



Machine-learning bias correction of passive-microwave sea ice concentration against a SAR-derived product: impact on Arctic sea ice forecasts

5 Davi Mignac¹, Matthew J. Martin¹

¹Met Office, Exeter, EX1 3PB, United Kingdom

Correspondence to: Davi Mignac (davi.carneiro@metoffice.gov.uk)

Abstract

Sea ice concentration (SIC) assimilation in operational systems relies predominantly on passive-microwave satellite retrievals, which present well-known limitations, such as smoothing of the ice edge and underestimation of SIC during the melt season. We develop a Random Forest model that bias-corrects AMSR2 SIC observations towards an existing deep learning, SAR-derived product (ML-SAR) in the Arctic. The bias-corrected AMSR2 observations, together with ML-SAR, are then assimilated into the 1/4° configuration of the Met Office Forecast Ocean Assimilation Model (FOAM) as the key experiment. Assimilating ML-SAR and bias-corrected AMSR2 produces a sharper ice edge during the ice growing season and reduces the 5-day forecast Integrated Ice Edge Error by 30% in summer, relative to a control experiment assimilating only AMSR2 alongside the standard ocean observations. It also largely mitigates the SIC underestimation in the melt season, raising the Critical Success Index against the ice charts in all SIC categories and at all 5-day forecast lead times, by about 25% on average, with respect to the control. Because the bias-corrected AMSR2 is assimilated wherever ML-SAR is absent, sea ice forecast improvements reach the whole Arctic, including regions of limited SAR coverage, and could not be obtained by assimilating ML-SAR alone. The additional ice volume produced by assimilating both ML-SAR and bias-corrected AMSR2, particularly in spring and summer, contributes to alleviating a known thin bias in the central Arctic in FOAM, but also leads to an overestimation of thickness in the Arctic marginal seas. Overall, the results demonstrate that, by combining the spatial completeness of passive-microwave observations with the fidelity of SAR-based data, this learned bias correction offers a computationally inexpensive, observation-based pathway towards improved Arctic sea ice prediction.

25 1 Introduction

In the face of substantial reductions in Arctic sea ice extent over the past few decades (Stroeve and Notz, 2018; Meredith et al., 2019), there is a growing need for accurate sea ice predictions across timescales ranging from days to seasons and beyond. The retreat of Arctic sea ice has wide-ranging implications: it may modulate weather and climate at lower latitudes (Pedersen et al., 2016), alter the polar ecosystem (Meier et al., 2014), and open new shipping routes (Smith and Stephenson, 2013; Wei et al., 2020) that expose maritime operations to an increasingly variable and mobile ice cover (Eicken, 2013).



Sea ice concentration (SIC) is among the few sea ice variables for which a long and homogeneous observational record exists, owing to passive-microwave satellite retrievals available since 1979. Therefore, SIC has been assimilated extensively into stand-alone sea ice models (Thomas et al., 1996), coupled ocean–sea ice models (Lisæter et al., 2003; Peterson et al., 2015; Posey et al., 2015) and fully coupled ocean–atmosphere–sea ice systems (Lea et al., 2015; Guiavarc'h et al., 2019; Skachko, et al., 2019; Barton et al., 2021; de Rosnay et al., 2022). As a result, SIC data assimilation (DA) is now well established and routine at many operational centres (Bertino and Lisæter, 2008; Posey et al., 2015; Lemieux et al., 2016; Mignac et al., 2025). Passive-microwave SIC retrievals, however, are subject to well-documented limitations that arise primarily from surface conditions. Foremost is a pronounced summertime low bias associated with melt ponds, whose microwave signature is difficult to distinguish from open water and which can induce concentration errors exceeding 20 % (Kern et al., 2016, 2020). Snow and slush (Andersen et al., 2007; Shokr and Kaleschke, 2012), the limited detectability of new and thin ice (Shokr and Kaleschke, 2012), land contamination near coastlines and weather-related effects over open water (Andersen et al., 2007) constitute additional error sources. Moreover, operational passive-microwave products are relatively coarse and spatially smooth, their resolution is limited by the large footprint size and they are typically gridded at ~10 km or coarser resolution (Spreen et al., 2008; Lavergne et al., 2019). They therefore tend to blur the sharp gradients that characterise the ice edge, where retrieval uncertainties are also largest (Ivanova et al., 2015).

Synthetic-aperture radar (SAR) resolves the ice cover at far higher spatial resolution, typically at ~40 m (Wang et al., 2021), and machine learning has rapidly become the central tool for exploiting it. Deep convolutional neural networks trained on manually produced ice charts can retrieve sea ice parameters directly from Sentinel-1 SAR imagery, overcoming the ambiguous backscatter signatures that had long hindered automated SAR interpretation (Stokholm et al., 2022). Malmgren-Hansen et al. (2021) showed that fusing SAR with coarser passive-microwave observations in a convolutional neural network that ingests both as inputs further improves accuracy and robustness. The passive-microwave signal acts as a low-resolution constraint that helps resolve the ambiguity of the SAR backscatter (e.g. wind-roughened open water that can mimic ice), while the SAR provides the fine-scale detail and edge position. Together they yield concentrations that agree closely with the ice charts, despite a difference in sensor footprint of roughly three orders of magnitude. Building on these developments, deep learning SIC products derived from Sentinel-1 and the Radarsat Constellation Mission (RCM) are now emerging operationally. They resolve the ice edge and the marginal ice zone (MIZ) in fine detail, approaching the accuracy of ice charts and providing calibrated uncertainties (Wulf et al., 2024, 2026). Their coverage, however, is fundamentally limited by the Sentinel-1 and RCM acquisition patterns. On any given day the SAR-derived fields are spatially incomplete and intermittent, with gaps that shift from day to day, in contrast to the near-complete daily coverage of passive-microwave retrievals.

Passive-microwave and SAR-derived SIC can thus be strongly complementary. The former provides near-complete, temporally consistent coverage, whereas the latter provides local accuracy and well-resolved gradients. This complementarity motivates the use of the SAR-derived product as a reference against which the passive-microwave observations can be bias-corrected, retaining Arctic-wide coverage while inheriting SAR-like accuracy. In addition to the commonly used variational schemes, machine learning techniques, such as Random Forests and neural networks, have been emerging as robust alternatives



65 to bias-correct observations, mostly in the atmospheric domain so far (Jin et al., 2019; Zhou & Grassotti, 2020; Zhang et al., 2024). In the sea ice domain, machine learning has been directed at model error, learning and correcting systematic sea ice model biases from DA increments (Gregory et al., 2023, 2024), rather than at the observations themselves. Even purely data-driven, machine-learning sea ice forecasting systems rely on passive-microwave SIC as their target and reference (e.g. Palerme et al., 2024), therefore inheriting the same observational limitations.

70 While the convolutional network of Malmgren-Hansen et al. (2021) fuses SAR and passive-microwave observations and may implicitly compensate for passive-microwave biases, it does not explicitly correct the passive-microwave SIC observations, nor does it yield a spatially complete, bias-corrected product. To our knowledge, the explicit machine-learning bias correction of passive-microwave sea ice observations towards a high-resolution SAR-derived product, and its impact once assimilated in a global forecasting system, has not yet been explored. In this study, a Random Forest model is trained to bias-correct SIC
75 from the Advanced Microwave Scanning Radiometer 2 (AMSR2) towards a machine-learning SAR product (hereafter ML-SAR; Wulf et al., 2024). The impacts of assimilating the bias-corrected AMSR2 and ML-SAR on Arctic forecasts are then evaluated in the Met Office Forecast Ocean Assimilation Model (FOAM; Mignac et al., 2025) through 10-month experiments assimilating the following SIC observations: AMSR2 Level-3 data (AMSR2-L3), as in FOAM's current operational configuration; AMSR2 Level-2 data (AMSR2-L2); AMSR2-L2 in conjunction with ML-SAR; and the bias-corrected AMSR2-
80 L2 in conjunction with ML-SAR.

The paper is outlined as follows. Section 2 describes the methodology: the FOAM system (Sect. 2.1), the new SIC observations (Sect. 2.2), the Random Forest bias correction of AMSR2 (Sect. 2.3) and the experiment setup and assessment metrics (Sect. 2.4). Section 3 presents the results, covering the impacts of the SIC assimilation on the sea ice state (Sect. 3.1), the validation of the 1-day forecasts against ice charts (Sect. 3.2), the impacts on the sea ice thickness (SIT) (Sect. 3.3) and the assessment of
85 5-day sea ice forecasts (Sect. 3.4). Section 4 summarises the main findings and discusses their implications.

2 Methods

2.1 The FOAM system

The Forecast Ocean Assimilation Model (FOAM; Mignac et al., 2025) is the Met Office's operational, coupled ocean–sea ice system. FOAM produces global analyses and 6-day forecasts of ocean and sea ice variables each day. FOAM sea ice and ocean
90 components are also employed in a coupled ocean–sea ice–land–atmosphere short-range forecasting system (Guiavarc'h et al., 2019) and analyses from this fully coupled system are used operationally to initialise the Met Office's seasonal prediction system, GloSea (Kettleborough et al., 2025). Any developments implemented in FOAM will therefore also benefit these coupled short-range and seasonal prediction systems.

The FOAM configuration used in this work is based on the Global Ocean and Sea Ice version 9 (GOSI9), the latest Met Office
95 global ocean–sea ice configuration, whose model and DA changes relative to the previous version are described in detail by



Mignac et al. (2025). GOSI9 employs version 4.0.4 of the NEMO ocean model (Nucleus for European Modelling of the Ocean; Madec et al., 2017), coupled to the native NEMO sea-ice model SI³, which supersedes the CICE model used previously and is derived from the Louvain-la-Neuve sea-ice model (LIM3) with functionality merged from CICE (Blockley et al., 2024). The operational FOAM system has recently been upgraded from 1/4° to 1/12° horizontal resolution (Barbosa Aguiar et al., 2024),
100 but the 1/4° version is used here. Both the ocean and sea ice components use the quasi-isotropic tripolar ORCA grid (Madec and Imbard, 1996) with the 1/4° version called ORCA025. NEMO is configured with a non-linear free surface and 75 vertical levels, with 1 m vertical resolution near the surface and decreasing resolution with depth (Storkey et al., 2018). SI³ is configured to use five ice-thickness categories (plus open water), an elastic–viscous–plastic rheology and multi-layer, energy-conserving thermodynamics, with the sea ice physics otherwise largely unchanged from the previous CICE-based configuration (Ridley
105 et al., 2018; Blockley et al., 2024). FOAM is forced at the surface by hourly winds and 3-hourly heat, freshwater and radiative fluxes from the operational Met Office Unified Model (MetUM; Willet et al., 2026) global Numerical Weather Prediction (NWP) system, at a spatial resolution of ~10 km over the study period (October 2024 to July 2025).

Sea ice concentration and ocean observations are assimilated using NEMOVAR, an incremental, multi-variate, multi-length-scale scheme configured in FOAM as a 3D-Var FGAT scheme (First Guess at Appropriate Time; Waters et al., 2015), with
110 the increments added over the analysis window through an Incremental Analysis Update (Bloom et al., 1996). A 24-hour window is used to assimilate observations of sea surface temperature (SST), sea level anomaly (SLA), temperature and salinity profiles, as well as SIC. SST observations come from in situ sources, such as ships, moored and drifting buoys, and from satellite sensors including AMSR2, AVHRR, SEVIRI, SLSTR and VIIRS. SLA observations are along-track altimetry from AltiKa, CryoSat-2, HaiYang-2A, Jason-2/3 and Sentinel-3A/3B. Temperature and salinity profiles are obtained from Argo
115 floats, moored and drifting buoys, gliders, CTDs, marine-mammal sensors and XBTs. Finally, satellite SIC from AMSR2-L3 are provided by the Ocean and Sea Ice Satellite Application Facility (OSI SAF) of EUMETSAT. The full list of instruments, references and data sources is given in Tab. 2 of Barbosa Aguiar et al. (2024). Note that Tab. 2 lists SSMIS as the SIC source, since the operational switch to AMSR2-L3 was only made in late 2025.

SIC observations are currently assimilated through a NEMOVAR minimisation that is separate from that of the ocean variables.
120 Within SI³, positive concentration increments are added to the thinnest ice category with an assumed SIT of 0.45 m—the upper limit of that category—whereas negative increments are first removed from the thinnest category until its concentration reaches zero and then progressively removed from thicker categories (Peterson et al., 2015). The resulting changes in SIC are then used to adjust all other SI³ prognostic variables proportionally, including ice and snow volume, ice and snow enthalpy, and melt pond fraction and volume, so that consistent values are maintained after the increments are added. See Mignac et al.
125 (2025) for further details. Note that SIT is not assimilated in the experiments presented here. SIT is adjusted only indirectly, through the SIC increments described above. A CryoSat-2 and SMOS SIT assimilation capability has been developed in FOAM (Mignac et al., 2022) but is not used in this study.



2.2 New SIC observations

Operationally, FOAM assimilates the L3 gridded product OSI-408-a (<https://osi-saf.eumetsat.int/products/osi-408-a>), a daily
 130 composite of AMSR2 sea-ice concentration provided on a 10 km polar-stereographic grid. In all the new DA experiments
 presented here we instead use, and further process, the L2 swath product OSI-410-a (<https://osi-saf.eumetsat.int/products/osi-410-a>),
 which is also sampled at 10 km. In both products the SIC is retrieved from AMSR2 brightness temperatures using the
 OSI SAF hybrid algorithm, which combines the 18.7 GHz (vertical) and 36.5 GHz (vertical and horizontal) channels, whose
 footprints are approximately 14×22 km and 7×12 km, respectively (OSI SAF, 2023). The retrieval is referenced to
 135 dynamically updated open-water and sea-ice tie-points, and the brightness temperatures are first corrected for the atmospheric
 contribution, using a radiative-transfer model driven by ECMWF NWP fields, and for land spillover near the coasts (Maaß
 and Kaleschke, 2010; Lavergne et al., 2019). A per-observation uncertainty is provided with each retrieval and is computed as
 the sum in quadrature of two independent components: a tie-point uncertainty, which accounts for residual atmospheric noise,
 sensor noise and the variability of the ice surface emissivity, and a representativeness (smearing) error arising from the
 140 mismatch between the sensor footprint and the target grid (OSI SAF, 2023). AMSR2-L2 is disseminated about 3 h 40 min
 after sensing and is updated up to ~15 times per day, whereas the daily AMSR2-L3 product is centred at 12:00 UTC and
 delivered about 5 h after the last contributing observation. For the 24-hour DA window used in FOAM this difference is of
 limited consequence. However, the sub-daily sampling and shorter latency of the L2 product are particularly advantageous for
 the Met Office coupled NWP system (Guiavarc'h et al., 2019), which employs 6-hourly DA windows and therefore benefits
 145 from observations delivered closer to real time and distributed throughout the day.

The high-resolution reference used to correct the AMSR2 retrievals in this work is a ML-SAR product, generated by the Danish
 Meteorological Institute's Automated Sea Ice Products system (DMI-ASIP; Wulf et al., 2024, 2026). ML-SAR retrieves sea
 ice concentration from SAR imagery in near-real time, with AMSR2 brightness temperatures as an auxiliary input, using an
 ensemble of convolutional neural networks based on a modified U-Net. Although the network was trained solely on Sentinel-
 150 1 scenes matched to regional ice charts over 2018–2021, RCM imagery is additionally incorporated in the near real-time
 product, increasing the SAR coverage. The SAR imagery is first resampled onto an 80 m grid, on which each pixel is classified
 into eleven concentration classes in 10 % steps from 0 % (open water) to 100 % (fully ice-covered). SIC is then taken as the
 class-probability-weighted mean of the bins and its per-pixel uncertainty as the standard deviation of that distribution. The
 class probabilities are obtained from the network ensemble and post-hoc calibrated by vector scaling on a held-out validation
 155 set (see Wulf et al., 2024 for more details).

Here, we use the L3 Arctic Ocean product from DMI-ASIP, distributed through the Copernicus Marine Service (CMEMS;
<https://doi.org/10.48670/mds-00343>). The L3 ML-SAR is a daily pan-Arctic mosaic assembled from individual SAR
 acquisitions without time averaging and provided on a 500 m polar-stereographic grid. The near real-time product used here
 covers October 2024 onwards and is delivered at 12:00 UTC on the following day. Because the mosaic is built only from the
 160 available SAR scenes from Sentinel-1 and RCM, its coverage is spatially incomplete and varies from day to day. Currently,



only the L3 gridded product is distributed through CMEMS, although a L2 version is planned to be made available by DMI in the future. As seen in Fig. 1, the resulting concentrations of ML-SAR agree closely with the ice charts and resolve the MIZ in greater detail than AMSR2, particularly the sharpness of the ice edge.

Ice charts are produced by expert analysts from SAR and other high-resolution imagery and are the reference product for sea ice navigation in the Arctic. Relative to them, passive-microwave retrievals are known to underestimate SIC, most markedly in the marginal ice zone and during the summer melt (Agnew and Howell, 2003; Meier et al., 2015). The AMSR2 underestimation is clear for the higher SIC categories, particularly in July (Fig. 1), potentially reflecting the impact of melt ponds on the passive-microwave retrievals (Kern et al., 2016). It should be noted, however, that ice charts may themselves be conservative, more likely to overestimate than to underestimate ice cover, as they are produced to safeguard navigation.

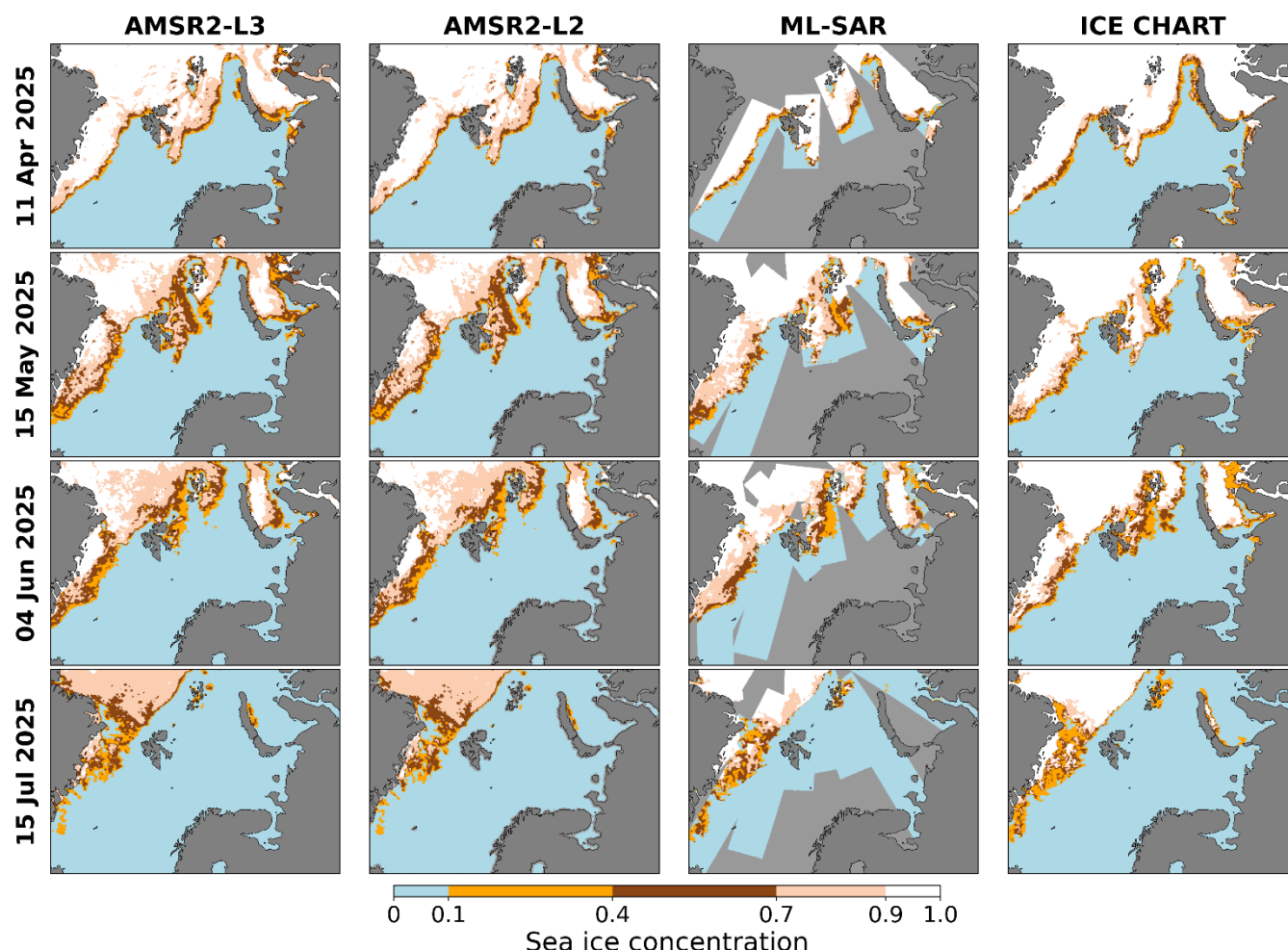


Figure 1. SIC over the Atlantic sector of the Arctic on 11 April, 15 May, 4 June and 15 July 2025 (rows), from AMSR2-L3, AMSR2-L2, ML-SAR, and the regional ice chart (columns, left to right).



Both the AMSR2-L2 and ML-SAR observations are further processed in this work into daily super-observations with their associated observation errors before being assimilated in FOAM. For these new SIC observations, we adopt the observation-error treatment used in GOSI9 for SST and SLA (Mignac et al., 2025), rather than the seasonally and spatially varying observation error estimates used for AMSR2-L3 DA. The total errors for the new SIC observations combine measurement errors and a representation error (RE) accounting for the model sub-grid scale variability, i.e. the SIC variability within an ORCA025 grid cell. It should be noted that the mean ORCA025 grid resolution is ~10 km in the Arctic due to the tripolar grid. As in Oke and Sakov (2008), for each model grid cell containing at least four observations that adequately span the cell, the RE is taken as the standard deviation of the differences between the individual observations and their within-cell average at a given time. Grid cells with fewer than four observations are masked. In practice, AMSR2-L2 and ML-SAR observations falling within the 24-hour assimilation window are gathered and, separately for each observation type, a daily RE per ORCA025 grid cell is obtained as described above. The super-observation for each observation type is taken as the median of the SIC observations within each grid cell, which is less sensitive to outliers than the mean. The measurement error associated with each super-observation is obtained by propagating the individual measurement uncertainties provided by the data providers to the cell average, as the standard error of the observations within that cell.

It is important to note that the RE approach of Oke and Sakov (2008) cannot be applied to the operationally assimilated AMSR2-L3, which is already a daily gridded SIC dataset at 10 km resolution and therefore does not provide enough sampling to consistently estimate the sub-grid variability within ORCA025. On the other hand, the multiple daily swath crossings of AMSR2-L2 and the 500 m resolution of the ML-SAR mosaic yield a high sampling density per ORCA025 grid cell (Fig. 2). The resulting RE for each observation type is also physically consistent, having larger values along the ice edge than over the less variable, more homogeneous ice pack. Conversely, seasonal observation errors used operationally to assimilate the AMSR2-L3 data are spatially very smooth, drawing only a marginal distinction between the ice edge and the pack (Fig. 2). This is expected because the seasonal observation errors are climatological and thus represent neither the short-term evolution of the sea ice, such as the daily changes in the position of the ice edge, nor the interannual variations in the ice edge. In grid cells where the AMSR2-L2 and ML-SAR SIC is zero, a constant observation error of 0.05 is applied in the assimilation, following the operational AMSR2-L3 DA. Since these cells are masked for AMSR2-L2 and ML-SAR in Fig. 2, this constant value is not shown there.

Because the new observation errors have far more spatial structure and a wider range of magnitudes between the ice pack and the ice edge, the convergence of the NEMOVAR cost function within the adopted tolerance for SIC DA required increasing the number of inner-loop iterations from 40 to 480. However, since SIC DA in FOAM is still univariate, the elapsed time for assimilating SIC was only increased from ~1 to 3 minutes, keeping the same computational resources. The SIC DA is run in parallel with the ocean DA in FOAM and since the elapsed time is still much shorter than that of the ocean DA, this increase in cost does not affect the overall production time.

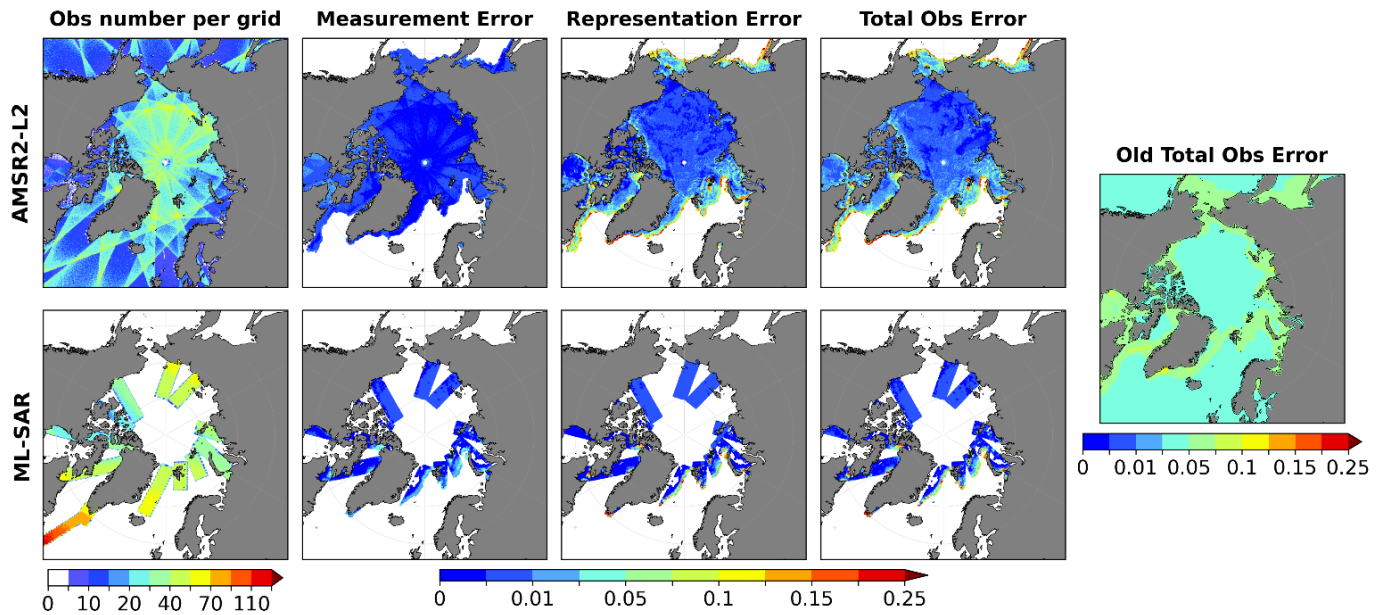


Figure 2. Observation number per grid cell and measurement, representation and total observation error for AMSR2-L2 and ML-SAR on 11 April 2025 over the Arctic. Grid cells with fewer than four observations or with SIC = 0 are masked. For comparison, the seasonal climatological total observation error used operationally for AMSR2-L3 is shown on the right.

2.3 Random Forest bias correction of AMSR2

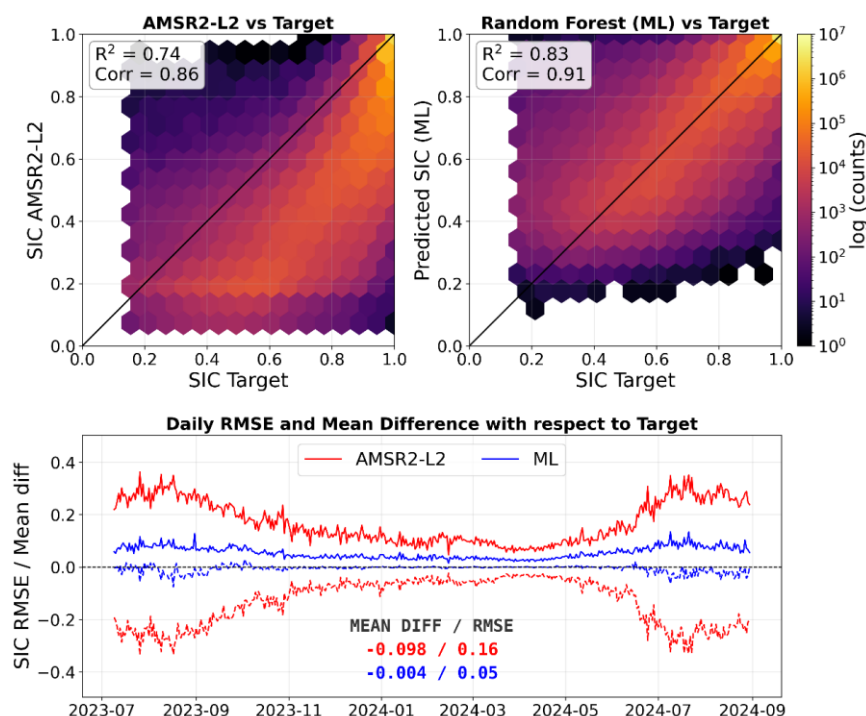
The AMSR2-L2 super-observations are bias-corrected towards ML-SAR using a Random Forest regressor (Breiman, 2001; Pedregosa et al., 2011). The model is trained to predict the ML-SAR super-observed SIC (the target) from the following predictors: the AMSR2-L2 super-observed SIC, the super-observation latitude and longitude, and the day of the year, which is encoded as its sine and cosine to preserve its cyclic nature. While the latitude and longitude help the Random Forest model learn the spatial behaviour of the ice pack and the MIZ, the day of the year supplies the seasonal information needed to represent the marked seasonality of the AMSR2 bias, particularly its growth over the summer months (Fig. 1). We also tested additional predictors, such as the solar zenith angle, but they brought no significant improvement to the training and were therefore left out, keeping the model as simple as possible. All quantities are taken from the co-located AMSR2 and ML-SAR super-observations on the ORCA025 grid (see Sect. 2.2). Only super-observations with SIC ≥ 15 % are used in the training and corrections are not applied below this threshold.

Although the ML-SAR historical product goes back to 2014, the OSI SAF AMSR2-L2 product has only been disseminated since July 2021. The training pairs therefore span three years of co-located super-observations (September 2021 to August 2024), divided into training and test sets by an 80/20 temporal split, so that the model is evaluated on the most recent, independent 20 % of the record. Because the AMSR2 underestimation relative to ML-SAR grows with concentration (Fig. 1 and Fig. 3), the training samples are weighted by concentration to focus the fit where the underestimation is largest. Since SIC below 15 % is already excluded, the weighting applies only above this threshold: samples with SIC between 15 % and 50 %

are given a relative weight of 1, and those above 50 % a relative weight of 100. This yields a more robust correction in the high-concentration range, where the AMSR2 underestimation is most pronounced.

230 The Random Forest comprises 100 trees grown to full depth. Training takes about 4 minutes on 32 CPU cores, and applying the trained model to a full field takes less than 30 seconds, making the correction inexpensive for operational use. Although the Random Forest could in principle be trained directly against the ice charts—as Wulf et al. (2024) did for their SAR retrieval—we deliberately target ML-SAR instead. Since ML-SAR is itself assimilated (Sect 2.4), training the correction towards it ensures that the corrected AMSR2 is as consistent as possible with the ML-SAR observations it complements, providing a spatially complete and corrected field that fills the daily gaps in the ML-SAR coverage.

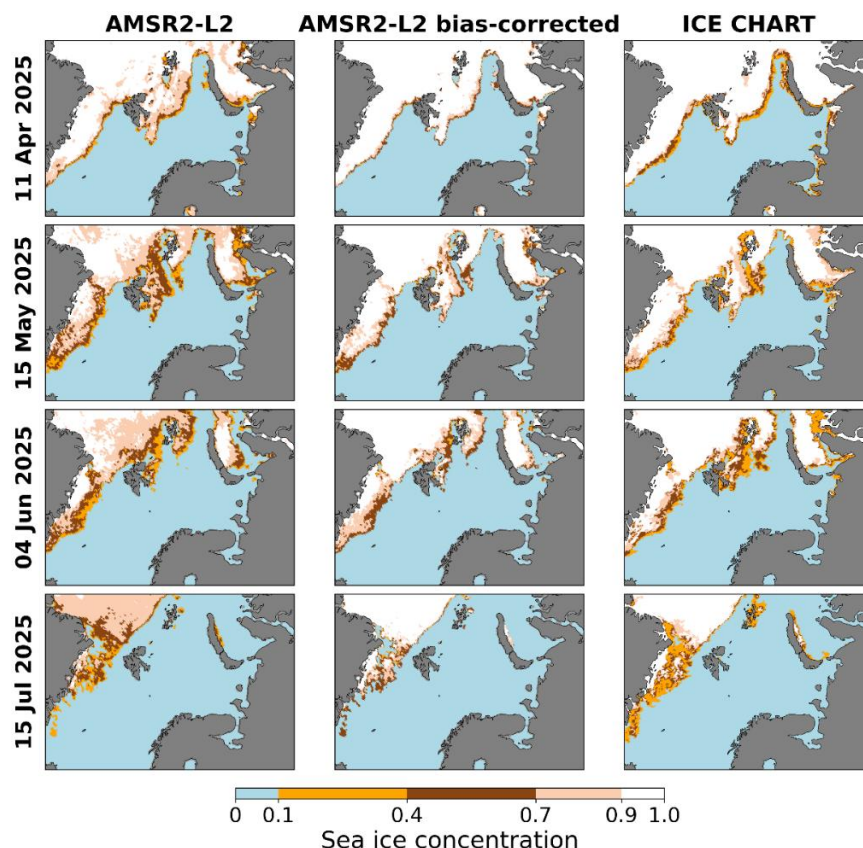
235 Figure 3 evaluates the Random Forest correction against the ML-SAR target over the independent test period (9 July 2023 to 30 August 2024). Consistent with Fig. 1, the uncorrected AMSR2-L2 exhibits a clear SIC underestimation. In the density scatter plot, the points lie predominantly below the 1:1 line, with a mean difference of -0.10 , a Root Mean Squared Error (RMSE) of 0.16 and a coefficient of determination (R^2) of 0.74 (correlation 0.86). After correction, the Random Forest brings the concentrations onto the 1:1 line, reducing the RMSE more than threefold to 0.05 , essentially removing the bias and raising R^2 to 0.83 (correlation 0.91). These statistics are computed for concentrations above 15 %, the range over which the correction is trained and applied.



245 **Figure 3. Evaluation of the Random Forest SIC bias correction against the ML-SAR target over the independent test period (9 July 2023 to 30 August 2024).** (Top) Density scatter of AMSR2-L2 super-observations (left) and the Random Forest SIC (right) versus the ML-SAR super-observations on a logarithmic scale. The black line is the 1:1 reference. (Bottom) daily RMSE (solid) and mean difference (dashed) of AMSR2-L2 (red) and the Random Forest (blue) with respect to the target. All statistics are computed for SIC above 15 %.



250 The daily time series of the statistics shows that this improvement is consistent throughout the year (Fig. 3). The AMSR2
 RMSE against the ML-SAR is strongly seasonal: its RMSE rises from around 0.10 in winter to 0.30–0.35 during the summer
 melt, and its mean difference becomes markedly negative in summer, reaching about -0.3 . This is consistent with the melt
 pond-driven underestimation of AMSR2 SIC retrievals (Kern et al., 2016). The Random Forest suppresses this seasonal cycle
 of the errors almost entirely, with SIC RMSE staying close to 0.05 and the mean difference near zero throughout, including
 255 the summer months. These effects are shown spatially in Fig. 4, which compares the AMSR2-L2 super-observations, their
 bias-corrected counterpart and the regional ice chart for specific dates. The Random Forest correction raises the AMSR2
 concentrations in the highest SIC categories and produces a sharper ice edge, bringing the corrected field into closer agreement
 with the ice chart. At lower concentrations, however, the corrected field agrees less closely with the ice chart (Fig. 4). This is
 expected since the sample weighting deliberately concentrates the fit on the higher concentrations (> 0.5), where the AMSR2
 260 underestimation is largest and most critical. For clarity, Fig. 4 focuses on the Atlantic sector of the Arctic to be able to show
 the ice edge details, but the correction behaves similarly across the whole Arctic (see Sect. 3).



265 **Figure 4.** Effect of the Random Forest bias correction over the Atlantic sector of the Arctic on 11 April, 15 May, 4 June and 15 July
 2025 (rows), from AMSR2-L2, their bias-corrected counterpart and the regional ice chart (columns, left to right).



2.4 Experiment setup and assessment metrics

Four data-assimilation experiments are run from 3 October 2024 to 23 July 2025, the ~10-month period that immediately follows the training and test datasets. All experiments are initialised from the same restart of a FOAM hindcast run using the operational configuration (i.e. already assimilating AMSR2-L3) and are forced at the surface by the operational MetUM global NWP fields at ~10 km resolution (see Sect. 2.1). They are identical, having the same model configuration, the same assimilation of SST, SLA and temperature/salinity profiles, and the same NEMOVAR configuration, except for the SIC DA. Their SIC DA differs in the SIC observations assimilated, in how their observation errors are specified, and in the number of inner-loop iterations (Tab. 1).

The DA experiments are designed to isolate the successive contributions of the AMSR2-L2 and new observation errors, the ML-SAR product, and the bias correction. The control (CTL) assimilates the operational gridded AMSR2-L3 with the seasonal observation errors currently used for SIC in FOAM. The AMSR2-L2 experiment replaces it with the super-observations derived from the multiple swath crossings: besides offering a lower latency for future operational use, this allows more consistent, daily observation errors to be estimated (Sect. 2.2). The AMSR2-ML-SAR experiment then adds the ML-SAR product to assess its potential to improve FOAM sea ice. Because AMSR2 and ML-SAR are markedly different datasets (Fig. 1), SIC discontinuities are expected after assimilating both. Finally, the AMSR2-BC-ML-SAR experiment combines the bias-corrected AMSR2-L2 with ML-SAR, aiming for further improvements and for a spatially homogeneous analysis that minimises the discontinuities between the two datasets. In the two experiments that combine AMSR2-L2 and ML-SAR (with or without the bias correction), ML-SAR takes precedence: where ML-SAR provides observations, only ML-SAR is assimilated, and the AMSR2-L2 is assimilated solely in the locations left uncovered by ML-SAR. In AMSR2-ML-SAR, giving ML-SAR precedence avoids assimilating two different and sometimes contrasting SIC estimates at the same location, especially over the summer months (Fig. 1). In AMSR2-BC-ML-SAR, the bias-corrected AMSR2-L2 is no longer independent of ML-SAR, having been trained towards it (albeit during a different period) and designed precisely to extend it into the SAR gaps. All four experiments produce ocean–sea ice 1-day forecasts. 5-day forecasts are additionally run each day during the whole experimental period for CTL and AMSR2-BC-ML-SAR, which are the operational baseline and the target configuration, respectively.

Regional ice charts from several providers are used to validate the assimilation experiments (Fig. 5). The Atlantic sector and Greenland charts are obtained from CMEMS (<https://doi.org/10.48670/moi-00128>), produced by MET Norway and DMI, respectively; the Hudson Bay, Baffin Bay, East Coast and Beaufort Sea charts are produced by the Canadian Ice Service (<https://iceweb1.cis.ec.gc.ca>); and the Pacific sector is extracted from the U.S. National Ice Center (NIC) Arctic charts (<https://doi.org/10.7265/46cc-3952>) to close the pan-Arctic gap. These charts are not produced on a regular temporal schedule, with their frequency varying from provider to provider, and they do not report continuous SIC values: some use 0.1 concentration bins, while others follow the World Meteorological Organization (WMO) concentration categories. To standardise across providers, the WMO categories are adopted for all charts (see Fig. 1 and Fig. 4): open water (< 0.1), very



open drift ice (0.1–0.3), open drift ice (0.4–0.7), close drift ice (0.7–0.8) and very close drift ice (0.9–1.0). Fast ice was merged with the very close drift ice category because the sea ice model does not explicitly represent landfast ice. The charts are finally averaged to 10 km, the mean resolution of the ORCA025 tripolar grid in the Arctic, for a fairer comparison with the model. Two U.S. NIC products are used to evaluate the ice edge position. The Integrated Ice Edge Error (IIEE; Goessling et al., 2016) is computed against the NIC Multisensor Analyzed Sea Ice Extent product (MASIE), provided at 4 km and subsampled to 10 km. For qualitative comparison of ice edge locations, the daily NIC product directly containing the latitude and longitude positions of the ice edge, is used (Helfrich et al., 2007). These NIC products are based on a wide range of satellite imagery—visible/infrared, SAR, scatterometer and passive microwave. They also incorporate NIC ice charts, which include in situ observations and ship reports, and are consistently checked by an ice analyst (Helfrich et al., 2007). The model ice edge is defined by the SIC at 15 %.

	Assimilated SIC observations	SIC observation errors and number of iterations
CTL	AMSR2-L3 (gridded)	Seasonal, spatially varying observation errors + 40 iterations
AMSR2-L2	AMSR2-L2 super-observations	Daily measurement + representation errors + 480 iterations
AMSR2-ML-SAR	ML-SAR where available; AMSR2-L2 super-observations in ML-SAR gaps	Daily measurement + representation errors per observation type + 480 iterations
AMSR2-BC-ML-SAR	ML-SAR where available; bias-corrected AMSR2-L2 super-observations in ML-SAR gaps	Daily measurement + representation error per observation type + 480 iterations

Table 1. Differences between the DA experiments.

To validate the results, the daily-mean SIC fields from the model are bilinearly interpolated to MASIE and the ice charts data point locations. Because the ice charts are discretised into concentration categories, rather than using conventional metrics such as RMSE and mean difference, we assess the model SIC performance with the Critical Success Index (CSI) for each WMO category. The CSI is defined as:

$$CSI (\%) = \frac{hits}{(hits+misses+false\ alarms)} \times 100 \quad (1)$$

where, for a given category, hits are points in which both the observation and the model fall in the category, misses are points in the category in the observation but not in the model, and false alarms are points in the category in the model but not in the observation.

SIT measurements from the bottom-anchored, upward-looking-sonar moorings of the Beaufort Gyre Exploration Project (BGEP; Krishfield et al., 2014) are also used to assess the impact of the different SIC experiments on the model's SIT (Fig.



5). The BGEP moorings measure the sea ice draft with an estimated accuracy of 0.05–0.10 m at 2-second intervals. These are processed by the data provider into daily averages with their standard deviations and are further converted here into thickness by dividing by a factor of 0.89 (Yu et al., 2004). In addition, the experiments are compared with the merged CryoSat-2/Sentinel-3/SMOS SIT product (hereafter SMOS-CS2-S3), which extends the optimal interpolation method of Ricker et al. (2017) to include Sentinel-3 and provides a gridded, weekly-averaged Arctic SIT field. It is used here at 12.5 km resolution (version 300), obtained from the Alfred Wegener Institute (AWI; ftp://ftp.awi.de/sea_ice/product/cryosat2_smos). To compare the model with these independent SIT datasets, the daily-mean (weekly-mean) model SIT fields are also bilinearly interpolated to the BGEP (SMOS-CS2-S3) locations for each date.

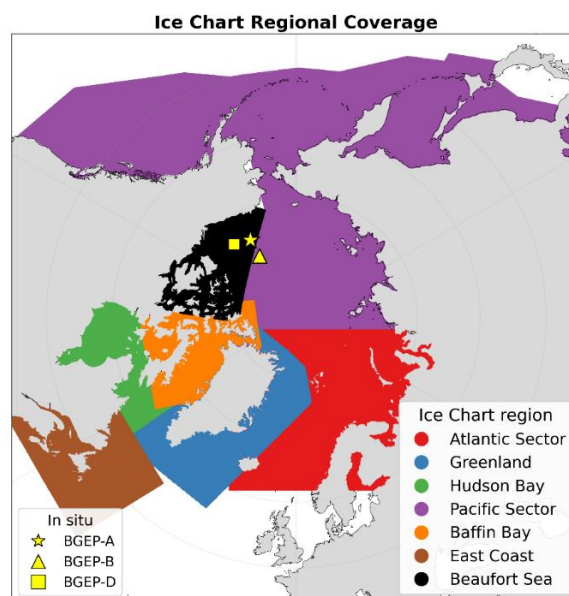


Figure 5. Spatial coverage of the regional ice charts used for experiment validation, each shown in a distinct colour. The in-situ BGEP moorings (BGEP-A, BGEP-B and BGEP-D) used for SIT validation are marked in yellow.

3 Results

The DA experiments are assessed in three steps. Sections 3.1, 3.2 and 3.3 use the 1-day forecasts of all four experiments to evaluate, respectively, the impact on the SIC state, the agreement with the ice charts and SIT datasets. Section 3.4 then evaluates the 5-day forecasts, which were generated only for CTL and the target experiment AMSR2-BC-ML-SAR.

3.1 Impacts on the SIC state

Figure 6 shows the seasonal SIC for CTL and the difference of each experiment with respect to CTL for three seasons: December–February, March–May and June–July. August is not shown because the experiments end in late July, and the first two months (October–November) are excluded to remove the initial model spin-up as it adjusts to the assimilation of new

observations. During the ice growing months, there is a marked increase in the ice edge SIC and, consequently, a clear sharpening of the ice edge in AMSR2-ML-SAR and AMSR2-BC-ML-SAR relative to CTL. This is expected since passive-microwave retrievals tend to smooth the ice edge gradients (Ivanova et al., 2015; Lavergne et al., 2019), so the experiments assimilating only AMSR2 keep a smoother ice edge, whereas the assimilation of ML-SAR and the bias-corrected AMSR2 sharpens it. The bias correction further enhances this sharpening in the Labrador Sea and, especially, in the Pacific Sector (outside the Bering Strait), where SAR coverage is essentially absent (Fig. S1). This illustrates one of the main purposes of the Random Forest corrections: to propagate ML-SAR patterns, such as a sharper ice edge, into AMSR2 observations, improving SIC in regions where SAR observations are less dense or do not reach. This is consistent with the SIC gradients in Fig. 7, which show markedly larger ice edge gradients in AMSR2-ML-SAR than in CTL. Near the ice edge, the change in gradient has a dipole structure, negative on the poleward side and positive on the equatorward side, consistent with the sharpening. For AMSR2-BC-ML-SAR, enhanced SIC gradients are also seen in ice edge regions of sparse SAR coverage, such as the Pacific Sector of the Arctic.

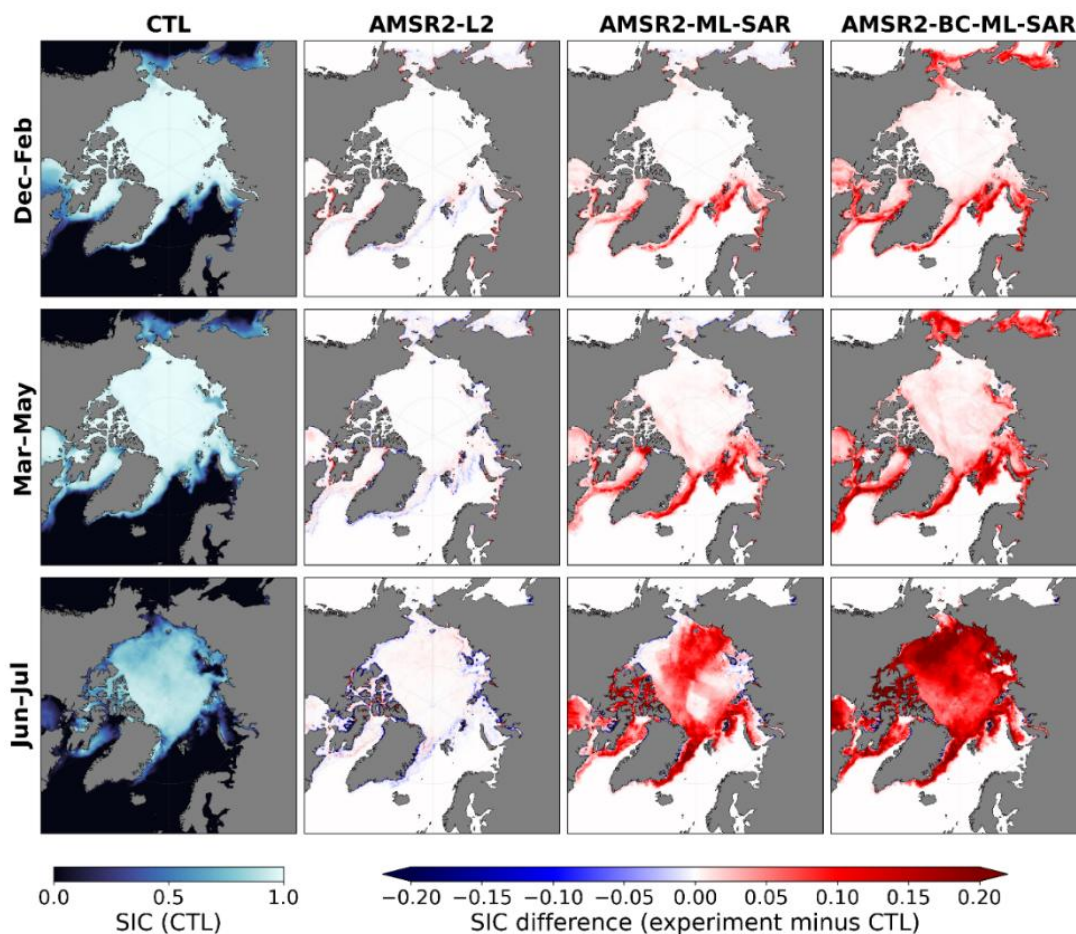


Figure 6. Seasonal SIC from the 1-day forecasts, for December–February, March–May and June–July (rows). The first column shows the CTL SIC and the remaining columns show the difference of each experiment relative to CTL.

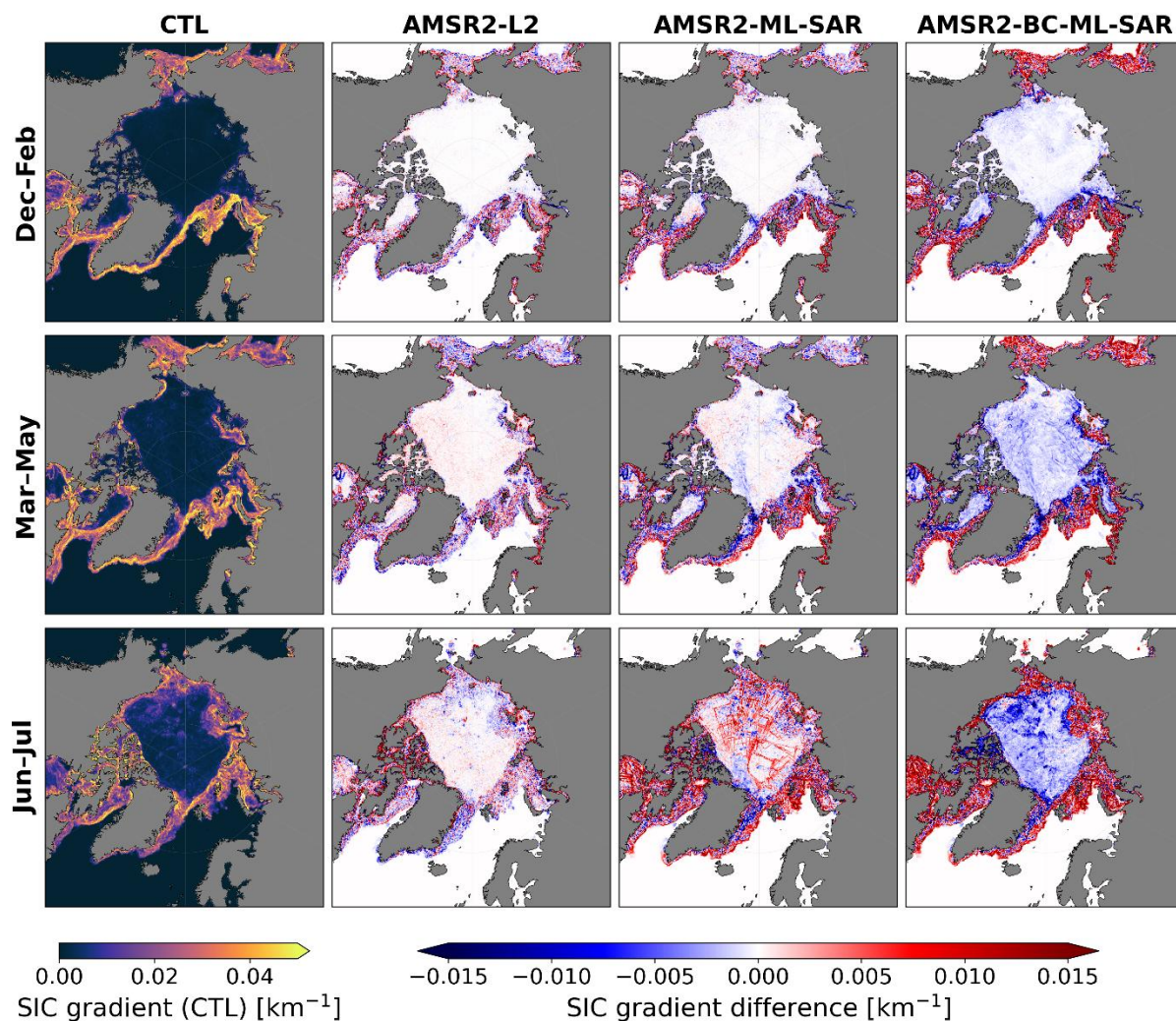


Figure 7. Seasonal maximum of the daily SIC gradients (km⁻¹) from the 1-day forecasts over the Arctic for December–February, March–May and June–July (rows). The first column shows the control (CTL) and the remaining columns show the difference of each experiment relative to CTL.

In the transition to summer, the SIC increase in both AMSR2-ML-SAR and AMSR2-BC-ML-SAR extends across the whole Arctic relative to CTL, including the interior ice pack (Fig. 6). This is explained due to the growing AMSR2 SIC underestimation associated with the presence of melt ponds. In the AMSR2-ML-SAR experiment, this SIC increase is confined to the SAR locations, creating clear discontinuities in the ice pack that show up even in the seasonal means. The spatial SIC gradients of Fig. 7 confirm this pattern, with the ice pack of AMSR2-ML-SAR showing artificial gradients in June and July. For example, artefacts in the SIC gradients are clearly seen at the edges of the ML-SAR due to the large differences between AMSR2 and ML-SAR in the melt season. In contrast, AMSR2-BC-ML-SAR SIC changes are spatially more homogeneous over the summer months, being extended to regions of sparse SAR coverage within the ice pack. As a result, its SIC gradients



370 in the ice pack weakens towards summer, since the bias correction restores the near-uniform, high concentrations of the consolidated ice pack, as also observed in the ice charts (see Fig. 1), thereby weakening the interior gradients. Finally, Fig. 6 and Fig. 7 show that the AMSR2-L2 experiment closely resembles CTL in the seasonal means. In terms of the observation values, this is expected, as the AMSR2-L2 super-observations projected onto the ORCA025 grid is very similar to the operational, gridded AMSR2-L3 (see Fig. 1). However, small differences do appear at the ice edge, where the changes
 375 in AMSR2-L2 are slightly negative relative to CTL. This is consistent with the larger observation errors that the new daily estimates assign to the ice edge with respect to seasonal observation errors used for AMSR2-L3 (Sect. 2.2). Therefore, for very similar innovations, the greater observation error there gives the observations less weight in the DA, so the SIC in the edge is slightly less sharp in AMSR2-L2 than in CTL.

3.2 Validation with ice charts

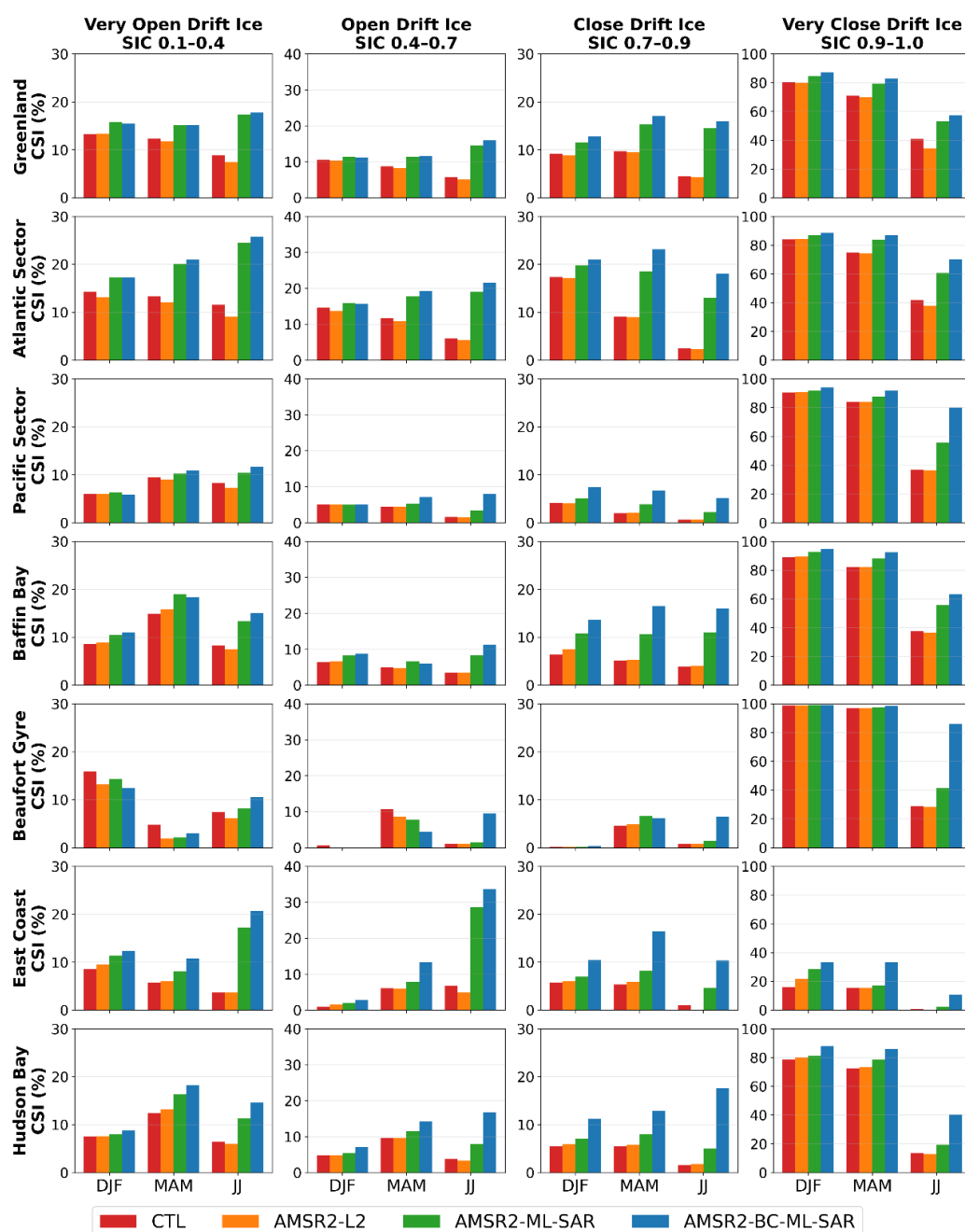
380 The performance of 1-day SIC forecasts against the ice charts is assessed by calculating the CSI (Sect. 2.4) for each WMO concentration category, Arctic region and season (Fig. 8). Although ML-SAR is trained with ice charts as the target, its training period is 2018–2021 (Wulf et al., 2024). The ice charts used here for validation span December 2024 to July 2025, a period not seen during that training. The same holds for the Random Forest correction, which was developed from ML-SAR data spanning September 2021 to August 2024 (Sect. 2.3), all of which precedes the December 2024–July 2025 period used here.
 385 The ice charts are not a fully independent dataset, given their historical link to ML-SAR, but they still provide an out-of-sample evaluation over an unseen period.

Before comparing the experiments, we note in Fig. 8 that the CSI in the three lower SIC categories is much smaller than in the 0.9–1.0 category. This reflects the strongly skewed SIC distribution, in which the consolidated pack (0.9–1.0) accounts for most of the samples — one to two orders of magnitude more grid points than the narrow, transitional 0.1–0.9 categories. In
 390 these sparsely populated SIC categories, a given number of misclassifications weighs far more heavily on the CSI, therefore explaining the lower CSI in the 0.1–0.9 when compared to 0.9–1.0 SIC category (Mason, 1989).

Regardless of the lower CSI in the transitional categories, the CSI is consistently highest for AMSR2-BC-ML-SAR across almost all regions, categories and seasons, with AMSR2-ML-SAR second and CTL and AMSR2-L2 close to each other and generally lowest (Fig. 8). The improvement is modest during winter, when all experiments already perform well in the higher
 395 concentration categories, but grows markedly towards summer. However, in winter the SIC changes are localised near the ice edge (Fig. 6), so the higher CSI shows that the ice edge in AMSR2-BC-ML-SAR agrees more closely with the ice charts than in the other experiments, particularly CTL and AMSR2-L2. In summer, the CSI in CTL and AMSR2-L2 collapses for the very close drift ice, falling from 80–90% to lower than ~40% in all Arctic regions, whereas AMSR2-BC-ML-SAR retains a much higher skill, ranging between 60 and 80%. A similar summer advantage is also seen in the lower categories and across MIZ
 400 regions, such as the Arctic East Coast, Hudson Bay and Baffin Bay. Finally, it is important to highlight that the gains in



AMSR2-BC-ML-SAR are largest precisely where the ML-SAR coverage is sparse, such as in the Beaufort Gyre and the Pacific Sector (Fig. S1).



405 **Figure 8.** CSI (%) from 1-day forecasts of the DA experiments evaluated against the regional ice charts, broken down by SIC category, region and season. Rows correspond to the seven ice chart regions (see Fig. 5), and columns to the four SIC categories following the WMO convention. Note that the y-axis is scaled independently for each SIC category.

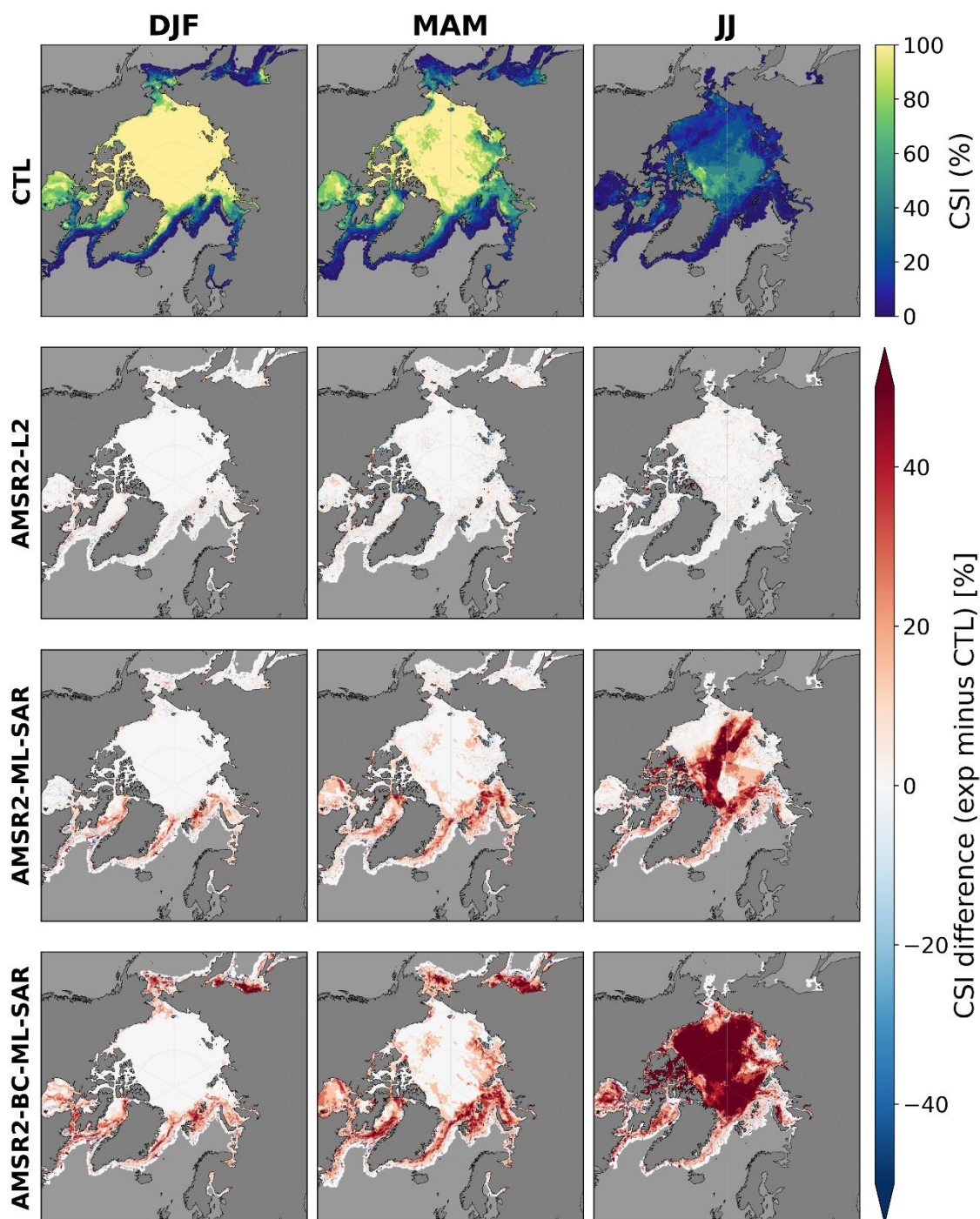


The components underlying these CSI gains differ by SIC category (see Figs. S2–S4). In the high concentration categories (0.7–0.9 and 0.9–1.0), the improvement is driven mainly by a large reduction in misses during summer. While CTL underestimates the consolidated pack relative to ice charts, the assimilation of ML-SAR and bias-corrected AMSR2 restore it, converting misses (Fig. S3) into hits (Fig. S2). In the lower categories (0.1 – 0.7), the CSI gain comes mainly from fewer false alarms, as the better placed, sharper ice edge in AMSR2-BC-ML-SAR reduces the SIC wrongly assigned to these transition categories in the CTL (Fig. S4). The only trade-off is a very small increase in false alarms for AMSR2-BC-ML-SAR in the 0.9–1.0 category (Fig. S4), reflecting a slight over-addition of SIC. This is small relative to the accompanying reduction in misses, so the CSI still substantially improves.

At each ice chart data point, the seasonal hits, misses and false alarms are accumulated into a local CSI in Fig. 9, mapped across the Arctic domain and thus expanding the regional CSI evaluation of Fig. 8. Although the regional CSI gains of AMSR2-BC-ML-SAR appear modest for the winter in Fig. 8, the map in Fig. 9 shows that local CSI gains in winter can reach 40–50% near the ice edge, such as in the Bering and Barents–Kara Seas, and the Labrador Sea and Baffin Bay. As the AMSR2 SIC underestimation grows towards summer relative to ice charts, CSI gains spread from the edge into the interior pack, exceeding 50% over large parts of the Arctic in AMSR2-ML-SAR and AMSR2-BC-ML-SAR. However, AMSR2-ML-SAR improves the CSI only in patches following the discontinuous ML-SAR coverage, whereas AMSR2-BC-ML-SAR consistently propagates the CSI improvement across the whole ice pack.

3.3 Impacts on the SIT

Expanding the evaluation to other sea ice variables, we first assess the impact of the DA experiments on SIT at the BGEP moorings (Fig. 10), with the SMOS-CS2-S3 product shown at these locations as an additional reference (see Sect. 2.4). All experiments reproduce the observed seasonal ice growth and remain within the large observational spread, so the differences between them are modest relative to daily BGEP uncertainties. However, in terms of the mean difference, CTL and AMSR2-L2 underestimate SIT the most, particularly at BGEP-A, and are nearly indistinguishable, which is consistent with the near-equivalence of the two AMSR2-only experiments for SIC (see Sect. 3.1). The same SIT underestimation pattern is seen for SMOS-CS2-S3, with negative differences relative to the BGEP measurements of approximately –0.2 to –0.3 m on average. In contrast, assimilating ML-SAR observations, and especially the bias-corrected AMSR2, consistently reduces the RMSE with respect to CTL at all three moorings, but at the same time AMSR2-BC-ML-SAR shifts the mean difference towards positive values. This behaviour is directly associated with the positive SIC increments produced by these experiments, particularly AMSR2-BC-ML-SAR, which are larger and broader in spring and summer, raising SIC relative to CTL (Fig. 6). A positive SIC increment adds ice to the thinnest category with a prescribed thickness of 0.45 m (see Sect. 2.1), so that higher SIC is accompanied by a greater ice volume and hence a larger SIT.



440 Figure 9. CSI (%) of the experiments against the combined pan-Arctic ice chart from Fig. 5, shown by season. The top row shows the absolute CSI of CTL (0–100 %). The other rows show the difference from CTL for AMSR2-L2, AMSR2-ML-SAR and AMSR2-BC-ML-SAR, with positive (red) values indicating closer agreement with ice charts than CTL.



Therefore, when the experiments are compared against the SMOS-CS2-S3 SIT product in Fig. 11, it is not surprising that AMSR2-ML-SAR and particularly AMSR2-BC-ML-SAR overestimate the SITs across the Arctic marginal seas. As a result, in these regions the SIT RMSE relative to SMOS-CS2-S3 is higher for these two experiments than for CTL. On the other hand, the CTL experiment has ice that is too thin over the central Arctic, particularly north of Greenland and the Canadian Archipelago, where it shows the largest negative differences and the highest RMSE relative to SMOS-CS2-S3. This thin bias of the FOAM system in the region has already been reported by Mignac et al. (2022). There, the additional SIT introduced by AMSR2-BC-ML-SAR reduces the thin bias and lowers the RMSE with respect to SMOS-CS2-S3.

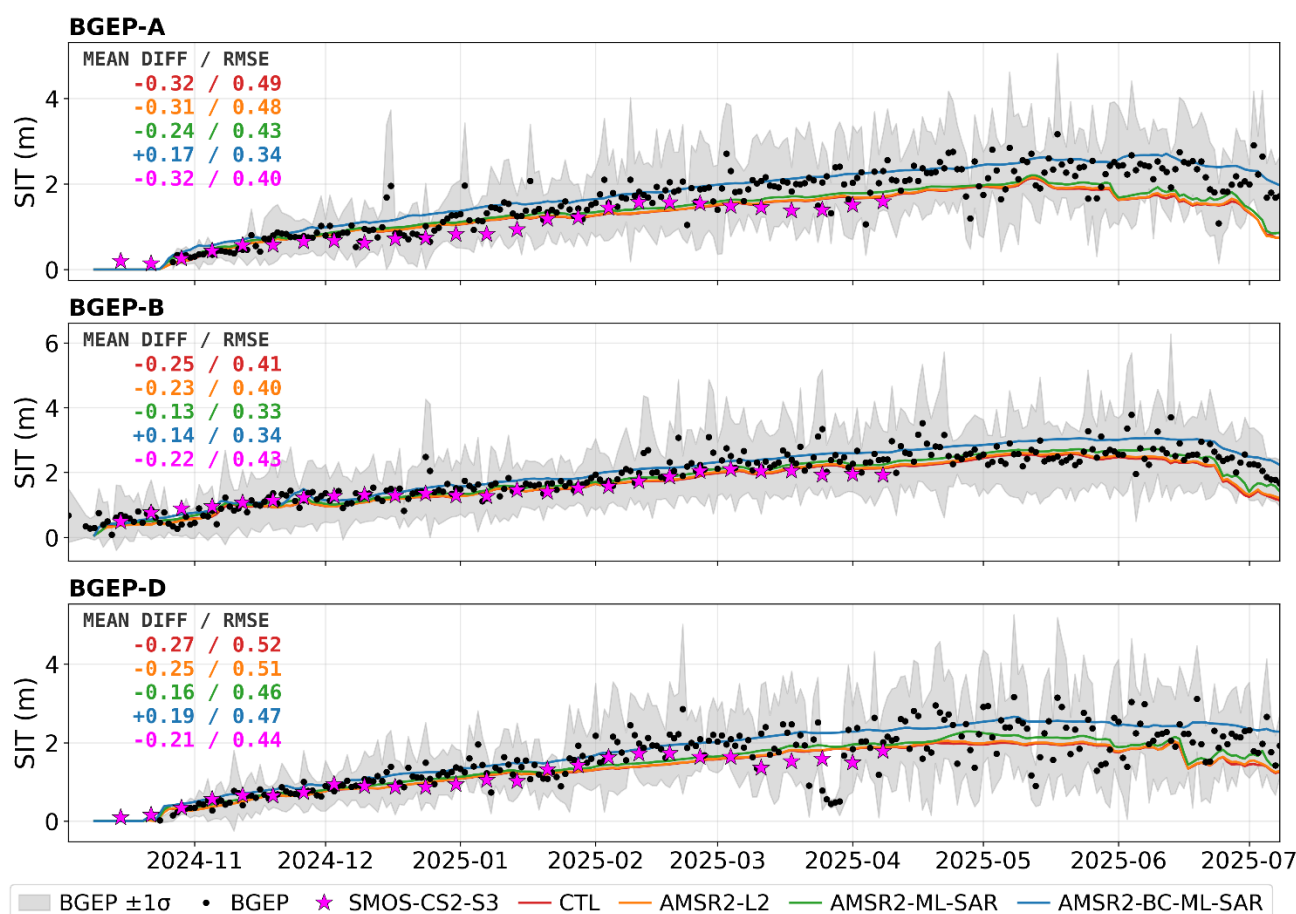


Figure 10. SIT (m) time series from 1-day forecasts of the DA experiments at the three BGEP moorings in the Beaufort Gyre (see Fig. 5). Black dots show daily BGEP SIT measurements with the grey band indicating daily measurement uncertainties. Magenta stars show the weekly SMOS-CS2-S3 product at each mooring location.



In the scheme of Peterson et al. (2015) used here, positive SIC increments are always added to the thinnest category. Therefore, two additional experiments were conducted, varying the thickness assigned to the ice with positive SIC increments. In the first experiment, the prescribed thickness value was reduced from 0.45 m to 0.15 m. In the second experiment, it was taken at each grid point from the model's own SIT in the thinnest category. Both experiments reduce the SIT overestimation of AMSR2-BC-ML-SAR in the Arctic marginal seas but degrade its SIC against the ice charts, particularly in summer, as the smaller volume of added ice by DA is not sustained and partially melts (not shown). This behaviour is consistent with the rationale for the 0.45 m value adopted here, established by Peterson et al. (2015), which corresponds to the upper bound of the thinnest category and is chosen to prevent the newly added ice from melting immediately. More consistent approaches for distributing the SIC increments across the model ice categories, such as the one recently proposed by Browne et al. (2026), will be evaluated in the future, but this is beyond the scope of this paper.

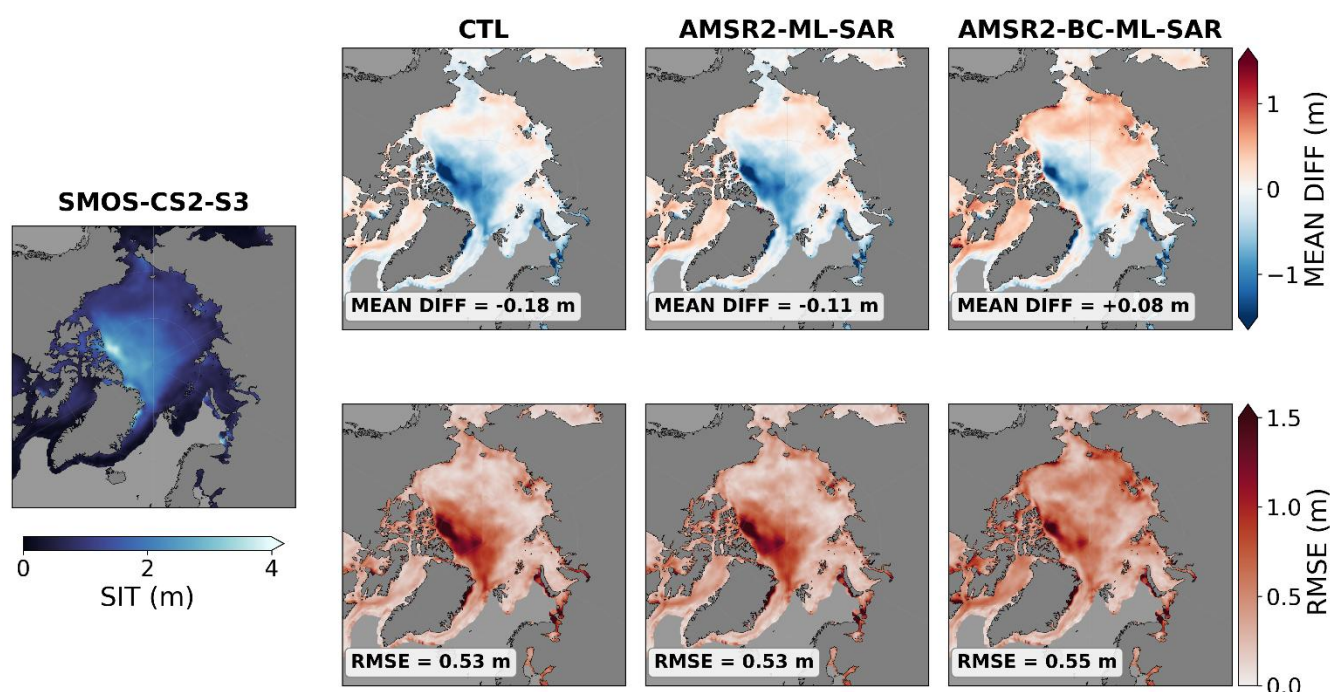


Figure 11. SIT (m) mean difference (top) and RMSE (bottom) from 1-day forecasts of the DA experiments with respect to the SMOS-CS2-S3 product (leftmost panel), spanning 01 December 2024 to 15 April 2025. Note that SMOS-CS2-S3 provides thickness only during the ice growing season, so the observations end on 15 April 2025. AMSR2-L2 is omitted as it is nearly identical to CTL. The domain-wide mean difference and RMSE are annotated in each panel. Mean difference is calculated as model minus SMOS-CS2-S3.

3.4 Assessment of short-range sea ice forecasts

The evaluation so far has focussed on the 1-day forecasts. We now assess whether the improvements brought by the assimilation of AMSR2 bias-corrected observations and ML-SAR are retained in the short-range sea ice forecasts. 5-day

forecasts were produced only for the CTL and AMSR2-BC-ML-SAR, which is the configuration proposed for operational use. Forecast skill is evaluated as a function of lead time (1–5 days) against the same references used in Sect. 3.2 and 3.3: the CSI per WMO SIC category against the regional ice charts, and the RMSE and mean difference of SIT against SMOS-CS2-S3 and the BGEP moorings (Fig. 12). Additionally, the IIEE is computed against the MASIE product and is also included in Fig. 12.

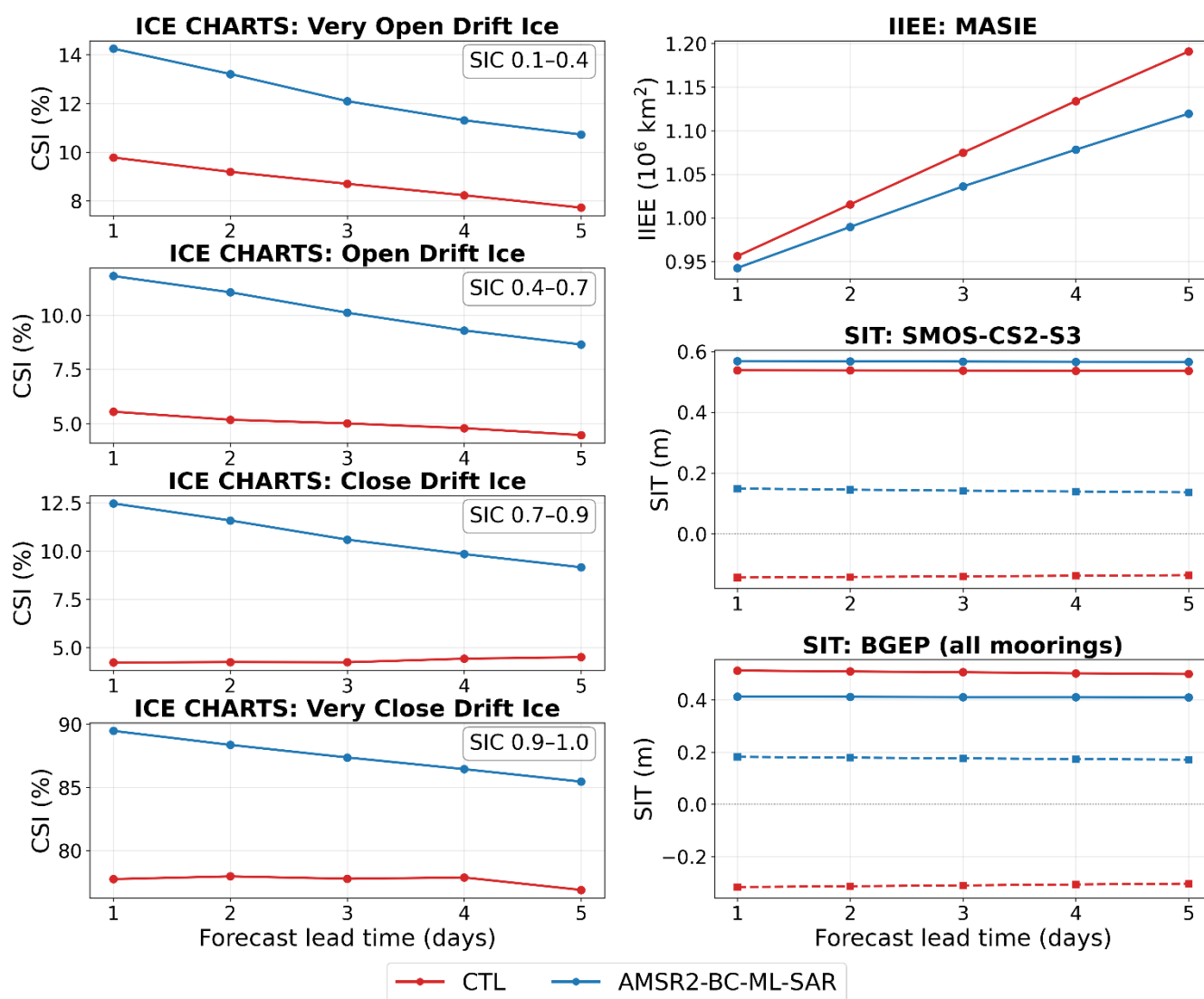


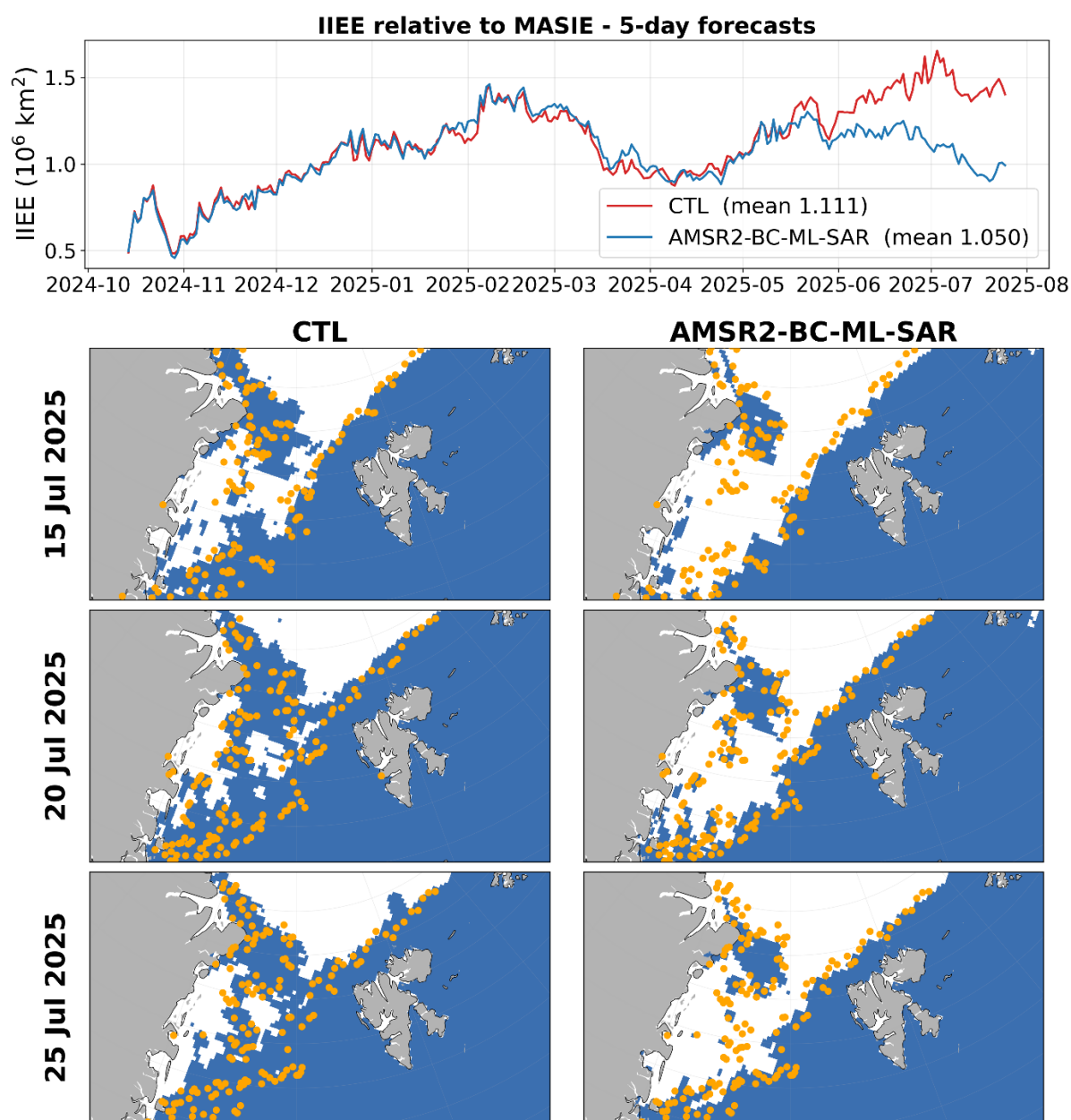
Figure 12. Sea ice forecast skill as a function of lead time (1–5 days) for CTL (red) and AMSR2-BC-ML-SAR (blue) over December 2024–July 2025. CSI is computed against the combined pan-Arctic ice chart for the four WMO SIC categories (left); IIEE against the MASIE product (top right); and RMSE and mean differences against SIT from the SMOS-CS2-S3 product (centre right) and the BGEP moorings (bottom right). In the two SIT panels, solid lines show the RMSE and dashed lines the mean difference (model minus observation). The comparison against SMOS-CS2-S3 extends only to April 2025, as this product is not available after the onset of the melt season.



AMSR2-BC-ML-SAR improves the sea ice edge and concentration throughout the forecast range, having consistently higher
490 CSI than CTL in every SIC category and at every lead time (Fig. 12). The CSI gains per forecast lead time in AMSR2-BC-
ML-SAR are even more evident over June and July (see. Fig. S5), with CSI improvements of ~40 % in the 0.9-1.0 SIC category
when compared to CTL. As a result of the SIC forecast improvements across all WMO categories, Fig. 12 also shows a lower
IIEE for AMSR2-BC-ML-SAR than for CTL at every lead time. Although the IIEE of both experiments grows with lead time
as forecast errors accumulate, the reduction achieved by AMSR2-BC-ML-SAR becomes larger at longer leads. In other words,
495 the better-constrained initial ice edge in AMSR2-BC-ML-SAR translates into a forecast of sea ice edge that drifts away from
the MASIE product more slowly than CTL.

The impact on SIT is, as expected, almost independent of lead time, since thickness evolves slowly over five days (Fig. 12).
Both the mean difference and the RMSE are essentially flat across lead times. Consistent with the SIT results in Sect. 3.3, CTL
underestimates SIT while AMSR2-BC-ML-SAR overestimates it, against both SMOS-CS2-S3 (-0.13 m vs $+0.15$ m) and the
500 BGEP moorings (-0.35 m vs $+0.18$ m). It is worth highlighting that there is a large spatial error cancellation against SMOS-
CS2-S3 (see Fig. 11), as AMSR2-BC-ML-SAR reduces (increases) the mean differences in the thick central Arctic (marginal
seas). Against the BGEP moorings, AMSR2-BC-ML-SAR also has a slightly lower RMSE than CTL, whereas against SMOS-
CS2-S3 the RMSE is marginally larger.

Finally, we assess the 5-day forecast IIEE in more detail (Fig. 13). Throughout autumn, winter and early spring, the IIEE of
505 CTL and AMSR2-BC-ML-SAR is almost identical. In this period, the assimilation of ML-SAR and the bias-corrected AMSR2
sharpens the ice edge and raises the concentration gradient across it (see Fig. 5 and Fig. 6), without displacing the edge itself.
As a result, the two forecasts place it in essentially the same location, and their IIEE against MASIE largely coincides. As the
melt season progresses, the MIZ becomes fragmented. For example, patches of open water appear within the pack and the ice
edge broadens into a more mobile, less sharply defined transition zone, whose position varies markedly from day to day. In
510 this regime, the 5-day forecast of ice edge is intrinsically more sensitive to the initial conditions, with AMSR2-BC-ML-SAR
keeping a consistently lower IIEE than CTL (Fig. 13). This summer behaviour is illustrated by the ice edge maps for 15, 20
and 25 July 2025 in Fig. 13. The CTL 5-day forecast develops exaggerated areas of open water in the Greenland Sea, retreating
well inside the observed ice edge. Conversely, AMSR2-BC-ML-SAR 5-day forecasts retain a more coherent ice-covered area
whose edge follows the observed ice edge contour much more closely. AMSR2-BC-ML-SAR therefore not only reduces the
515 IIEE on average but also yields a summer ice edge that is better positioned with respect to the independent ice edge products.
It is important to highlight that 5-day ocean forecasts in the Arctic were also verified down to 500 m against 2530 Argo profiles
from 218 floats over December 2024–July 2025. Despite the additional ice formed in AMSR2-BC-ML-SAR, no significant
differences were found between CTL and AMSR2-BC-ML-SAR for temperature and salinity (not shown). Although in-situ
observations are relatively sparse in the Arctic, this lack of significant differences is consistent with the univariate nature of
520 the SIC DA performed here, whose minimisation is carried out separately from the ocean variables and therefore does not
directly change ocean temperature and salinity.



525 **Figure 13.** 5-day forecast time series of IIEE (10⁶ km²) for CTL (red) and AMSR2-BC-ML-SAR (blue) against MASIE, followed by ice edge maps of 5-day forecasts for three summer dates (15, 20 and 25 July 2025). In each map, sea-ice covered areas (SIC > 0.15) are shown in white, whereas ocean covered areas are shown in blue. The orange dots mark the daily positions from the NIC ice edge contour product.



530 4. Discussion and conclusions

Bias-corrected AMSR2 SIC observations applying a Random Forest model, which was trained towards a deep-learning SAR-derived SIC product (ML-SAR; Wulf et al., 2024), have been successfully assimilated into the 1/4° (ORCA025) configuration of the Met Office's Forecast Ocean Assimilation Model (FOAM). To demonstrate that assimilating these bias-corrected observations improves Arctic sea ice forecasts, four FOAM experiments were conducted, each assimilating a different set of
 535 SIC observations: the operational AMSR2 L3 data (CTL); the AMSR2 L2 product (AMSR2-L2); AMSR2-L2 in conjunction with ML-SAR (AMSR2-ML-SAR); and the bias-corrected AMSR2-L2 in conjunction with ML-SAR (AMSR2-BC-ML-SAR). In the last two experiments, the AMSR2-L2 observations, with or without bias correction, are assimilated only where ML-SAR is absent, as ML-SAR is taken as the reference product.

Assimilating the bias-corrected AMSR2 observations together with ML-SAR significantly improves two aspects of the FOAM
 540 sea ice forecasts, in both cases by overcoming well-known limitations of the AMSR2 retrievals. The first aspect is the representation of the ice edge. Because passive-microwave products are spatially smooth, they tend to blur the sharp concentration gradients that define the ice edge (Ivanova et al., 2015; Lavergne et al., 2019; Wulf et al., 2024). Assimilating ML-SAR and the bias-corrected AMSR2 instead produces a sharper ice edge during the ice growing season, which is also better positioned in summer when the marginal ice zone is more mobile and its location harder to constrain. The second aspect
 545 is the SIC underestimation in passive-microwave retrievals during the ice melt season, mainly attributed to the presence of melt ponds (Titchner and Rayner, 2014; Kern et al., 2016, 2020). The Random Forest bias correction largely mitigates this AMSR2 underestimation, so that AMSR2-BC-ML-SAR yields summer SIC forecasts that retain more of the ice pack within the highest concentration range (i.e., 0.9–1.0) relative to CTL. The improvements in both aspects are confirmed against regional ice charts, where AMSR2-BC-ML-SAR achieves a higher CSI than CTL in all SIC categories and in nearly all Arctic regions.
 550 These CSI improvements in AMSR2-BC-ML-SAR also persist throughout the 5-day forecasts relative to CTL, together with a consistently lower IIEE against MASIE, most notably during summer.

Owing to the sparse SAR coverage, it is important to highlight that these Arctic-wide gains could not have been achieved without the bias-corrected AMSR2 observations. In AMSR2-ML-SAR, where ML-SAR is assimilated but AMSR2 is left uncorrected, the improvements remain confined to the SAR locations, producing spatially discontinuous SIC fields with
 555 artificial gradients and only localised CSI gains. In contrast, the bias-corrected AMSR2 spreads these gains consistently across the whole Arctic, combining the spatial completeness of the passive-microwave observations with the fidelity of SAR in resolving fine-scale features, such as the ice edge.

The only trade-off in AMSR2-BC-ML-SAR is a SIT overestimation in the Arctic marginal seas. Because positive SIC increments always add ice to the model's thinnest category with a prescribed thickness of 0.45 m (Peterson et al., 2015), the
 560 additional ice introduced by AMSR2-BC-ML-SAR relative to CTL, which is largest in spring and summer, translates into a greater ice volume and hence a higher SIT. Although this is beneficial where FOAM is known to have ice that is too thin, such as in the central Arctic (Mignac et al., 2022), the mean differences against the SMOS-CS2-S3 product in the Arctic marginal



seas are degraded in AMSR2-BC-ML-SAR with respect to CTL. However, at the BGEP moorings, even though the mean difference turns from negative (CTL) to positive (AMSR2-BC-ML-SAR), the SIT RMSE of AMSR2-BC-ML-SAR is still smaller than the CTL at all three sites.

Mitigating the SIT overestimation in the Arctic marginal seas is a natural direction for future work. The Met Office recently developed a hybrid ensemble-variational ocean and sea-ice data assimilation framework (Lea et al., 2022). In this context, the ensemble could be used to distribute the SIC increment across the model's ice categories in a more physically consistent, flow-dependent way, instead of adding it to the thinnest category with a fixed thickness. Moreover, although SIT was not assimilated in the present experiments, this capability already exists in FOAM through the combined CryoSat-2 and SMOS SIT assimilation (Mignac et al. 2022). Assimilating SIT in conjunction with the bias-corrected SIC should further constrain the SIT field and alleviate its overestimation in the marginal seas, at least during the ice growing season, when thickness observations are available, i.e. until mid-April.

It is also worth mentioning that, although the AMSR2-L2 experiment had little impact on the sea ice forecasts, closely resembling CTL, the transition from the gridded AMSR2-L3 to the L2 product is nonetheless valuable for operations. The L2 product has a lower latency and sub-daily availability, enabling more consistent daily representation errors to be estimated for SIC and a better use of the AMSR2 observations, particularly in the Met Office coupled NWP system that employs 6-hour assimilation windows. In the same context, future experiments will test a L2 ML-SAR product from DMI, which can be provided with reduced latency compared to the L3 SAR-derived observations used here.

Finally, the experiments presented here were performed with the ORCA025 FOAM configuration. As a next step, similar experiments will be carried out with the 1/12° (ORCA12) FOAM configuration used at the Met Office, to confirm the positive impacts of assimilating ML-SAR in conjunction with bias-corrected AMSR2 observations at a much higher model resolution. This will not only support the transition of this new sea ice DA approach towards operations but also contribute to FOAM's goal of improving the accuracy of its Arctic sea ice forecasts against the ice charts, which are the reference products used for navigation.

Author contributions. DM trained the Random Forest model, ran the DA experiments and wrote the paper. MM provided expertise on data assimilation, contributed to planning the work, and interpreting the results. MM also commented on the manuscript.

Competing interests. The authors declare no competing interests.

Data availability. Details of how to download the NEMO and SI³ source codes can be found at <https://doi.org/10.5281/zenodo.6334656>. Random Forest code is available at <https://doi.org/10.5281/zenodo.21793412>. Due to intellectual property copyright restrictions, we cannot provide the source code of NEMOVAR.



OSI SAF AMSR2-L3 and AMSR2-L2 files were downloaded from <https://osi-saf.eumetsat.int/products/osi-408-a> and <https://osi-saf.eumetsat.int/products/osi-410-a>, respectively. ML-SAR files were downloaded from CMEMS (<https://doi.org/10.48670/mds-00343>). The Atlantic sector and Greenland ice charts were also obtained from CMEMS (<https://doi.org/10.48670/moi-00128>). The Hudson Bay, Baffin Bay, East Coast and Beaufort Sea charts were obtained from the Canadian Ice Service (<https://iceweb1.cis.ec.gc.ca>). Arctic charts were also downloaded from U.S. NIC (<https://doi.org/10.7265/46cc-3952>). The NIC MASIE product was also obtained from U.S. NIC (<https://doi.org/10.7265/N5GT5K3K>). Finally, 2024-2025 BGEP SIT dataset was extracted from <https://www2.whoi.edu/site/beaufortgyre/data/mooring-data/2024-2025-mooring-data-from-the-bgep-project> and version 300 of SMOS-CS2-S3 product was downloaded from AWI (ftp://ftp.awi.de/sea_ice/product/cryosat2_smos). Due to the size of FOAM runs, which require a large storage space of more than 3 TB, the full model output fields are not made available. They can be shared for non-commercial research use by contacting the authors.

Acknowledgements. We would like to acknowledge Jonah Roberts-Jones, Ed Blockley and Emma Fiedler from the Met Office for their contributions. Jonah contributed with ancillary and observation files needed for the model runs. Ed and Emma contributed with their sea ice expertise to improve the selection of predictors for the Random Forest model. Moreover, we would like to acknowledge Tore Wulf and Matilde Kreiner from DMI for their contribution with ML-SAR data.

Financial support. This work and its contributors were funded by the Met Office Advancing Arctic meteorological and oceanographic Capabilities & services project under the International Science Partnerships Fund (ISPF).

References

- Agnew, T. and Howell, S.: The use of operational ice charts for evaluating passive microwave ice concentration data, *Atmosphere-Ocean*, 41, 317–331, <https://doi.org/10.3137/ao.410405>, 2003.
- Andersen, S., Tonboe, R., Kaleschke, L., Heygster, G., and Pedersen, L. T.: Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration Arctic sea ice, *J. Geophys. Res. Oceans*, 112, C08004, <https://doi.org/10.1029/2006JC003543>, 2007.
- Barbosa Aguiar, A., Bell, M. J., Blockley, E., Calvert, D., Crocker, R., Inverarity, G., King, R., Lea, D. J., Maksymczuk, J., Martin, M. J., Price, M. R., Siddorn, J., Smout-Day, K., Waters, J., and While, J.: The Met Office Forecast Ocean Assimilation Model (FOAM) using a 1/12-degree grid for global forecasts, *Q. J. R. Meteorol. Soc.*, 150, 3827–3852, <https://doi.org/10.1002/qj.4798>, 2024.
- Barton, N., Metzger, E. J., Reynolds, C. A., Ruston, B., Rowley, C., Smedstad, O. M., Ridout, J. A., Wallcraft, A., Frolov, S., Hogan, P., Janiga, M. A., Shriver, J. F., McLay, J., Thoppil, P., Huang, A., Crawford, W., Whitcomb, T., Bishop, C. H.,



- Zamudio, L., and Phelps, M.: The Navy's Earth System Prediction Capability: A new global coupled atmosphere–ocean–sea ice prediction system designed for daily to subseasonal forecasting, *Earth Space Sci.*, 8, <https://doi.org/10.1029/2020EA00119>, 2021.
- 630 Bertino, L. and Lisæter, K. A.: The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans, *J. Oper. Oceanogr.*, 1, 15–18, <https://doi.org/10.1080/1755876X.2008.11020098>, 2008.
- Blockley, E., Fiedler, E., Ridley, J., Roberts, L., West, A., Copsey, D., Feltham, D., Graham, T., Livings, D., Rousset, C., Schroeder, D., and Vancoppenolle, M.: The sea ice component of GC5: coupling SI3 to HadGEM3 using conductive fluxes, *Geosci. Model Dev.*, 17, 6799–6817, <https://doi.org/10.5194/gmd-17-6799-2024>, 2024.
- 635 Bloom, S. C., Takacs, L. L., da Silva, A. M., and Ledvina, D.: Data assimilation using incremental analysis updates, *Mon. Weather Rev.*, 124, 1256–1271, 1996.
- Breiman, L.: Random forests, *Mach. Learn.*, 45, 5–32, <https://doi.org/10.1023/A:1010933404324>, 2001.
- Browne, P., de Boisseson, E., Keeley, S., Pelletier, C., and Zuo, H.: Sea ice data assimilation in ORAS6, *The Cryosphere*, 20, 3299–3311, <https://doi.org/10.5194/tc-20-3299-2026>, 2026.
- 640 de Rosnay, P., Browne, P., de Boissès, E., Fairbairn, D., Hirahara, Y., Ochi, K., Schepers, D., Weston, P., Zuo, H., Balmaseda, M. A., Balsamo, G., Bonavita, M., Bormann, N., Brown, A., Chrust, M., Dahoui, M., De Chiara, G., English, S., Geer, A., Healy, S., Hersbach, H., Laloyaux, P., Magnusson, L., Massart, S., McNally, A., Pappenberger, F., and Rabier, F.: Coupled data assimilation at ECMWF: current status, challenges and future developments, *Q. J. R. Meteorol. Soc.*, 148, 2672–2702, <https://doi.org/10.1002/qj.4330>, 2022.
- 645 Eicken, H.: Ocean science: Arctic sea ice needs better forecasts, *Nature*, 497, 431–433, 2013.
- Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T.: Predictability of the Arctic sea ice edge, *Geophys. Res. Lett.*, 43, 1642–1650, <https://doi.org/10.1002/2015GL067232>, 2016.
- Gregory, W., Bushuk, M., Adcroft, A., Zhang, Y., and Zanna, L.: Deep learning of systematic sea ice model errors from data assimilation increments, *J. Adv. Model. Earth Syst.*, 15, e2023MS003757, <https://doi.org/10.1029/2023MS003757>, 2023.
- 650 Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., and Zanna, L.: Machine learning for online sea ice bias correction within global ice–ocean simulations, *Geophys. Res. Lett.*, 51, e2023GL106776, <https://doi.org/10.1029/2023GL106776>, 2024.
- Guiavarc'h, C., Roberts-Jones, J., Harris, C., Lea, D. J., Ryan, A., and Ascione, I.: Assessment of ocean analysis and forecast from an atmosphere-ocean coupled data assimilation operational system, *Ocean Sci.*, 15, 1307–1326, <https://doi.org/10.5194/os-15-1307-2019>, 2019.



- 655 Helfrich, S. R., McNamara, D., Ramsay, B. H., Baldwin, T., and Kasheta, T.: Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS), *Hydrol. Process.*, 21, 1576–1586, <https://doi.org/10.1002/hyp.6720>, 2007.
- Ivanova, N., Pedersen, L. T., Tonboe, R. T., Kern, S., Heygster, G., Lavergne, T., Sørensen, A., Saldo, R., Dybkjær, G., Brucker, L., and Shokr, M.: Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using passive microwave observations, *The Cryosphere*, 9, 1797–1817, <https://doi.org/10.5194/tc-9-1797-2015>, 2015.
- 660 Jin, J., Lin, H. X., Segers, A., Xie, Y., and Heemink, A.: Machine learning for observation bias correction with application to dust storm data assimilation, *Atmos. Chem. Phys.*, 19, 10009–10026, <https://doi.org/10.5194/acp-19-10009-2019>, 2019.
- Kern, S., Rösel, A., Pedersen, L. T., Ivanova, N., Saldo, R., and Tonboe, R. T.: The impact of melt ponds on summertime microwave brightness temperatures and sea-ice concentrations, *The Cryosphere*, 10, 2217–2239, <https://doi.org/10.5194/tc-10-2217-2016>, 2016.
- 665 Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., and Tonboe, R. T.: Satellite passive microwave sea-ice concentration data set intercomparison for Arctic summer conditions, *The Cryosphere*, 14, 2469–2493, <https://doi.org/10.5194/tc-14-2469-2020>, 2020.
- 670 Kettleborough, J. A., Davis, P. J., Comer, R. E., Martin, M. J., Barbosa Aguiar, A., Collier, T., Scaife, A. A., Clark, A., Mancell, J., Smout-Day, K. A., and Hardiman, S. C.: Global Seasonal Forecasting System 6 (GloSea6): A Large Ensemble Seasonal Forecasting System, *Mon. Weather Rev.*, 154, 39–57, <https://doi.org/10.1175/MWR-D-25-0007.1>, 2026.
- Krishfield, R. A., Proshutinsky, A., Tateyama, K., Williams, W. J., Carmack, E. C., McLaughlin, F. A., and Timmermans, M.-L.: Deterioration of perennial sea ice in the Beaufort Gyre from 2003 to 2012 and its impact on the oceanic freshwater cycle, *J. Geophys. Res. Oceans*, 119, 1271–1305, <https://doi.org/10.1002/2013JC008999>, 2014.
- 675 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 ESA CCI sea-ice concentration climate data records, *The Cryosphere*, 13, 49–78, <https://doi.org/10.5194/tc-13-49-2019>, 2019.
- 680 Lea, D. J., Mirouze, I., Martin, M. J., King, R. R., Hines, A., Walters, D., and Thurlow, M.: Assessing a new coupled data assimilation system based on the Met Office coupled atmosphere–land–ocean–sea ice model, *Mon. Weather Rev.*, 143, 4678–4694, <https://doi.org/10.1175/MWR-D-15-0174.1>, 2015.
- Lea, D. J., While, J., Martin, M. J., Weaver, A., Storto, A., and Chrust, M.: A new global ocean ensemble system at the Met Office: Assessing the impact of hybrid data assimilation and inflation settings, *Q. J. R. Meteorol. Soc.*, 148, 1996–2030, <https://doi.org/10.1002/qj.4292>, 2022.
- 685



- Lemieux, J.-F., Beaudoin, C., Dupont, F., Roy, F., Smith, G. C., Shlyueva, A., Buehner, M., Caya, A., Chen, J., Carrieres, T., Pogson, L., DeRepentigny, P., Plante, A., Pestieau, P., Pellerin, P., Ritchie, H., Garric, G., and Ferry, N.: The Regional Ice Prediction System (RIPS): Verification of forecast sea ice concentration, *Q. J. R. Meteorol. Soc.*, 142, 632–643, <https://doi.org/10.1002/qj.2526>, 2016.
- 690 Lisæter, K. A., Rosanova, J., and Evensen, G.: Assimilation of ice concentration in a coupled ice-ocean model, using the Ensemble Kalman filter, *Ocean Dynam.*, 53, 368–388, <https://doi.org/10.1007/s10236-003-0049-4>, 2003.
- Maaß, N. and Kaleschke, L.: Improving passive microwave sea ice concentration algorithms for coastal areas: applications to the Baltic Sea, *Tellus A*, 62, 393–410, <https://doi.org/10.1111/j.1600-0870.2010.00452.x>, 2010.
- MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal Forecast System version 5 (GloSea5): A high resolution seasonal forecast system, *Q. J. R. Meteorol. Soc.*, 141, 1072–1084, <https://doi.org/10.1002/qj.2396>, 2014.
- 695 Madec, G.: NEMO Ocean Engine, Tech. Rep. 27, Pôle de modélisation de l'Institut Pierre-Simon Laplace, <http://hdl.handle.net/2122/13309>, 2017.
- Madec, G. and Imbard, M.: A global ocean mesh to overcome the North Pole singularity, *Clim. Dynam.*, 12, 381–388, 1996.
- 700 Malmgren-Hansen, D., Pedersen, L. T., Nielsen, A. A., Kreiner, M. B., Saldo, R., Skriver, H., Lavelle, J., Buus-Hinkler, J., and Krane, K. H.: A convolutional neural network architecture for Sentinel-1 and AMSR2 data fusion, *IEEE Trans. Geosci. Remote Sens.*, 59, 1890–1902, <https://doi.org/10.1109/TGRS.2020.3004539>, 2021.
- Mason, I.: Dependence of the critical success index on sample climate and threshold probability, *Aust. Meteorol. Mag.*, 37, 75–81, 1989.
- 705 Meier, W. N., Hovelsrud, G. K., van Oort, B. E. H., Key, J. R., Kovacs, K. M., Michel, C., Haas, C., Granskog, M. A., Gerland, S., Perovich, D. K., Makshtas, A., and Reist, J. D.: Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity, *Rev. Geophys.*, 52, 185–217, <https://doi.org/10.1002/2013RG000431>, 2014.
- Meier, W. N., Fetterer, F., Stewart, J. S., and Helfrich, S.: How do sea-ice concentrations from operational data compare with passive microwave estimates? Implications for improved model evaluations and forecasting, *Ann. Glaciol.*, 56, 332–340, <https://doi.org/10.3189/2015AoG69A694>, 2015.
- 710 Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M., Ottersen, G., Pritchard, H., and Schuur, E.: Polar Regions, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, 2019.



- Mignac, D., Martin, M., Fiedler, E., Blockley, E., and Fournier, N.: 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, Q. J. R. Meteorol. Soc., 148, 1144–1167, <https://doi.org/10.1002/qj.4252>, 2022.
- Mignac, D., Waters, J., Lea, D. J., Martin, M. J., While, J., Weaver, A. T., Vidard, A., Guiavarc'h, C., Storkey, D., Ford, D., Blockley, E. W., Baker, J., Haines, K., Price, M. R., Bell, M. J., and Renshaw, R.: Improvements to the Met Office's global ocean–sea ice forecasting system including model and data assimilation changes, Geosci. Model Dev., 18, 3405–3425, <https://doi.org/10.5194/gmd-18-3405-2025>, 2025.
- Oke, P. R. and Sakov, P.: Representation error of oceanic observations for data assimilation, J. Atmos. Oceanic Technol., 25, 1004–1017, <https://doi.org/10.1175/2007JTECHO558.1>, 2008.
- OSI SAF: Product User Manual for Global Sea Ice Concentration Level 2 and Level 3 (OSI-410-a, OSI-401-d, OSI-408-a), Version 1.2, prepared by Baordo, F., Vargas, L. F., and Howe, E., EUMETSAT Ocean and Sea Ice Satellite Application Facility, 2023.
- Palermé, C., Lavergne, T., Rusin, J., Melsom, A., Brajard, J., Kvanum, A. F., Sørensen, A. M., Bertino, L., and Müller, M.: Improving short-term sea ice concentration forecasts using deep learning, The Cryosphere, 18, 2161–2176, <https://doi.org/10.5194/tc-18-2161-2024>, 2024.
- Pedersen, R. A., Cvijanovic, I., Langen, P. L., and Vinther, B. M.: The impact of regional Arctic sea ice loss on atmospheric circulation and the NAO, J. Clim., 29, 889–902, <https://doi.org/10.1175/JCLI-D-15-0315.1>, 2016.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.
- Peterson, K. A., Arribas, A., Hewitt, H. T., Keen, A. B., Lea, D. J., and McLaren, A. J.: Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system, Clim. Dynam., 44, 147–162, <https://doi.org/10.1007/s00382-014-2190-9>, 2015.
- Posey, P. G., Metzger, E. J., Wallcraft, A. J., Hebert, D. A., Allard, R. A., Smedstad, O. M., Phelps, M. W., Fetterer, F., Stewart, J. S., Meier, W. N., and Helfrich, S. R.: Improving Arctic sea ice edge forecasts by assimilating high horizontal resolution sea ice concentration data into the US Navy's ice forecast systems, The Cryosphere, 9, 1735–1745, <https://doi.org/10.5194/tc-9-1735-2015>, 2015.
- 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.



- Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., and Schroeder, D.: The sea ice model component of
 745 HadGEM3-GC3.1, *Geosci. Model Dev.*, 11, 713–723, <https://doi.org/10.5194/gmd-11-713-2018>, 2018.
- Skachko, S., Buehner, M., Laroche, S., Lapalme, E., Smith, G., Roy, F., Surcel-Colan, D., Bélanger, J.-M., and Garand, L.:
 Weakly coupled atmosphere–ocean data assimilation in the Canadian global prediction system (v1), *Geosci. Model Dev.*, 12,
 5097–5112, <https://doi.org/10.5194/gmd-12-5097-2019>, 2019.
- Shokr, M. and Kaleschke, L.: Impact of surface conditions on thin sea ice concentration estimate from passive microwave
 750 observations, *Remote Sens. Environ.*, 121, 36–50, <https://doi.org/10.1016/j.rse.2012.01.005>, 2012.
- Smith, L. C. and Stephenson, S. R.: New Trans-Arctic shipping routes navigable by mid-century, *P. Natl. Acad. Sci. USA*,
 110, E1191–E1195, <https://doi.org/10.1073/pnas.1214212110>, 2013.
- Spreen, G., Kaleschke, L., and Heygster, G.: Sea ice remote sensing using AMSR-E 89-GHz channels, *J. Geophys. Res.*
Oceans, 113, C02S03, <https://doi.org/10.1029/2005JC003384>, 2008.
- 755 Stokholm, A., Wulf, T., Kucik, A., Saldo, R., Buus-Hinkler, J., and Pedersen, L. T.: AI4SeaIce: toward solving ambiguous
 SAR textures in convolutional neural networks for automatic sea ice concentration charting, *IEEE Trans. Geosci. Remote*
Sens., 60, 1–13, <https://doi.org/10.1109/TGRS.2022.3149323>, 2022.
- Storkey, D., Blaker, A. T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., Calvert, D., Graham, T., Hewitt, H. T.,
 Hyder, P., Kuhlbrodt, T., Rae, J. G. L., and Sinha, B.: UK Global Ocean GO6 and GO7: A traceable hierarchy of model
 760 resolutions, *Geosci. Model Dev.*, 11, 3187–3213, <https://doi.org/10.5194/gmd-11-3187-2018>, 2018.
- Stroeve, J. and Notz, D.: Changing state of Arctic sea ice across all seasons, *Environ. Res. Lett.*, 13, 103001,
<https://doi.org/10.1088/1748-9326/aade56>, 2018.
- Thomas, D. R., Martin, S., Rothrock, D. A., and Steele, M.: Assimilating satellite concentrations into an Arctic mass balance
 model: 1979–1985, *J. Geophys. Res.*, 101, 20849–20868, 1996.
- 765 Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1.
 Sea ice concentrations, *J. Geophys. Res. Atmos.*, 119, 2864–2889, <https://doi.org/10.1002/2013JD020316>, 2014.
- Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J.,
 Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S.,
 Willett, M., et al.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations,
 770 *Geosci. Model Dev.*, 12, 1909–1963, <https://doi.org/10.5194/gmd-12-1909-2019>, 2019.
- Wang, Y.-R. and Li, X.-M.: Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning, *Earth Syst.*
Sci. Data, 13, 2723–2742, <https://doi.org/10.5194/essd-13-2723-2021>, 2021.



- Waters, J., Lea, D., Martin, M. J., Mirouze, I., Weaver, A., and While, J.: Implementing a variational data assimilation system in an operational 1/4 degree global ocean model, *Q. J. R. Meteorol. Soc.*, 141, 333–349, 2015.
- 775 Wei, T., Yan, Q., Qi, W., Ding, M., and Wang, C.: Projections of Arctic sea ice conditions and shipping routes in the twenty-first century using CMIP6 forcing scenarios, *Environ. Res. Lett.*, 15, <https://doi.org/10.1088/1748-9326/abb2c8>, 2020.
- Willett, M., Brooks, M., Bushell, A., Earnshaw, P., Smith, S., Tomassini, L., Best, M., Boutle, I., Brooke, J., Edwards, J. M., Elvidge, A. D., Furtado, K., Hardacre, C., Hartley, A. J., Hewitt, A. J., Johnson, B., Lock, A., Malcolm, A., Mulcahy, J., Müller, E., Renfrew, I. A., Rumbold, H., Rooney, G. G., Sellar, A., Ujiie, M., van Niekerk, A., Wiltshire, A., and Whittall, M.:
 780 The Met Office Unified Model Global Atmosphere 8.0 and JULES Global Land 9.0 configurations, *Geosci. Model Dev.*, 19, 1473–1517, <https://doi.org/10.5194/gmd-19-1473-2026>, 2026.
- Wulf, T., Buus-Hinkler, J., Singha, S., Shi, H., and Kreiner, M. B.: Pan-Arctic sea ice concentration from SAR and passive microwave, *The Cryosphere*, 18, 5277–5300, <https://doi.org/10.5194/tc-18-5277-2024>, 2024.
- Wulf, T., Buus-Hinkler, J., Singha, S., Dasgupta, N., Athanasiadis, A., and Kreiner, M. B.: A decade of sea ice concentration
 785 retrieved from Sentinel-1, *Remote Sens. Environ.*, 337, 115252, <https://doi.org/10.1016/j.rse.2026.115252>, 2026.
- Yu, Y., Maykut, G. A., and Rothrock, D. A.: Changes in the thickness distribution of Arctic sea ice between 1958–1970 and 1993–1997, *J. Geophys. Res. Oceans*, 109, C08004, <https://doi.org/10.1029/2003JC001982>, 2004.
- Zhang, X., Xu, D., Shen, F., and Min, J.: Impacts of offline nonlinear bias correction schemes using the machine learning technology on the all-sky assimilation of cloud-affected infrared radiances, *J. Adv. Model. Earth Syst.*, 16, e2024MS004281,
 790 <https://doi.org/10.1029/2024MS004281>, 2024.
- Zhou, Y. and Grassotti, C.: Development of a machine learning-based radiometric bias correction for NOAA's Microwave Integrated Retrieval System (MiRS), *Remote Sens.*, 12, 3160, <https://doi.org/10.3390/rs12193160>, 2020.