Sensitivity of short-range forecasts to sea ice thickness data assimilation parameters in a coupled ice-ocean syste

Carmen Nab[1,2], Davi Mignac[2], Jack Landy[3], Matthew Martin[2], Julienne Stroeve[1,4,5], and Michel Tsamados[1]

[1]Centre for Polar Observation and Modelling, University College London
[2]Ocean Forecasting Research & Development
[3]Department of Physics and Technology, UiT The Arctic University of Norway
[4]Centre for Earth Observation Science, University of Manitoba
[5]National Snow and Ice Data Center, University of Colorado Boulder

February 28, 2024

Abstract

Sea ice thickness (SIT) estimates derived from CryoSat-2 radar freeboard measurements are assimilated into the Met Office's global ocean–sea ice forecasting system, FOAM. We test the sensitivity of short-range forecasts to the snow depth, radar freeboard product and assumed radar penetration through the snowpack in the freeboard-to-thickness conversion. We find that modifying the snow depth has the biggest impact on the modelled SIT, changing it by up to 0.88 m (48%), compared to 0.65 m (33%) when modifying the assumed radar penetration through the snowpack and 0.55 m (30%) when modifying the freeboard product. We find a doubling in the thermodynamic volume change over the winter season when assimilating SIT data, with the largest changes seen in the congelation ice growth. Next, we determine that the method used to calculate the observation uncertainties of the assimilated data products can change the mean daily model SIT by up to 36%. Compared to measurements collected at upward-looking sonar moorings and during the Operation IceBridge campaign, we find an improvement in the SIT forecasts' variability representation when assuming partial radar penetration through the snowpack and when improving the method used to calculate the CryoSat-2 observation uncertainties. This paper highlights a concern for future SIT data assimilation and forecasting, with the chosen parameterisation of the freeboard-to-thickness conversion having a substantial impact on model results.

Sensitivity of short-range forecasts to sea ice thickness data assimilation parameters in a coupled ice-ocean system

Carmen Nab[1,2], Davi Mignac[2], Jack Landy[3], Matthew Martin[2], Julienne Stroeve[1,4,5], Michel Tsamados[1]

[1]Centre for Polar Observation and Modelling, University College London, London, UK
[2]Ocean Forecasting Research & Development, Met Office, Exeter, UK
[3]Department of Physics and Technology, UiT The Arctic University of Norway, Norway
[4]Centre for Earth Observation Science, University of Manitoba, Canada
[5]National Snow and Ice Data Center, University of Colorado Boulder, USA

Key Points:

- We test the sensitivity of short-range forecasts to the parameters used in the CryoSat-2 radar freeboard-to-sea ice thickness conversion
- The snow depth, radar penetration and retracker used change the thickness by up to 0.88 m (48%), 0.65 m (33%) and 0.55 m (30%), respectively
- Changing the method used to characterize the observation uncertainties can change the mean daily model sea ice thickness by up to 36%

Corresponding author: Carmen Nab, `carmen.nab.18@ucl.ac.uk`

Abstract

Sea ice thickness (SIT) estimates derived from CryoSat-2 radar freeboard measurements are assimilated into the Met Office's global ocean–sea ice forecasting system, FOAM. We test the sensitivity of short-range forecasts to the snow depth, radar freeboard product and assumed radar penetration through the snowpack in the freeboard-to-thickness conversion. We find that modifying the snow depth has the biggest impact on the modelled SIT, changing it by up to 0.88 m (48%), compared to 0.65 m (33%) when modifying the assumed radar penetration through the snowpack and 0.55 m (30%) when modifying the freeboard product. We find a doubling in the thermodynamic volume change over the winter season when assimilating SIT data, with the largest changes seen in the congelation ice growth. Next, we determine that the method used to calculate the observation uncertainties of the assimilated data products can change the mean daily model SIT by up to 36%. Compared to measurements collected at upward-looking sonar moorings and during the Operation IceBridge campaign, we find an improvement in the SIT forecasts' variability representation when assuming partial radar penetration through the snowpack and when improving the method used to calculate the CryoSat-2 observation uncertainties. This paper highlights a concern for future SIT data assimilation and forecasting, with the chosen parameterisation of the freeboard-to-thickness conversion having a substantial impact on model results.

Plain Language Summary

Satellite altimeters can be used to estimate sea ice thickness, by estimating how far the sea ice sticks out above the surrounding waterline. This is done by measuring the time taken for radar waves to reach the sea ice and ocean surfaces and return to the altimeter. These radar waves are processed using a retracking algorithm, to calculate the radar freeboard. Several assumptions are used to convert this radar freeboard into sea ice thickness, including values for snow depth and the ability of the altimeter's radar waves to penetrate through the snow overlying the sea ice. We test the sensitivity of modelled sea ice thickness to the retracking algorithm used, the snow depth and assumed radar penetration. We do this by assimilating our sea ice thickness estimates, converted from radar freeboard using varying values for these parameters, into the Met Office's global ocean–sea ice forecasting system, FOAM. We then determine the sensitivity of modelled sea ice thickness to the observation uncertainties assigned to the assimilated data. We find that snow depth has the biggest influence on modelled sea ice thickness, followed by the uncertainty calculation, assumed radar penetration and the retracking algorithm.

1 Introduction

Accurate monitoring and modelling of Arctic sea ice thickness (SIT) is integral to determining its implications on regional and global climates, safe travel for Arctic coastal communities, shipping routes, the marine ecosystem and wildlife dependent on the ice for hunting and traveling. While the observation and modelling of sea ice concentration is well developed, the same is not true for SIT. This is because estimating SIT from satellite data requires knowledge about the hydrostatic properties of the sea ice and its overlying snow cover. SIT can be estimated using satellite radar altimeters by measuring the time taken for the radar pulse to travel to the sea ice surface and return to the altimeter. The same can then be done for the sea surface found in open water areas between the ice floes, with the difference between these two measurements referred to as the *radar freeboard*. By factoring in assumptions on the ability of the altimeter's radar waves to penetrate the snowpack, this can then be converted into the *sea ice freeboard* (the height of the ice above the surrounding sea water). The radar range to the assumed ice surface is calculated through the retracking of the radar waveform to obtain the presumed range to a single scattering surface, with the choice of retracking algorithm affecting the radar freeboard retrieved. Currently, two types of retracking algorithm are used in SIT products. The most common is the threshold algorithm, which applies a fixed percentage threshold to the waveform's first maximum power return (e.g. Guerreiro et al., 2017; Laxon et al., 2013). Alternatively, a physical algorithm can be used, which varies the percentage threshold used according to the physical properties of the sea ice (e.g. Landy et al., 2020). Using assumptions on snow depth, as well as sea ice and snow density, SIT can then be calculated assuming hydrostatic equilibrium.

The method for converting radar freeboard estimates to SIT requires an assumed value for the fractional depth of the snowpack where the retracker detects the backscattered radar echo, α (as per Nab et al., 2023). An assumption of $\alpha =$ 1 means that the radar goes entirely through the snowpack, such that the pulse path length through snow has no impact on the radar freeboard estimate. $\alpha < 1$ represents a height for the mean radar scattering intensity within the snowpack (or at its surface if $\alpha = 0$). As α is proportional to SIT (Kwok & Cunningham, 2015), a reduction in the former will reduce the latter. All current SIT estimates from Ku-band altimeters rely on the assumption of full radar penetration of the snowpack, since it is not yet possible to measure α directly from space. However, studies have shown that this assumption may not be the case in reality, as snow properties affect the ability of radar waves to penetrate the snowpack. During Arctic field campaigns in 2006 and 2008, Willatt et al. (2010) found that the proportion of radar returns appearing closer to the air-snow interface than the snow-ice interface increased with temperature. Similarly, Nandan et al. (2017) predicted an upward shift in the dominant scattering surface of CryoSat-2 waves with increased snow salinity over first-year ice using a radiative transfer model. Nandan et al. (2023) found a sensitivity of Ku-band radar waves to previous air-snow interfaces, buried within the snowpack after new snowfall. On satellite footprint-scales, Nab et al. (2023) showed synoptic timescale correlations between radar freeboard estimates and snow accumulation, revealing radar echoes returning from within the snowpack. Over Arctic sea ice, previous studies have calculated α to be between 0.40 - 0.96 for Ku-band radar waves (Armitage & Ridout, 2015; Kilic et al., 2019; Shen et al., 2020; Nab et al., 2023).

The fast-changing Arctic climate means that accurate short- and long-term predictions of its sea ice cover are becoming increasingly important. Data assimilation of sea ice variables can be used to improve model estimates of sea ice concentration, extent and thickness (e.g. Fritzner et al., 2019; Y.-F. Zhang et al., 2018; Williams et al., 2023; Y. Zhang et al., 2023; Gregory et al., 2023). Studies have shown that SIT has a longer memory than sea ice concentration, such that the change in the initial model state caused by assimilating the former will persist for longer than by assimilating the latter (Guemas et al., 2016). When assimilating CryoSat-2-derived SIT into the Met Office's coupled sea ice-ocean system (FOAM), Fiedler et al. (2022) found an improvement in FOAM SIT results of 0.61 m mean difference (0.42 m root mean square difference) relative to a control experiment without SIT assimilation, when both model runs were compared to Operation IceBridge (OIB) estimates. When compared to buoy SIT data in the Beaufort region, no improvement was found in the modelled SIT when assimilating CryoSat-2 SIT data. When adding the assimilation of data from the Soil Moisture and Ocean Salinity (SMOS) satellite radiometer over thin ice to this setup, Mignac et al. (2022) found a reduction in SIT over first-year ice, which better matched the OIB and buoy data than when only assimilating CryoSat-2-derived SIT or not assimilating any thickness information.

Uncertainties in the snow depth, radar freeboard product and assumed radar penetration all contribute to the overall error in satellite-derived SIT. To account for these errors, data assimilation methods use observational error estimates to determine how much weight observational data should carry in the assimilation process, with observations deemed to have a lower error having a larger influence on the assimilation results. The accuracy of the resulting modelled SIT estimates is thus expected to improve with improved error estimates for the data assimilated. The lack of error estimates provided with currently operational CryoSat-2-derived radar freeboard products means that the observation uncertainty calculation in CryoSat-2 freeboard and thickness assimilation research has been relatively simplified so far. For example, Fritzner et al. (2019) and Chen et al. (2017) used fixed uncertainties of 0.5 m and 1.5 m, respectively, for all CryoSat-2 observations when assimilating CryoSat-2-derived SIT. Fiedler et al. (2022) and Mignac et al. (2022) used a parameterisation method, assigning high uncertainty values (0.5 - 8 m) to SIT values below 1.5 m and above 4 m, with SIT values between 1.5 - 4 m assigned uncertainty values < 0.5 m.

We test the sensitivity of 1-day sea-ice forecasts in FOAM to the parameters used in the freeboard-to-thickness conversion, focusing on the snow depth, radar freeboard product and assumed radar penetration (α). This allows us to determine how much of the sensitivity in the radar freeboard-to-thickness conversion is carried through into the model forecasts. Previous studies have shown the ability of CryoSat-2-derived SIT assimilation to improve modelled SIT (e.g. Fritzner et al., 2019; Fiedler et al., 2022; Mignac et al., 2022; Sievers et al., 2023) but, to our knowledge, there have been no studies on the sensitivity of these results to the parameterisation of the freeboard-to-thickness conversion. We also test the sensitivity of the 1-day SIT forecasts to the SIT observation uncertainties used in the data assimilation. We do this by comparing data assimilation results obtained using a simple parameterisation scheme to derive SIT observation uncertainties, as described in Fiedler et al. (2022) and Mignac et al. (2022), to one using Gaussian error propagation to derive SIT uncertainties from each individual radar freeboard measurement, as described in Ricker et al. (2014).

2 Methods

2.1 UK Met Office Forecast Ocean Assimilation Model (FOAM)

FOAM is the UK Met Office's global, coupled ocean-sea ice forecasting system. It is used to produce daily analyses and short-range forecasts of ocean temperature, salinity, velocities and various sea ice parameters. FOAM's operational ocean-sea ice analysis is also used by the Met Office's Global Seasonal (GloSea) coupled ensemble prediction system to initialise its ocean and sea ice components daily. FOAM is forced at the surface using output from the Met Office Numerical Weather Prediction (NWP) system. The ocean model component of FOAM, NEMO (Nucleus for European Modelling of the Ocean; Madec et al., 2023) is coupled to the recently developed SI^3 (Sea Ice modelling Integrated Initiative; Vancoppenolle et al., 2023). SI^3 merges the capabilities of three sea ice models formerly coupled to NEMO: CICE, GELATO and LIM (Madec et al., 2023). The SI^3 configuration includes five thickness categories (plus open water), multi-layer thermodynamics and prognostic melt ponds. We run the FOAM system with a 1/4 degree tripolar grid (ORCA025) for both sea ice and ocean components, with 75 vertical levels in the latter (Storkey et al., 2018).

Data assimilation is a procedure for producing a complete estimate of the current state of the ocean and sea ice by combining information from observations with the model. A previous model forecast is compared with newly acquired observations and the two sets of data, together with information about their respective errors, are combined to produce a new, more accurate model state from which to launch a forecast. The observation uncertainties are important in this procedure since erroneous observations with low prescribed uncertainties would unduly influence the analysis. FOAM is set up to assimilate in situ and satellite-derived observational data using the three-dimensional variation assimilation scheme NEMOVAR (Waters et al., 2015). A 24-hour assimilation window is used to assimilate observations of sea-surface temperature (SST), sea-level anomaly, sea ice concentration, and temperature and salinity profiles. The in situ SST data consists of data from buoys (drifting and moored) and ships. This is supplemented by satellite-derived SSTs from NOAA's Advanced Very High Resolution Radiometer (AVHRR), the Advanced Microwave Scanning Radiometer 2 (AMSR2) and the Sea and Land Surface Temperature Radiometer (SLSTR), as well as data from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor data from the Suomi-NPP (National Polar-orbiting Partnership) satellite. The sea-level anomaly data are in the form of along-track satellite data from the Jason-2, Jason-3, Sentinel-3A and -3B, CryoSat-2 and AltiKa satellites. Sea ice concentration data are provided by the Special Sensor Microwave Imager/Sounder (SSMI/S) instruments from the Defense Meteorological Satellites Program (DMSP) satellites. These are processed by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea-Ice Satellite Application Facility (OSI-SAF). The temperature and salinity profiles are taken from moored arrays, gliders, Argo floats and research Conductivity, Temperature and Depth (CTD) instruments. Additional temperature profiles from Expendable Bathythermographs (XBTs) and marine mammal sensors are also used.

NEMOVAR has previously been used to assimilate SIT observations from CryoSat-2 and SMOS into FOAM (Fiedler et al., 2022; Mignac et al., 2022), although this is currently not done operationally. SMOS is an L-band microwave radiometer, able to estimate SIT by evaluation of the ice's surface brightness temperatures (Tbs) using a radiative transfer model (Tian-Kunze et al., 2014). Over thin ice (<1 m), its relative uncertainties are believed to be significantly lower than altimeter-based techniques. Over thicker ice, the Tbs become insensitive to ice thickness. While SMOS has a much higher temporal resolution than altimeters, providing daily pan-Arctic coverage (of the thinner ice), it has coarser spatial resolution, with a larger effective footprint of $\sim$ 40 km (Kaleschke et al., 2012). In this study, SIT derived from both SMOS and CryoSat-2 is assimilated, alongside the operationally assimilated variables mentioned previously. The SMOS data is only assimilated where SIT $\leq$ 1 m (Figure 1). SMOS SIT uncertainties are provided with the data and used in the assimilation of this dataset. Although SMOS SITs are assimilated in all experiments performed here, our sensitivity study is purely focused on the CryoSat-2 data assimilation: Only the CryoSat-2 freeboard-to-thickness conversion parameters and uncertainty calculation are changed in the experiments performed in this study. See Mignac et al. (2022) for a full description of the methodology for assimilating CryoSat-2 and SMOS SIT estimates into FOAM.

2.2 Calculation of sea ice thickness with varying parameters

To derive SIT from CryoSat-2 radar freeboard estimates (F_r), we first calculate sea ice freeboard (F_i) as follows:

$$F_i = F_r + (\alpha\frac{c}{c_s} - 1)h_s \tag{1}$$

where α is the radar penetration, c is the speed of light in a vacuum, c_s is the speed of light in snow and h_s is the snow depth.

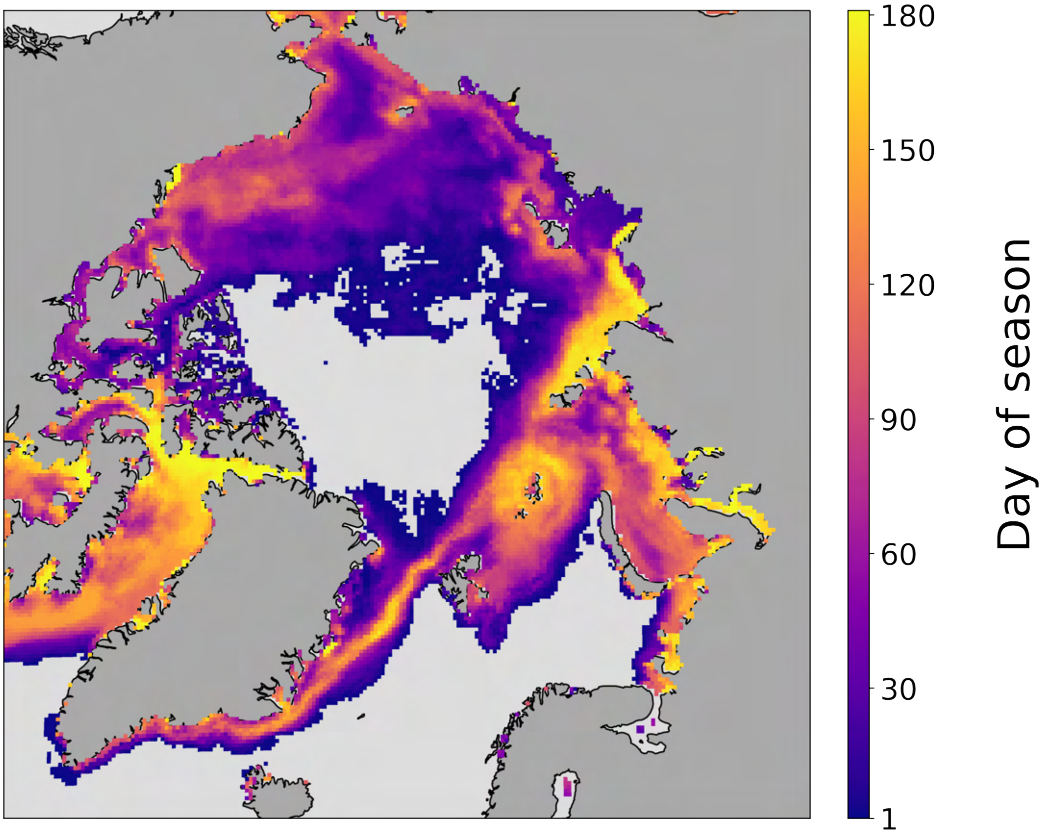

Figure 1. Day of 2016-2017 winter season until which SMOS data is assimilated into FOAM (Day 1 = 17 October 2016). Areas where SMOS is never assimilated are shown in grey.

We then use this to calculate SIT:

$$SIT = \frac{F_i \rho_w + h_s \rho_s}{\rho_w - \rho_i} \tag{2}$$

where ρ_w is the density of seawater, taken from the model density field at the surface, ρ_i is the bulk density of sea ice and ρ_s is the snow density, evolving temporally following Eq. 11 of Mallett et al. (2020):

$$\rho_s = 6.5t + 274.51 \tag{3}$$

where t is the number of months since October.

2.3 Description of experiments

We start with a control experiment (CTRL), running FOAM with the assimilation of its operationally assimilated ocean and sea-ice observations (see subsection 2.1), which does not include SIT. We then replicate the setup of Mignac et al. (2022), who assimilated CryoSat-2 and SMOS SIT into FOAM (using CICE as the sea ice component). Since the publication of that paper, the sea ice component of FOAM has been replaced with SI^3 (see Blockley et al., 2023). We assimilate CryoSat-2 and SMOS into FOAM following the methods of Mignac et al. (2022), using SI^3 as the sea ice model. We use the AWI-derived CryoSat-2 radar freeboard data for this, as this is the CryoSat-2-derived product the Met Office is currently testing for operational assimilation. Mignac et al. (2022) assumed a pan-Arctic fixed sea ice density (916.7 kgm^{-3}), taken from CICE. We instead assume a sea ice density based on ice type: Using the OSI-SAF daily ice type product (OSI-403-c; Aaboe et al., 2021), we set the bulk density of sea ice (ρ_i) to 916.7 kgm^{-3} for first-year ice and 882 kgm^{-3} for multi-year ice (as per Ricker et al., 2014). We use this second experiment, following the setup of Mignac et al. (2022) in SI^3, with ice type-dependent sea ice density, as our baseline experiment (BASE). We perform four sets of experiments in relation to this BASE experiment: Three of these are performed to test the sensitivity of modelled SIT to changing the values of one parameter in the freeboard-to-thickness conversion, with the last performed to test the impact of changing the CryoSat-2 SIT uncertainties. A summary of the experiments is given in Table 1, with the following sub-sections describing their set-up in more detail.

Table 1. Configuration of the experiments: Parameters for conversion of CryoSat-2 radar freeboard estimates to assimilated SIT.

Experiment	Snow depth	Freeboard product	Radar penetration (α)	SIT uncertainty
CTRL	N/A	N/A	N/A	N/A
BASE	FOAM	AWI	1.0	Parameterised
SN_SMLG	SM-LG	AWI	1.0	Parameterised
SN_AWI	AWI	AWI	1.0	Parameterised
FB_LARM	FOAM	LARM	1.0	Parameterised
FB_CPOM	FOAM	CPOM	1.0	Parameterised
α**_0.9**	FOAM	AWI	0.9	Parameterised
α**_0.6**	FOAM	AWI	0.6	Parameterised
UNC	FOAM	AWI	1.0	Derived

We run all experiments from 17 October 2016 to 15 April 2017. Whilst conducting the experiments for the full CryoSat-2 period would be better, the computing time required per experiment per season makes this unfeasible. We choose the 2016-17 winter season to maximise the amount of independent data available for evaluation. This means that the results found in this paper are an example of the potential impacts of the changes we are making, and should not be taken as a quantification of the impacts of CryoSat-2 assimilation in general.

2.3.1 Snow depth

We test the sensitivity of short-range SIT forecasts to the choice of snow depth product (h_s), used in Equations 1 and 2 to convert the CryoSat-2-derived radar freeboard estimates into SIT. The BASE experiment uses FOAM's model snow depths for this. We perform two experiments, replacing the FOAM snow depths used in the conversion with:

- **AWI:** Monthly snow depth parameterisation used by the Alfred Wegener Institute (AWI), based on merging the Warren climatology with daily snow depth values derived from AMSR2 over first-year ice (Hendricks et al., 2021).
- **SM-LG:** Daily estimates from SnowModel-LG (SM-LG; Liston et al., 2020, 2021), a snow evolution model that provides daily, pan-Arctic snow property distributions for snow on sea ice. These snow depths have been evaluated using a range of in situ observations and were bias-corrected using OIB snow depths, such that the average modeled snow depth is equal to the average observed OIB snow depth over OIB observation tracks. The model is forced using ERA5 reanalysis data including air temperature, precipitation and wind variables.

The snow depth products vary noticeably, with the FOAM product showing a consistently lower snow depth than the SM-LG and AWI products (Figure 2). Regionally, the difference between the snow depth products is largest around Greenland, where the daily SM-LG snow depth is about three times as high as the AWI and FOAM snow depths, and in the Laptev region, where the AWI snow depth is often double the FOAM and SM-LG snow depth.

Figure 2. Regional and pan-Arctic daily mean snow depth for each of the snow products used in the freeboard-to-thickness conversion.

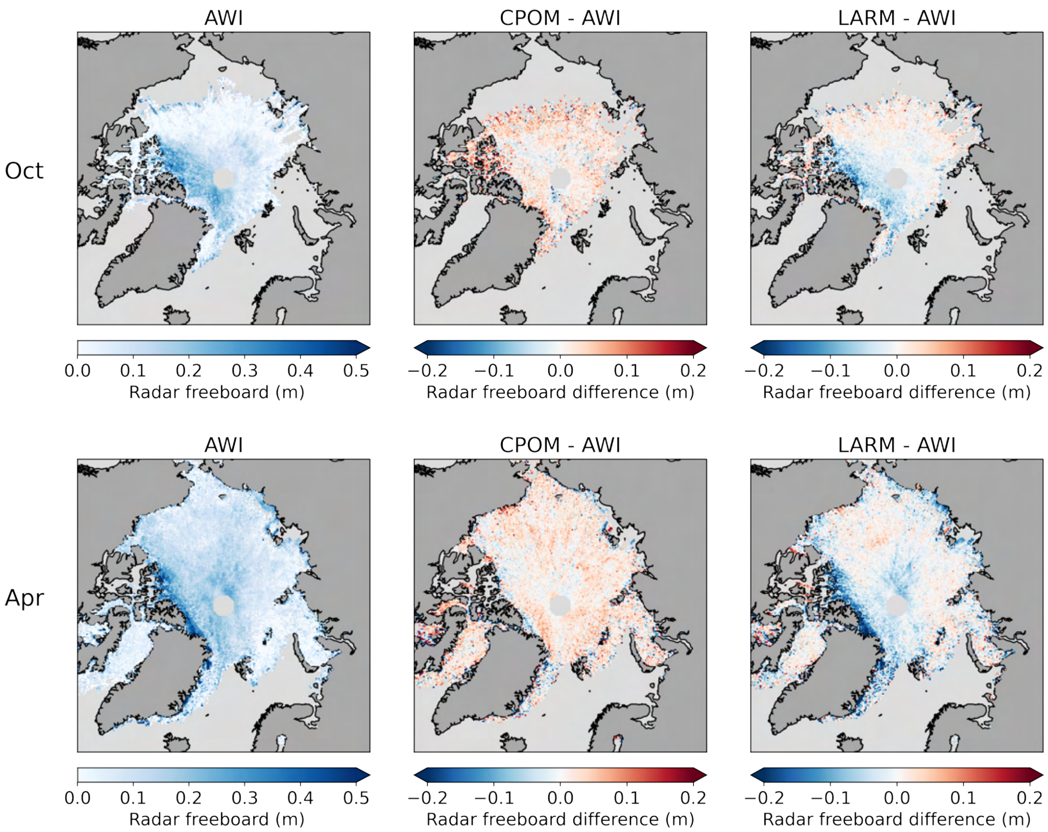

Figure 3. Monthly average radar freeboard for the assimilated CryoSat-2 SIT products, and the difference between them: AWI (left), CPOM-AWI (middle) and LARM-AWI (right). Top row shows October 2016, bottom row shows April 2017.

2.3.2 Freeboard product

We test the sensitivity of short-range forecasts to the use of different CryoSat-2-derived along-track radar freeboard products (F_r in equation (1)). The BASE experiment uses the AWI freeboard product, derived using the AWI threshold retracker (Hendricks et al., 2021). We perform two additional experiments, using CryoSat-2 radar freeboard data created using the physical LARM retracker (Landy et al., 2020) and the CPOM threshold retracker (e.g. Tilling et al., 2018).

The radar freeboard products vary noticeably, with the CPOM product showing a consistently higher radar freeboard than the LARM and AWI products on average, particularly in the marginal seas. In the Central Arctic, the LARM product shows consistently lower radar freeboards than the AWI and CPOM products (Figure 3).

2.3.3 Assumed mean penetration of radar waves into the snowpack

In the conversion of radar freeboard to SIT in the BASE experiment, we assume the radar waves from CryoSat-2 penetrate all the way through the snowpack, reflecting off the snow-ice interface ($\alpha = 1.0$ in Equation 1). We perform two additional experiments, assuming different mean scattering depths of the radar waves within the snowpack:

- $\alpha = \mathbf{0.9}$**:** as for summer sea ice, following Landy et al. (2022)
- $\alpha = \mathbf{0.6}$**:** the mean pan-Arctic value calculated by Nab et al. (2023)

Using a constant value for α is a first approach: Several studies have shown that α varies depending on wind distribution, temperature and snow properties such as salinity (Nandan et al., 2023; Nab et al., 2023; Nandan et al., 2017). This means that α is not a constant in reality - it varies spatially and temporally. However, the exact value of α and its determinants is still the subject of much research. We thus assign a constant pan-Arctic α value to determine the sensitivity of the data assimilation to this parameter, rather than trying to determine the perfect α to use to represent reality.

2.3.4 Freeboard uncertainty

The current method for assimilating CryoSat-2 along-track SIT into FOAM involves weighting the CryoSat-2 observations dependent on their SIT, as per Figure 3 of Fiedler et al. (2022). This involves assigning high uncertainty values (0.5 - 8 m) to SIT values below 1.5 m and above 4 m, such that they are weighted lower in the assimilation than SIT values between 1.5 - 4 m, which are assigned uncertainty values $<$ 0.5 m. Although this works well for thick ice, Mignac et al. (2022) found this method to lead to an overestimation in SIT in areas of thin ice, as higher SIT estimates in these areas are given a higher weighting in the data assimilation, despite not being more accurate than thinner CryoSat-2 SIT estimates that are down-weighted. Instead of this parameterised uncertainty, the assimilation weighting can be based on the uncertainty calculated for each freeboard observation. This is currently only possible for the AWI freeboard product, as the LARM and CPOM products do not include pre-calculated freeboard uncertainties, providing only uncertainties on the interpolated sea surface elevation at ice floes. We run an experiment (UNC) using SIT uncertainties derived from each individual freeboard measurement uncertainty, by applying a Gaussian propagation of the freeboard uncertainties through Equation 2 (as per Ricker et al., 2014). Uncertainties for the other parameters used in the freeboard-to-thickness conversion, such as the snow depth, sea-ice and snow densities, are taken from Figure 4 of Ricker et al. (2014).

2.4 Datasets used for evaluation

The 1-day SIT forecasts are compared to SIT derived from the Beaufort Gyre Exploration Project (BGEP; Krishfield et al., 2014) mooring observations, as well as airborne measurements from NASA's OIB campaign (Kurtz et al., 2013). The location of these data and mean SIT estimated by the airborne campaign are shown in Figure 4. It is worth noting that the sea ice model used here has an approximate resolution of 12 km in the Arctic, with the SIT computed as the model grid cell average. Therefore, there are limitations in the model's ability to represent the variability found at point measurements.

2.4.1 Mooring data

SIT measurements from the BGEP campaign were derived from bottom-anchored moorings equipped with upward-looking sonars. The BGEP data were collected continuously at three locations (BGEP-A, BGEP-B and BGEP-D) in the Beaufort Sea between 17 October 2016 to 15 April 2017 (Krishfield et al., 2014). The buoys estimate the sea ice draft at 2-second intervals, which are processed into daily averages. The daily drafts are converted into SIT by dividing them by 0.89, as per Rothrock (2003). We choose this simplified method for converting draft to SIT, instead of per Equation 2 of Kern et al. (2015), to avoid the use of snow depth data in this conversion and keep this as an independent dataset. For the evaluation, the daily model SITs are interpolated to the mooring locations.

2.4.2 Airborne data

Total freeboard (snow depth + ice freeboard) measurements were taken during the OIB campaign using an Airborne Topographic Mapper (Krabill, 2013), a 532 nm wavelength conically scanning laser altimeter with a 1 m footprint. The altimeter measures the elevation of the top of the snowpack and the elevation of the nearby ocean surface with an accuracy of <10 cm, such that the total freeboard can be estimated by taking the difference between these two (Krabill et al., 1995). At the same time, the snow depth was measured using a snow radar, which is assumed to reflect off the snow-ice interface. The sea ice freeboard is then estimated by taking the difference between these two, and SIT calculated under the assumption of hydrostatic equilibrium as per Equation 2. Multiple data products have been derived from the OIB measurements. The National Snow and Ice Data Center (NSIDC) Quick-Look product is used here, which is known to underestimate snow depth by up to 8.8 cm (Kwok et al., 2017). In this product, SIT point measurements are averaged over 50 km clusters. Point measurements with a standard deviation greater than 1 m are not used in the cluster for ice thinner than 1 m, and point measurements with a standard deviation greater than 2 m are not used in the cluster for ice thicker than 4 m (Kurtz et al., 2013). Further processing is conducted here to remove cluster observations with a standard deviation greater than 2 m, as per Mignac et al. (2022). For the evaluation, the daily model SITs are interpolated to the OIB cluster locations.

3 Results

3.1 Data assimilation influence on model sea ice thickness

To determine regional patterns, we use the Arctic regions defined by the NSIDC (Meier et al., 2007) to plot the evolution of the regional mean SIT for the different experiments (Figure 5). In the Baffin region, the CTRL experiment shows significantly higher daily mean SIT than the other experiments, with this difference increasing as the season progresses. By April, the mean daily SIT in the CTRL experiment in this region is double the SIT found in the other experiments. In the Central Arctic and East Siberian region, the CTRL experiment shows a lower daily mean SIT than the other experiments, particularly between November - March. In the Greenland region, the experiments with CryoSat-2 assimilation show a sudden drop in daily mean SIT in February, which is not shown in the CTRL experiment. Compared to the CTRL experiment, all experiments show a higher daily mean SIT in all regions except for the Beaufort Sea, Baffin Bay and Barents Sea. The increase in SIT after CryoSat-2 assimilation is particularly evident in the Central Arctic, where the increase is largest in the SN_SMLG and SN_AWI experiments and occurs immediately after data assimilation begins in October. In the Central Arctic, the $\alpha\_0.6$ experiment shows the smallest difference compared to the CTRL experiment, but the SIT is still 10s cm higher early in the winter season.

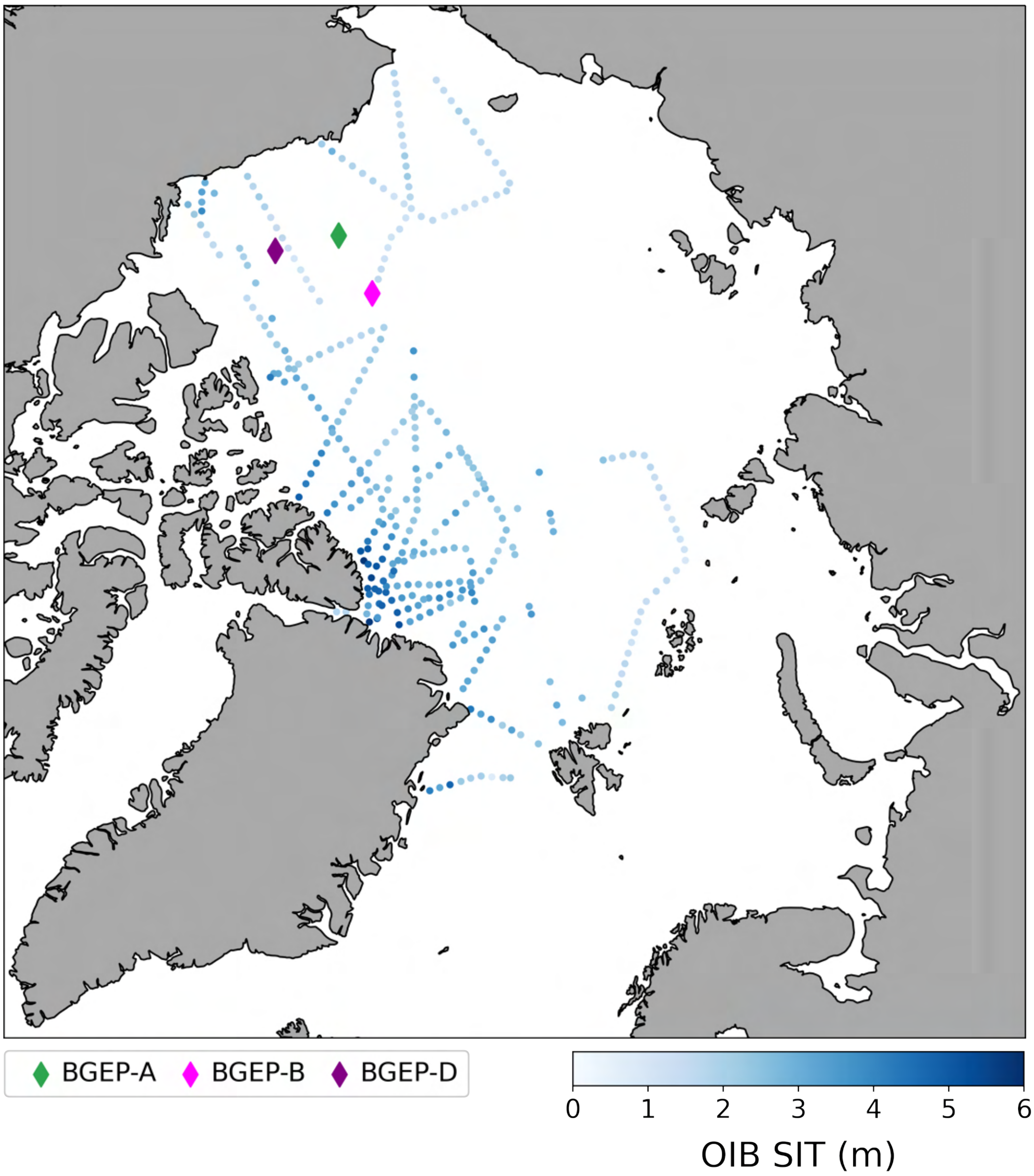

Figure 4. Location of bottom-anchored moorings (BGEP-A, BGEP-B, BGEP-D) and SIT derived from airborne OIB measurements.

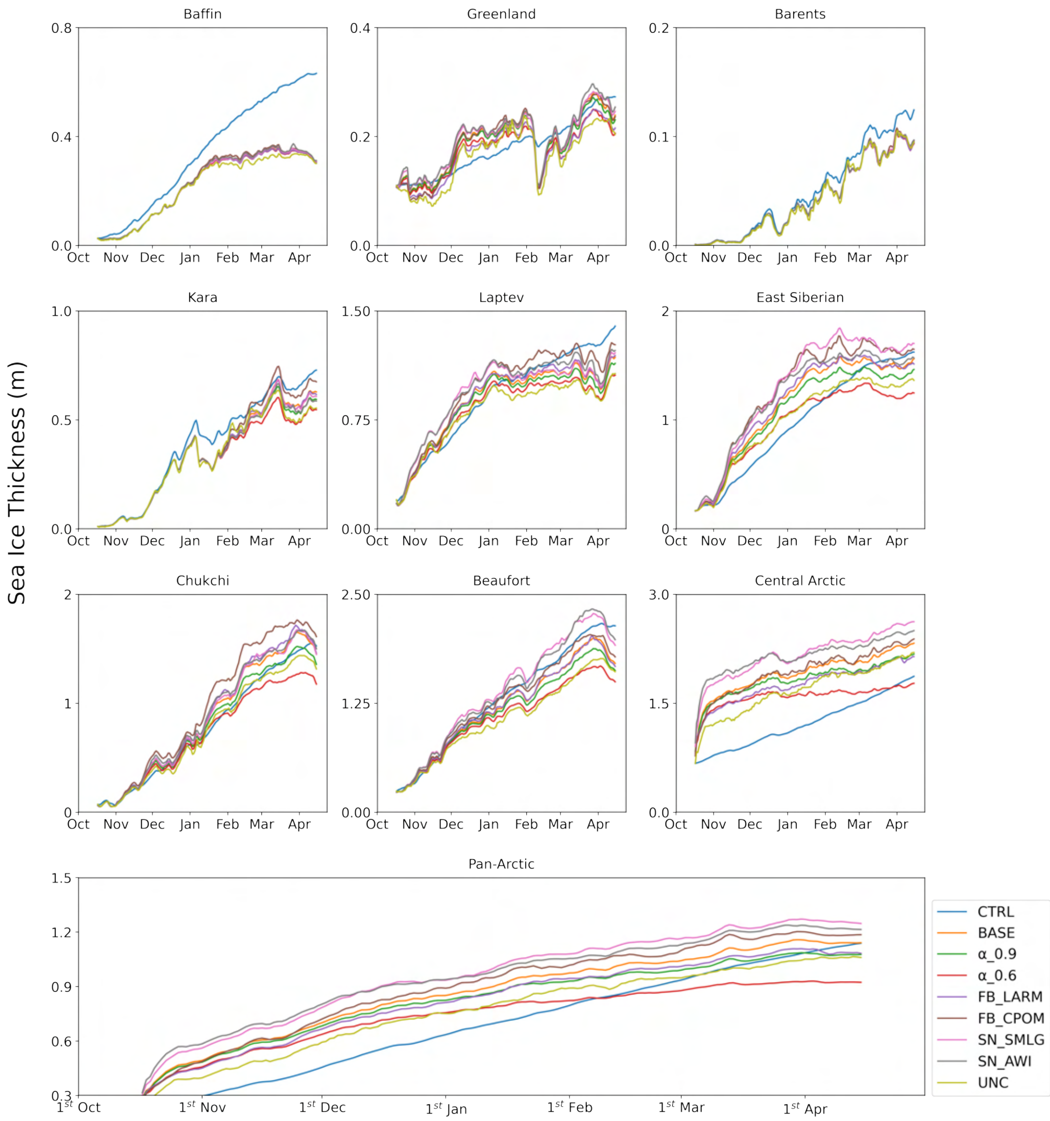

Figure 5. Regional mean daily SIT. Note the different y-axes.

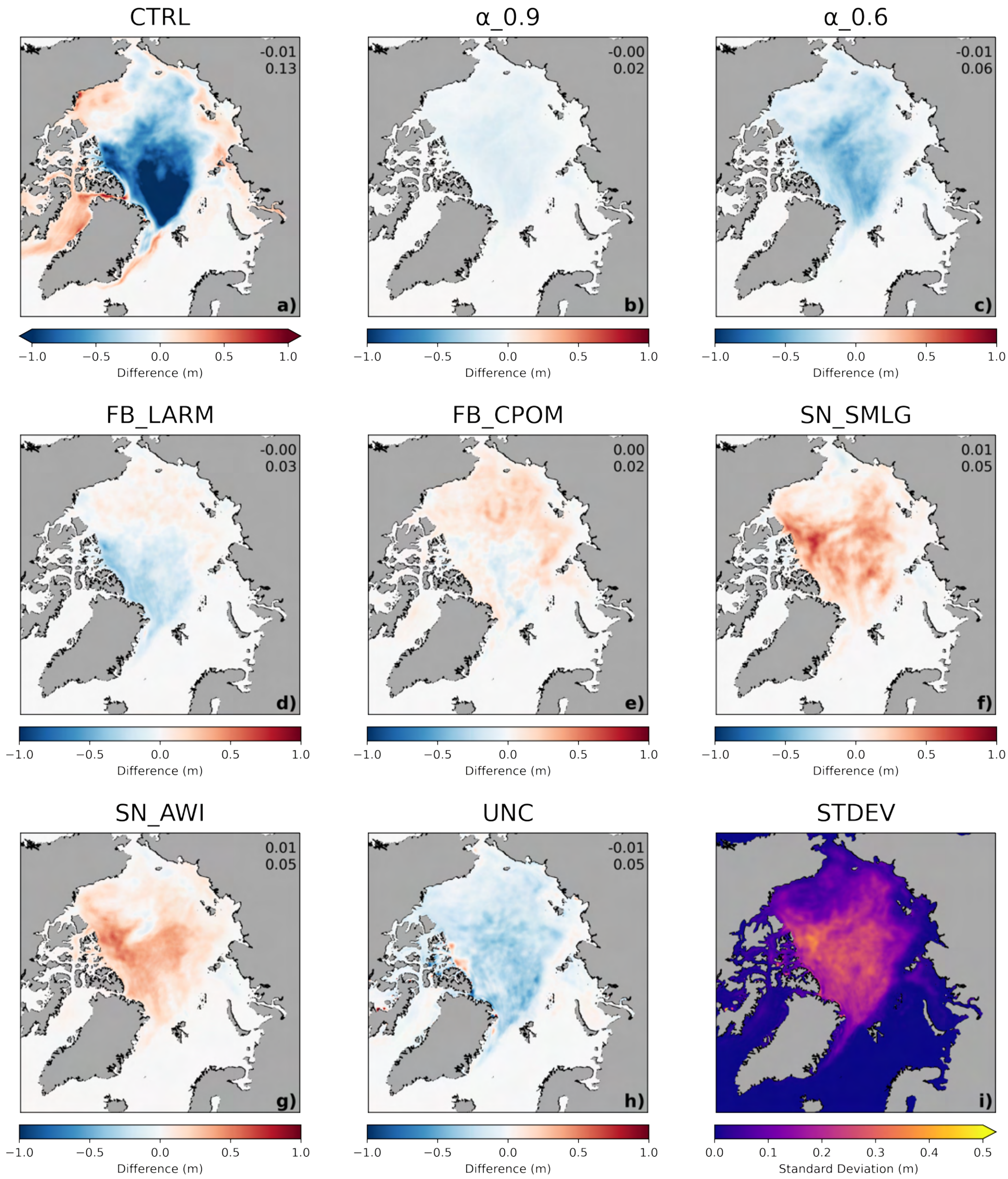

Figure 6. **a-h)** Mean daily difference in SIT between each experiment and the BASE experiment (experiment - BASE). Text shows the pan-arctic mean (top) and standard deviation (bottom). **i)** Standard deviation between the mean daily SIT for all the assimilation experiments (excluding CTRL).

Compared to the BASE experiment, the α_0.6 and α_0.9 experiments show a pan-Arctic decrease in SIT on average, while the SN_SMLG and SN_AWI experiments show a pan-Arctic increase in SIT (Figure 6). The FB_LARM experiment shows a decrease in SIT over multi-year ice and an increase in SIT over first-year ice, while the FB_CPOM experiment shows an increase in SIT in almost all regions of the Arctic, except for some parts of the Central Arctic. The UNC experiment shows a pan-Arctic decrease in SIT, with an increase over a small part of the Central Arctic. The standard deviation (STDEV) of the SIT in these experiments is higher over the Central Arctic and Russian Arctic ($>$ 0.4 m), where there are more CryoSat-2 observations, decreasing to $<$0.2 m towards the Beaufort Sea and in Baffin Bay (Figure 6), where a much larger number of SMOS observations dominate the SIT assimilation results (see Figure 1). The large spread in the Central Arctic and Russian Arctic is supported by the large differences between the assimilation experiments as the winter season progresses (Figure 5).

3.2 Relative impact of changing each parameter

In relation to the BASE experiment, we find that changing the assumed radar penetration (α), freeboard product and snow depth product used in the freeboard-to-thickness conversion can change the daily regional model SIT by up to 0.65 m, 0.55 m and 0.88 m, respectively, with each of these maximum values found in the Central Arctic. For the α experiments, the largest relative differences were seen in the Central Arctic, where reducing the assumed radar penetration decreased the daily mean SIT by up to 33%. Large differences were also seen in East Siberian and Chukchi regions, where decreasing the assumed radar penetration decreased SIT by up to 28%. When changing the freeboard product used, we find the largest differences in these same three regions, with SIT changing by up to 30%. When changing the snow depth product used, the largest differences where found in the Central Arctic and Beaufort Sea, with SIT changing by up to 48%. When considering absolute values, changing the snow depth product, freeboard product and assumed radar penetration resulted in changes in modelled SIT of 0.11 m (σ = 0.18 m), 0.07 m (σ = 0.11 m) and 0.08 m (σ = 0.15 m) on average, respectively (Figure 7).

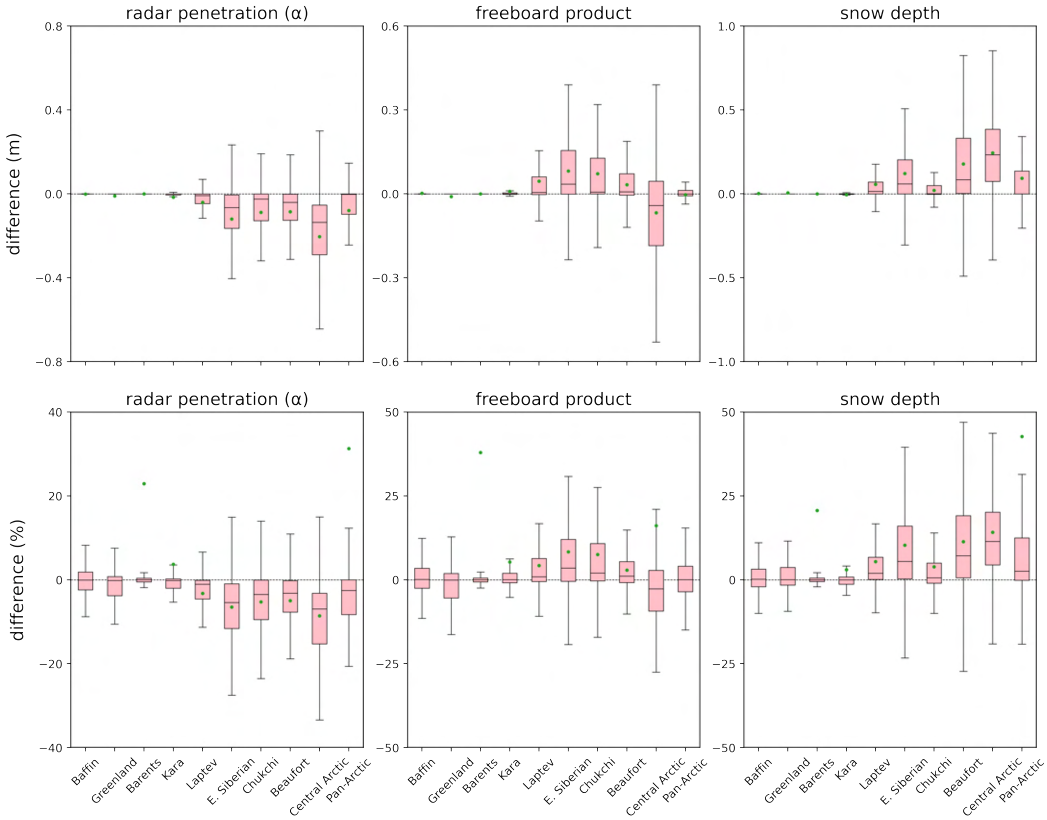

Figure 7. Difference in the daily regional SIT compared to the BASE experiment caused by changing the assumed α, freeboard product and snow depth product used in the freeboard-to-thickness conversion. Top row shows value difference (m), bottom row shows percentage difference. Values are calculated as the daily difference between the BASE experiment and the experiments where the parameter of interest is changed, in each grid cell. Boxes extend from the 25th to 75th percentiles of the values, with a black line showing the median and a green dot showing the mean. Whiskers extend to the 5th and 95th percentiles. Note the different y-axes. Outliers are not shown. The radar penetration (α) set contains the α_0.9 and α_0.6 experiments. The freeboard product set contains the FB_CPOM and FB_LARM experiments. The snow depth set contains the SN_SMLG and SN_AWI experiments. The CTRL and UNC experiments are not included in this analysis.

3.3 Impact of sea ice thickness data assimilation changes on the thermodynamic sea ice budget

We now investigate the impact of changing the freeboard-to-thickness conversion parameters and uncertainties on the thermodynamic sea ice budgets, in order to determine the sensitivity of the model processes to these changes.

When assimilating CryoSat-2 data, we find a decrease in thermodynamic volume change in the Central Arctic and an increase in the marginal seas of the BASE experiment compared to the CTRL experiment (Figure 8a). When reducing the assumed radar penetration, we find a decrease in the thermodynamic volume change in all regions of the Arctic, when compared to the BASE experiment. When changing the snow depth product used in the freeboard-to-thickness conversion, we find an increase in thermodynamic volume change in almost all regions of the Arctic compared to the BASE experiment, particularly in the Beaufort Sea. The spatial distribution of these changes is very similar when using SM-LG and AWI snow depths. When changing the freeboard product used, we find a decrease in thermodynamic volume change in the Central Arctic and an increase in the marginal seas, compared to the BASE experiment. The UNC experiment shows a decrease in thermodynamic volume change over most of the Central Arctic, with an increase in the marginal seas and Baffin Bay. The standard deviation of the runs where CryoSat-2 is assimilated is up to 1 x 10 ^{6} m^3 in parts of the Central Arctic and East Siberian regions, decreasing down to 0 m^3 in parts of the Canadian Archipelago region (Figure 8).

Figure 9 shows the difference in ice volume change due to thermodynamic processes when using differing parameters in the freeboard-to-thickness conversion. In relation to the BASE experiment, we find that modifying the assumed radar penetration, freeboard product and snow depth product used can shift the daily model thermodynamic volume change by up to 1.6 x 10^6 m^3, 1.2 x 10^6 m^3 and 1.4 x 10^6 m^3 respectively. For the α experiments, the largest relative differences were seen in the East Siberian and Central Arctic regions, where modifying the assumed radar penetration shifted the daily ice volume change by up to 125%. When modifying the freeboard product used, we also find the largest relative differences in the Central Arctic, with thermodynamic volume change shifting by up to 140%. Similarly, when modifying the snow depth product used, the largest differences in thermodynamic volume change were found in the Central Arctic and the Beaufort Sea, with differences in ice volume change of up to 150%. When considering absolute values, changing the snow depth product, freeboard product and assumed radar penetration used changed the model ice volume change by 0.32 x 10 ^{6} m^3 ($\sigma = 0.93$ x 10 ^{6} m^3), 0.36 x 10 ^{6} m^3 ($\sigma = 0.87$ x 10 ^{6} m^3) and 0.31 x 10 ^{6} m^3 ($\sigma = 0.80$ x 10 ^{6} m^3) on average, respectively.

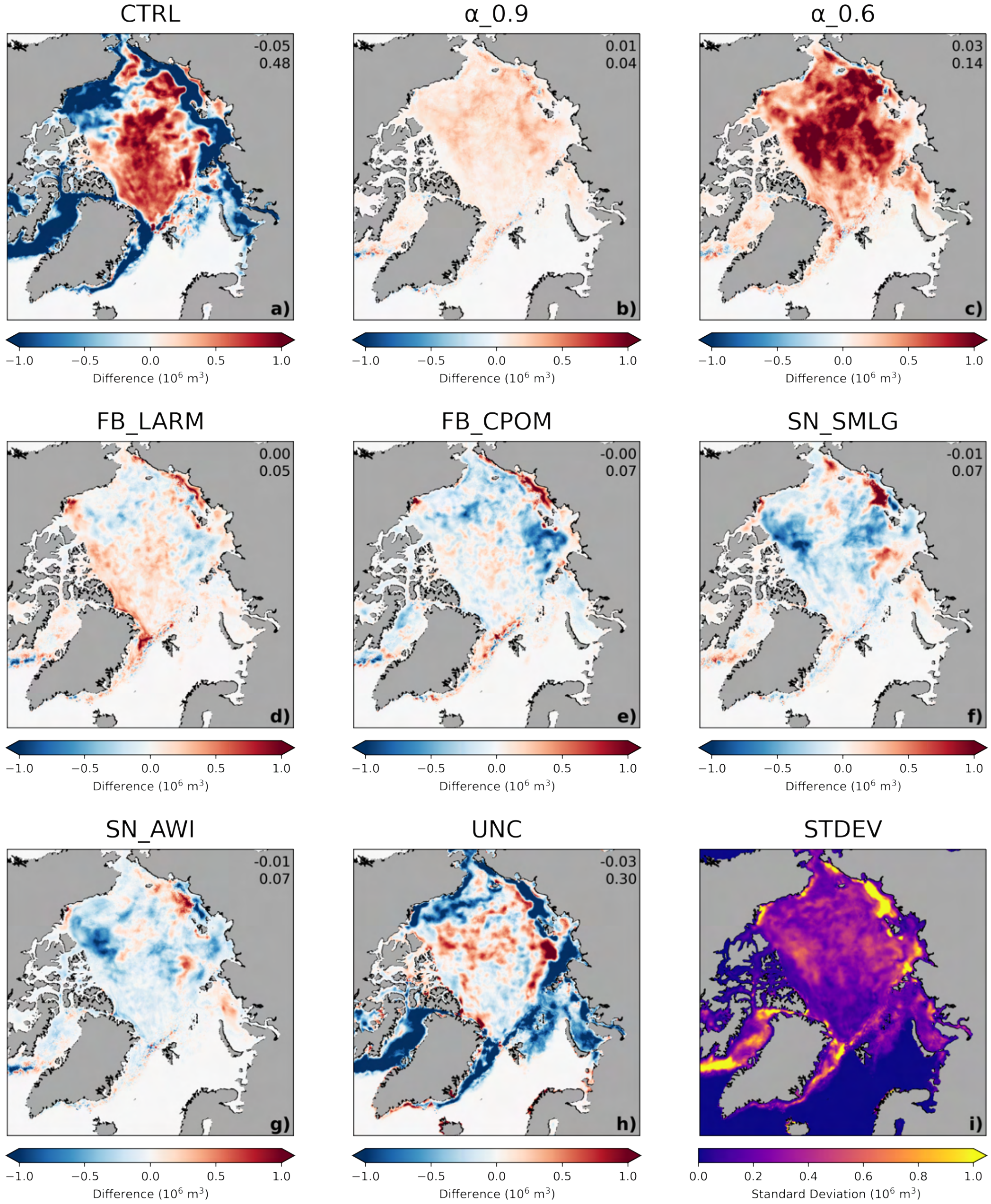

Figure 8. Difference between mean daily ice volume change due to thermodynamic processes for each experiment and the BASE experiment (experiment - BASE). Text shows the pan-Arctic mean (top) and standard deviation (bottom).

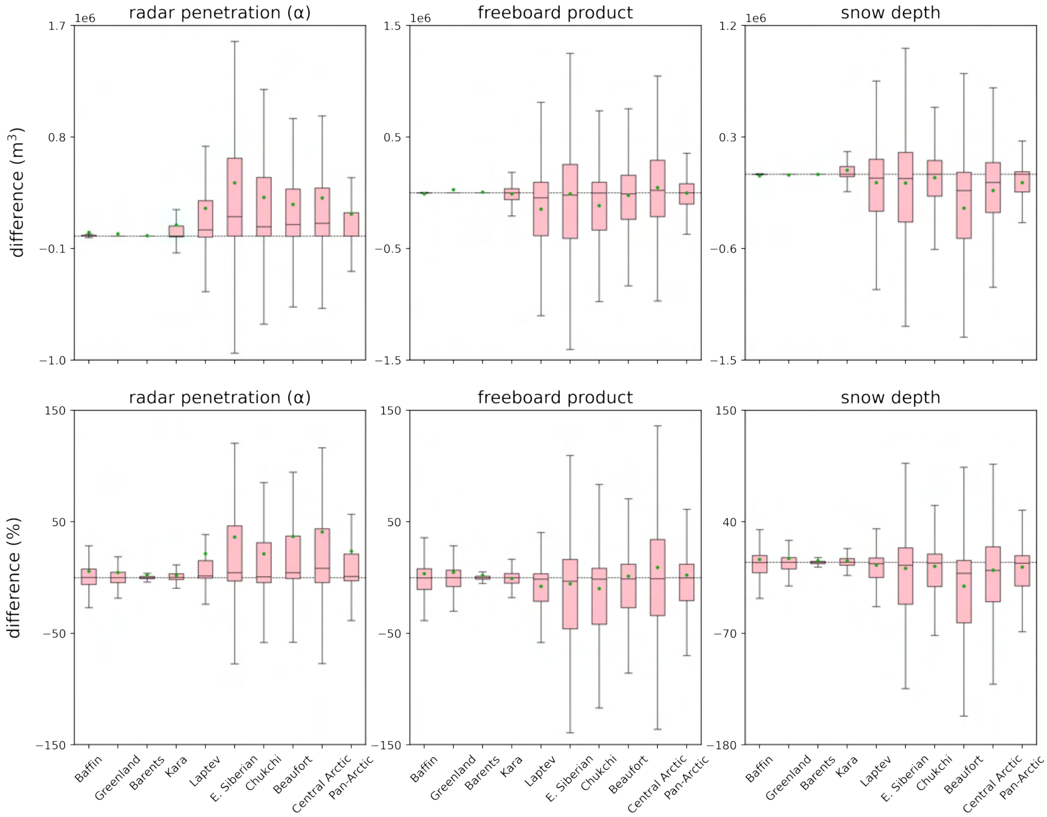

Figure 9. Difference in the daily ice volume change due to thermodynamic processes, caused by changing the assumed α, freeboard product and snow depth product used in the freeboard-to-thickness conversion. Top row shows value difference (m), bottom row shows percentage difference. Values are calculated as the daily difference between the BASE experiment and the experiments where the parameter of interest is changed, in each grid cell. Boxes extend from the 25th to 75th percentiles of the values, with a black line showing the median and a green dot showing the mean. Whiskers extend to the 5th and 95th percentiles. Note the different y-axes. Outliers are not shown. The radar penetration (α) set contains the α_0.9 and α_0.6 experiments. The freeboard product set contains the FB_CPOM and FB_LARM experiments. The snow depth set contains the SN_SMLG and SN_AWI experiments. The CTRL and UNC experiments are not included in this analysis.

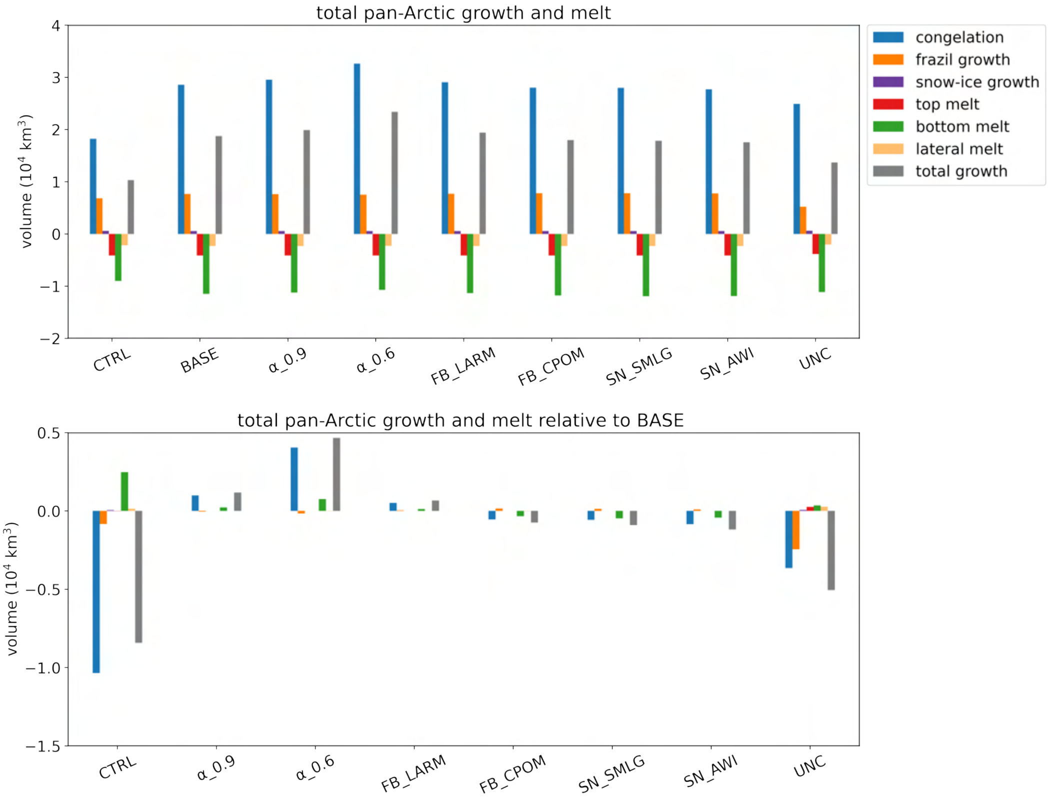

Figure 10. Total pan-Arctic volume of ice gained or lost through thermodynamic processes between October 2017 - April 2017 in each experiment. Top shows absolute values, bottom shows values relative to BASE experiment (experiment - BASE). Note the different y-axes.

Figure 10 decomposes the thermodynamic ice mass changes into separate components that take place during ice growth (congelation growth, frazil ice formation and snow ice formation) and ice melt (surface melt, bottom melt and lateral melt), as per Figure 4 of Tsamados et al. (2015). On a pan-Arctic scale, we find the biggest differences between the experiments come from congelation ice growth, followed by bottom melt and frazil growth. This is consistent with Figure 5, which shows that ice generally grows much faster in the assimilation experiments relative to CTRL. We find a positive total growth in all experiments over the winter season, as expected, with the largest increase in sea ice volume found in the α_0.6 experiment and the smallest in the CTRL experiment. Compared to the BASE experiment, we find an increase in congelation growth, and consequently total growth, in the α_0.6, α_0.9 and FB_LARM experiments, with decreases in the other experiments. This means that reducing the assumed radar penetration and changing the freeboard to inherently thinner products increase the rate of winter ice growth, whereas changing the snow depth to inherently thicker products reduces the rate of winter ice growth.

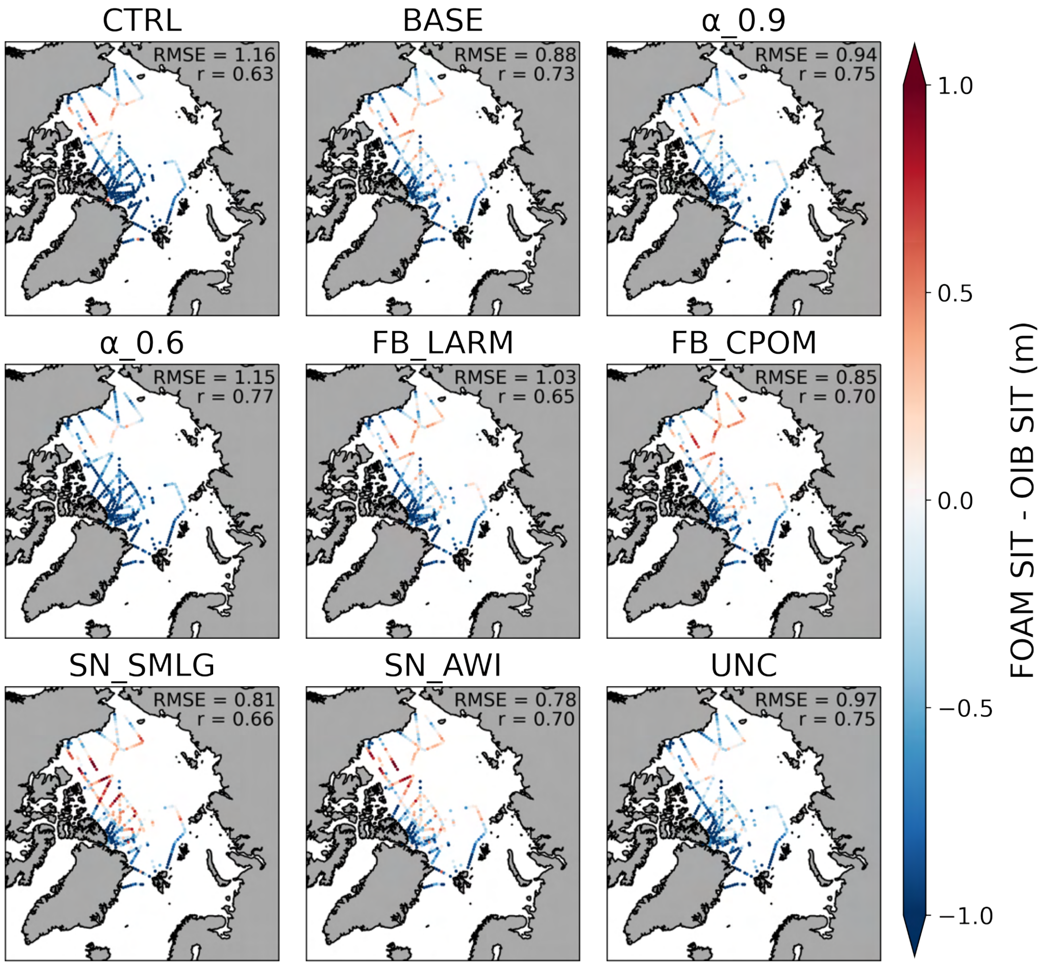

Figure 11. Difference between FOAM SIT and OIB SIT, between 9 March and 19 April 2017.

3.4 Comparison with buoy- and airborne-derived sea ice thickness

We now compare our experiments to independent SIT datasets, to determine the impact of changing the freeboard-to-thickness conversion parameters and uncertainties on the model's ability to represent buoy- and airborne-derived SIT.

3.4.1 Operation IceBridge (OIB)

The spatiotemporal patterns of modelled SIT showed noticeable differences between the CTRL experiment and the assimilation experiments. This is particularly evident to the north of Greenland, where there was a consistent increase in SIT by up to 1 m, bringing the model SIT closer to the OIB SIT (Figure 11). Despite this, all experiments still show an underestimation of SIT north of Greenland, with the SN_SMLG and SN_AWI experiments showing an overestimation in the Central Arctic and marginal seas.

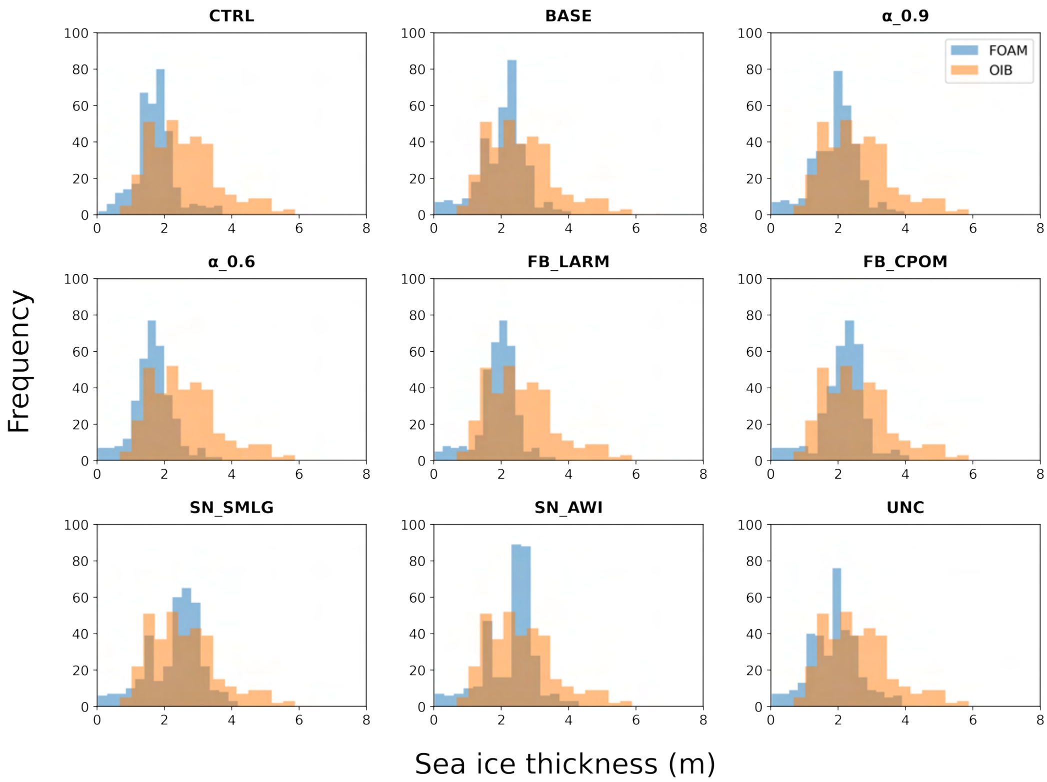

Figure 12. Histograms of FOAM SIT and OIB SIT, between 9 March and 19 April 2017.

Between the CTRL and BASE experiments, there is an improvement in both the r and RMSE values. When decreasing the assumed value for α, we find an improvement in FOAM's ability to represent the variability in the OIB SIT data, compared to the CTRL and BASE experiments. However, we find an increase in the RMSE with a decrease in assumed α compared to the BASE experiment, as the model SIT becomes too low. When changing the CryoSat-2 freeboard product, we find a degradation in the model's ability to represent the SIT variability in the OIB data, with the r value decreasing from 0.73 in the BASE experiment (r = 0.73), to 0.7 for the FB_CPOM experiment and 0.65 for the FB_LARM experiment. However, we find an improvement in the magnitude of the modelled SIT in the FB_CPOM experiment (RMSE = 0.85 m) compared to the BASE (0.88 m) and FB_LARM (1.03 m) experiments. When changing the snow depth used in the SIT conversion, we find a decrease in r value, from 0.77 for the BASE experiment to 0.70 for the SN_AWI experiment and 0.66 for the SN_SMLG experiment. In terms of RMSE, changing the snow depth shows an improvement, with the SN_AWI experiment (0.78 m) performing better than the SN_SMLG (0.81 m) and BASE (0.88 m) experiments.

We find a wider range of values in the OIB SIT data than in the FOAM data, particularly in the CTRL, α and freeboard product experiments, where values of 4 - 6 m are much more frequently found in the OIB data than in the FOAM data (Figure 12). Additionally, areas of thin ice (<1 m) are present in all the FOAM experiments, but not in the OIB data. In terms of variability representation, the α_0.6 experiment (r = 0.77) performs the best out of all the experiments, followed by the α_0.9 and UNC experiments (r = 0.75). In terms of RMSE, the SN_AWI experiment performs the best (0.78 m), followed by the SN_SMLG experiment (0.81 m).

3.5 Beaufort Gyre Exploration Project (BGEP)

We find an overestimation in SIT in each of the experiments in the January - February period at BGEP-B, with the exception of the CTRL and α_0.6 experiments (Figure 13). We find an overestimation at BGEP-A in the CTRL experiment in January - February and at BGEP-D in CTRL in February, which is not found in the experiments where CryoSat-2 is assimilated. We find the lowest RMSE in the UNC and CTRL experiments at BGEP-A and BGEP-B, respectively. At BGEP-D, we find similar RMSE values in all experiments (0.42-0.47 m), with SN_AWI performing the best. Overall, we find similar RMSE values in all experiments (0.35-0.37 m) except FB_CPOM and SN_SMLG (0.41-0.42 m). We find that the FOAM experiments do not represent the daily variability measured by the BGEP buoys well, finding an improvement when taking a one-week running mean of the BGEP data (Figure S1). This is not surprising considering the repeat sub-cycle of CryoSat-2 (30 days) and the model grid size of ∼12 km, compared to the high temporal sampling resolution of the moorings.

As we are only using one season of data, the linear correlation coefficient is not a suitable method of assessing how well the model is able to represent the variability measured at stand-alone buoys. Instead, we calculated an error by taking the difference between the model and the daily mean BGEP SIT, before dividing this by the daily BGEP standard deviation. This tells us how many standard deviations the model prediction is away from the buoy measurement on each day. We find that the UNC and CTRL experiments perform best at BGEP-A and BGEP-B, respectively, with errors of 0.17. At BGEP-D, the α_0.9 and UNC experiments perform the best with an error of 0.22. When the buoys are combined, the α_0.9 and UNC experiments perform the best with an error of 0.14. The CTRL, FB_CPOM and SN_SM-LG experiments perform the worst at the individual buoys, with the FB_CPOM experiment performing the worst when the buoys are combined. At each buoy, the experiments are normally within two standard deviations of the buoy measurements, decreasing to one standard deviation when the buoys are combined (Figure S2).

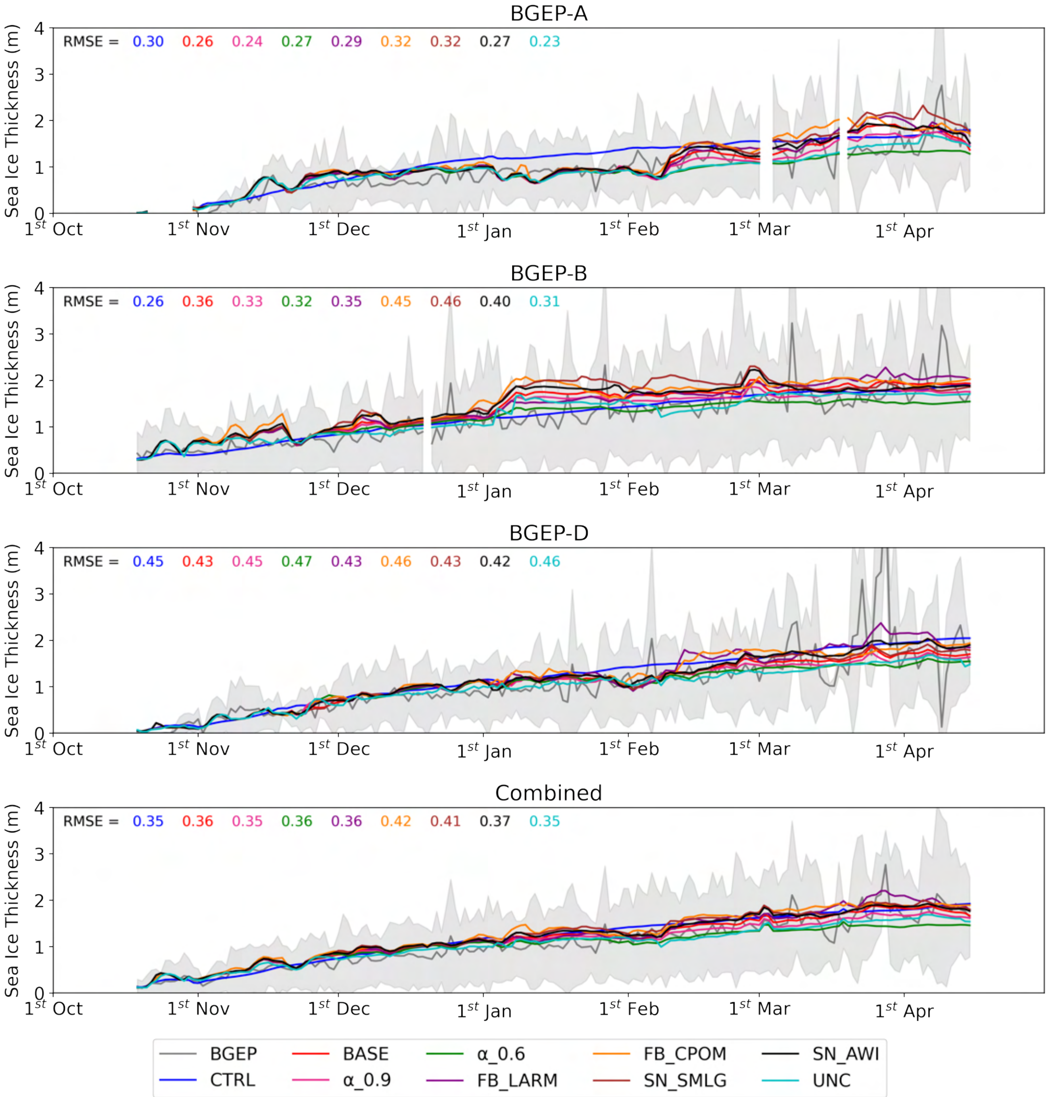

Figure 13. FOAM SIT and BGEP SIT at BGEP-A, BGEP-B and BGEP-D mooring locations from 17 October 2016 to 15 April 2017. FOAM SIT is only shown on days where BGEP data is available. The grey shaded area shows the BGEP uncertainty, represented by the daily standard deviation. Coloured text shows the mean root-mean-squared error for the season for each experiment.

4 Discussion and Conclusions

4.1 Relative influence of each parameter

We find large differences in the modelled SIT when changing the freeboard and snow product used, as well as the assumed radar penetration. However, each parameter has a strongly regional impact, which differs across the experiments. Our results suggest that differences in the Central Arctic are dominated by changes in the snow depth, whereas the use of different freeboard products plays a more dominant role in the marginal seas. The use of different observation uncertainties also plays a significant role in the marginal seas, as it increases the relative weighting of the assimilated CryoSat-2-derived SIT data compared to the SMOS data.

For the snow depth product used, the largest changes occur in the Beaufort, East Siberian and Central Arctic regions. Using snow depth observations from SnowModel-LG or passive microwave coupled to the Warren snow climatology, instead of the FOAM snow depths, increases the modelled SIT in these regions by 9-25 cm (7-16%) on average. This is because both snow depth products are thicker than FOAM and therefore produce thicker estimates of sea ice for a given freeboard. Our snow depth experiments imply that models with a thin spring snow cover ($\sim$20 cm) in the Central Arctic will not produce reliable SIT after assimilating CryoSat-2 freeboards. When assimilating CryoSat-2-derived SIT data, Mignac et al. (2022) found a decrease in the underestimation of SIT in this region compared to OIB-derived SIT, with the root-mean-squared difference decreasing from 0.95 m to 0.65 m, although the underestimation bias remained to the north of Greenland, where the FOAM snow depths are too low. Previous modelling studies have found a large underestimation of SIT in the Central Arctic. For example, Wang et al. (2016) found an underestimation of up to 2 m in February - March. Xia & Xie (2018) found that CryoSat-2 freeboards increasingly underestimated those from OIB as the OIB freeboard increased, up to -44 cm mean bias when the airborne laser scanner freeboards were >66 cm thick. Since the thickest sea ice north of Greenland is also the roughest in the Arctic, this suggests that the CryoSat-2 radar returns might be biased towards smoother areas of the topography, underestimating the true mean freeboard. A freeboard bias caused by the radar penetrating only a fraction of the snow layer would act in the other direction, generating thinner SIT estimates compared to OIB, so this appears to be a less important factor over the multi-year ice region north of Greenland.

We find differences in SIT of up to 35% when decreasing the assumed mean radar penetration of the snowpack from 100 to 90 and 60%, with the largest differences seen in the Western and Russian Arctic. In comparison to the OIB data, which mostly cover multi-year ice north of Greenland and Canada, the 60% penetration experiment led to more of an underestimation of the ice thickness than other experiments. However, in comparison to thinner ice measured by the Beaufort Gyre mooring ice drafts, the 60% penetration experiment was not found to underestimate the in situ-measured ice thickness. Nab et al. (2023) found a response in CryoSat-2-derived radar freeboard estimates to snow accumulation, wind speed and air temperature, with the magnitude of the response varying between tested freeboard products, which suggests that the radar penetration depth can be <100% under certain conditions. Results from our bias assessment versus in situ data suggest these conditions may occur more frequently over first-year than multi-year ice; however, there could be competing biases that obscure this finding.

CryoSat-2 measures thick ice with relatively high accuracy but struggles with very thin ice (Ricker et al., 2014), so improvements in forecast accuracy caused by changing the assimilation parameters are mainly found in the Central Arctic where the observations are more accurate than the free-model run. We find a larger spread in areas that have more CryoSat-2 observations, such as the Central Arctic and East Siberian regions. In regions where there are fewer CryoSat-2 observations, such as the Baffin and Barents regions, there is a much smaller spread in the data, as the SIT assimilation is dominated by the assimilation of SMOS data. The assimilated CryoSat-2 freeboard products have noticeable differences over thin ice regions, with the use of the CPOM product leading to a higher SIT in these regions than the LARM and AWI products.

We find that using observation errors derived from the individual freeboard measurements, rather than parameterising them based on the SIT value, leads to a noticeably different SIT and thermodynamic growth distribution. The BASE and UNC experiments show a 36% daily mean absolute difference in SIT, despite the use of the same snow depth, assumed radar penetration and radar freeboard product in the freeboard-to-thickness conversion. Compared to the BASE experiment, the UNC experiment shows thinner ice on average on a pan-Arctic scale, except for a small region north of the Canadian Archipelago. This is likely due to the high observation errors assigned to CryoSat-2-derived SIT values below 1 m (σ = 8 m), compared to values between 1 - 4 m (σ<0.5 m) in the BASE experiment with an empirical parameterisation of the errors. This means that measurements of thicker ice have a substantially higher weighting in the data assimilation, which can lead to an overestimation in SIT in areas where low-value CryoSat-2-derived SIT values are removed from the assimilation despite being accurate.

The differences in thermodynamic volume change between the experiments and the BASE run show similar spatial patterns as those for SIT, but of the opposite sign. This is expected, as thin ice grows faster due to steeper temperature gradients, leading to a negative feedback between SIT and thermodynamic volume change (Haas, 2003). Therefore, when the assimilated CryoSat-2 observations are thinner than the model (e.g. in the α_0.6 and α_0.9 experiments), the model reacts by growing more sea ice thermodynamically. This is particularly evident in the FB_CPOM and FB_LARM experiments, where areas that show a thinning sea ice cover as a result of the data assimilation show an increase in thermodynamic growth, while areas with a thickening ice cover show a decrease in thermodynamic growth. We find a decrease in mean daily thermodynamic volume change in the SN_SMLG and SN_AWI experiments. This is likely due to the increase in snow depth in these runs, with the two snow depth products generally thicker than modelled snow from FOAM, which leads to an increased derived SIT, insulating the ice and thus reducing winter ice growth. The UNC results are slightly different from the others, with a pan-Arctic thinning in SIT resulting in decreases in thermodynamic growth in the marginal seas and increases over the Central Arctic. We find this to be due to a seasonal change in the assimilation impact on the SIT in this experiment: in October - November, the SIT in the UNC experiment increases by up to 40% more than the BASE experiment in the marginal seas, while the rest of the ice cover thins by more than 40% compared to the BASE experiment. As the season progresses, this increase in SIT in the UNC experiment decreases, with decreases in SIT seen in almost all regions by April (Figure S3). The difference in thermodynamic volume growth between these experiments shows the inverse pattern: the UNC experiment shows a much lower thermodynamic volume growth in the marginal seas in the early winter months than the BASE experiment, with the difference decreasing as the season progresses (Figure S4).

In this study, we have tested a variation of parameterisation combinations for the freeboard-to-thickness conversion. We find a large spread in the model results between these different experiments, particularly over thicker ice (e.g. the Central Arctic), that increases as the winter season progresses. This highlights a concern for future SIT data assimilation and forecasting, with the chosen parameterisation of the freeboard-to-thickness conversion having a substantial impact on model SIT results, with a difference in model SIT of up to 30-48%. Additionally, our results call for an improved, consistent definition of CryoSat-2 radar freeboard uncertainties, with our derived SIT uncertainties causing a change in mean model SIT of up to 36%.

4.2 Comparison with independent datasets

We find a consistent underestimation in SIT and total thickness north of Greenland in all our experiments, likely due to the consistent underestimation of the FOAM model snow depth in this area. When replacing this snow depth with the SM-LG and AWI products, we find a significant improvement in the model SIT results in this area, compared to independent datasets; however, the SIT is still underestimated. These snow depth values are generally higher than those of FOAM, leading to an increase in SIT on a pan-Arctic scale. The SIT is also improved when using derived uncertainties from the AWI product, rather than uncertainties parameterized for specific ranges of ice thickness, which suggests the uncertainties for thick multi-year ice are well-constrained in the product. Over thinner ice areas, the CryoSat-2 data should either be entirely discarded or the derived freeboard uncertainties used, since down-weighting the assimilation of the thinnest freeboards can increase the impact of erroneously thick freeboards in these areas.

In terms of variability, compared to the OIB estimates, each of our experiments performs better than the CTRL experiment, with the α_0.6, α_0.9 and UNC experiments also performing better than the BASE experiment. Compared to OIB estimates, decreasing the assumed α used in the freeboard-to-thickness conversion leads to an increase in RMSE. As the OIB SIT estimates are calculated using an assumption of full C-S-band snowradar penetration through the snowpack, this is unsurprising. We find the lowest RMSE in the SN_AWI experiment (0.78 m), followed by the SN_SMLG experiment (0.81 m). Compared to the BGEP SITs, the experiments perform equally well in terms of RMSE (0.35 - 0.37 m when the buoys are combined), with the exception of the FB_CPOM and SN_SMLG experiments which perform worse (0.42 and 0.41 m, respectively). The BGEP data are limited to the Beaufort Sea region, while the OIB covers large parts of the Arctic, with the exception of the Russian Arctic. As a result, the BGEP datasets largely cover areas of thin ice (<1.5 m), which CryoSat-2 and FOAM struggle with compared to thicker ice. The UNC experiment was found to perform better than the BASE experiment over areas of thin ice: when isolating areas where the OIB-derived SIT is lower than 1 m, the UNC experiment was found to have a lower RMSE (0.32 m) than the BASE experiment (0.41 m), with the same r value (0.82). Similarly, compared to the BGEP-derived SIT data, the UNC experiment performs better than the BASE experiment at BGEP-A and BGEP-B, and when the buoys are combined. This highlights the importance of assigning realistic observation errors when assimilating CryoSat-2 data. This suggests that improving the observation errors used when assimilating CryoSat-2 SIT data could improve model performance over areas of thin ice, as the number of measurement over thin ice that are assimilated can be increased by more accurately separating the inaccurate ones.

We repeated the above analysis with airborne estimates taken during the IceBird campaign series. However, we found these results to give very limited insight into the ability of the FOAM experiments to produce accurate SIT forecasts. We believe this is due to the large amount of pre-processing required to make this data comparable to the model data, due to the former's much higher spatial and temporal resolution. The results are shown in Supplementary Section 1.

Our results show that, while improving the representation of one parameter in the freeboard-to-thickness conversion does often lead to an improved model SIT result in comparison to independent observations, this is not always the case. This is likely due to competing biases, which have been created by tuning the model to observational data using the current parameterisation, such that changing one parameter without changing the others can often bring the model results further away from the observations, despite an improvement in that parameter's values. This means that changing the parameters altogether is often the only way to bring the model results closer to the observations, as competing biases can be corrected for at the same time. The challenge is optimizing the combination of parameter changes, when they have different impacts (negative or positive) in different regions of the Arctic or part of the winter season. However, the purpose of this study is to determine the sensitivity of the model SIT results to the parameterisation of the radar freeboard-to-thickness conversion, rather than to find an optimal model configuration. Optimization is complex and will likely require a trade-off between accurately simulating the mean ice state characteristics and capturing the inter-annual ice state variability. This is a key point, since the quality of a seasonal sea ice forecasts relies on the ability to capture interannual variations in the anomalies of the SIT which translate into later ice extent forecasting skill (e.g. Landy et al., 2022).

4.3 Future Work

The sensitivity of the FOAM system's SIT forecasts to the parameterisation of the freeboard-to-thickness conversion and other data assimilation parameters informs us how best to configure the Met Office's operational system when implementing SIT assimilation. A near-real time CryoSat-2 product is required for operational assimilation of SIT into the Met Office systems. Currently, FOAM assimilates data from the previous 24 hours into its systems. The AWI product is available near-real time, so its SIT and snow depth values are currently being tested for operational assimilation. This product also comes with freeboard uncertainties, allowing for the use of derived SIT uncertainties in the operational assimilation. Additionally, testing the SIT assimilation on a fully coupled ocean-sea ice-land-atmosphere system would allow us to explore the impacts of SIT assimilation on land and atmosphere variables, to determine how SIT uncertainties affect parameters outside of the Arctic ocean.

Tests with other satellite-derived SIT products, such as those derived from ICESat-2 and Sentinel-3, would allow us to increase the temporal and spatial resolution of the assimilated SIT data. This would require a consistent definition of uncertainties such that the different SIT products can be combined. Assimilation of ICESat-2 SIT would remove the need for radar scattering assumptions, reducing uncertainty in the SIT. However, as ICESat-2 estimates total freeboard (snow depth + ice freeboard), using it to calculate SIT still requires the use of a snow product, and the uncertainties coming from the snow loading for a laser altimeter product are higher than for a radar altimeter product (Kaminski et al., 2018). Additionally, tests with spatially and temporally varying radar penetration assumptions would improve the representation of changes in radar penetration in response to meteorological conditions found by previous studies.

Author Contribution Statement

Carmen Nab: Conceptualization, Formal Analysis, Investigation, Writing - Original Draft Preparation, Writing - Review & Editing. Davi Mignac: Conceptualization, Methodology, Supervision, Writing - Review & Editing. Jack Landy: Data Curation, Writing - Review & Editing. Matthew Martin: Writing - Review & Editing. Julienne Stroeve: Writing - Review & Editing. Michel Tsamados: Supervision, Writing - Review & Editing.

Acknowledgments

CN acknowledges support from NERC (#NE/S007229/1) and the UK Met Office (CASE Partnership). JL acknowledges support from the Research Council of Norway (#328957), ERC project SI/3D (#101077496), and ESA (#ESA/AO/1-10061/19/I-EF). JS acknowledges support from the Canada 150 Chair Program (#50296) and Horizon 2020 (#101003826). MT acknowledges support from ESA (#ESA/AO/1-9132/17/NL/MP & #ESA/AO/1-10061/19/I-EF,SINX'S,CLEV2ER) and NERC (#NE/T000546/1 & #NE/X004643/1).

Open Research

The NEMO ocean model and SI^3 sea ice model are available at `https://forge.nemo-ocean.eu/nemo` (Madec et al., 2023; Vancoppenolle et al., 2023). The AWI snow depth and radar freeboard data are available at `https://doi.org/10.5281/zenodo.10044553` (Hendricks et al., 2021). The LARM radar freeboard data are available at `https://doi.org/10.5285/cbd2cf78-462a-4968-be20-05f9c125ad10` (Landy & Stroeve, 2020). The CPOM radar freeboard data are available at `http://www.cpom.ucl.ac.uk/csopr/seaice.php` (Laxon et al., 2013). The SM-LG snow depth data are available at `https://doi.org/10.5067/27A0P5M6LZBI` (Liston et al., 2021). The OIB data are available at `https://doi.org/10.5067/19SIM5TXKPGT` (Krabill, 2013). The BGEP data are available at `https://www2.whoi.edu/site/beaufortgyre/data/mooring-data/` (BGEP, 2022). The IceBird total thickness and snow depth data are available at `https://doi.org/10.1594/PANGAEA.924848` (Hendricks et al., 2020) and `https://doi.org/10.1594/PANGAEA.932668` (Jutila et al., 2021) respectively. All code required to reproduce this analysis can be found at `https://github.com/carmennab/foam_sensitivity`.

References

Aaboe, S., Down, E., Sørensen, A., Lavergne, T., & Eastwood, S. (2021). *Sea-ice type climate data record Oct1978-Aug2023, v2.0. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).* ECMWF. doi: 10.24381/CDS.29C46D83

Armitage, T. W. K., & Ridout, A. L. (2015). Arctic sea ice freeboard from AltiKa and comparison with CryoSat-2 and Operation IceBridge. *Geophysical Research Letters*, *42*(16), 6724–6731. doi: 10.1002/2015GL064823

BGEP. (2022). *Mooring Data from the Beaufort Gyre Exploration Project (BGEP).* Retrieved from `https://www2.whoi.edu/site/beaufortgyre/data/mooring-data/`

Blockley, E., Fiedler, E., Ridley, J., Roberts, L., West, A., Copsey, D., ... Vancoppenolle, M. (2023). The sea ice component of GC5: coupling SI ^{3} to HadGEM3 using conductive fluxes. *EGUSphere*. doi: 10.5194/egusphere-2023-1731

Chen, Z., Liu, J., Song, M., Yang, Q., & Xu, S. (2017). Impacts of Assimilating Satellite Sea Ice Concentration and Thickness on Arctic Sea Ice Prediction in the NCEP Climate Forecast System. *Journal of Climate*, *30*(21), 8429–8446. doi: 10.1175/JCLI-D-17-0093.1

Fiedler, E. K., Martin, M. J., Blockley, E., Mignac, D., Fournier, N., Ridout, A., ... Tilling, R. L. (2022). Assimilation of sea ice thickness derived from CryoSat-2 along-track freeboard measurements into the Met Office's Forecast Ocean Assimilation Model (FOAM). *The Cryosphere*, *16*(1), 61–85. doi: 10.5194/tc-16-61-2022

Fritzner, S., Graversen, R., Christensen, K. H., Rostosky, P., & Wang, K. (2019). Impact of assimilating sea ice concentration, sea ice thickness and snow depth in a coupled ocean–sea ice modelling system. *The Cryosphere*, *13*(2), 491–509. doi: 10.5194/tc-13-491-2019

Gregory, W., Bushuk, M., Adcroft, A., Zhang, Y., & Zanna, L. (2023). Deep Learning of Systematic Sea Ice Model Errors From Data Assimilation Increments. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003757. doi: 10.1029/2023MS003757

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., ... Tietsche, S. (2016). A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 546–561. doi: 10.1002/qj.2401

Guerreiro, K., Fleury, S., Zakharova, E., Kouraev, A., Rémy, F., & Maisongrande, P. (2017). Comparison of CryoSat-2 and ENVISAT radar freeboard over Arctic sea ice: toward an improved Envisat freeboard retrieval. *The Cryosphere*, *11*(5), 2059–2073. doi: 10.5194/tc-11-2059-2017

Haas, C. (2003, May). Dynamics versus Thermodynamics: The Sea Ice Thickness Distribution. In D. N. Thomas & G. S. Dieckmann (Eds.), *Sea Ice* (1st ed., pp. 82–111). Wiley. doi: 10.1002/9780470757161.ch3

Hendricks, S., Ricker, R., Haas, C., & Herber, A. (2020). *Airborne sea ice plus snow thickness during the PAMARCMIP 2017 aircraft campaign in the Arctic Ocean.* doi: 10.1594/PANGAEA.924848

Hendricks, S., Ricker, R., & Paul, S. (2021). *Product User Guide & Algorithm Specification: AWI CryoSat-2 Sea Ice Thickness (version 2.4).* Retrieved from `https://epic.awi.de/id/eprint/54733/`

Jutila, A., King, J., Ricker, R., Hendricks, S., Helm, V., Binder, T., & Herber, A. (2021). *Airborne snow depth on sea ice during the PAMARCMIP2017 campaign in the Arctic Ocean, Version 1.* doi: 10.1594/PANGAEA.932668

Kaleschke, L., Tian-Kunze, X., Maaß, N., Mäkynen, M., & Drusch, M. (2012). Sea ice thickness retrieval from SMOS brightness temperatures during the

Arctic freeze-up period. *Geophysical Research Letters*, *39*(5), L05501. doi: 10.1029/2012GL050916

Kaminski, T., Kauker, F., Toudal Pedersen, L., Voßbeck, M., Haak, H., Niederdrenk, L., … Gråbak, O. (2018). Arctic Mission Benefit Analysis: impact of sea ice thickness, freeboard, and snow depth products on sea ice forecast performance. *The Cryosphere*, *12*(8), 2569–2594. doi: 10.5194/tc-12-2569-2018

Kern, S., Khvorostovsky, K., Skourup, H., Rinne, E., Parsakhoo, Z. S., Djepa, V., … Sandven, S. (2015). The impact of snow depth, snow density and ice density on sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise. *The Cryosphere*, *9*(1), 37–52. doi: 10.5194/tc-9-37-2015

Kilic, L., Tonboe, R. T., Prigent, C., & Heygster, G. (2019). Estimating the snow depth, the snow–ice interface temperature, and the effective temperature of Arctic sea ice using Advanced Microwave Scanning Radiometer 2 and ice mass balance buoy data. *The Cryosphere*, *13*(4), 1283–1296. doi: 10.5194/tc-13-1283-2019

Krabill, W. B. (2013). *IceBridge ATM L1B Elevation and Return Strength, Version 2.* NASA National Snow and Ice Data Center DAAC. doi: 10.5067/19SIM5TXKPGT

Krabill, W. B., Thomas, R. H., Martin, C. F., Swift, R. N., & Frederick, E. B. (1995). Accuracy of airborne laser altimetry over the Greenland ice sheet. *International Journal of Remote Sensing*, *16*(7), 1211–1222. doi: 10.1080/01431169508954472

Krishfield, R. A., Proshutinsky, A., Tateyama, K., Williams, W. J., Carmack, E. C., McLaughlin, F. A., & Timmermans, M.-L. (2014). Deterioration of perennial sea ice in the Beaufort Gyre from 2003 to 2012 and its impact on the oceanic freshwater cycle: SEA ICE IN THE BG FROM 2003 TO 2012. *Journal of Geophysical Research: Oceans*, *119*(2), 1271–1305. doi: 10.1002/2013JC008999

Kurtz, N. T., Farrell, S. L., Studinger, M., Galin, N., Harbeck, J. P., Lindsay, R., … Sonntag, J. G. (2013). Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data. *The Cryosphere*, *7*(4), 1035–1056. doi: 10.5194/tc-7-1035-2013

Kwok, R., & Cunningham, G. F. (2015). Variability of Arctic sea ice thickness and volume from CryoSat-2. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2045), 20140157. doi: 10.1098/rsta.2014.0157

Kwok, R., Kurtz, N. T., Brucker, L., Ivanoff, A., Newman, T., Farrell, S. L., … Tschudi, M. (2017). Intercomparison of snow depth retrievals over Arctic sea ice from radar data acquired by Operation IceBridge. *The Cryosphere*, *11*(6), 2571–2593. doi: 10.5194/tc-11-2571-2017

Landy, J. C., Dawson, G. J., Tsamados, M., Bushuk, M., Stroeve, J. C., Howell, S. E. L., … Aksenov, Y. (2022). A year-round satellite sea-ice thickness record from CryoSat-2. *Nature*, *609*(7927), 517–522. doi: 10.1038/s41586-022-05058-5

Landy, J. C., Petty, A. A., Tsamados, M., & Stroeve, J. C. (2020). Sea Ice Roughness Overlooked as a Key Source of Uncertainty in CryoSat-2 Ice Freeboard Retrievals. *Journal of Geophysical Research: Oceans*, *125*(5). doi: 10.1029/2019JC015820

Landy, J. C., & Stroeve, J. (2020). *Arctic sea ice and physical oceanography derived from CryoSat-2 Baseline-C Level 1b waveform observations, Oct-Apr 2010-2018.* UK Polar Data Centre, Natural Environment Research Council, UK Research & Innovation. doi: 10.5285/CBD2CF78-462A-4968-BE20-05F9C125AD10

Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., … Davidson, M. (2013). CryoSat-2 estimates of Arctic sea ice thickness and volume. *Geophysical Research Letters*, *40*(4), 732–737. doi: 10.1002/grl.50193

Liston, G., Itkin, P., Stroeve, J., Tschudi, M., Stewart, J. S., Pedersen, S. H., …

Elder, K. (2020). A Lagrangian Snow-Evolution System for Sea-Ice Applications (SnowModel-LG): Part I—Model Description. *Journal of Geophysical Research: Oceans*, *125*(10). doi: 10.1029/2019JC015913

Liston, G., Stroeve, J., & Itkin, P. (2021). *Lagrangian Snow Distributions for Sea-Ice Applications, Version 1. ERA5 subset.* NASA National Snow and Ice Data Center DAAC. doi: 10.5067/27A0P5M6LZBI

Madec, G., Bell, M., Blaker, A., Bricaud, C., Bruciaferri, D., Castrillo, M., … Wilson, C. (2023). *NEMO Ocean Engine Reference Manual* (Tech. Rep.). (Version Number: v4.2.1) doi: 10.5281/ZENODO.8167700

Mallett, R. D. C., Lawrence, I. R., Stroeve, J. C., Landy, J. C., & Tsamados, M. (2020). Brief communication: Conventional assumptions involving the speed of radar waves in snow introduce systematic underestimates to sea ice thickness and seasonal growth rate estimates. *The Cryosphere*, *14*(1), 251–260. doi: 10.5194/tc-14-251-2020

Meier, W. N., Stroeve, J., & Fetterer, F. (2007). Whither Arctic sea ice? A clear signal of decline regionally, seasonally and extending beyond the satellite record. *Annals of Glaciology*, *46*, 428–434. doi: 10.3189/172756407782871170

Mignac, D., Martin, M., Fiedler, E., Blockley, E., & Fournier, N. (2022). Improving the Met Office's Forecast Ocean Assimilation Model (FOAM) with the assimilation of satellite-derived sea-ice thickness data from CryoSat-2 and SMOS in the Arctic. *Quarterly Journal of the Royal Meteorological Society*, qj.4252. doi: 10.1002/qj.4252

Nab, C., Mallett, R., Gregory, W., Landy, J., Lawrence, I. R., Willatt, R., … Tsamados, M. (2023). Synoptic Variability in Satellite Altimeter-Derived Radar Freeboard of Arctic Sea Ice. *Geophysical Research Letters*. doi: 10.1029/2022GL100696

Nandan, V., Geldsetzer, T., Yackel, J., Mahmud, M., Scharien, R., Howell, S., … Else, B. (2017). Effect of Snow Salinity on CryoSat-2 Arctic First-Year Sea Ice Freeboard Measurements: Sea Ice Brine-Snow Effect on CryoSat-2. *Geophysical Research Letters*, *44*(20), 10,419–10,426. doi: 10.1002/2017GL074506

Nandan, V., Willatt, R., Mallett, R., Stroeve, J., Geldsetzer, T., Scharien, R., … Hoppmann, M. (2023). Wind redistribution of snow impacts the Ka- and Ku-band radar signatures of Arctic sea ice. *The Cryosphere*, *17*(6), 2211–2229. doi: 10.5194/tc-17-2211-2023

Ricker, R., Hendricks, S., Helm, V., Skourup, H., & Davidson, M. (2014). Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation. *The Cryosphere*, *8*(4), 1607–1622. doi: 10.5194/tc-8-1607-2014

Rothrock, D. A. (2003). The arctic ice thickness anomaly of the 1990s: A consistent view from observations and models. *Journal of Geophysical Research*, *108*(C3), 3083. doi: 10.1029/2001JC001208

Shen, X., Ke, C.-Q., Xie, H., Li, M., & Xia, W. (2020). A comparison of Arctic sea ice freeboard products from Sentinel-3A and CryoSat-2 data. *International Journal of Remote Sensing*, *41*(7), 2789–2806. doi: 10.1080/01431161.2019.1698078

Sievers, I., Rasmussen, T. A. S., & Stenseng, L. (2023). Assimilating CryoSat-2 freeboard to improve Arctic sea ice thickness estimates. *The Cryosphere*, *17*(9), 3721–3738. doi: 10.5194/tc-17-3721-2023

Storkey, D., Blaker, A. T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., … Sinha, B. (2018). UK Global Ocean GO6 and GO7: a traceable hierarchy of model resolutions. *Geoscientific Model Development*, *11*(8), 3187–3213. doi: 10.5194/gmd-11-3187-2018

Tian-Kunze, X., Kaleschke, L., Maaß, N., Mäkynen, M., Serra, N., Drusch, M., & Krumpen, T. (2014). SMOS-derived thin sea ice thickness: algorithm baseline, product specifications and initial verification. *The Cryosphere*, *8*(3), 997–1018.

doi: 10.5194/tc-8-997-2014
Tilling, R. L., Ridout, A., & Shepherd, A. (2018). Estimating Arctic sea ice thickness and volume using CryoSat-2 radar altimeter data. *Advances in Space Research*, *62*(6), 1203–1225. doi: 10.1016/j.asr.2017.10.051
Tsamados, M., Feltham, D., Petty, A., Schroeder, D., & Flocco, D. (2015, October). Processes controlling surface, bottom and lateral melt of Arctic sea ice in a state of the art sea ice model. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2052), 20140167. doi: 10.1098/rsta.2014.0167
Vancoppenolle, M., Rousset, C., Blockley, E., Aksenov, Y., Feltham, D., Fichefet, T., ... Tietsche, S. (2023, January). *SI3, the NEMO Sea Ice Engine.* doi: 10.5281/ZENODO.7534900
Wang, Q., Ilicak, M., Gerdes, R., Drange, H., Aksenov, Y., Bailey, D. A., ... Yeager, S. G. (2016). An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part I: Sea ice and solid freshwater. *Ocean Modelling*, *99*, 110–132. doi: 10.1016/j.ocemod.2015.12.008
Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A., & While, J. (2015). Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 333–349. doi: 10.1002/qj.2388
Willatt, R., Giles, K., Laxon, S. W., Stone-Drake, L., & Worby, A. (2010). Field Investigations of Ku-Band Radar Penetration Into Snow Cover on Antarctic Sea Ice. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(1), 365–372. doi: 10.1109/TGRS.2009.2028237
Williams, N., Byrne, N., Feltham, D., Van Leeuwen, P. J., Bannister, R., Schroeder, D., ... Nerger, L. (2023). The effects of assimilating a sub-grid-scale sea ice thickness distribution in a new Arctic sea ice data assimilation system. *The Cryosphere*, *17*(6), 2509–2532. doi: 10.5194/tc-17-2509-2023
Xia, W., & Xie, H. (2018). Assessing three waveform retrackers on sea ice freeboard retrieval from Cryosat-2 using Operation IceBridge Airborne altimetry datasets. *Remote Sensing of Environment*, *204*, 456–471. doi: 10.1016/j.rse.2017.10.010
Zhang, Y., Bushuk, M., Winton, M., Hurlin, B., Gregory, W., Landy, J., & Jia, L. (2023). Improvements in September Arctic Sea Ice Predictions Via Assimilation of Summer CryoSat-2 Sea Ice Thickness Observations. *Geophysical Research Letters*, *50*(24), e2023GL105672. doi: 10.1029/2023GL105672
Zhang, Y.-F., Bitz, C. M., Anderson, J. L., Collins, N., Hendricks, J., Hoar, T., ... Massonnet, F. (2018). Insights on Sea Ice Data Assimilation from Perfect Model Observing System Simulation Experiments. *Journal of Climate*, *31*(15), 5911–5926. doi: 10.1175/JCLI-D-17-0904.1

Supporting Information for: Sensitivity of short-range forecasts to sea ice thickness data assimilation parameters in a coupled ice-ocean system

Carmen Nab[1,2], Davi Mignac[2], Jack Landy[3], Matthew Martin[2], Julienne Stroeve[1,4,5], Michel Tsamados[1]

[1]Centre for Polar Observation and Modelling, University College London, London, UK

[2]Ocean Forecasting Research & Development, Met Office, Exeter, UK

[3]Department of Physics and Technology, UiT The Arctic University of Norway, Norway

[4]Centre for Earth Observation Science, University of Manitoba, Canada

[5]National Snow and Ice Data Center, University of Colorado Boulder, USA

Contents of this file

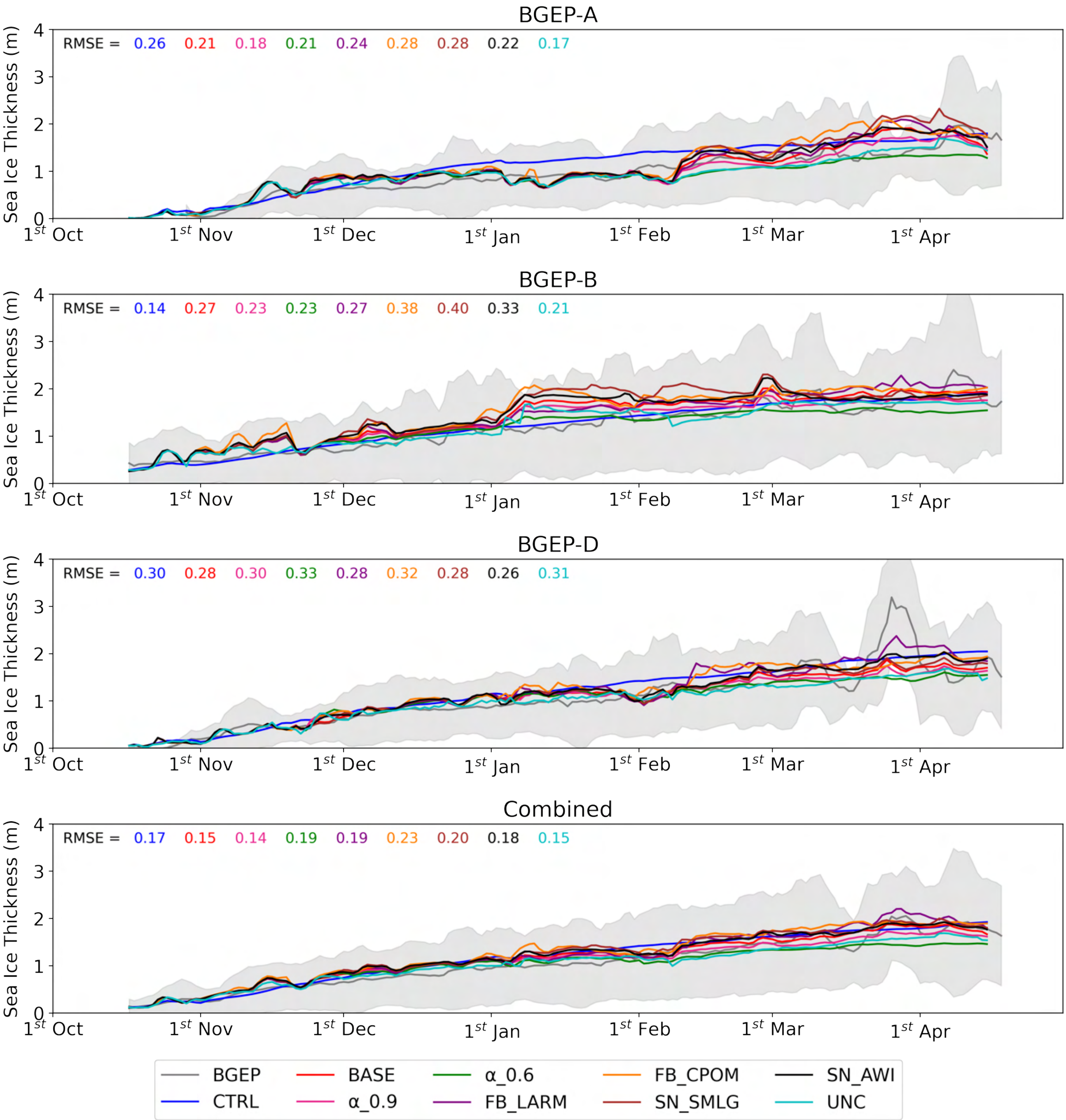

Figure S1. FOAM SIT and one-week running mean of BGEP SIT at BGEP-A, BGEP-B and BGEP-D mooring locations from 25 November 2016 to 15 April 2017. The grey shaded area shows the BGEP uncertainty, represented by the daily standard deviation. Coloured text shows the mean root-mean-squared error for the season for each experiment.

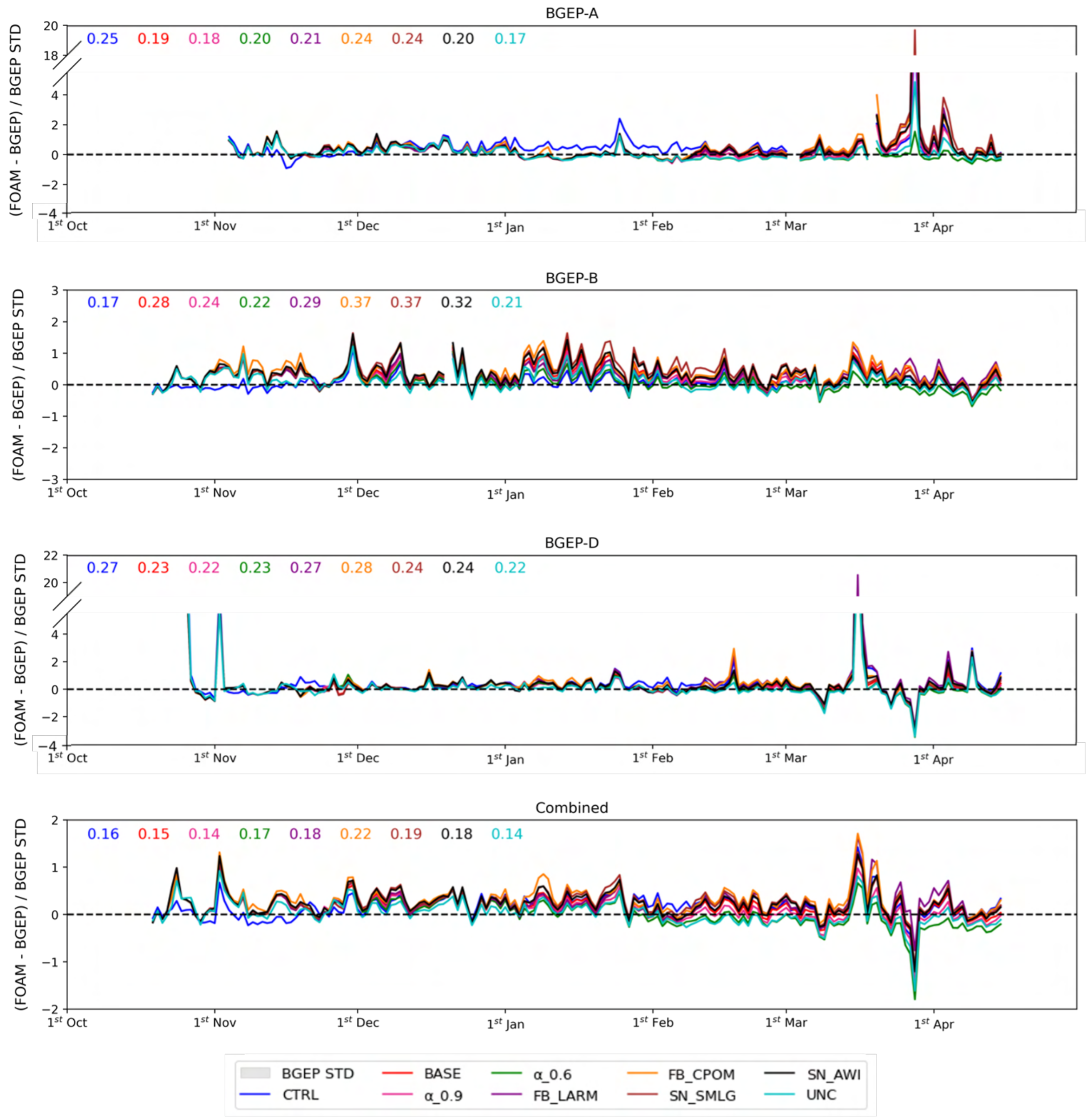

Figure S2. Daily difference between FOAM SIT and BGEP SIT at BGEP-A, BGEP-B and BGEP-D mooring locations, normalised by the BGEP standard deviation. Coloured text shows the mean absolute value for each experiment. Note the different y-axes.

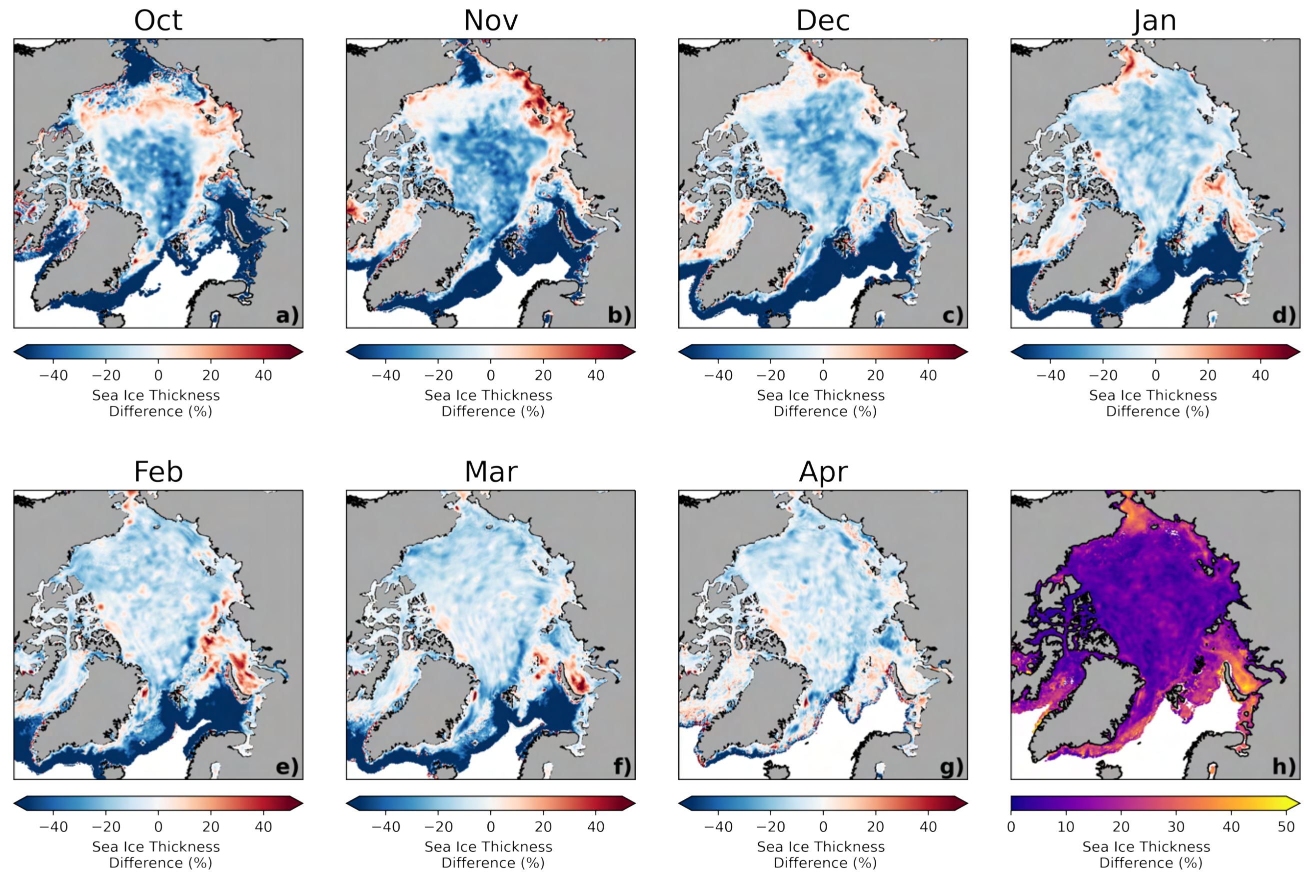

Figure S3. **a-g)** Monthly difference between mean daily sea ice thickness in the UNC experiment and the BASE experiment (UNC-BASE) **h)** standard deviation of the monthly differences.

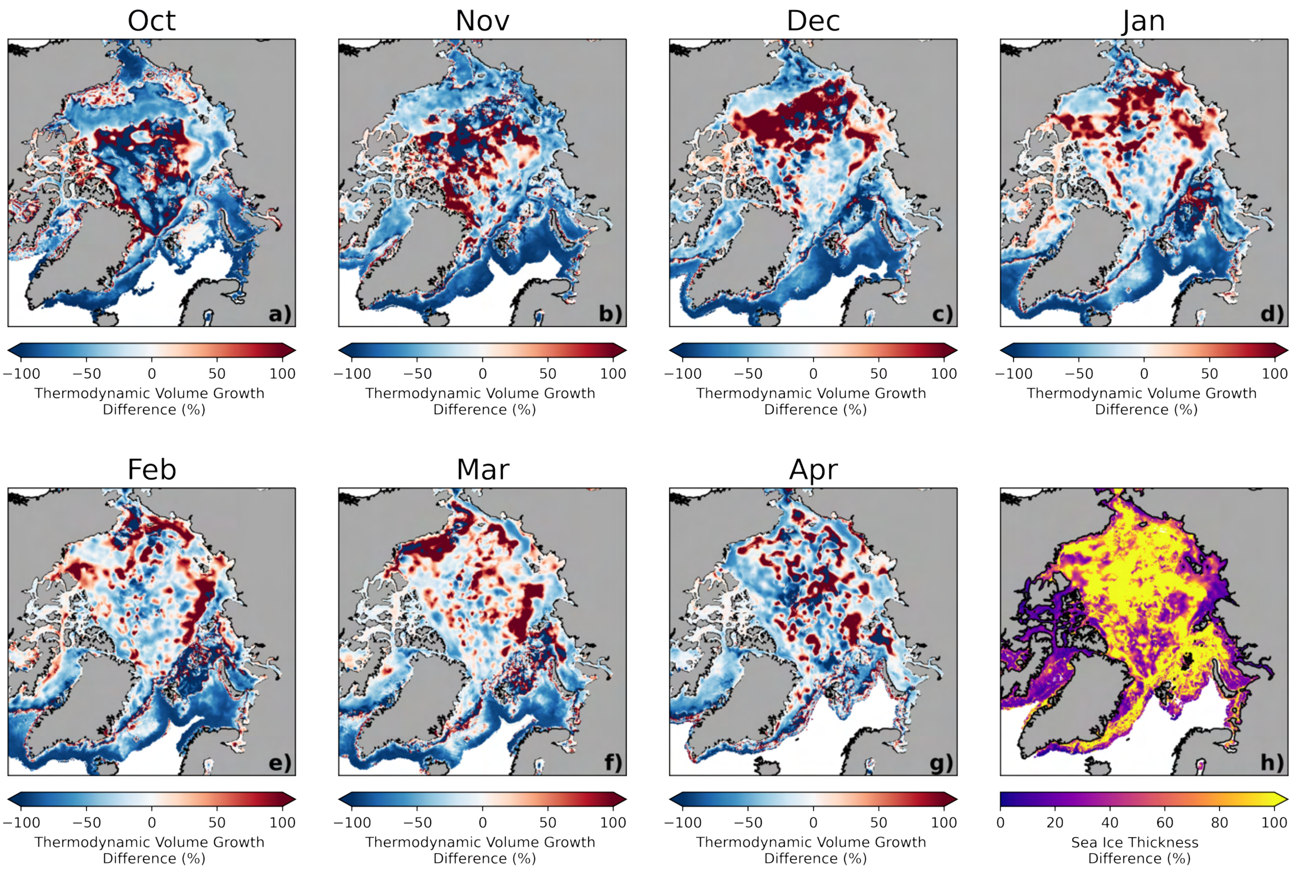

Figure S4. **a-g)** Monthly difference between mean daily thermodynamic volume change in the UNC experiment and the BASE experiment (UNC-BASE) **h)** standard deviation of the monthly differences.

1. IceBird

As part of the IceBird campaign series, total thickness (snow depth + sea ice thickness) measurements were collected during the PAMARCMIP campaign in March and April 2017 (Hendricks et al., 2020). The data was gathered using the AWI EM-Bird, a helicopter-borne electromagnetic (EM) induction system on flights in the Beaufort Sea and the Chukchi Sea. The system works by sending a low-frequency EM field through the sea ice, generating eddy currents in the water below. This induces a secondary EM field that propagates back upwards through the sea ice, the strength of which is directly related to the distance between the EM system and the underside of the sea ice. Using a laser, the distance between the EM system and the snow surface is then measured, such that the total thickness can be derived as the difference between the laser-determined snow surface and the EM-determined underlying water surface. See Haas et al. (2009) for a full description of this method. On the same flights, snow depth measurements were taken using the AWI Snow Radar (Jutila et al., 2021), an ultrawideband microwave radar able to estimate snow depth on sea ice with a mean bias of 0.86 cm. See Jutila et al. (2022) for a full description of this method. The high-frequency IceBird snow depth and total thickness data are binned into 12 km grid cells (equivalent to the FOAM grid size in the Arctic), with the median taken to avoid the impact of outliers. To determine the sea ice thickness, we take the difference between the coincident total thickness and snow depth measurements. Only points where data for both the total thickness and snow depth are available are used. For the evaluation, the model SITs are then interpolated to these observation locations.

With the exception of the $\alpha\_0.6$ and UNC experiments, we find a consistent overestimation in the model SIT compared to the IceBird SIT everywhere except at the ice edge (Figure S5). Similarly to the OIB comparison, we find a wider range of values in the IceBird data than in the model results, particularly at SIT values below 1 m and above 3 m (Figure S6). We find a negative correlation between the IceBird data and each of the experiments (Figure S6). Each of the assimilation experiments performs worse than the CTRL experiment in terms of the magnitude of the r value. The experiments perform similarly in terms of RMSE (0.80 - 0.85 m), with the exception of the SN_SMLG and SN_AWI experiments (0.91 and 0.96 m, respectively).

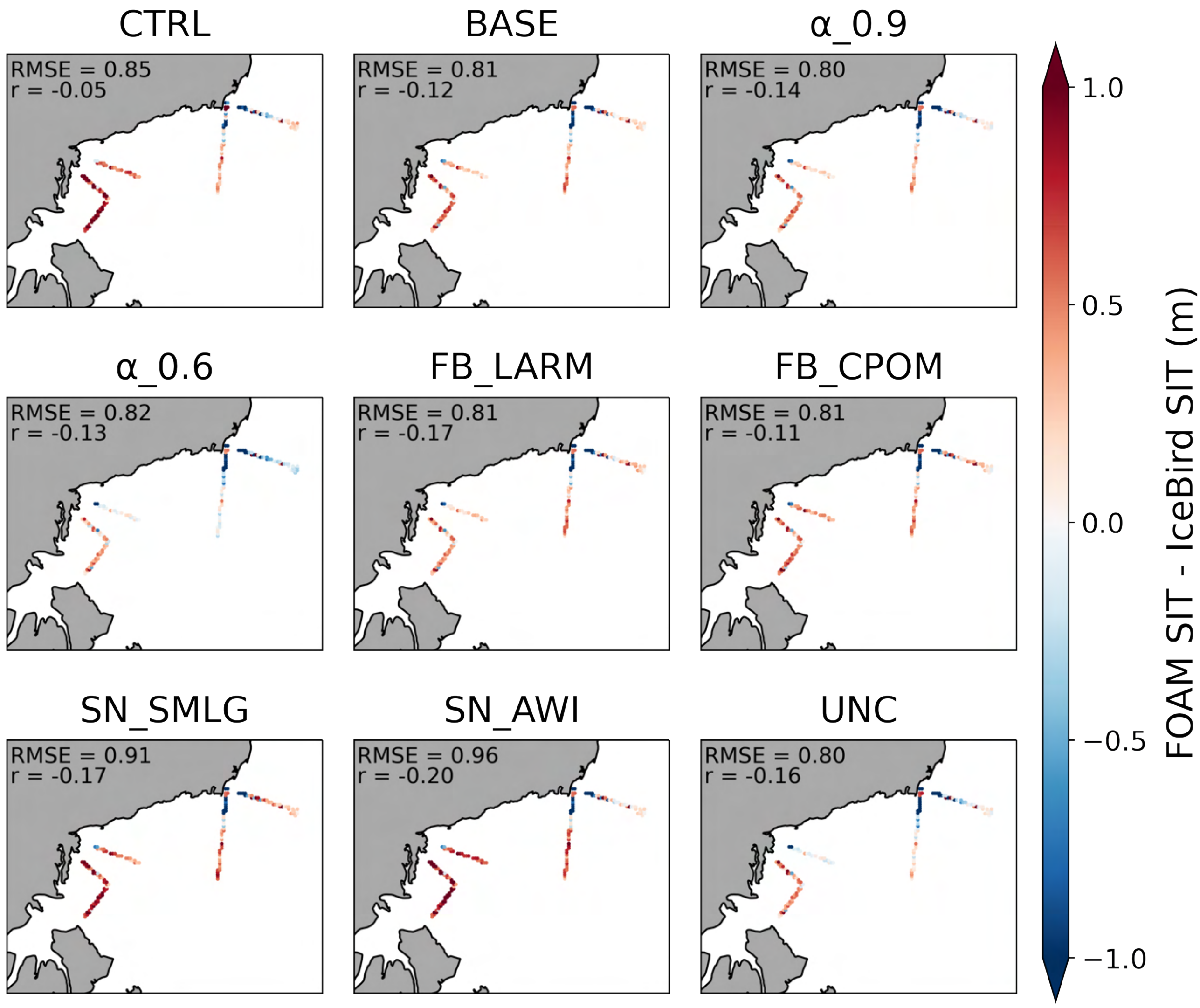

Figure S5. Difference between FOAM SIT and IceBird SIT, between 9 March and 19 April 2017.

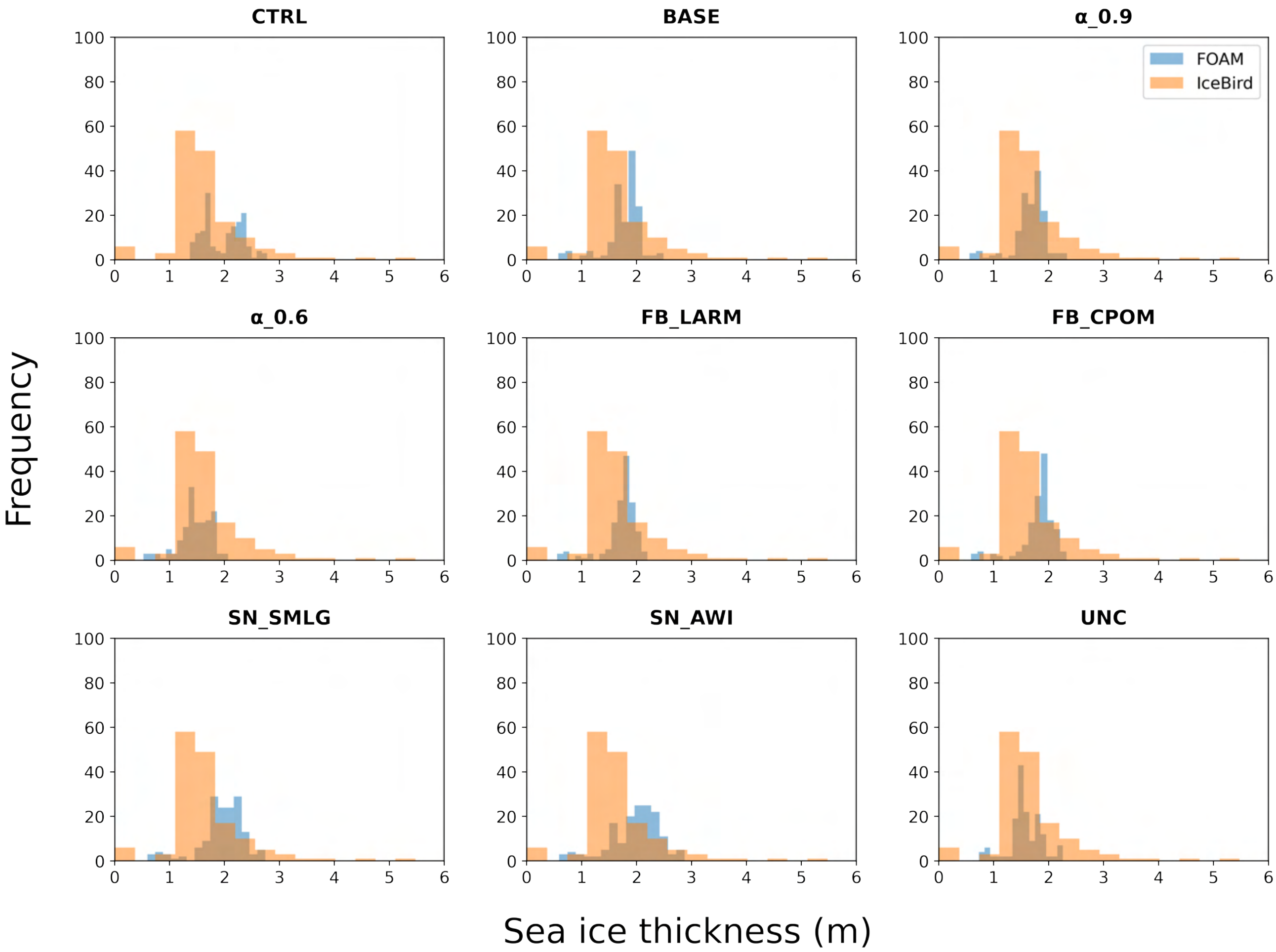

Figure S6. Histograms of FOAM SIT and IceBird SIT in the Beaufort and Chukchi Seas, between 30 March and 8 April 2017.

References

Haas, C., Lobach, J., Hendricks, S., Rabenstein, L., & Pfaffling, A. (2009). Helicopter-borne measurements of sea ice thickness, using a small and lightweight, digital EM system. *Journal of Applied Geophysics*, *67*(3), 234–241. doi: 10.1016/j.jappgeo.2008.05.005

Hendricks, S., Ricker, R., Haas, C., & Herber, A. (2020). *Airborne sea ice plus snow thickness during the PAMARCMIP 2017 aircraft campaign in the Arctic Ocean.* doi: 10.1594/PANGAEA.924848

Jutila, A., King, J., Paden, J., Ricker, R., Hendricks, S., Polashenski, C., … Haas, C. (2022). High-Resolution Snow Depth on Arctic Sea Ice From Low-Altitude Airborne Microwave Radar Data. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–16. doi: 10.1109/TGRS.2021.3063756

Jutila, A., King, J., Ricker, R., Hendricks, S., Helm, V., Binder, T., & Herber, A. (2021). *Airborne snow depth on sea ice during the PAMARCMIP2017 campaign in the Arctic Ocean, Version 1.* doi: 10.1594/PANGAEA.932668