. For example, having a reanalysis system based on a fully-coupled NWP system would open a possibility of representing sea ice cover as a drift medium with a much higher level of detail. However, practical considerations might not allow applying a fully coupled atmospheric model with a strongly coupled data assimilation system in a reanalysis project due to great computational and development costs of such a solution. Thus, constraining the state
660   of a simplified sea ice model by means of external data sets or data assimilation (see, e.g., Batrak, 2021; Scott et al., 2014) may be still a viable approach to representing sea ice state in future atmospheric reanalyses.

Data availability. The CARRA and ERA5 reanalysis products are available through the Copernicus Climate Data Store portal (https://cds.climate.copernicus.eu/). The Level-2 sea ice surface temperature products from the MODIS instrument (MOD29 and MYD29) are available from the National Snow and Ice Data Center (https://nsidc.org/). The combined CryoSat-2 and SMOS weekly ice thickness product is available from the Alfred Wegener Institute (https://awi.de/). Monthly snow depth retrievals of Lee et al. (2021) are publicly available from the authors of the original publication (https://doi.org/10.5281/zenodo.5081765). The Observation Ice Bridge snow depths data are available from the National Snow and Ice Data Center (https://doi.org/10.5067/G519SHCKWQV6 and https://doi.org/10.5067/GRIXZ91DE0L9). CRREL ice mass balance buoy data are available from the CRREL-Dartmouth Mass Balance Buoy Program portal (https://imb-crrel-dartmouth.org). Data from the SIMBA buoys deployed during the N-ICE2015 campaign are available from the Norwegian Polar Institute (https://doi.org/10.21334/npolar.2015.6ed9a8ca). Data access links for the MOSAiC SIMBA buoys are provided in Table S6. The CLARA-A2 albedo product is available from the EUMETSAT dissemination portal (http://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002_01).

Author contributions. This study was devised by YB and BC. YB performed initial data retrieval and processing of CARRA and ERA5, validated ice surface temperature, ice thickness and snow depth fields against satellite retrievals, contributed to the analysis of albedo intercomparisons and to validation of CARRA and ERA5 against ice mass balance buoys. BC validated the reanalysis products against ice mass balance buoys. VKM validated sea ice albedo fields in the reanalysis products against a satellite retrieval. YB wrote the paper with contributions from BC and VKM. All authors contributed to the analysis and interpretation of the obtained results, the paper preparation and editing.

Competing interests. At least one of the (co-)authors is a member of the editorial board of The Cryosphere.

Acknowledgements. This work has been funded by the Copernicus Climate Change Service. ECMWF implements this Service on behalf of the European Commission. BC was supported by the European Commission, Horizon 2020 (PolarRES, grant no. 101003590)

References

Aagaard, K. and Coachman, L. K.: The East Greenland Current North of Denmark Strait: Part I, Arctic, 21, 181–200, https://doi.org/10.14430/arctic2927, 1968.

Anttila, K., Manninen, T., Jääskeläinen, E., and Riihelä, A.: Validation Report (VAL) CLARA-A2 Surface Albedo Product, Tech. rep., EU-METSAT Satellite Application Facility on Climate Monitoring, https://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V001, 2016.

Arduini, G., Keeley, S., Day, J. J., Sandu, I., Zampieri, L., and Balsamo, G.: On the Importance of Representing Snow Over Sea-Ice for Simulating the Arctic Boundary Layer, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 777, https://doi.org/10.1029/2021MS002777, 2022.

Batrak, Y.: Implementation of an Adaptive Bias-Aware Extended Kalman Filter for Sea-Ice Data Assimilation in the HARMONIE-AROME Numerical Weather Prediction System, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 533, https://doi.org/10.1029/2021MS002533, 2021.

Batrak, Y. and Müller, M.: On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice, Nature Communications, 10, 4170, https://doi.org/10.1038/s41467-019-11975-3, 2019.

Batrak, Y., Kourzeneva, E., and Homleid, M.: Implementation of a simple thermodynamic sea ice scheme, SICE version 1.0-38h1, within the ALADIN–HIRLAM numerical weather prediction system version 38h1, Geoscientific Model Development, 11, 3347–3368, https://doi.org/10.5194/gmd-11-3347-2018, 2018.

Bengtsson, L. and Shukla, J.: Integration of Space and In Situ Observations to Study Global Climate Change, Bulletin of the American Meteorological Society, 69, 1130–1143, https://doi.org/10.1175/1520-0477(1988)069<1130:IOSAIS>2.0.CO;2, 1988.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K.-I., Lenderink, G., Niemelä, S., Nielsen, K. P., Onvlee, J., Rontu, L., Samuelsson, P., Muñoz, D. S., Subias, A., Tijm, S., Toll, V., Yang, X., and Køltzow, M. Ø.: The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System, Monthly Weather Review, 145, 1919–1935, https://doi.org/10.1175/MWR-D-16-0417.1, 2017.

Boone, A.: Modeling hydrological processes in the land surface scheme ISBA: inclusion of a hydrological reservoir, ice and a snow model, Ph.D. thesis, Paul Sabatier University, 2000.

Boone, A. and Etchevers, P.: An Intercomparison of Three Snow Schemes of Varying Complexity Coupled to the Same Land Surface Model: Local-Scale Evaluation at an Alpine Site, Journal of Hydrometeorology, 2, 374–394, https://doi.org/10.1175/1525-7541(2001)002<0374:AIOTSS>2.0.CO;2, 2001.

Chen, S., Liu, J., Ding, Y., Zhang, Y., Cheng, X., and Hu, Y.: Assessment of Snow Depth over Arctic Sea Ice in CMIP6 Models Using Satellite Data, Advances in Atmospheric Sciences, 38, 168–186, https://doi.org/10.1007/s00376-020-0213-5, 2021.

Cheng, B., Launiainen, J., and Vihma, T.: Modelling of Superimposed Ice Formation and Sub-Surface Melting in the Baltic Sea, Geophysica, 39, 31–50, 2003.

Cheng, B., Zhang, Z., Vihma, T., Johansson, M., Bian, L., Li, Z., and Wu, H.: Model experiments on snow and ice thermodynamics in the Arctic Ocean with CHINARE 2003 data, Journal of Geophysical Research: Oceans, 113, https://doi.org/10.1029/2007JC004654, 2008.

Cheng, B., Vihma, T., Palo, T., Nicolaus, M., Gerland, S., Rontu, L., Haapala, J., and Perovich, D.: Observation and modelling of snow and sea ice mass balance and its sensitivity to atmospheric forcing during spring and summer 2007 in the Central Arctic, Advances in Polar Science, 32, 312–326, https://doi.org/10.13679/j.advps.2021.0047, 2021.

Chung, C. E., Cha, H., Vihma, T., Räisänen, P., and Decremer, D.: On the possibilities to use atmospheric reanalyses to evaluate the warming structure in the Arctic, Atmospheric Chemistry and Physics, 13, 11 209–11 219, https://doi.org/10.5194/acp-13-11209-2013, 2013.

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., and Jones, J.: Recent Arctic amplification and extreme mid-latitude weather, Nature Geoscience, 7, 627–637, https://doi.org/10.1038/ngeo2234, 2014.

Comiso, J. C. and Hall, D. K.: Climate trends in the Arctic as observed from space, WIREs Climate Change, 5, 389–409, https://doi.org/10.1002/wcc.277, 2014.

Curry, J. A., Schramm, J. L., and Ebert, E. E.: Sea Ice-Albedo Climate Feedback Mechanism, Journal of Climate, 8, 240 – 247, https://doi.org/10.1175/1520-0442(1995)008<0240:SIACFM>2.0.CO;2, 1995.

Douville, H., Royer, J.-F., and Mahfouf, J.-F.: A new snow parameterization for the Météo-France climate model, Climate Dynamics, 12, 21–35, https://doi.org/10.1007/BF00208760, 1995.

Dunbar, M.: Ice Regime and Ice Transport in Nares Strait, Arctic, 26, 282–291, https://doi.org/10.14430/arctic2927, 1973.

Ebert, E. E. and Curry, J. A.: An intermediate one-dimensional thermodynamic sea ice model for investigating ice-atmosphere interactions, Journal of Geophysical Research: Oceans, 98, 10 085–10 109, https://doi.org/10.1029/93JC00656, 1993.

ECMWF: IFS Documentation CY41R2 – Part IV: Physical Processes, 4, ECMWF, https://doi.org/10.21957/tr5rv27xu, 2016.

Frank, C. W., Pospichal, B., Wahl, S., Keller, J. D., Hense, A., and Crewell, S.: The added value of high resolution regional reanalyses for wind power applications, Renewable Energy, 148, 1094–1109, https://doi.org/10.1016/j.renene.2019.09.138, 2020.

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., Davis, S., Gerber, E. P., Harvey, V. L., Hegglin, M. I., Homeyer, C. R., Knox, J. A., Krüger, K., Lambert, A., Long, C. S., Martineau, P., Molod, A., Monge-Sanz, B. M., Santee, M. L., Tegtmeier, S., Chabrillat, S., Tan, D. G. H., Jackson, D. R., Polavarapu, S., Compo, G. P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J. S., and Zou, C.-Z.: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems, Atmospheric Chemistry and Physics, 17, 1417–1452, https://doi.org/10.5194/acp-17-1417-2017, 2017.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Graham, R. M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R. G., Rinke, A., Walden, V. P., Granskog, M. A., and Hudson, S. R.: Evaluation of Six Atmospheric Reanalyses over Arctic Sea Ice from Winter to Early Summer, Journal of Climate, 32, 4121–4143, https://doi.org/10.1175/JCLI-D-18-0643.1, 2019.

Greenwald, M. and Khanna, S.: Space-Efficient Online Computation of Quantile Summaries, SIGMOD Rec., 30, 58–66, https://doi.org/10.1145/376284.375670, 2001.

Hall, D., Key, J., Casey, K., Riggs, G., and Cavalieri, D.: Sea ice surface temperature product from MODIS, IEEE Transactions on Geoscience and Remote Sensing, 42, 1076–1087, https://doi.org/10.1109/TGRS.2004.825587, 2004.

Hall, D. K. and Riggs, G.: MODIS/Terra Sea Ice Extent 5-Min L2 Swath 1km, Version 6. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/MODIS/MOD29.006, 2015a.

Hall, D. K. and Riggs, G.: MODIS/Aqua Sea Ice Extent 5-Min L2 Swath 1km, Version 6. Boulder, Colorado USA. NASA National Snow
755   and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/MODIS/MYD29.006, 2015b.

Hansen, E., Gerland, S., Granskog, M. A., Pavlova, O., Renner, A. H. H., Haapala, J., Løyning, T. B., and Tschudi, M.: Thinning of Arctic sea
       ice observed in Fram Strait: 1990–2011, Journal of Geophysical Research: Oceans, 118, 5202–5221, https://doi.org/10.1002/jgrc.20393,
       2013.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Sim-
760   mons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren,
       P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J.,
       Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Vil-
       laume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049,
       https://doi.org/10.1002/qj.3803, 2020.

765 IHO (International Hydrographic Organization): Limits of Oceans and Seas, 3rd edition. Special Publication No. 23 (S-23), Tech. rep.,
       International Hydrographic Organization, Monaco, 1953.

Isaksen, K., Nordli, Ø., Ivanov, B., Køltzow, M. A. Ø., Aaboe, S., Gjelten, H. M., Mezghani, A., Eastwood, S., Førland, E., Benestad, R. E.,
       Hanssen-Bauer, I., Brækkan, R., Sviashchennikov, P., Demin, V., Revina, A., and Karandasheva, T.: Exceptional warming over the Barents
       area, Scientific Reports, 12, 9371, https://doi.org/10.1038/s41598-022-13568-5, 2022.

770 Jackson, K., Wilkinson, J., Maksym, T., Meldrum, D., Beckers, J., Haas, C., and Mackenzie, D.: A Novel and Low-Cost Sea Ice Mass
       Balance Buoy, Journal of Atmospheric and Oceanic Technology, 30, 2676–2688, https://doi.org/10.1175/JTECH-D-13-00058.1, 2013.

Kaiser-Weiss, A. K., Borsche, M., Niermann, D., Kaspar, F., Lussana, C., Isotta, F. A., van den Besselaar, E., van der Schrier, G., and
       Undén, P.: Added value of regional reanalyses for climatological applications, Environmental Research Communications, 1, 071 004,
       https://doi.org/10.1088/2515-7620/ab2ec3, 2019.

775 Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP-DOE AMIP-II reanalysis (R-2),
       Bull. Amer. Math. Soc., 83, 1631–1644, https://doi.org/10.1175/BAMS-83-11-1631, 2002.

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Fokke Meirink, J., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E.,
       Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N., and Hollmann, R.: CLARA-A2: the
       second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data, Atmospheric Chemistry and Physics,
780   17, 5809–5828, https://doi.org/10.5194/acp-17-5809-2017, 2017.

Koo, Y., Lei, R., Cheng, Y., Cheng, B., Xie, H., Hoppmann, M., Kurtz, N. T., Ackley, S. F., and Mestas-Nuñez, A. M.: Estimation of
       thermodynamic and dynamic contributions to sea ice growth in the Central Arctic using ICESat-2 and MOSAiC SIMBA buoy data,
       Remote Sensing of Environment, 267, 112 730, https://doi.org/10.1016/j.rse.2021.112730, 2021.

Kurtz, N., Studinger, M., Harbeck, J., Onana, V., and Yi, D.: IceBridge L4 Sea Ice Freeboard, Snow Depth, and Thickness, Version 1,
785   https://doi.org/10.5067/G519SHCKWQV6, 2015.

Kurtz, N., Studinger, M., Harbeck, J., Onana, V., and Yi, D.: IceBridge Sea Ice Freeboard, Snow Depth, and Thickness Quick Look, Version
       1, https://doi.org/10.5067/GRIXZ91DE0L9, 2016.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G.,
       Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and
790   ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, https://doi.org/10.5194/tc-13-49-2019, 2019.

Lawrence, H., Bormann, N., Sandu, I., Day, J., Farnan, J., and Bauer, P.: Use and impact of Arctic observations in the ECMWF Numerical Weather Prediction system, Quarterly Journal of the Royal Meteorological Society, 145, 3432–3454, https://doi.org/10.1002/qj.3628, 2019.

Lee, S.-M., Shi, H., Sohn, B.-J., Gasiewski, A. J., Meier, W. N., and Dybkjær, G.: Winter Snow Depth on Arctic Sea Ice From Satellite Radiometer Measurements (2003–2020): Regional Patterns and Trends, Geophysical Research Letters, 48, e2021GL094 541, https://doi.org/10.1029/2021GL094541, 2021.

Lei, R., Cheng, B., Heil, P., Vihma, T., Wang, J., Ji, Q., and Zhang, Z.: Seasonal and Interannual Variations of Sea Ice Mass Balance From the Central Arctic to the Greenland Sea, Journal of Geophysical Research: Oceans, 123, 2422–2439, https://doi.org/10.1002/2017JC013548, 2018.

Lei, R., Cheng, B., Hoppmann, M., Zhang, F., Zuo, G., Hutchings, J. K., Lin, L., Lan, M., Wang, H., Regnery, J., Krumpen, T., Haapala, J., Rabe, B., Perovich, D. K., and Nicolaus, M.: Seasonality and timing of sea ice mass balance and heat fluxes in the Arctic transpolar drift during 2019–2020, Elementa: Science of the Anthropocene, 10, https://doi.org/10.1525/elementa.2021.000089, 2022.

Liao, Z., Cheng, B., Zhao, J., Vihma, T., Jackson, K., Yang, Q., Yang, Y., Zhang, L., Li, Z., Qiu, Y., and Cheng, X.: Snow depth and ice thickness derived from SIMBA ice mass balance buoy data using an automated algorithm, International Journal of Digital Earth, 12, 962–979, https://doi.org/10.1080/17538947.2018.1545877, 2018.

Lindsay, R. W., Zhang, J., Schweiger, A., Steele, M., and Stern, H.: Arctic Sea Ice Retreat in 2007 Follows Thinning Trend, Journal of Climate, 22, 165–176, https://doi.org/10.1175/2008JCLI2521.1, 2009.

Liu, J., Zhang, Z., Inoue, J., and Horton, R. M.: Evaluation of snow/ice albedo parameterizations and their impacts on sea ice simulations, International Journal of Climatology, 27, 81–91, https://doi.org/10.1002/joc.1373, 2007.

Nielsen-Englyst, P., Høyer, J. L., Kolbe, W. M., Dybkjær, G., Lavergne, T., Tonboe, R. T., Skarpalezos, S., and Karagali, I.: A combined sea and sea-ice surface temperature climate dataset of the Arctic, 1982–2021, Remote Sensing of Environment, 284, 113 331, https://doi.org/10.1016/j.rse.2022.113331, 2023.

Parkinson, C. L., Ward, A., and King, M. D., eds.: Earth science reference handbook: a guide to NASA's Earth science program and Earth observing satellite missions, Washington, D.C.: National Aeronautics and Space Administration, 2006.

Perovich, D. and Richter-Menge, J.: Regional variability in sea ice melt in a changing Arctic, Philosophical Transactions of the Royal Society A-Mathematical Physical Engineering Sciences, 373, https://doi.org/10.1098/rsta.2014.0165, 2015.

Pistone, K., Eisenman, I., and Ramanathan, V.: Observational determination of albedo decrease caused by vanishing Arctic sea ice, Proceedings of the National Academy of Sciences, 111, 3322–3326, https://doi.org/10.1073/pnas.131820111, 2014.

Pohl, C., Istomina, L., Tietsche, S., Jäkel, E., Stapf, J., Spreen, G., and Heygster, G.: Broadband albedo of Arctic sea ice from MERIS optical data, The Cryosphere, 14, 165–182, https://doi.org/10.5194/tc-14-165-2020, 2020.

Provost, C., Sennéchael, N., Miguet, J., Itkin, P., Rösel, A., Koenig, Z., Villacieros-Robineau, N., and Granskog, M. A.: Observations of flooding and snow-ice formation in a thinner Arctic sea-ice regime during the N-ICE2015 campaign: Influence of basal ice melt and storms, Journal of Geophysical Research: Oceans, 122, 7115–7134, https://doi.org/10.1002/2016JC012011, 2017.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.: The Arctic has warmed nearly four times faster than the globe since 1979, Communications Earth & Environment, 3, 168, https://doi.org/10.1038/s43247-022-00498-3, 2022.

Renner, A. H. H., Gerland, S., Haas, C., Spreen, G., Beckers, J. F., Hansen, E., Nicolaus, M., and Goodwin, H.: Evidence of Arctic sea ice thinning from direct observations, Geophysical Research Letters, 41, 5029–5036, https://doi.org/10.1002/2014GL060369, 2014.

Richter-Menge, J., Perovich, D., Elder, B. C., Claffey, K., Rigor, I., and Ortmeyer, M.: Ice mass-balance buoys: a tool
830   for measuring and attributing changes in the thickness of the Arctic sea-ice cover, Annals of Glaciology, 44, 205–210,
       https://doi.org/10.3189/172756406781811727, 2006.

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., and Haas, C.: A weekly Arctic sea-ice thickness data record from merged
       CryoSat-2 and SMOS satellite data, The Cryosphere, 11, 1607–1623, https://doi.org/10.5194/tc-11-1607-2017, 2017.

Riihelä, A., Bright, R. M., and Anttila, K.: Recent strengthening of snow and ice albedo feedback driven by Antarctic sea-ice loss, Nature
835   Geoscience, 14, 832–836, https://doi.org/10.1038/s41561-021-00841-x, 2021.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine,
       R., Gayno, G., Wang, J., Hou, Y.-T., ya Chuang, H., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P. V., Keyser,
       D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue,
       Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y.,
840   Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, Bulletin of the American
       Meteorological Society, 91, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1, 2010.

Schmith, T. and Hansen, C.: Fram Strait Ice Export during the Nineteenth and Twentieth Centuries Reconstructed from
       a Multiyear Sea Ice Index from Southwestern Greenland, Journal of Climate, 16, 2782–2791, https://doi.org/10.1175/1520-
       0442(2003)016<2782:FSIEDT>2.0.CO;2, 2003.

845   Scott, K. A., Buehner, M., and Carrieres, T.: An Assessment of Sea-Ice Thickness Along the Labrador Coast From AMSR-E
       and MODIS Data for Operational Data Assimilation, IEEE Transactions on Geoscience and Remote Sensing, 52, 2726–2737,
       https://doi.org/10.1109/TGRS.2013.2265091, 2014.

Screen, J. A. and Simmonds, I.: Increasing fall-winter energy loss from the Arctic Ocean and its role in Arctic temperature amplification,
       Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL044136, 2010.

850   Serreze, M. C. and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Global and Planetary Change, 77,
       85–96, https://doi.org/10.1016/j.gloplacha.2011.03.004, 2011.

Serreze, M. C., Clark, M. P., and Bromwich, D. H.: Monitoring Precipitation over the Arctic Terrestrial Drainage System: Data Requirements,
       Shortcomings, and Applications of Atmospheric Reanalysis, Journal of Hydrometeorology, 4, 387–407, https://doi.org/10.1175/1525-
       7541(2003)4<387:MPOTAT>2.0.CO;2, 2003.

855   Shi, H., Sohn, B.-J., Dybkjær, G., Tonboe, R. T., and Lee, S.-M.: Simultaneous estimation of wintertime sea ice thickness and snow depth
       from space-borne freeboard measurements, The Cryosphere, 14, 3761–3783, https://doi.org/10.5194/tc-14-3761-2020, 2020.

Tang, C. C., Ross, C. K., Yao, T., Petrie, B., DeTracey, B. M., and Dunlap, E.: The circulation, water masses and sea-ice of Baffin Bay,
       Progress in Oceanography, 63, 183–228, https://doi.org/10.1016/j.pocean.2004.09.005, 2004.

Tonboe, R. T., Eastwood, S., Lavergne, T., Sørensen, A. M., Rathmann, N., Dybkjær, G., Pedersen, L. T., Høyer, J. L., and Kern, S.: The
860   EUMETSAT sea ice concentration climate data record, The Cryosphere, 10, 2275–2290, https://doi.org/10.5194/tc-10-2275-2016, 2016.

Toudal Pedersen, L., Dybkjær, G., Eastwood, S., Heygster, G., Ivanova, N., Kern, S., Lavergne, T., Saldo, R., Sandven, S.,
       Sørensen, A., and Tonboe, R.: ESA Sea Ice Climate Change Initiative (Sea_Ice_cci): Sea Ice Concentration Climate Data
       Record from the AMSR-E and AMSR-2 instruments at 25km grid spacing, version 2.1. Centre for Environmental Data Analysis,
       https://doi.org/10.5285/f17f146a31b14dfd960cde0874236ee5, 2017.

865   Vinje, T. and Kvambekk, Å. S.: Barents Sea drift ice characteristics, Polar Research, 10, 59–68, https://doi.org/10.3402/polar.v10i1.6728,
       1991.

Wang, C., Graham, R. M., Wang, K., Gerland, S., and Granskog, M. A.: Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution, The Cryosphere, 13, 1661–1679, https://doi.org/10.5194/tc-13-1661-2019, 2019.

870 Webster, M. A., Rigor, I. G., Nghiem, S. V., Kurtz, N. T., Farrell, S. L., Perovich, D. K., and Sturm, M.: Interdecadal changes in snow depth on Arctic sea ice, Journal of Geophysical Research: Oceans, 119, 5395–5406, https://doi.org/10.1002/2014JC009985, 2014.

Yang, X., Nielsen, K. P., Amstrup, B., Peralta, C., Høyer, J., Nielsen Englyst, P., Schyberg, H., Homleid, M., Køltzow, M. A. Ø., Randriamampianina, R., Dahlgren, P., Støylen, E., Valkonen, T., Palmason, B., Thorsteinsson, S., Bojarova, J., Körnich, H., Lindskog, M., Box, J., and Mankoff, K.: C3S Arctic regional reanalysis – Full system documentation, Tech. rep., Danish Meteorological Institute, 2020.

875 Zhou, L., Stroeve, J., Xu, S., Petty, A., Tilling, R., Winstrup, M., Rostosky, P., Lawrence, I. R., Liston, G. E., Ridout, A., Tsamados, M., and Nandan, V.: Inter-comparison of snow depth over Arctic sea ice from reanalysis reconstructions and satellite retrieval, The Cryosphere, 15, 345–367, https://doi.org/10.5194/tc-15-345-2021, 2021.
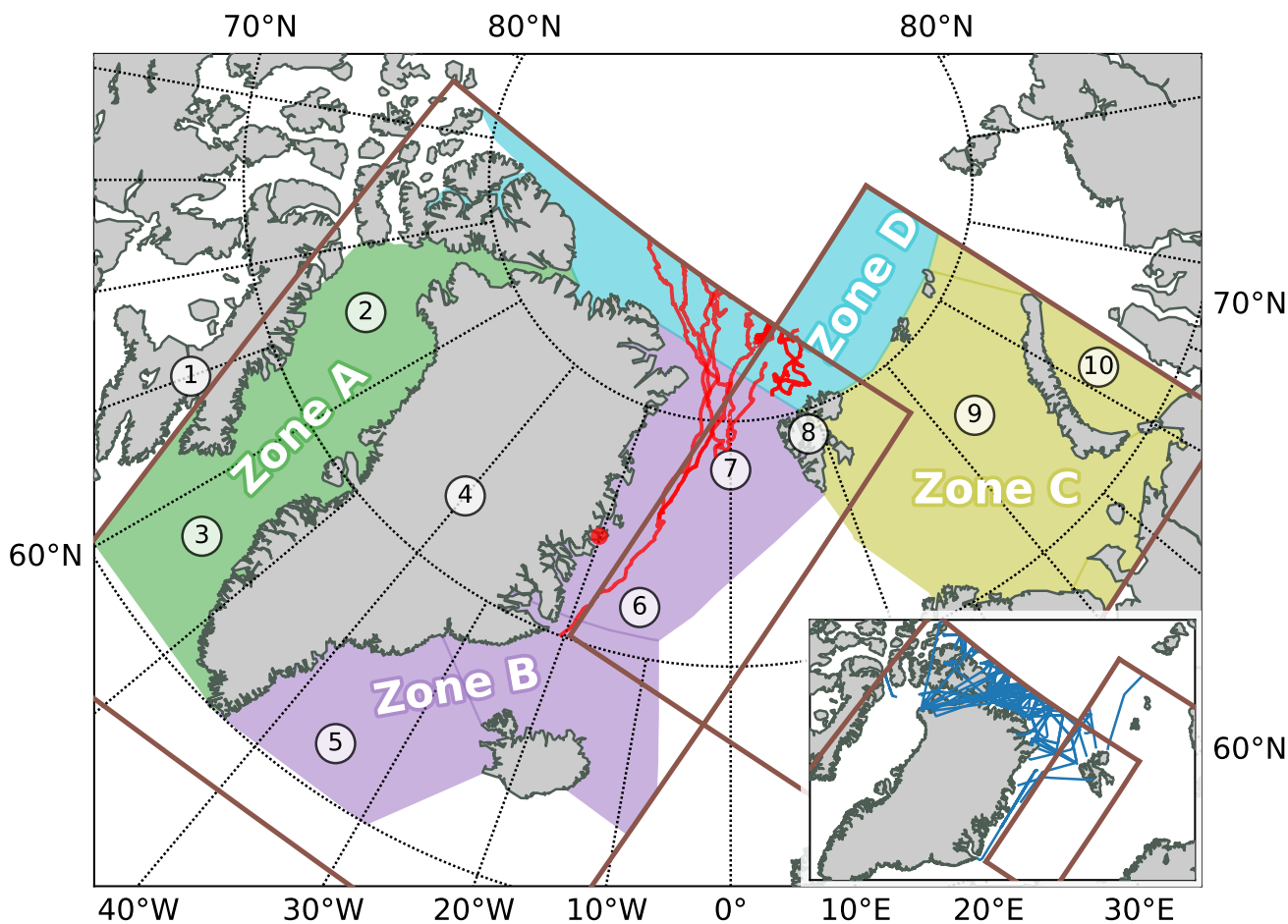
Figure 1. Overview of the CARRA model domains and locations of the areas of interest discussed in the present study. Also on the figure, drift trajectories of individual ice mass balance buoys (IMB) are shown (position of the IMB deployed on the land fast ice is marked with a dot). Marked on the map: 1 – Baffin Island; 2 – Baffin Bay; 3 – Davis Strait; 4 – Greenland; 5 – North Atlantic Ocean adjacent to the Greenland coast; 6 – Greenland Sea; 7 – Fram Strait; 8 – Svalbard archipelago; 9 – Barents Sea; 10 – Kara Sea. Inset: tracks of the Operation Ice Bridge flights.
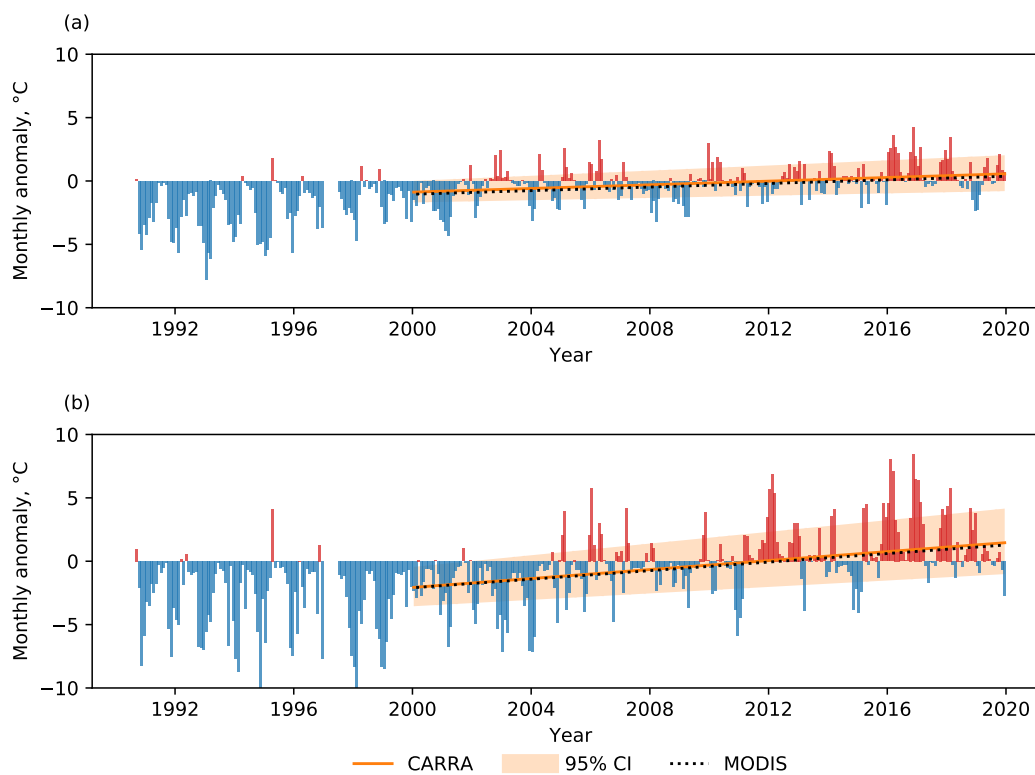
Figure 2. Monthly ice surface temperature anomalies in the CARRA product and fitted ice temperature anomaly trend. Multiyear monthly means computed over the time period from 2000 to 2020 are used as reference data. (a) Western CARRA model domain; (b) eastern CARRA model domain. Also on the panels, the 95% confidence interval of the CARRA anomaly trend, and the anomaly trend of the MODIS ice surface temperature product are shown.
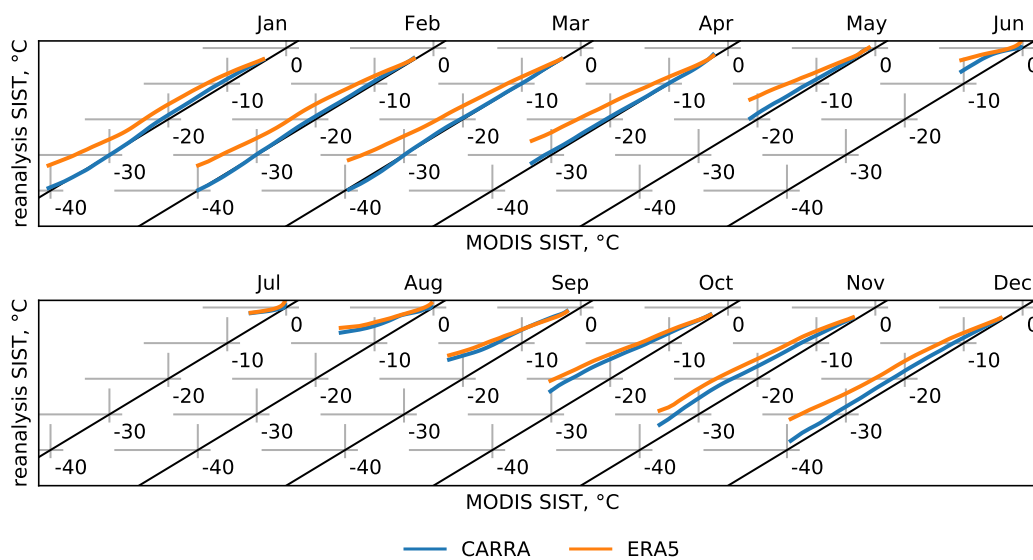
Figure 3. Estimated monthly quantiles of the ice surface temperature, $q \in [0.01; 0.99]$, in atmospheric reanalysis products compared against the MODIS product over the eastern CARRA domain. Quantiles are computed for the period from 2000 to 2020. Diagonals represent reference 1:1 match lines between observational and reanalyis quantiles. Numerical values of the computed quantiles can be found in Table S2.
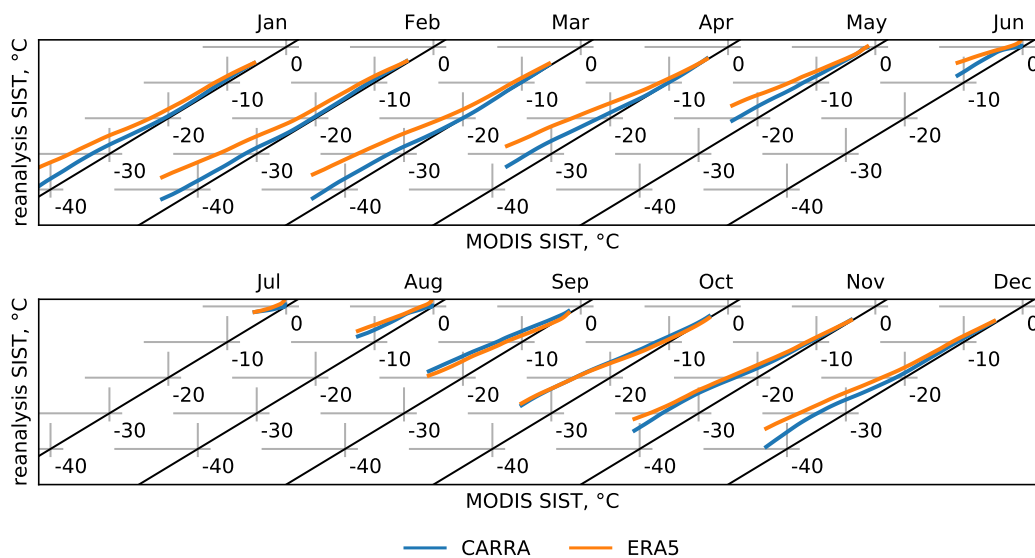
Figure 4. Same as Fig. 3 but for the western CARRA domain. Numerical values of the computed quantiles can be found in Table S1.
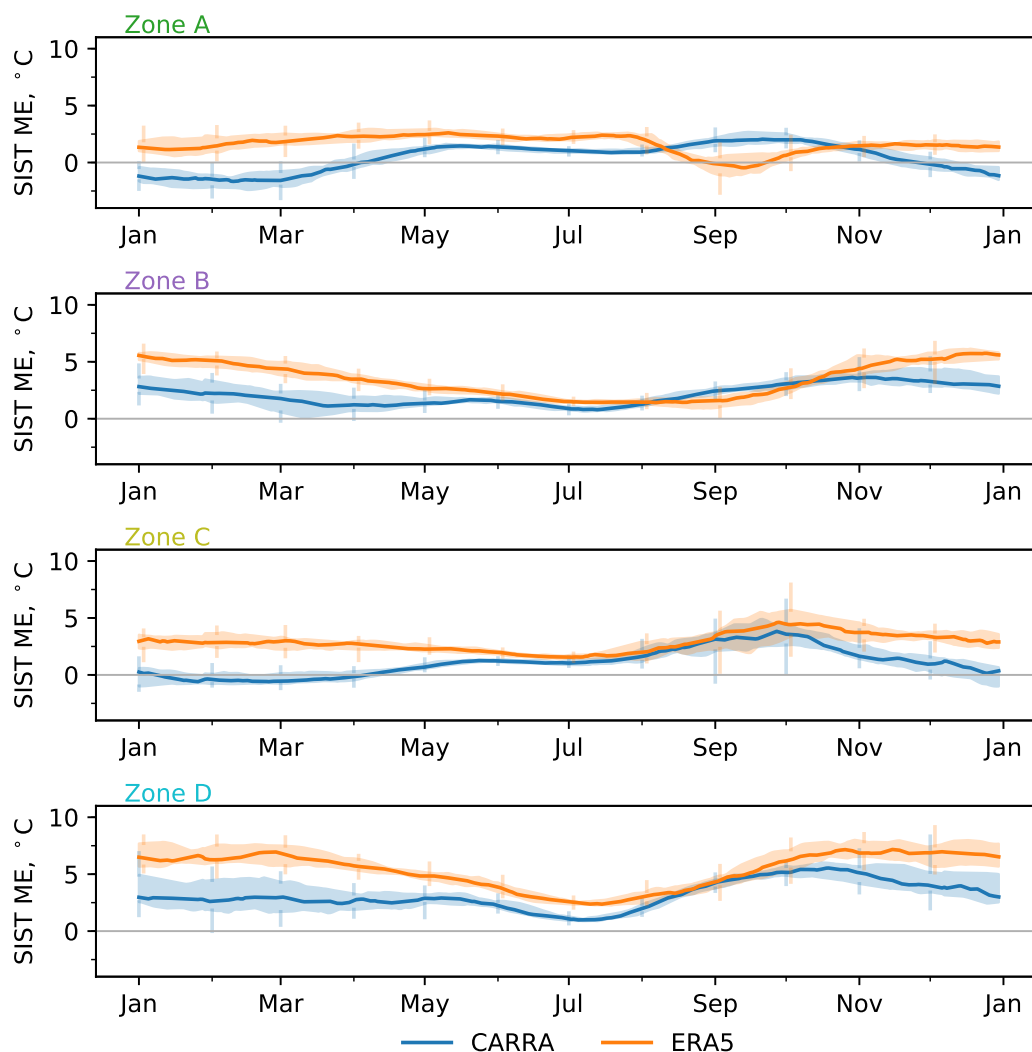
Figure 5. Average annual cycle of the mean error (ME) of sea ice surface temperature in CARRA and ERA5 computed over the period from 2000 to 2020 for the four areas of interest. On the panels, centre lines show the median ME value, shading bands show the corresponding interquartile ranges, and spikes show the $q \in [0.05; 0.95]$ quantile range. Values are obtained by aggregating four-week series of mean error computed against the MODIS product for each individual year in the reanalysis data sets. Numerical monthly values of the computed scores can be found in Table S3.
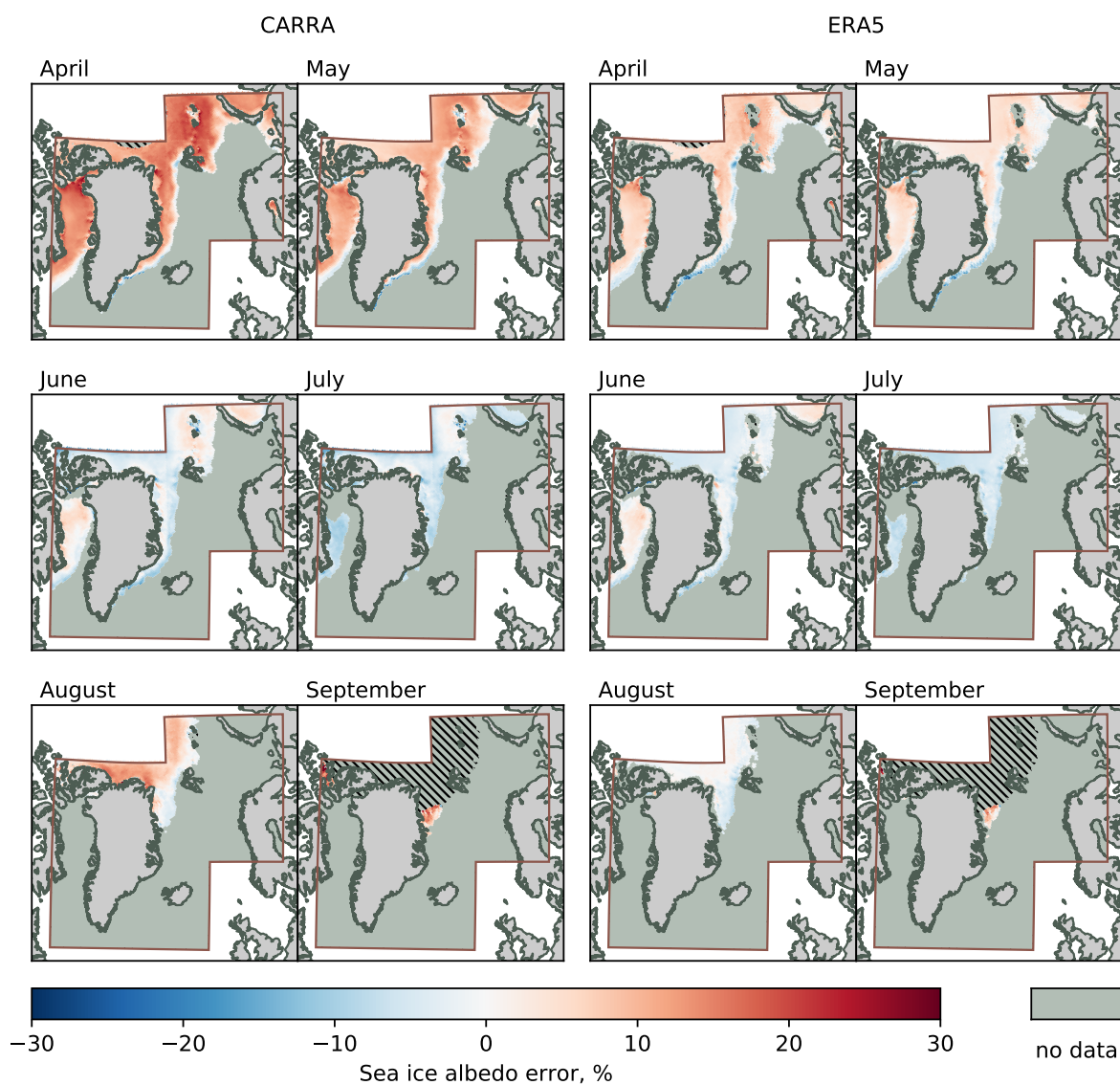
Figure 6. Monthly mean errors of the modelled surface albedo over sea-ice covered regions in CARRA and ERA5, computed against the CLARA-A2 SAL product over the time period from 2000 to 2015. Note that in September the observational product has considerably reduced coverage in the northern-most parts of the CARRA model domains due to challenging light conditions. Areas with missing SAL data are marked with hatches.
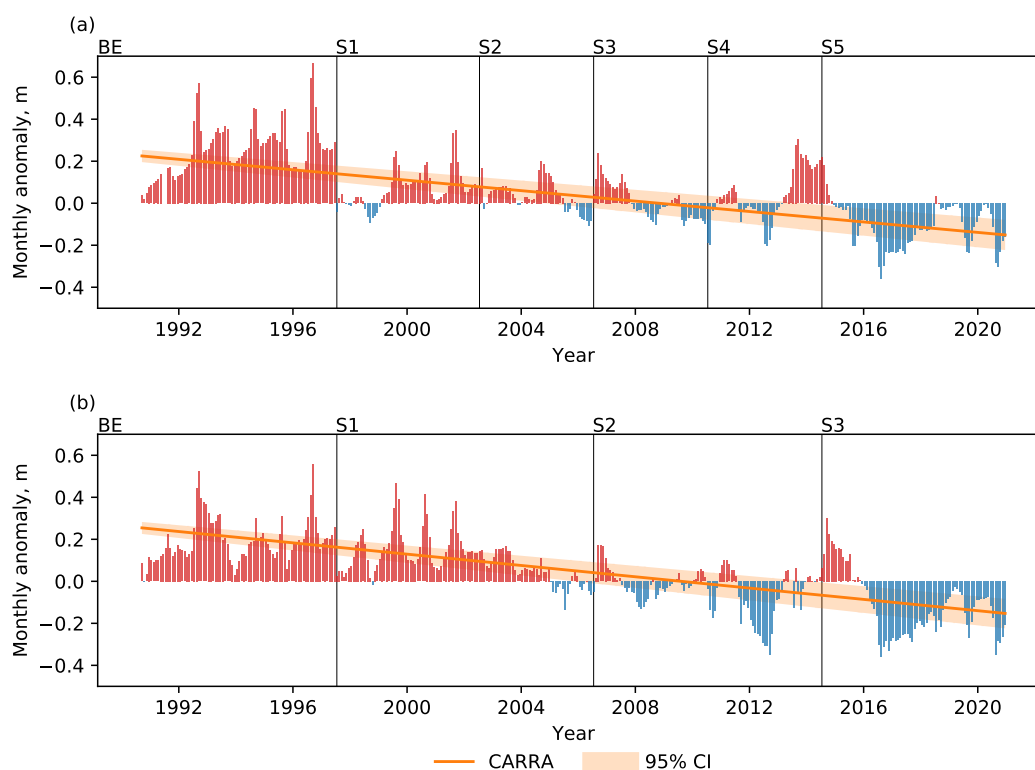
Figure 7. Monthly ice thickness anomalies in the CARRA product and fitted ice thickness anomaly trend. Multiyear monthly means for the time period from 2000 to 2020 are used as a reference when computing anomalies. (a) Western CARRA model domain; (b) eastern CARRA model domain. Also on the panels, the 95% confidence interval of the CARRA anomaly trend is shown, and separate productions streams S1–S5 of the CARRA system and the back extension stream BE are marked.
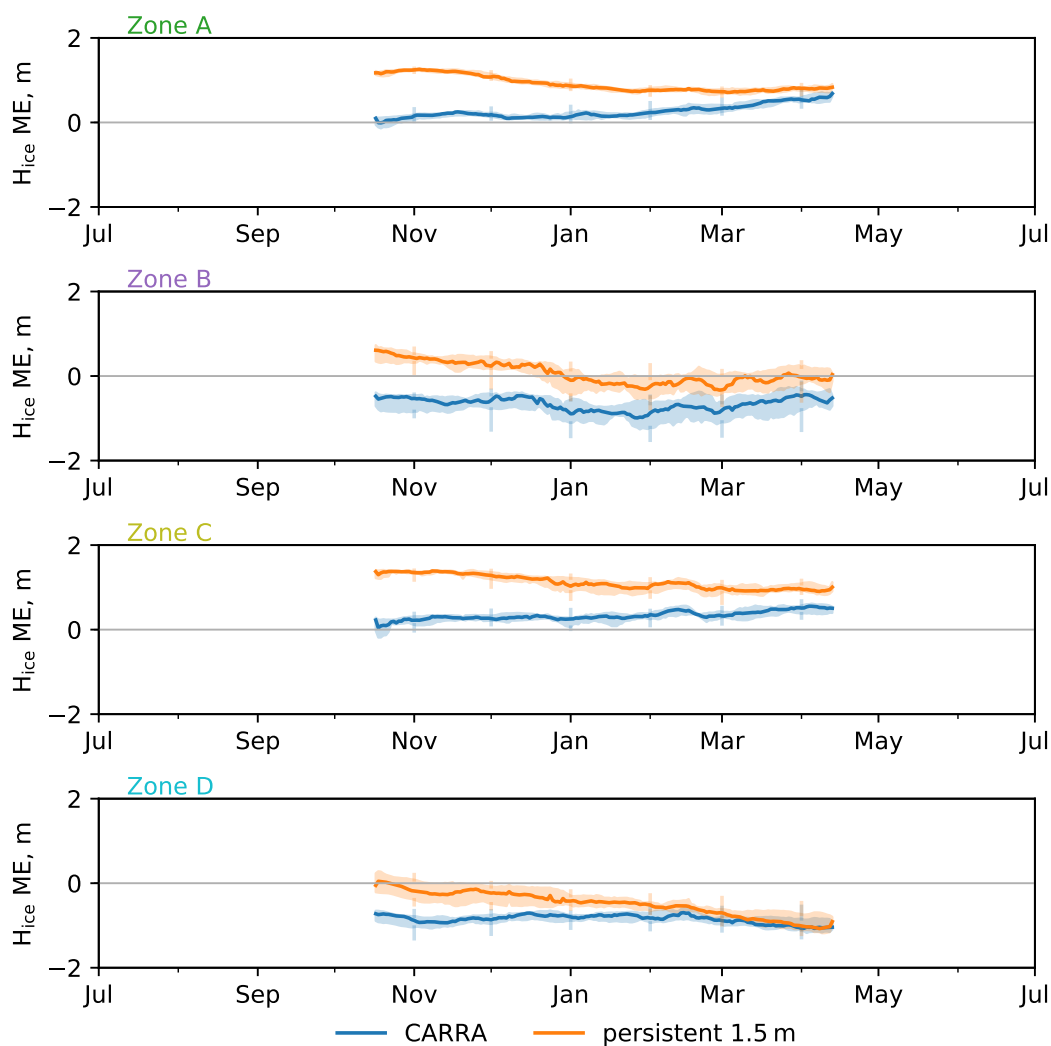
Figure 8. Average annual cycle of the mean error (ME) of sea ice thickness in CARRA over the period from 2010 to 2020 for the four areas of interest. On the panels, centre lines show the median ME value, shading bands show the corresponding interquartile ranges, and spikes show the $q \in [0.05; 0.95]$ quantile range. Series of mean error are computed against the combined CryoSat-2 SMOS ice thickness product. Also on the figure, ice thickness errors obtained using constant and uniform ice thickness of 1.5 m as in the ERA5 product are shown. Numerical monthly values of the computed scores can be found in Table S4.
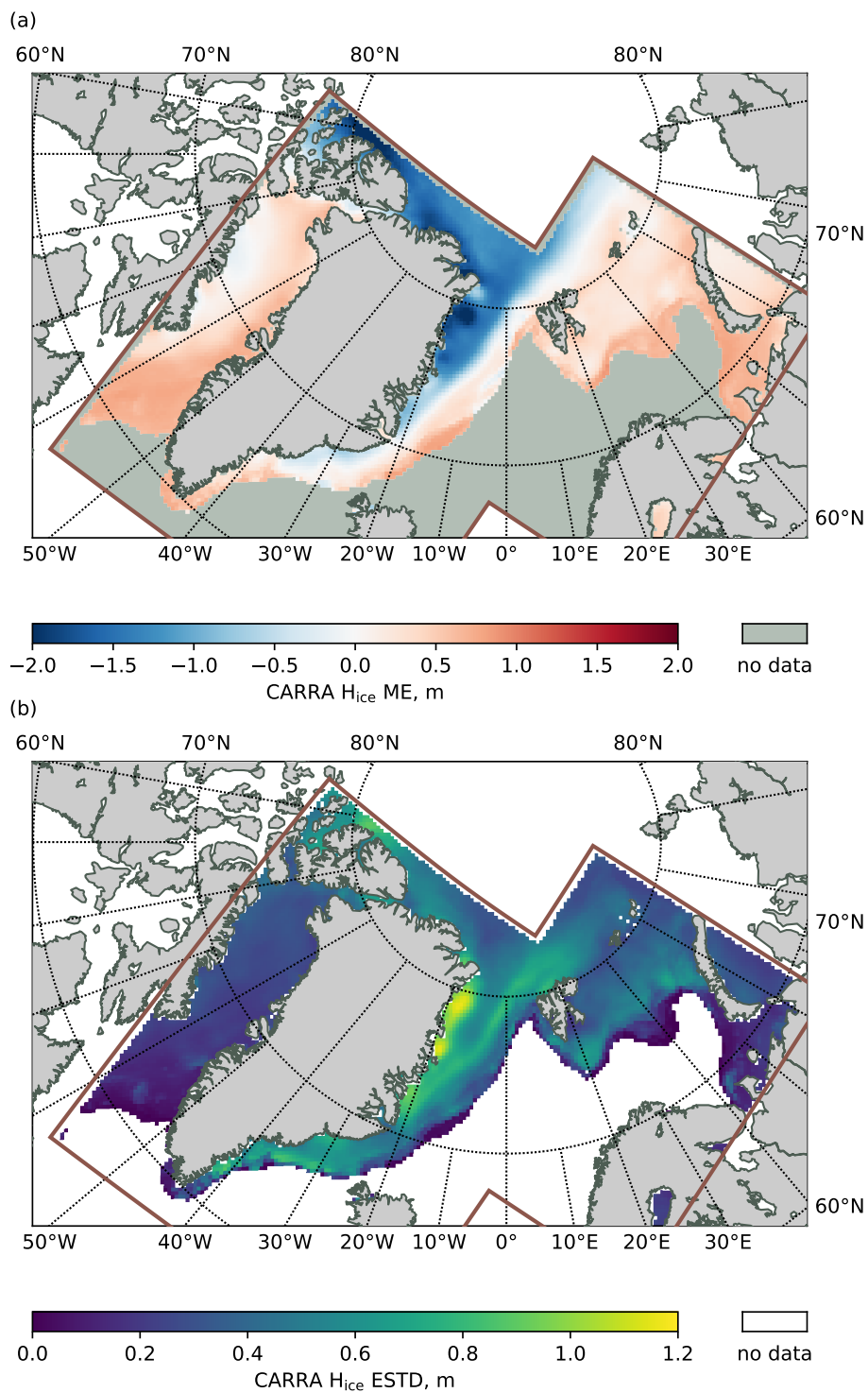
Figure 9. Mean error (ME) and standard deviation of errors (ESTD) of sea ice thickness in the CARRA product computed against the combined CryoSat-2 SMOS satellite ice thickness product.
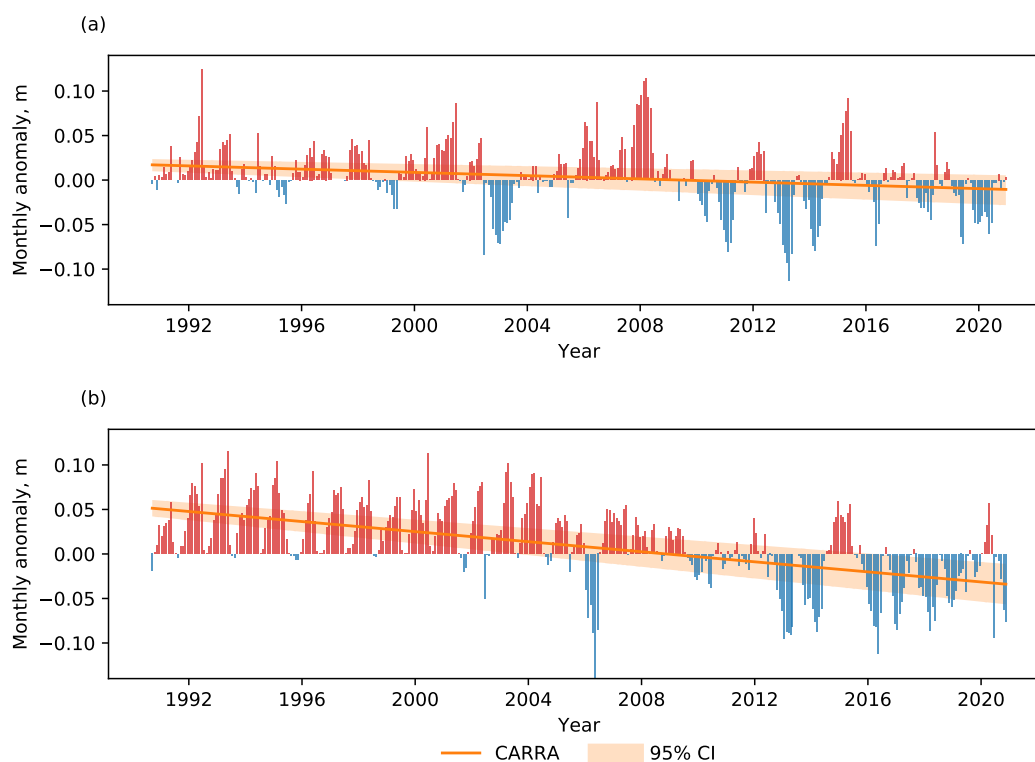
Figure 10. Monthly snow depth anomalies in the CARRA product and fitted snow depth anomaly trend. Multiyear monthly means for the time period from 2000 to 2020 are used as a reference when computing anomalies. (a) Western CARRA model domain; (b) eastern CARRA model domain. Also on the panels, the 95% confidence interval of the CARRA anomaly trend is shown.
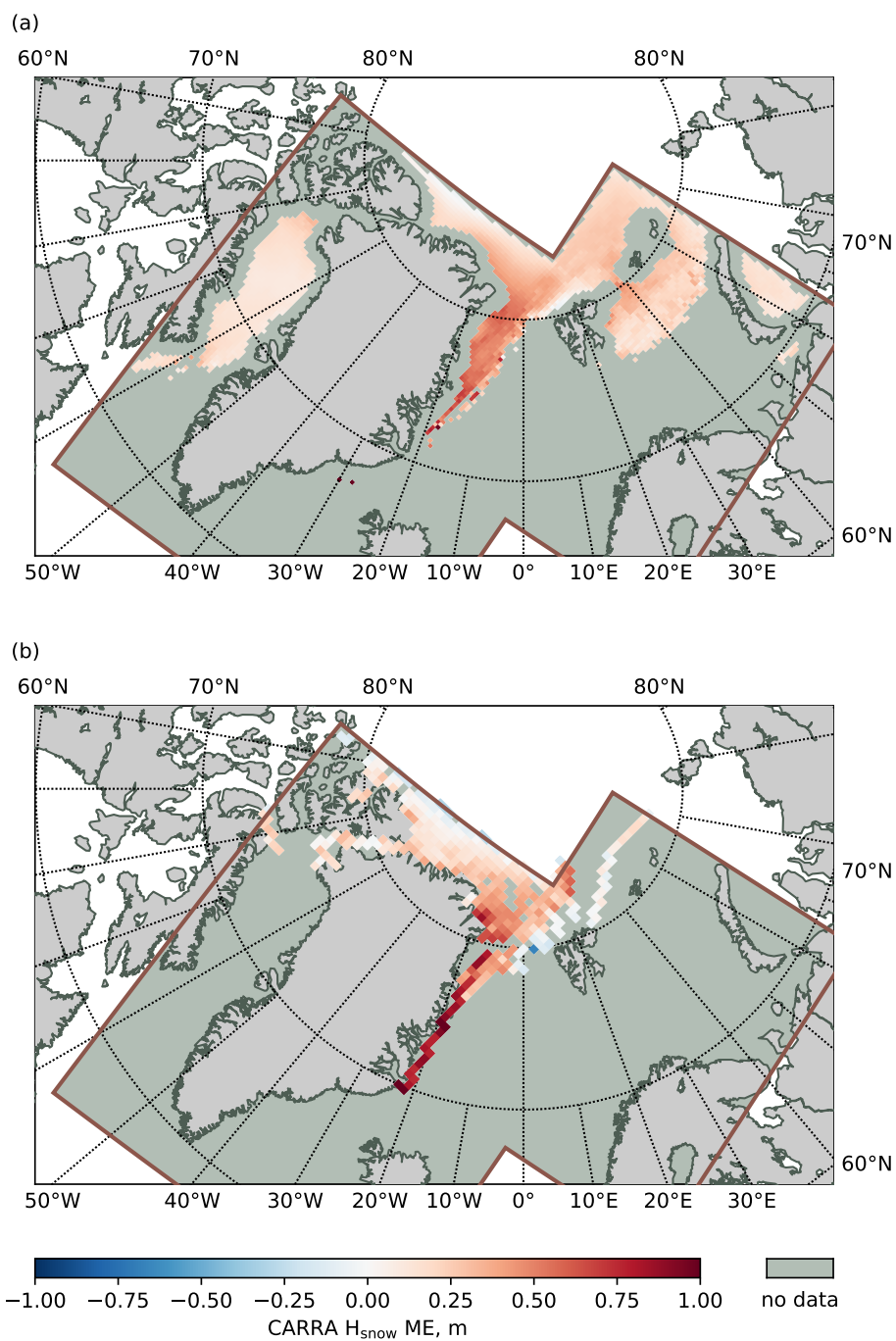
Figure 11. Mean error (ME) of snow depth over sea ice in the CARRA product computed against the satellite ice thickness retrieval product and Operation IceBridge flight campaign data. (a) January-March ME computed against the satellite product over the time period from 2003 to 2020; (b) March-April computed against the Operation IceBridge data over the time period from 2009 to 2019 and presented on a 50 km grid.
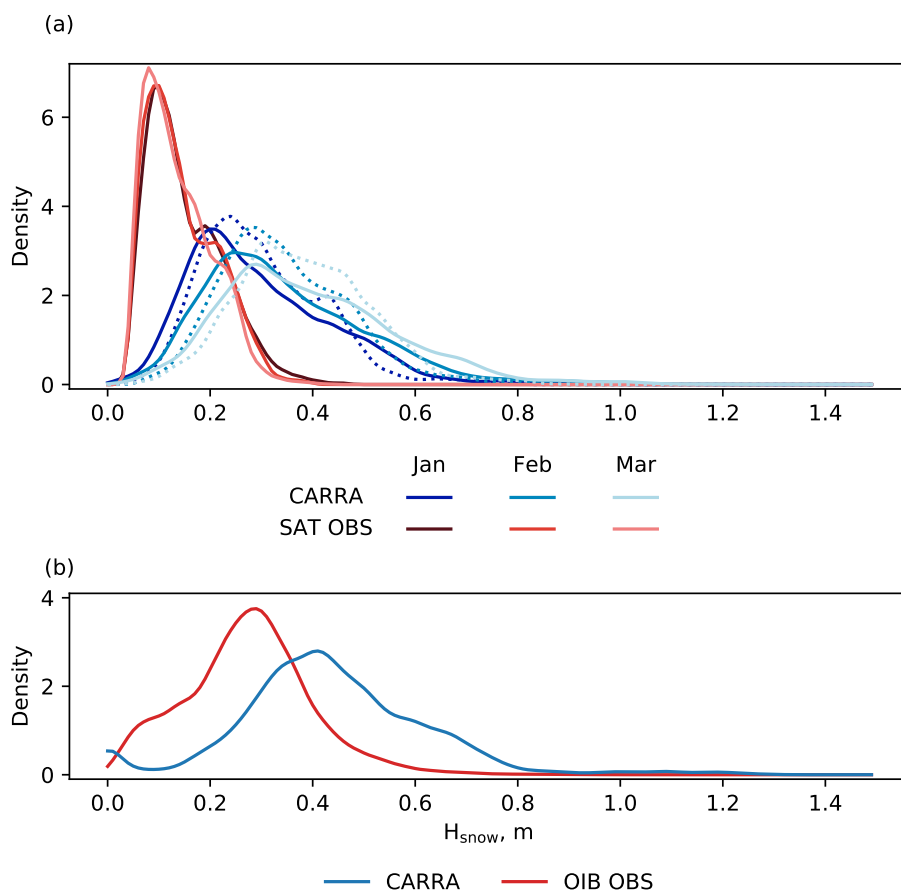
Figure 12. Estimated probability density functions (PDF) of snow depth in the CARRA product and in the observational products. (a) monthly PDFs of snow depth in CARRA and in the satellite snow depth retrieval; (b) PDFs of snow depth in CARRA and in the Operation IceBridge flight campaign data. Also on (a), the PDFs of corrected snow depth in CARRA are shown with dotted lines.
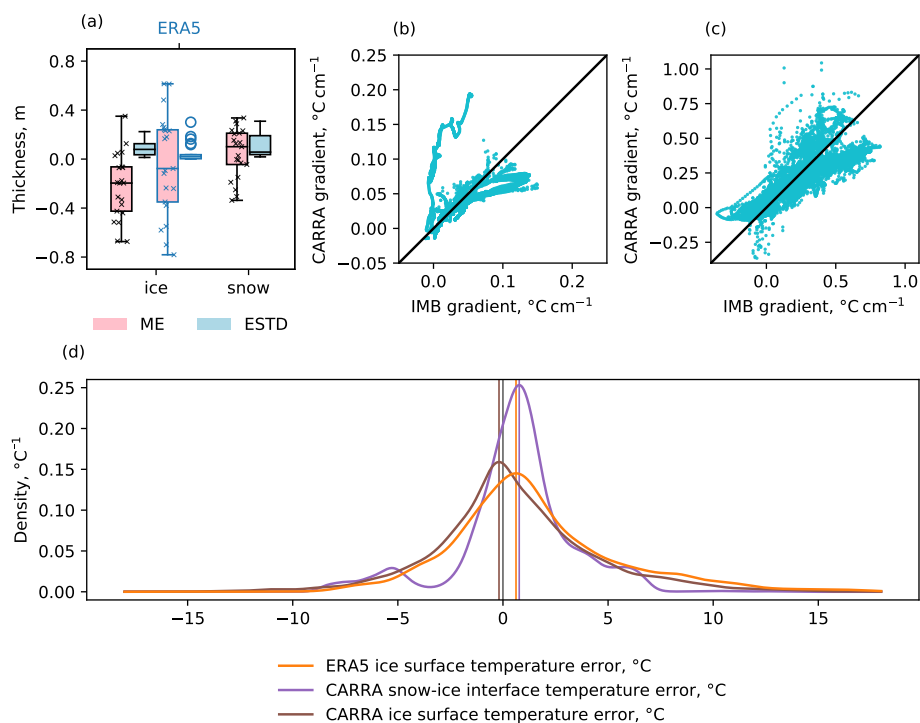
Figure 13. Representation of sea ice in the CARRA and ERA5 reanalysis products as compared against buoy observations. (a) Box plots of the per-buoy mean error (ME) and standard deviation of errors (ESTD) of snow depth and ice thickness; (b) temperature gradient within the ice layer computed from buoy data and computed from the CARRA product; (c) temperature gradient within the snow layer computed from buoy data and computed from the CARRA product; (d) estimated probability density functions of ice surface temperature error and snow-ice interface temperature error in CARRA, and ice surface temperature error in ERA5. Also on (d), major modes of the modelling errors in CARRA and ERA5 are marked with vertical bars of corresponding colours.