CLIMATE-SPACE - THEME I - B. ADDITIONAL ESSENTIAL CLIMATE VARIABLES (ECVS) - NEW ECV PRODUCTS

SAGE CCI

(Sea Ice Age and Drift)

Product Validation and Algorithm Selection Report (PVASR)

Prime & Science Lead: Signe Aaboe
Met Norway, Norway

Technical Officer: Sarah Connors
ESA ECSAT, United Kingdom

Consortium: Norwegian Meteorological Institute (MET Norway)
Science and Technology AS (S&T)
Nansen Environmental and Remote Sensing Center (NERSC)
University of Bremen (UB)
Université catholique de Louvain (UCLouvain)
University of Hamburg (UH)
University of Manitoba (UM)

Document Change Log

Issue	Author	Date	Change	Status
0.5	D. Fantin (S&T)	2025.09.16	Document created	
1.0	Hannah Niehaus, Gunnar Spreen (UB), Anton Korosov (NERSC) Signe Aaboe, Thomas Lavergene (MET Norway), Stefan Kern (UH)	2025.04.01	V1.0 consolidated	Released to ESA
1.1	Hannah Niehaus, Gunnar Spreen (UB), Anton Korosov (NERSC) Signe Aaboe, Thomas Lavergene (MET Norway), Stefan Kern (UH)	2025.05.18	Update addressing ESA feedback	Released to ESA

Document Approval

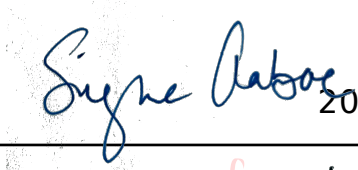
Prepared by	Hannah Niehaus, Lead author, University of Bremen	
Issued by	Daniele Fantin, Project Manager, S&T	 Digitally signed by Daniele Fantin Date: 2026.05.20 13:27:07 CEST
Checked by	Signe Aaboe Science Leader, MET Norway	 2026-05-20
Approved by	Sarah Connors ESA Technical Officer	 Digitally signed by Sarah Connors Date: 2026.05.27 09:59:47 +01'00'

Table of Contents

Acronyms and Abbreviations

AARI	Arctic and Antarctic Research Institute (RUS)
ADP	Algorithm Development Plan
AMSR2	Advanced Microwave Scanning Radiometer 2
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
AR	Annual Review
ASCAT	Advanced Scatterometer
ASCAT_KNMI	ASCAT algorithm from the Royal Netherlands Meteorological Institute
ASIMIR	Atmosphere, Surface, Ice Microwave Inversion Retrieval
ATBD	Algorithm Theoretical Basis Document
AWI	Alfred-Wegener-Institute for Marine and Polar Research
BGEP	Beaufort Gyre Exploration Project
C3S	Copernicus Climate Change Service
C3Sp	C3S sea-ice type product given in probability estimates
CAR	Climate Assessment Report
CCI	ESA's Climate Change Initiative
CDR	Climate Data Records
CFOSAT	Chinese-French Oceanography Satellite
CMEMS	Copernicus Marine Service
CMIP	Coupled Model Intercomparison Project
CM SAF	The Climate Monitoring Satellite Application Facility
CMUG	Climate Modelling User Group
CP	Communication Package
CRDP	Climate Research Data Package
CRG	Climate Research Group
DAL	Distance Along the Line
DMI	Danish Meteorological Institute (DK)
DNN	Diffusion Neural Network
DOI	Digital Object Identifier
E3UB	End-to-End ECV Uncertainty Budget
ECICE	Enhanced Classification for Sea Ice Concentration
ECV	GCOS Essential Climate Variable
ECCC	Environment and Climate Change Canada (CA)
ECMWF	European Center for Medium-Range Weather Forecasts
EO	Earth Observation
ERA5	ECMWF Reanalysis ver. 5
ERS	European Remote-Sensing Satellite
ES	Executive Summary
ESA	European Space Agency
EUMETSAT	European Organization for the Exploration of Meteorological Satellites
FCDR	Fundamental Climate Data Record
FM	Final Meeting
FP	Final Presentation
FR	Final Report
FYI	First-Year Ice
GCOS	WMO/ICO/UNEP Global Climate Observing System
GR	Gradient Ratio

HY-2	Haiyang-2
IABP	International Arctic Buoy Programme
ICESat-2	Ice, Cloud and land Elevation Satellite 2
ICDR	Interim Climate Data Record
IPCC	Intergovernmental Panel on Climate Change
IPS	Ice Profiling sonar
ITT	Invitations to Tender
JAMSTEC	Japan Agency for Marine-Earth Science and Technology (JP)
JAXA	Japan Aerospace Exploration Agency (JP)
KNMI	Royal Netherlands Meteorological Institute
KO	Kick Off
LM	Lagrangian Mesh
METNO or MET Norway	Norwegian Meteorological Institute
MIZ	Marginal Ice Zone
ML	Machine Learning
MPR	Monthly Progress Report
MS	MileStone
MYI	Multiyear Ice
NaN	Not a Number
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
NERSC	Nansen Environmental and Remote Sensing Center
NH	Northern Hemisphere
NIC	National Ice Center
NorESM	Norwegian Earth System Model
NSIDC	National Snow and Ice Data Center (US)
Obs4MIPS	Observations for Model Intercomparison Projects
OSI SAF	The Ocean and Sea Ice Satellite Application Facility
OSISAFp	OSI SAF sea-ice type product given in probability estimates
PM	Progress Meeting, Project Manager
PMP	Project Management Plan
PMW	Passive Microwave
PSD	Product Specification Document
PR	Polarisation Ratio
PSH	Project Scientific Highlights
PUG	Product User Guide
PVASR	Product Validation and Algorithm Selection Report
PVIR	Product Validation and Intercomparison Report
PVP	Product Validation Plan
QRS	Quarterly Status Reports
QuikSCAT	Quick Scatterometer Mission
RCM	Radarsat Constellation Mission
RID	Review Item Discrepancy
RMSD	Root Mean Square Difference
RMSE	Root Mean Square Error
RRDP	Round Robin Data Package
SAGE	Sea Ice Age and Drift

SAR	Synthetic Aperture Radar
SCAT	Scatterometer
SH	Southern Hemisphere
SIC	Sea Ice Concentration
SID	Sea Ice Drift
SIMIP	Sea Ice Model Intercomparison Project
SMMR	Scanning Multichannel Microwave Radiometer
SRD	System Requirement Document
SoW	Statement of Work
SSD	System Specification Document
SSMI,SSM/I	Special Sensor Microwave - Imager
SSMIS	Special Sensor Microwave - Imager/Sounder
SYI	Second-Year Ice
S&T	S&T Norway AS
T2m	2 Meter Temperature
TB	Brightness Temperature
UB	University of Bremen
UCLouvain	Université Catholique de Louvain
UH	University of Hamburg
ULS	Upward-Looking Sonar
UM	University of Manitoba
UNK	Unknown
URD	User Requirement Document
UWR	User Workshop Report
WAI	Warm Air Intrusion
WBS	Work Breakdown Structure
WMO	World Meteorological Organization
WP	Work Package
WPD	Work Package Description
YI	Young Ice

1 Introduction

1.1 Purpose and Scope

This document contains the Product Validation and Algorithm Selection Report (PVASR) for the SAGE project for CLIMATE-SPACE - THEME I - B. ADDITIONAL ESSENTIAL CLIMATE VARIABLES (ECVS) - NEW ECV PRODUCTS, in accordance with the contract [AD1], SoW [AD2] and proposal [AD3-AD10].

The report summarises the initial results of the algorithm Round-Robin intercomparison exercise carried out to evaluate candidate algorithms for the generation of climate data records (CDRs) of the sea-ice age and type in both the Arctic and Antarctic.

1.2 Document Overview

This document is structured as follows:

- Chapter 1 introduces this document.
- Chapter 2 describes the round-robin exercise
- Chapter 3 details the results of the round-robin exercise
- Chapter 4 presents the final ranking of the sea ice age algorithms

1.3 Applicable Documents

No	Doc. Id	Doc. Title	Date	Issue/ Revision/ Version
AD-1	4000147560/25/I-LR	ESA Contract No. 4000147560/25/I-LR	12/03/2025	NA
AD-2	ESA-EOP-SC-AMT-2024-36	Statement of Work and Annexes and Appendexes	31/07/2024	1.2
AD-3	METNO-ESA-SAGE-CL-001	SAGE Cover Letter	8/11/2024	1.0
AD-4	METNO-ESA-SAGE-TPROP-001	SAGE Technical Proposal	8/11/2024	1.0
AD-5	METNO-ESA-SAGE-IPROP-001	SAGE Implementation Proposal	8/11/2024	1.0
AD-6	METNO-ESA-SAGE-MPROP-001	SAGE Management Proposal	8/11/2024	1.0
AD-7	METNO-ESA-SAGE-FPROP-001	SAGE Financial Proposal	8/11/2024	1.0
AD-8	METNO-ESA-SAGE-CPROP-001	SAGE Contractual Proposal	8/11/2024	1.0
AD-9	METNO-ESA-SAGE-BF-001	SAGE Background and Facilities	8/11/2024	1.0
AD-10	METNO-ESA-SAGE-CV-001	SAGE Curricula Vitae	8/11/2024	1.0
RD-1	METNO-ESA-SAGE-CCI-URD-001	SAGE User Requirement Document	6/11/2025	1.0
RD-2	METNO-ESA-SAGE-CCI-PVP-001	SAGE Product Validation Plan	xx/04/2026	1.0

Note: If not provided, the reference applies to the latest released Issue/Revision/Version

2 Description of the round-robin exercise

The sea-ice age and type algorithm intercomparison selection was organised in a round-robin exercise to ensure maximum transparency and reproducibility of our results. A core element of the round-robin exercise was the Round-Robin Data Package (RRDP), which holds a collection of high-quality reference data collocated with the EO data needed to run the algorithms. The RRDP is introduced in detail later in the section. The RRDP was built iteratively with frequent reviews and interactions within the group.

Starting from the RRDP, results from the selected sea-ice age and type algorithms were collected and collocated in the same grid and format. Some of the simpler sea-ice type/age algorithms were specifically re-implemented and re-run for this exercise to gain better control on their results. Other, more complex algorithms, were not re-implemented in the project but only run on the EO data contained in the RRDP. Additionally, some already processed sea-ice age/type classification products were downloaded from external data portals and collocated on the same grid and format. This allowed our exercise to also compare existing products without re-implementing them. Some algorithms require multiyear ice (MYI) and first-year ice (FYI) tie-points. To improve compatibility of the results, a common set of tiepoints based on the RRDP was prepared in the project and used in running the algorithms.

Finally, a set of software and computational notebooks was created to implement the selected accuracy metrics. A significant part of this software deals with the harmonisation of the results from different types of algorithm output: probabilities, partial concentrations, and binary classifications had to be harmonised in order to be compared. These notebooks also create the summary graphs and tables included in this report.
All software developed for the round-robin exercise was developed in the GitHub organisation for the project (https://github.com/esa-sage) and will be made open at the end of the exercise.

The purpose of the RRDP is to collect a reference dataset of high-certainty sea-ice type and age, along with relevant auxiliary information. This enables the evaluation of different sea-ice age and type algorithms within the same framework, and improves the understanding of error contributions from different sources. The RRDP consists of two complementary components: snapshot maps and a buoy-tracking dataset, which differ in their input data types and structures. The two RRDP components are described in detail in the following sections.

2.1 Snapshot-RRDP

The snapshot-RRDP is created by integrating daily ice-type maps sourced from national ice-charting services (Stage of Development ice charts) and automated sea-ice-type charts derived from Sentinel-1 SAR data.

2.1.1 Source data

The sources and descriptions of each ice-type map in the snapshot-RRDP are listed below.

Navigational (manually interpreted) ice charts from the national ice charting services:

- Arctic and Antarctic Research Institute (AARI): http://wdc.aari.ru/datasets/d0015/
- Canadian Ice Service(CIS): https://noaadata.apps.nsidc.org/NOAA/G02171/
- Danish Meteorological Institute (DMI): https://data.marine.copernicus.eu/product/SEAICE_ARC_SEAICE_L4_NRT_OBSERVATIONS_011_002/services
- U.S. National Ice Center (NIC): https://noaadata.apps.nsidc.org/NOAA/G10013/

Automated sea-ice type ice charts for the Arctic:

- Automated ice charts from the DMI (autoDMI, https://doi.org/10.48670/mds-00343): https://data.marine.copernicus.eu/product/SEAICE_ARC_PHY_AUTO_L3_MYNRT_011_023/files
- Sentinel-1 SAR ice-type classification (S1, https://doi.org/10.3390/rs12132165) from NERSC/MET Norway: https://thredds.met.no/thredds/catalog/cmems/si-tac/cmems_obs-si_arc_phy_nrt_L4-auto_P1D/cmems_obs-si_arc_phy-icetype_nrt_L4-auto_P1D/catalog.html
  (Previously accessed from CMEMS https://data.marine.copernicus.eu/product/SEAICE_ARC_PHY_AUTO_L4_NRT_011_015/files)

Table 2.1: Overview of datasets, their spatial resolution, and temporal and spatial coverage.

Dataset	Spatial resolution	Temporal coverage	Spatial coverage
AARI	1 km, shape files	Weekly since 2008 (Arctic), bi-weekly since 2007 (Antarctica)	Both hemispheres
CIS	1 km, shape files	Weekly since 2008	Canadian Arctic
DMI	1 km, shape files	Twice weekly since September 2020	Greenland Waters
NIC	1 km, shape files	Daily since 2003 (Arctic), changing between weekly and bi-weekly since 2003 until March 2023 (Antarctica)	Both hemispheres
autoDMI	0.5 km pixel	Daily since October 2014	Arctic, limited to Sentinel-1 coverage
S1	1 km pixel	Daily since 2021	European Arctic, limited to Sentinel-1 coverage

The data package distinguishes four ice types: young ice, first-year ice, second-year ice and multiyear ice, and is provided in the form of daily NetCDF files and for the hemispheres separately. If no ice-type matches between the source datasets could be identified, either due to different ice-type classification or because of lacking data sets, no file is produced for that day. Figure 2.1 shows how often the different data sources contribute to the files produced, by month and hemisphere. We note that for any future upgrade of the Snapshot-RRDP, a wider time window could be used to find matching ice maps. This will allow for more overlapping sources and will largely expand the data package. Especially the Southern Hemisphere would gain from this, since different national ice services appear to supplement each other on different days rather than overlap on the exact same days (see Figure 2.1, the gap in SH ice-type matches from 2022).

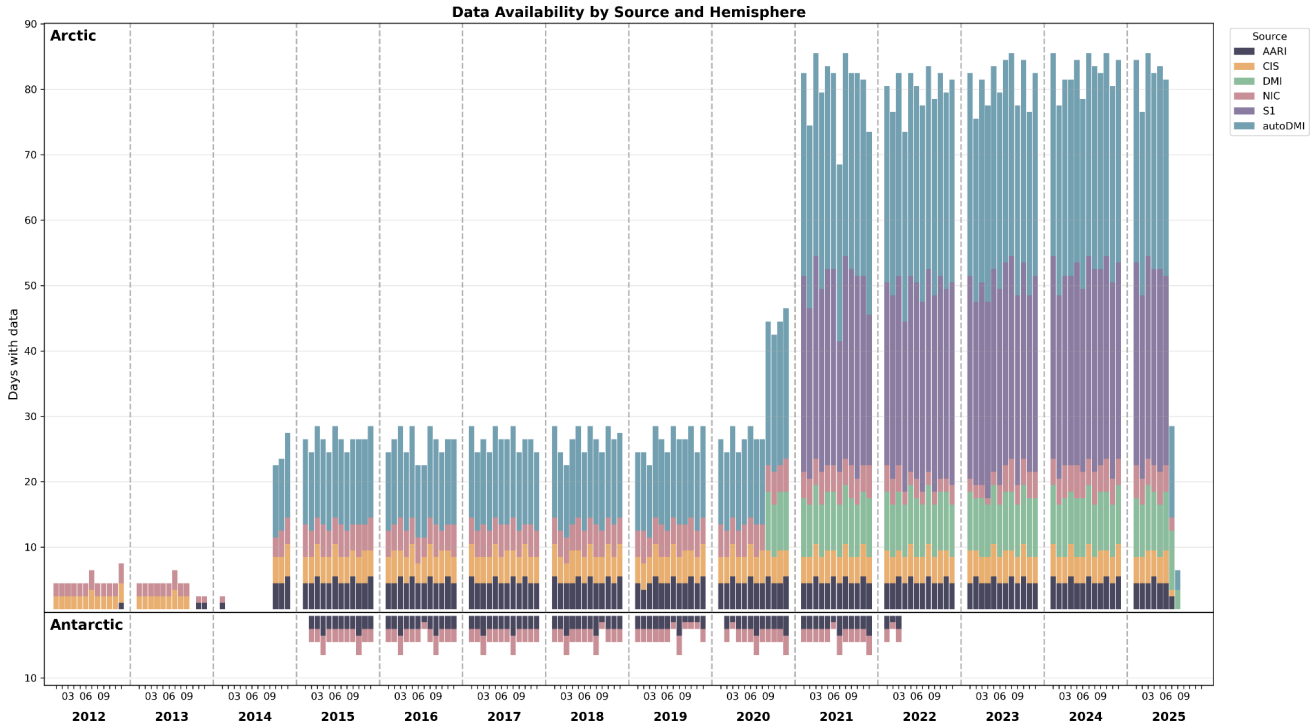


Figure 2.1: Monthly contribution of the different data sources to the snapshot-RRDP data package, for the Northern (upward facing bars) and Southern (downward facing bars) hemisphere. Different bar colours represent different origins of the source data.

2.1.2 Data structure

The snapshot-RRDP is provided as daily NetCDF files containing two-dimensional maps on the EASE-grid 2 (Brodzik et al., 2012) projection with a spatial resolution of 25 km. The data are produced separately for each hemisphere.

Each file contains the following layers:

- Ice type if consistent between datasets (icetype_matches)
- MYI concentration if consistent between datasets (myi_concentration_matches)
- For each available dataset of the day, the dominant ice type, its concentration and the MYI concentration; for icecharts, the total sea ice concentration is additionally provided
- AMSR brightness temperatures in horizontal and vertical polarisation
- ERA5 variables, including 2 m air temperature, skin temperature, total column water vapour, total column cloud liquid water, total column cloud ice water
- ASCAT backscatter data converted to 40-degree incidence angle

If no agreement in ice type between the source datasets can be identified for a given day, no file is produced.

2.1.3 Mapping data to a 25 km EASE grid

The source data have different original formats, structures, and resolutions. The ice charts come as shapefiles in SIGRID formats with “egg codes” (WMO & IOC, 2010) for the ice concentration and stage of development. The polygon geometries are first reprojected from their source coordinate reference system to the EASE-Grid 2.0 projection. To accurately capture polygon boundaries and minimise aliasing artefacts, the data is initially rasterised at a 1 km spatial resolution (25× oversampling of the target 25 km grid). For each of the four ice-type classes, the polygon concentration values are rasterised separately as pixel intensities. The 1 km concentration values are then aggregated into 25×25 blocks and averaged to produce the final 25 km-resolution ice-type concentrations. The dominant ice type at each pixel is defined as the ice type with the highest concentration, along with its corresponding concentration. A geometric coverage mask is calculated independently by rasterising all polygon boundaries at 1 km resolution, then averaging to determine what fraction of each 25 km pixel is covered by input data. Pixels are masked (set to invalid) where less than 95% of the pixel area is covered by polygon data or where the total sea ice concentration is below 1%.

The automated ice charts (autoDMI) and the Sentinel-1 derived sea-ice types (S1) are already provided as gridded products and at spatial resolutions of 0.5 km and 1 km, respectively. To obtain the dominant ice type and its concentration, as well as the total sea-ice concentration in the desired grid and resolution, the valid source pixels are sorted into the target grid cells and counted. Ice type-specific concentrations are calculated, and the dominant ice type and its concentration are determined. Again, pixels are masked if < 95% of the expected pixel coverage is valid.

2.1.4 Method for deriving ice type matches

If at least two data sources provide maps for a day, the ice type maps are rasterised and reprojected to the EASE grid at 25 km spatial resolution. Herein, ice type classes of more detailed distinction are reduced to the major WMO ice classes: multiyear ice (MYI, 4), second-year ice (SYI, 3), first-year ice (FYI, 2) and young ice (YI, 1). For ice charts that provide discrimination of (maximum three) different ice types per pixel, the dominant ice type and its concentration are determined per pixel and used for further processing.

Two kinds of matches are identified:

- Pure ice types matches: concentration of dominant ice type > 80%
- MYI concentration matches: available products are within +-20% of the mean MYI concentration

For each product, if the concentration of the dominant ice type is below 80%, the pixel is considered low certainty for “pure” ice type representation, and its value is thus set to invalid. The remaining products are compared, excluding invalid values. If all of them, but at least two, agree on the ice type, the output array of matched ice types is set to this value. The threshold of 80% originates from the discrete concentration intervals used in ice charts (typically 10–20%), making 80% the highest practical threshold that avoids losing a large number of matching points.

To find matches in the MYI concentration, the mean of all finite MYI concentration values is calculated pixelwise. Pixels where only one or no finite value is available are masked. The remaining data are filtered for pixels where the MYI concentration of all products is within ±20% of the mean.

2.1.5 Data access and format

The snapshot-RRDP is provided in the form of daily NetCDF files, which are accessible from https://data.seaice.uni-bremen.de/IceAge_RRDP/Snapshot/.

The naming convention is as follows:

YYYYmmdd_<hemisphere>_<version number>_<file version number>.nc

where the first 8 digits contain the date, the hemisphere can be 'N' or 'S' for Northern or Southern Hemisphere, respectively, and the two version numbers represent the upgrade version of the RRDP data and the file version (changes related to e.g. metadata, etc., but not the data itself), respectively.

The files each contain the variables listed in Table 2.2 mapped onto the 25 km-resolution EASE grid, provided that the auxiliary datasets are available. In particular, for ASCAT data before 2016 and for AMSR2 data before mid-July 2012, this is not the case.

Table 2.2: Description of variables in daily snapshot-RRDP files.

Variable name	**Description**
icetype_matches	Dominant ice type where all available datasets agree on ice type and its concentration >=80%
myi_concentration_matches	Mean multiyear ice concentration if all available datasets agree within ±20%
Ice chart section 'XXX' may be replaced by "AARI", "CIS", "DMI", "NIC", "autoDMI" or "S1", depending on the daily availability of datasets	
XXX_dom_icetype	Dominant ice type according to the product XXX
XXX_dom_icetype_conc	Concentration of the dominant ice type according to the product XXX
XXX_myi_conc	Multiyear ice concentration according to the product XXX
XXX_total_ice_conc	Total sea-ice concentration according to the product XXX
AMSR data section Every variable exists with "_M" and "_E" for the morning and evening overpass of the satellite	
AMSR2_TB6.9V	AMSR brightness temperature at 6.9 GHz, vertical/horizontal polarisation (K)

AMSR2_TB6.9H	
AMSR2_TB10.7V AMSR2_TB10.7H	AMSR brightness temperature at 10.7 GHz, vertical/horizontal polarisation (K)
AMSR2_TB18.7V AMSR2_TB18.7H	AMSR brightness temperature at 18.7 GHz, vertical/horizontal polarisation (K)
AMSR2_TB23.8V AMSR2_TB23.8H	AMSR brightness temperature at 23.8 GHz, vertical/horizontal polarisation (K)
AMSR2_TB36.5V AMSR2_TB36.5H	AMSR brightness temperature at 36.5 GHz, vertical/horizontal polarisation (K)
AMSR2_TB89V AMSR2_TB89H	AMSR brightness temperature at 89 GHz, vertical/horizontal polarisation (K)
ERA5 data section	
ERA5_t2m_min ERA5_t2m_max ERA5_t2m_mean	Minimum, maximum and mean values of ERA5 2 m air temperature within the day
ERA5_skt_min ERA5_skt_max ERA5_skt_miean	Minimum, maximum and mean values of ERA5 skin temperature within the day
ERA5_tcwv_min ERA5_tcwv_max ERA5_tcwv_miean	Minimum, maximum and mean values of ERA5 total column water vapour within the day
ERA5_tclw_min ERA5_tclw_max ERA5_tclw_miean	Minimum, maximum and mean values of ERA5 total column liquid water within the day
ERA5_tciw_min ERA5_tciw_max ERA5_tciw_miean	Minimum, maximum and mean values of ERA5 total column ice water within the day
OSISAF data section	
OSISAF_sic	Total sea-ice concentration from EUMETSAT OSI SAF
ASCAT data section	
ASCAT_S0	ASCAT 5.3 GHz VV, incidence angle-adjusted backscatter at 40 degree incidence angle

2.1.6 Mismatches in the ice-charts

The ice charts are manually interpreted and drawn as polygons for larger regions. Therefore, the exact positions of the boundaries between ice types (e.g., multiyear ice and first-year ice) often differ. Figure 2.2 below shows an example of a mismatch between the NIC and AARI ice charts for 31 Dec 2012. A clear difference between the locations of MYI on NIC and AARI ice charts is visible (red colour shows area where MYI is present in both NIC and AARI, blue colour - only in NIC, green colour - only AARI). A confusion matrix is constructed by comparing these two ice charts, and balanced accuracy is computed (reported in the title) as specified in the Product Validation Plan (PVP) [RD-2]. The accuracy of the match is not very high due to the differences in the FYI/MYI fields.

Figure 2.3 shows the time series of ice chart comparison accuracy, balanced accuracy and number of overlapping pixels from two or more ice charts in one snapshot RRDP file. Accuracies are highest in winter, when most ice charts show similar MYI distributions, and lowest in summer, when ice charts differ. The balanced accuracy shows much lower values in summer, indicating that sampling of MYI and FYI is not balanced and that the accuracy values are biased.

Sometimes, more than two ice charts are present in a snapshot RRDP file, and the number of collocated pixels indicates the number of pairwise collocations of all ice charts in the file. The average balanced accuracy computed from a confusion matrix (Tharwat, 2020), aggregated from the intercomparison of all ice charts, is 0.856. This value is close to the upper limit of the accuracy of ice-age algorithms. However, the ice charts were intercompared only when they overlapped, whereas the algorithms were compared against the entire ice charts. Therefore, in theory, the accuracy of individual algorithms may exceed the accuracy of the ice chart intercomparison.

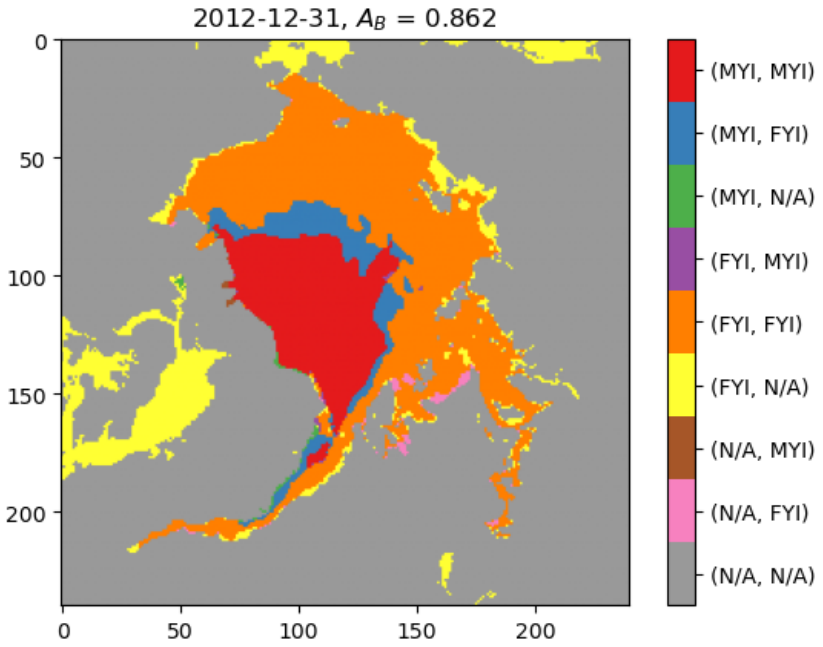


Figure 2.2. Comparison of MYI and FYI classification on NIC and AARI ice charts for 31/12/2012. Colours indicate various combinations of not available (N/A), FYI or MYI classification from NIC (left column of the colorbar labels and AARI, right column on the colorbar labels). The balanced accuracy of the match between FYI/MYI classification on these ice charts is only 0.862, as shown in the title.

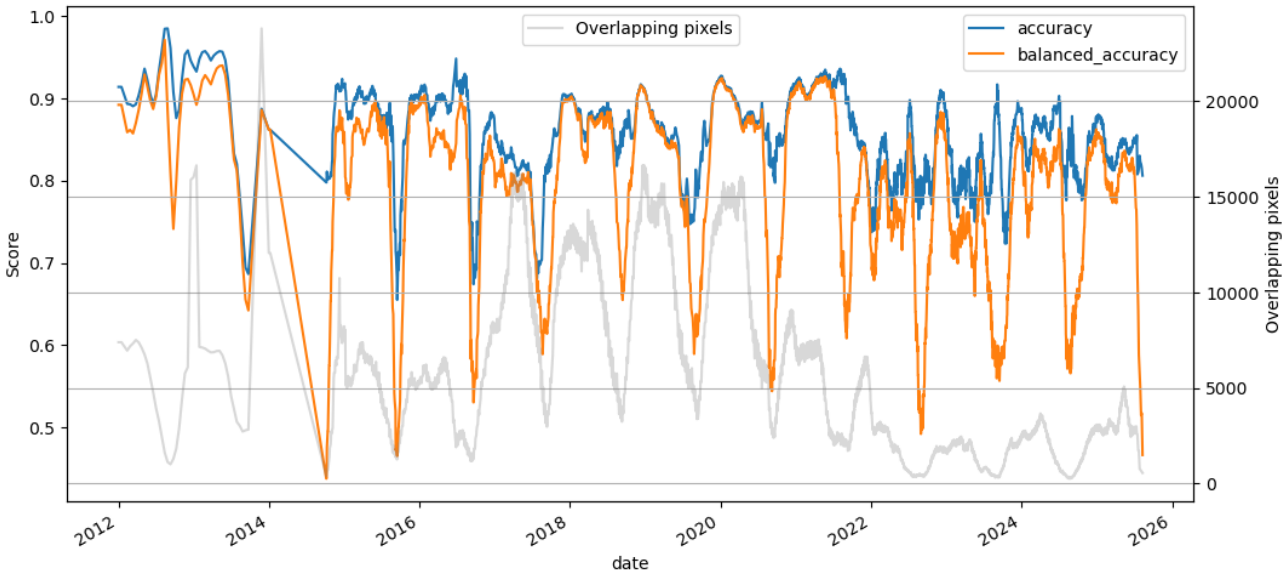


Figure 2.3. Time series of accuracy (blue line, left axis), balanced accuracy (orange line, left axis) and number of overlapping pixels (grey line, right axis) from comparison of the reference ice charts in the snapshot RRDP dataset.

This analysis of accuracy metrics within the ice charts illustrates the challenge of finding reliable reference data for the algorithm intercomparison conducted in CCI SAGE.

2.2 Buoy-RRDP

The buoy-RRDP uses drift trajectories of buoys deployed on sea ice with and without information on the sea-ice type upon deployment.

2.2.1 Source data

Buoy drift trajectory data are obtained from two primary sources, each with varying amounts of metadata information:

- Buoys listed on Meereisportal (https://data.meereisportal.de/relaunch/buoy?lang=en) are accompanied by deployment sheets, which give insight into the ice types at the time of buoy deployment.
- Buoy data obtained from the International Arctic Buoy Programme (IABP, https://iabp.apl.uw.edu/data.html) is missing this information. In these cases, the ice type upon buoy deployment is unknown, and the assignment of ice types along the drift follows a different approach.

2.2.2 Data structure

The buoy-RRDP is provided as two complementary NetCDF files:

One file contains a list of buoys with general information:

- Unique identifier of the buoy (buoy_id)
- Hemisphere ('N' or 'S') of the buoy deployment
- Release date of the buoy
- Buoy type
- Expedition name, if available
- Ice type at deployment (FYI, SYI, MYI or mixture of them), if available

The second file contains the actual drift trajectory data. The data points of each trajectory are flattened, and they can be identified by the buoy_id. Each data point is characterised by

- The geographical position of the buoy (latitude and longitude coordinates)
- and the time of the measurement.

In addition, the following sea ice and auxiliary information are included for each data point:

- Ice type class, if known
- Estimate of minimum and maximum possible sea-ice age
- Number of days since buoy deployment

- OSISAF sea-ice concentration in a 25 km grid cell of the buoy location
- ERA5 variables
- AMSR brightness temperature data (with time, lat, lon, angles)
- ASCAT backscatter data (converted to 40 degree incidence angle)

2.2.3 Naming convention of ice types

Compared to the snapshot-RRDP, which infers ice types only momentarily, the tracking of sea ice allows more detail on ice types and age in the buoy-RRDP. To preserve as much information as possible and allow for a comparison with differently detailed sea-ice type products, a new labelling has been constructed. This label counts the number of melting seasons the ice has survived as this melting period significantly alters the physical properties and microstructure of the sea ice. For buoys deployed without information on sea-ice type, the minimum number of melting seasons the sea ice has endured is recorded. These buoys are marked by '>0s' to illustrate that the sea ice experienced at least one (more than zero) melting season, but there is no information on whether they actually might have survived more melting seasons. Table 2.3 shows the new indices and labels with descriptions and comparisons to the traditional sea ice classes.

Table 2.3: Overview and definition of the new sea-ice age labelling with comparison to traditional classes.

Index	Label	Description	Traditional class naming
0	UNK	Ice of unknown type at deployment before a melting season has passed.	NaN
1	=0s	Ice that has not yet undergone a melting season yet, i.e. ice formed after the most recent summer melt and observed before the end of the next melting season. This is also known as first-year ice or FYI.	FYI
2	>0s	Ice that survived at least one melting season, but possibly more (related to buoys deployed on unknown ice type). This is sometimes referred to in the literature (not by WMO) as multiyear ice (MYI), often associated with remote-sensing sea-ice-type products.	MYI, as defined by many ice type products
3	=1s	Ice that survived exactly one melting season, i.e. ice formed before the most recent summer melt and observed after that melt season. This is also known as second-year ice or SYI.	SYI
4	>1s	Ice that survived at least two melting seasons. Again, related to buoys deployed on unknown ice types.	MYI, as defined by WMO

		Following the WMO sea-ice nomenclature, this is defined as multiyear ice (MYI).	
5	=2s	Ice that survived exactly two melting seasons.	MYI
6	>2s	Ice that survived at least three melting seasons.	MYI
…	…	…	MYI

The dates chosen to mark the end of the yearly melt season are set to September 30 for the Arctic and March 14 for Antarctica. When turning from September 30 to October 1 and from March 14 to 15, respectively, the labels are switched. However, a minimum of 30 days of observation before these switch dates is required to switch to the next higher label.

Additionally, the number of days since deployment is counted, and the possible minimum and maximum age in months is estimated. This estimate assumes that ice could not have formed between May 1 and September 1 in the Arctic or between October 15 and February 15 in Antarctica due to prevalent melting conditions.

2.2.4 Temporal resolution

In order to reduce the amount of data while maintaining the quality of the information, the measurement series are reduced to hourly resolution. If the temporal resolution was already below that, all data is kept, but not interpolated to hourly resolution.

When removing data from the high-resolution measurement series, the measurements closest to full hours are kept, and no interpolation is applied.

2.2.5 Filtering drift trajectories

To clean the measurement series from unrealistic data, the following thresholds are applied:

- maximum drift speed of 5 km/hour allowed
- maximum drift distance of 100 km between measurements allowed
- maximum time gap of 30 days allowed
- minimum of 50 data points per track required

The drift speed and distance thresholds are applied in five steps to remove outliers without discarding entire buoy tracks. If the drift thresholds are exceeded or the time gap threshold is triggered after this process, the measurement series is split. Any resulting series with fewer than 50 data points is discarded. When a measurement series is split, the resulting segments are given the original buoy ID with a suffix (A, B, C, etc.) appended. All segments except the first in terms of time are assigned an unknown initial ice type (NaN), and the age counter is reset to zero.

This splitting and renaming approach is also applied when using OSI SAF sea-ice concentration (SIC) data to ensure that the buoys remain in sea-ice areas throughout the measurement series. While the counters are also reset to zero, the ice type assigned to the later segments is FYI. The following thresholds are applied here:

- minimum SIC of 30%
- Maximum 1 day of no SIC data or SIC < 30%

For this purpose, the Global Sea-Ice Concentration climate data record (SMMR/SSMI/SSMIS), release 3 (OSI-450-a1), was used up to and including 2020, and the Global Sea-Ice Concentration interim climate data record, release 3 (OSI-430-a), was used from 2021 onwards, both provided daily and at a spatial resolution of 25 km.

2.2.6 Data access and format

The buoy-RRDP is provided in the form of two complementary NetCDF files, both accessible from https://data.seaice.uni-bremen.de/IceAge_RRDP/Buoy/. They each contain 2-dimensional tables.

The file B_RRDP_info.nc lists the 2620 buoys with the following information:

Table 2.4: Description of variables in B_RRDP_info.nc. Details on the buoy types can be found at: https://spaces.awi.de/spaces/MB/pages/516390626/Buoy+descriptions

Variable name	Description
buoy_id	IMEI, unique identifier of the buoy
release_date	Start of buoy measurements
hemisphere	'N' for Arctic, 'S' for Antarctica
ini_ice_type	Reported ice type at time of deployment, if information available
type	buoy type
station	Cruise or measurement campaign name of the buoy output

The file B_RRDP_data.nc contains the measurement series for all buoys, with one data point per row (i.e., multiple rows per buoy). Additionally, AMSR, ASCAT, ERA5 and ice chart information are collocated and provided per row/data point. The following table shows a list of the column names in the data file and a description of the variables.

Table 2.5: Description of variables in B_RRDP_data.nc.

Variable name	**Description**
time	Buoy measurement time
buoy_id	IMEI, unique identifier of the buoy
buoy_lat	Measurement latitude
buoy_lon	Measurement longitude
x	X coordinate in EASE grid projection
y	Y coordinate in EASE grid projection
OSISAF_sic	Sea-ice concentration in 25 km grid cell of buoy measurement
days_since_deployment	Days counted since start of measurement series
ice_age_index	Ice age index following the description in Table 2.1
Ice_age_label	Ice age label following the description in Table 2.1
WMO_ice_type_index	Indices referring to traditional WMO ice types: 1: young ice 2: first-year ice (FYI) 3: second-year ice (SYI) 4: multiyear ice (MYI)
ice_age_min_months	Estimate of minimum ice age in months
ice_age_max_months	Estimate of maximum possible ice age in months
ASCAT data section	
ASCAT_S0	ASCAT 5.3 GHz VV, incidence angle-adjusted backscatter at 40 degree incidence angle
ERA5 data section	
ERA5_t2m	2 meter air temperature from ERA5 (K)
ERA5_skt	Skin temperature from ERA5 (K)
ERA5_tcwv	Total column water vapor from ERA5 (kg/m2)

ERA5_tclw	Total column liquid water from ERA5 (kg/m2)
ERA5_tciw	Total column ice water from ERA5 (kg/m2)
ERA5_u10	10 m u-component of wind from ERA5 (m/s)
ERA5_v10	10 m v-component of wind from ERA5 (m/s)
AMSR data section	
AMSR_s_TB6.9V AMSR_s_TB6.9H	Spatially closest observed AMSR brightness temperature at 6.9 GHz, vertical/horizontal polarisation (K)
AMSR_s_TB10.7V AMSR_s_TB10.7H	Spatially closest observed AMSR brightness temperature at 10.7 GHz, vertical/horizontal polarisation (K)
AMSR_s_TB18.7V AMSR_s_TB18.7H	Spatially closest observed AMSR brightness temperature at 18.7 GHz, vertical/horizontal polarisation (K)
AMSR_s_TB23.8V AMSR_s_TB23.8H	Spatially closest observed AMSR brightness temperature at 23.8 GHz, vertical/horizontal polarisation (K)
AMSR_s_TB36.5V AMSR_s_TB36.5H	Spatially closest observed AMSR brightness temperature at 36.5 GHz, vertical/horizontal polarisation (K)
AMSR_s_TB89V AMSR_s_TB89H	Spatially closest observed AMSR brightness temperature at 89 GHz, vertical/horizontal polarisation (K)
AMSR_s_lat	Latitude of spatially closest AMSR measurement (°)
AMSR_s_lon	Longitude of spatially closest AMSR measurement (°)
AMSR_s_time	Timestamp of spatially closest AMSR measurement
AMSR_s_sun_azimuth	Solar azimuth angle of spatially closest AMSR measurement (°)
AMSR_s_sun_elevation	Sun elevation of spatially closest AMSR measurement (°)
AMSR_s_earth_azimuth	Azimuth angle of satellite viewing direction of spatially closest AMSR measurement (°)
AMSR_s_earth_incidence	Incidence angle of satellite viewing direction of spatially closest AMSR measurement (°)
AMSR_s_dist	Distance between spatially closest AMSR and buoy measurement (km)
AMSR_s_timediff	Time difference between the spatially closest AMSR and buoy measurement (minutes)
AMSR_t_TB6.9V AMSR_t_TB6.9H	Temporally closest observed AMSR brightness temperature at 6.9 GHz, vertical/horizontal polarisation (K)

AMSR_t_TB10.7V AMSR_t_TB10.7H	Temporally closest observed AMSR brightness temperature at 10.7 GHz, vertical/horizontal polarisation (K)
AMSR_t_TB18.7V AMSR_t_TB18.7H	Temporally closest observed AMSR brightness temperature at 18.7 GHz, vertical/horizontal polarisation (K)
AMSR_t_TB23.8V AMSR_t_TB23.8H	Temporally closest observed AMSR brightness temperature at 23.8 GHz, vertical/horizontal polarisation (K)
AMSR_t_TB36.5V AMSR_t_TB36.5H	Temporally closest observed AMSR brightness temperature at 36.5 GHz, vertical/horizontal polarisation (K)
AMSR_t_TB89V AMSR_t_TB89H	Temporally closest observed AMSR brightness temperature at 89 GHz, vertical/horizontal polarisation (K)
AMSR_t_lat	Latitude of temporally closest AMSR measurement (°)
AMSR_t_lon	Longitude of temporally closest AMSR measurement (°)
AMSR_t_time	Timestamp of temporally closest AMSR measurement
AMSR_t_sun_azimuth	Solar azimuth angle of temporally closest AMSR measurement (°)
AMSR_t_sun_elevation	Sun elevation of temporally closest AMSR measurement (°)
AMSR_t_earth_azimuth	Azimuth angle of satellite viewing direction of temporally closest AMSR measurement (°)
AMSR_t_earth_incidence	Incidence angle of satellite viewing direction of temporally closest AMSR measurement (°)
AMSR_t_dist	Distance between temporally closest AMSR and buoy measurement (km)
AMSR_t_timediff	Time difference between the temporally closest AMSR and buoy measurement (minutes)
Icechart data section	
AARI_dom_icetype	Dominant ice type (1 = YI, 2 = FYI, 3 = SYI, 4 = MYI, -77 = invalid) according to AARI ice charts
AARI_dom_icetype_conc	Concentration of the dominant ice type with respect to the total area of the polygon the data point lies in for AARI ice charts
AARI_MYI_conc	Concentration of the multiyear ice with respect to the total area of the polygon the data point lies in for AARI ice charts
AARI_total_ice_conc	Total sea-ice concentration with respect to the total area of the polygon the data point lies in for AARI ice charts
CIS_dom_icetype	Dominant ice type (1 = YI, 2 = FYI, 3 = SYI, 4 = MYI, -77 = invalid) according to CIS ice charts

CIS_dom_icetype_conc	Concentration of the dominant ice type with respect to the total area of the polygon the data point lies in for CIS ice charts
CIS_MYI_conc	Concentration of the multiyear ice with respect to the total area of the polygon the data point lies in for CIS ice charts
CIS_total_ice_conc	Total sea-ice concentration with respect to the total area of the polygon the data point lies in for CIS ice charts
DMI_dom_icetype	Dominant ice type (1 = YI, 2 = FYI, 3 = SYI, 4 = MYI, -77 = invalid) according to DMI ice charts
DMI_dom_icetype_conc	Concentration of the dominant ice type with respect to the total area of the polygon the data point lies in for DMI ice charts
DMI_MYI_conc	Concentration of the multiyear ice with respect to the total area of the polygon the data point lies in for DMI ice charts
DMI_total_ice_conc	Total sea-ice concentration with respect to the total area of the polygon the data point lies in for DMI ice charts
NIC_dom_icetype	Dominant ice type (1 = YI, 2 = FYI, 3 = SYI, 4 = MYI, -77 = invalid) according to NIC ice charts
NIC_dom_icetype_conc	Concentration of the dominant ice type with respect to the total area of the polygon the data point lies in for NIC ice charts
NIC_MYI_conc	Concentration of the multiyear ice with respect to the total area of the polygon the data point lies in for NIC ice charts
NIC_total_ice_conc	Total sea-ice concentration with respect to the total area of the polygon the data point lies in for NIC ice charts

2.2.7 Insights into the dataset

Figure 2.4 shows a histogram of the number of buoy measurements for each of the new sea-ice age labels, depending on the initial ice type at the buoy deployment. The data is presented on a logarithmic scale and for both hemispheres side by side, highlighting that there is an order of magnitude more data points for the Arctic. The majority of deployments with ice type information are on FYI. However, they turn into SYI and MYI eventually if the ice survives the summer melt seasons. In both hemispheres, but especially in the Arctic, a major part of the data points comes from buoys that do not contain information about the ice type at the buoy deployment (grey). These buoys will never provide information on FYI, even though they might initially be deployed on this ice type. Only after enduring two summer melt seasons, they are considered to be on MYI, but only a few buoys do. The mean survival time of a buoy is 5.6 months in the Northern Hemisphere and 3.3 months in the Southern Hemisphere, while the spread is high and the maximum duration of a measured track is close to three years.

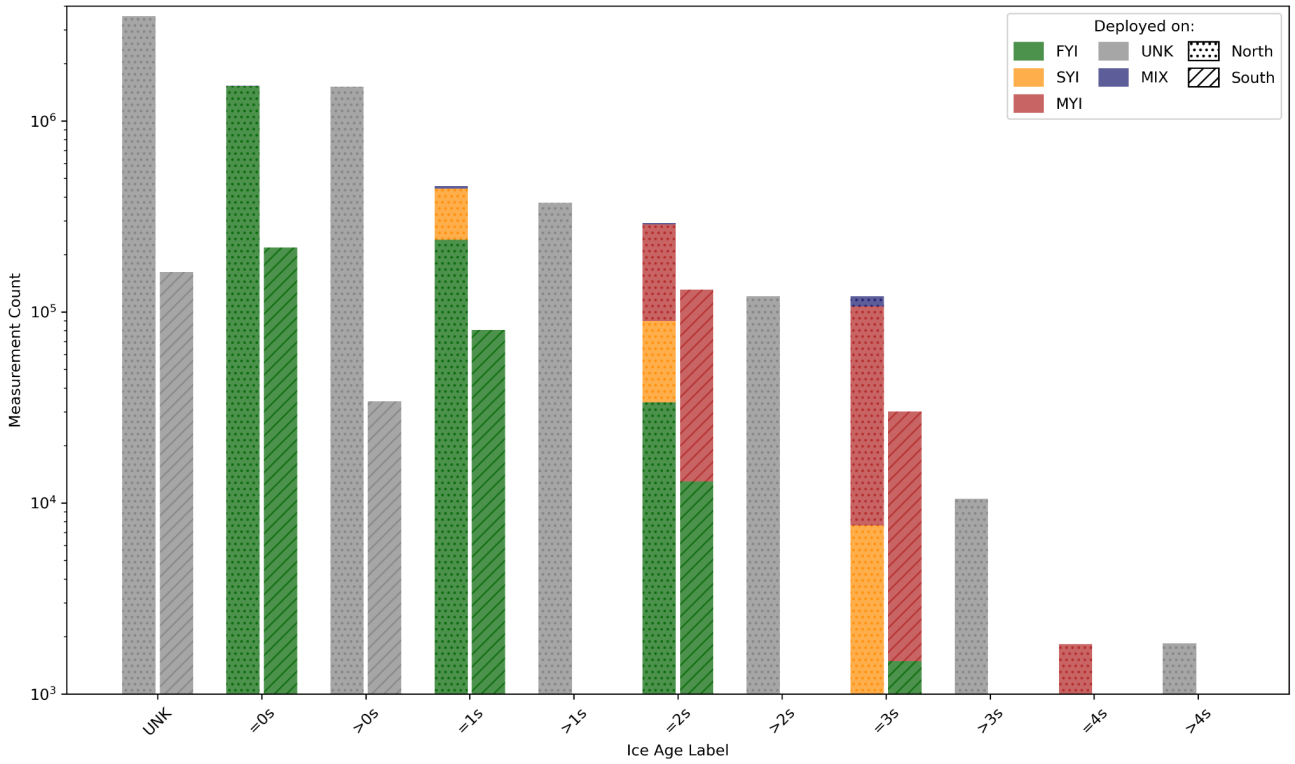


Figure 2.4: Number of measurement points per ice age category dependent on the initial deployment ice type. "MIX" means the deployment sheet of a buoy contains a combination of ice types. Dotted bars show the data for the Northern Hemisphere, striped bars for the Southern Hemisphere. Note the logarithmic scale.

The maps in Figure 2.5 underline the bias to observe FYI with buoy tracking measurements. For both hemispheres, an overview of buoy survival time in months (left panels), the associated new ice labels (middle panels), and WMO ice types (right panels) is shown, accumulated for all years. Even in the central Arctic, many buoys indicate the presence of FYI. That is not impossible, but it shows that buoy observations are problematic to use as a representative for a larger region. If buoys were deployed to investigate FYI processes they might be preferably stationed in a lead with FYI between two older floes. This is not wrong, but leads to a large amount of FYI data points in a region where FYI might not be representative at all. Also, the measurement data points presented in Figure 2.5 are highly correlated due to the continuous observation of a few floes and often at a high repetition rate of one hour.

For these reasons, the buoy-RRDP dataset will not be used for the evaluation of sea-ice type algorithms in the next chapter and in this first version of the PVASR. However, the dataset will be very valuable for future analysis of sea-ice age (rather than sea-ice type), and when investigating the tracking performance of Lagrangian ice-age algorithms.

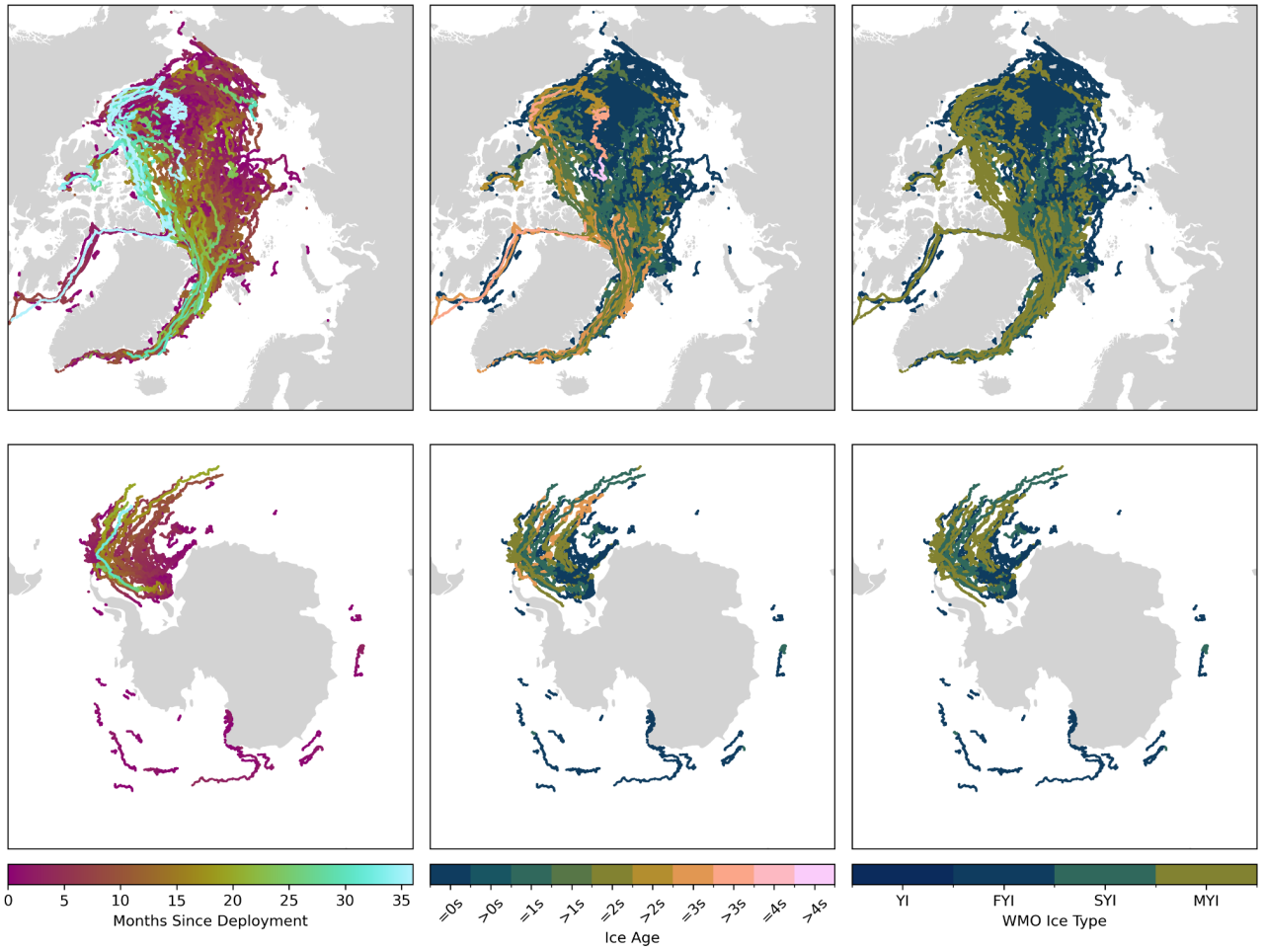


Figure 2.5: Overview of buoy observation tracks for the Northern (upper row) and the Southern (bottom row) Hemisphere accumulated for all years. The left panels show the drifting period of buoys since the start of observation in months. The middle panels show the assigned ice-age label, following the new labels described in Table 2.3. The right panels show the conversion of these new labels to the traditional WMO ice type classes.

2.3 Selection criteria

The criteria for selecting algorithms serve two purposes. Some criteria, such as spatial coverage, temporal coverage, and seasonal coverage, are considered eligibility criteria that determine whether a candidate algorithm is suitable for the generation of a climate data record. Algorithms that do not meet these requirements are considered less suitable for long-term CDR production.

Other criteria are used to compare the classification performance of the candidate algorithms. In the round-robin exercise, the comparison focuses on the ability to correctly discriminate between FYI and MYI, as

distinguishing between these two primary classes represents the minimum level of classification consistent with the recent GCOS threshold measurement uncertainty requirements (GCOS report no. 244, 2022, GCOS report no. 245, 2022). In particular, the balanced accuracy score is used, which accounts for class imbalance between FYI and MYI. Rankings are therefore primarily derived from these performance metrics, while the other criteria, listed below, provide context on the algorithms' suitability for long-term CDR production.

Spatial coverage and resolution

There are several sea-ice type/age products for the Northern Hemisphere, but only a few for the Southern Hemisphere. A criterion for algorithm selection is therefore the ability to provide consistent spatial coverage across both hemispheres.

The preferred spatial resolution should optimally be below 50 km, in line with GCOS requirements and user feedback collected during the SAGE webinar. Both the GCOS requirements and the outcome from the SAGE webinar are described in more detail in the User Requirement Document [RD-1]. Some user groups, in particular operational users, expressed a need for spatial resolution on the order of 10 km (or below). Conversely, climate model users indicated that coarser resolutions (~50-100 km) may still be valuable for their climate and modelling applications.

Temporal coverage

For climate applications, the products should provide the longest possible temporal coverage across the satellite era in order to support the generation of consistent Climate Data Records (CDRs).

In addition to long-term coverage, algorithms should demonstrate robust performance across all seasons. Sea-ice retrievals are known to be more challenging during the melt season due to changes in surface emissivity, melt ponds, and increased atmospheric influence. For the same reason, most existing sea-ice-type products are only provided during the winter months.

User feedback collected in the URD [RD-1] (and from users in general) highlights the importance of maintaining product availability during summer months, as the seasonal evolution of sea ice is important for both process studies and climate monitoring.

Temporal resolution

Products should provide temporal resolution consistent with user needs and GCOS Breakthrough requirements (RD-1, GCOS report no. 245, 2022). Daily resolution is desirable and is achievable using the traditional satellite products considered in this study.

Performance

Reliable discrimination between FYI and MYI is essential as these categories are widely used in observational and modelling communities. Users also strongly highlighted the importance of representing the mixture of ice types within a grid cell, for example, through multiyear ice fraction or age-class distributions.

Focus criteria of the round-robin exercise

Because of limitations in the products and algorithms being assessed, the quantitative intercomparison at the core of this PVASR cannot fully address all requirements listed above. For example, some of the tested products are valid only in the central Arctic Ocean; others are limited by other specific constraints; while several provide full hemispheric coverage. To ensure a consistent and fair comparison, we applied a common spatial mask for the Arctic Ocean (see 2.5.3).

Similarly, most ice-type products are limited primarily to winter-month classification, whereas Lagrangian ice-age products generally provide year-round coverage. For most of the comparison results, we therefore applied a temporal mask restricting the analysis to the overlapping period and winter months (see 2.5.3).

A major complexity in the intercomparison study is the large variation in how age/type classes are represented across different products. Some products provide the concentration of or fraction of one or more age classes within each grid cell; while others provide the probability of a given class, independent of the potential fraction or concentration. See more discussion on the topic of age definition in the URD [RD-1]. To enable a consistent comparison, all products were therefore reduced to two classes of FYI and MYI within each grid cell, based on product-specific threshold criteria (see 2.5.3).

Despite the inherent focus of the quantitative evaluation at the core of this round-robin exercise, the overall criteria translating user requirements remain important when interpreting and discussing the performance of each product.

2.4 Specifications of tested algorithms

Table 2.6: Overview of selected algorithms for the round-robin comparison exercise.

Name	Input	References
LM	OSISAF Drift, OSISAF Concentration	Korosov et al., 2026, https://doi.org/10.5194/essd-18-721-2026
NSIDC1	NSIDC Drift, NSIDC Concentration	Tschudi et al., 2020 https://doi.org/10.5194/tc-14-1519-2020
NSIDC2	OSISAF Drift, OSISAF Concentration	Korosov et al., 2018, https://doi.org/10.5194/tc-12-2073-2018 Tschudi et al., 2020, https://doi.org/10.5194/tc-14-1519-2020
ASIMIR	PMW	Rückert et al. (2023) https://doi.org/10.1029/2023EA003177 Scarlat et al. (2020) https://doi.org/10.1029/2019JC015749
NasaTeam	PMW	Cavalieri et al. (1984) https://doi.org/10.1029/JD089iD04p05355
C3S	PMW	Copernicus Climate Change Service, Climate Data Store, (2026): Global Sea ice type version 4 from 1978 to present derived from satellite observations. https://doi.org/10.24381/cds.29c46d83
ECICE (adapted)	PMW & SCAT	Melsheimer et al. (2023) https://doi.org/10.5194/tc-17-105-2023 Ye et al. (2016b) https://doi.org/10.3390/rs8050397 Ye et al. (2016a) https://doi.org/10.1109/TGRS.2015.2503884 Shokr et al. (2008) https://doi.org/10.1109/TGRS.2008.2000624
OSISAF	PMW & SCAT	EUMETSAT OSI SAF, Global Sea Ice Type (2005), OSI-403-d, https://doi.org/10.15770/EUM_SAF_OSI_NRT_2006
ASCAT-KNMI	SCAT	Belmonte et al. (2018) https://doi.org/10.5194/tc-12-2941-2018
Channel combinations	PMW or SCAT	Not algorithms per se. An easy way to grow the list of methods tested is to quickly add all the GRs, PRs, or other channel combinations we can think of (DAL). Linearly scaled between the two provided tie-points. This can support ground-work discussion on which channels have the best sensitivity to ice type. And maybe we learn something. In CCI SIC we ran 1-channel algorithms that we made up for demonstration purposes.

2.4.1 Radiometric algorithms

Discrimination between different ice types using passive microwave or scatterometer data exploits the contrast in their imagery brightness: multiyear ice generally appears darker (in brightness temperature) and lighter in backscatter, and vice versa for first-year ice. There are several well-established approaches for converting these microwave imagery brightness values into ice-type categories, notably linear distance and Bayes classification. In addition, these contrast-based approaches can be applied with different microwave frequencies and polarisation channels.

Some traditional radiometric sea-ice type classification retrievals rely on selected combinations of TB channels, typically expressed through ratios or gradients between frequency or polarisation channels. One commonly used parameter is the spectral gradient ratio (GR) of the brightness temperature (TB) at two different frequencies (freq1, freq2) using the same polarisation:

$$GR(freq1, freq2) = (TB(freq2) - TB(freq1)) / (TB(freq2) + TB(freq1))$$

A related ratio parameter is the polarisation ratio (PR), denoting the vertically (V) and horizontally (H) polarised TBs at the same frequency:

$$PR(freq, pol) = (TB(freq, V) - TB(freq, H)) / (TB(freq, V) + TB(freq, H))$$

2.4.1.1 The ASIMIR total sea ice and multiyear ice concentrations

The ASIMIR (Arctic Sea Ice and Atmosphere Multi-parameter Retrieval) algorithm, also known as the Multi-parameter Retrieval (MPR) algorithm, uses the Optimal Estimation Method (OEM) to invert a physical forward model and retrieve sea ice and atmospheric variables. Unlike traditional algorithms, which derive parameters sequentially, ASIMIR simultaneously solves for nine parameters, including total sea-ice concentration (SIC), multiyear ice (MYI) fraction, snow depth, and integrated water vapor, to ensure physical self-consistency across all variables. The retrieval process uses microwave brightness temperatures from AMSR-E and AMSR2 sensors at frequencies ranging from 6.9 to 89.0 GHz. These frequencies are resampled to a common spatial resolution before processing.

The core of the methodology is the iterative minimisation of a cost function that accounts for the difference between observed and simulated brightness temperatures. The forward model integrates the Microwave Emission Model of Layered Snowpacks (MEMLS) to simulate the emission and scattering characteristics of the ice surface. To accurately partition total sea-ice concentration and multiyear ice fraction, the algorithm treats each satellite footprint as a mixture of open water, first-year ice, and multiyear ice. The high-frequency channels (36.5 and 89 GHz) are sensitive to volume scattering in the porous structure of multiyear ice and the on average higher surface roughness of multiyear ice; the lower frequencies constrain the total ice fraction and surface temperature.

The retrieval requires specific prior information, including a climatological a priori state and an a priori covariance matrix, to guide the optimisation toward physically realistic solutions. Alongside the surface parameters, atmospheric effects such as liquid water path and water vapour are retrieved to mitigate weather-induced errors that often plague traditional sea ice algorithms. If the iterative procedure fails to converge or reaches unphysical limits (e.g., a maximum snow depth of 50 cm), the data points are excluded to maintain product quality. The algorithm is specifically designed for Arctic freezing conditions and typically operates from October through May. The resulting dataset provides daily, gridded fields of total SIC and MYI fraction on a 25 km EASE grid, each accompanied by a mathematically derived retrieval uncertainty, which is provided as a variance field in the final NetCDF product. More information on this algorithm is available here in Rückert et al. (2023).

The algorithm was developed and evaluated explicitly for the Arctic. However, the foundation and data for its application in the Southern Hemisphere are generally available and show potential for future work.

2.4.1.2 The NASA Team ice types

The NASA Team (NT or NasaTeam) algorithm (Cavalieri et al., 1984) is an empirical approach that retrieves total sea-ice concentration (SIC) and partitions it into first-year (FYI) and multiyear (MYI) ice fractions using brightness temperatures at 19 and 37 GHz. The algorithm's core mechanism relies on two radiometric ratios designed to minimise the impact of physical temperature variations: the polarisation ratio (PR), which distinguishes ice from open water based on polarisation differences at 19 GHz; and the spectral gradient ratio (GR), which exploits the tendency of MYI to scatter high-frequency radiation (37 GHz) more effectively than FYI. Concentrations are calculated relative to fixed reference tie points representing 100% open water, FYI, and MYI. These points form a triangle in PR-GR space, and a pixel's position within this triangle determines its fractional ice composition.

2.4.1.3 The Copernicus Climate Change Service (C3S) ice type product

The Copernicus Climate Change Service (C3S) sea-ice type version 4.0 product provides a daily global classification of sea ice into FYI, MYI, and ambiguous ice type. The climate data record spans October 1978 to 2020 and is derived from passive microwave radiometer observations from SMMR, SSM/I, and SSMIS. It is extended by an interim climate data record (ICDR) covering 2021 to the present, based on AMSR2. Note, however, at the time of performing the round-robin exercise, a temporary ICDR based on SSMIS was used because the AMSR-based ICDR was under implementation.

Sea-ice type retrieval is based on a Bayesian classification framework. The primary classification parameter is the spectral gradient ratio GR3719V, which exploits frequency-dependent emissivity contrasts between FYI and MYI. Brightness temperatures are corrected for atmospheric effects using a radiative transfer model and auxiliary Numerical Weather Prediction (ERA5) fields, and land spill-over corrections are applied in the swath

domain prior to gridding. Probability density functions for FYI and MYI are estimated daily using training data from a ±7-day window and predefined target regions, with iterative refinement based on recent classifications and independent ice-type information. This dynamic approach avoids fixed thresholds and ensures temporal consistency across sensors and decades.
For each grid cell, the ice type with the highest Bayesian probability is selected. Pixels with a classification probability below 75% are assigned to an ambiguous class. Several post-processing corrections are applied, including a warm-air intrusion correction based on near-surface air temperature anomalies, a statistical filter to suppress spurious MYI along the marginal ice zone, and additional geographic and climatological masks. Sea-ice drift and concentration data are further used through backtracking approaches to constrain training data and reduce physically implausible MYI emergence in dynamically active regions.

The final product is provided on a 25 km EASE2 grid with daily fields of ice type, uncertainty, and a bitwise status flag documenting applied corrections. Ice-type classification is restricted to the cold season (October-April in the Northern Hemisphere; March-August in the Southern Hemisphere), with summer periods flagged as ambiguous due to melt-induced signal overlap.

2.4.1.4 Exploring ice types from various channel combinations

In addition to the well-established algorithms, our round-robin intercomparison includes a systematic sensitivity analysis using a broad range of contrast-based methods tested with a broad range of passive microwave channel combinations.

Several of these channel combinations are at the core of the established algorithms. E.g. GR is at the core of the C3S and OSI-SAF approach, while both PR and GR are used in the NasaTeam algorithm. See definitions of GR and PR in the introduction of Section 2.4.1.

Distance along the line (DAL) is another parameter. It was introduced by Lavergne et al. (2019). The parameter is computed as the scalar product between the TB vector and the unit vector sustaining the consolidated “ice line” in a multidimensional TB channel space. As its name indicates, DAL measures the distance along the ice line, with lower values for MYI conditions and higher values for FYI conditions.

Several modulations of the standard parameters have been suggested (M. Huntemann, T. Lavergne, personal communication, 2026) to reduce the influence of the open-water fraction in the Marginal Ice Zone. Here, the observed TBs are assumed to represent a linear mixture of sea ice and open water. The contribution from the open water is estimated by including the open-water tiepoints. In the present round-robin exercise, we include both GRice and DALice, which are respectively the water-compensated versions of GR and DAL.

The multiyear ice fraction (MYIf) expresses the observed signal as the normalised distance between tiepoints representing pure FYI and MYI. The method assumes that the observed parameter varies approximately linearly between these two states. In a GR-based implementation (MYIf_GR), the distance is computed using

the GR values of the FYI and MYI tiepoints. For the round-robin exercise, the MYIf approach is applied to both the GR and the DAL, and to the water-compensated variants GRice and DALice.

The Bayesian classification estimates the probability of FYI and MYI (and potentially more classes) from one or more observed parameters. In this framework, the distributions of the selected parameters for FYI and MYI are represented by Gaussian class-conditional probability density functions derived from the tiepoint dataset. For a given observation, the likelihood of belonging to each ice class is computed and combined with prior probabilities using Bayes' theorem, yielding posterior probabilities for FYI and MYI. In the round-robin exercise, this approach is applied to the radiometric parameters of GR or DAL.

In addition, it is applied in a similar manner to the ASCAT backscatter (σ^0) values.

Note that in the result sections below, the outcomes from the Bayesian computation are named as "BAYES_<frequency combination>" (e.g BAYES_GR_freq1-freq2), whereas the outcomes from the MYIfraction are named solely as <frequency combination> (e.g. GR_freq1-freq2, not MYIf_GR_freq1-freq2).

2.4.2 Backscatter-based algorithms

2.4.2.1 The KNMI scatterometer ice-type product

The KNMI sea-ice type product is based on a purely scatterometer-driven ice-type classification algorithm, following the physical framework introduced by Belmonte et al. (2018). The method exploits the systematic dependence of radar backscatter (σ^0) on sea-ice age under winter conditions. As sea ice survives successive melt seasons, reductions in salinity and the development of internal air inclusions lead to enhanced volume scattering and, consequently, higher σ^0 compared to FYI. This physical contrast provides the basis for distinguishing FYI from older ice using spaceborne scatterometer observations. The resulting sea-ice age proxy data are distributed via the KNMI data portal (https://scatterometer.knmi.nl/ice_extents/). The product is provided independently for each scatterometer mission, ensuring sensor-consistent processing and long-term continuity. Both Ku-band (e.g., QuikSCAT and Oceansat-2 OSCAT) and C-band (e.g., ERS and ASCAT) scatterometers are included, with all products mapped onto a 12.5 km polar-stereographic grid for both hemispheres.

According to Belmonte et al. (2018), wintertime σ^0 distributions of FYI, second-year ice (SYI), and older MYI are sufficiently separated to allow threshold-based classification. Thresholds are derived empirically from late-winter reference periods, typically March, when thermodynamic conditions are stable and melt-related effects are absent. FYI is characterised by relatively low σ^0 values, while SYI occupies an intermediate regime between FYI and older MYI. In practice, SYI is identified using a secondary threshold, while MYI is associated with the highest σ^0 values. At Ku-band (QuikSCAT and OSCAT), FYI is characterised by $\sigma^0 < -14.5$ dB, SYI occupies the range $-14.5 \leq \sigma^0 < -12.5$ dB, and MYI exhibits $\sigma^0 \geq -12.5$ dB. At C-band (ERS and ASCAT), the

thresholds are shifted to lower values: FYI is defined as $\sigma^0 < -17$ dB, SYI as $-17 \leq \sigma^0 < -15$ dB, and MYI as $\sigma^0 \geq -15$ dB.

To prevent spurious classification of older ice in seasonal or marginal ice zones, the KNMI algorithm applies a geographical mask. This mask excludes regions where multiyear ice is climatologically unlikely, such as parts of the Greenland, Barents, Kara, and Chukchi seas in the Arctic, thereby reducing misclassification caused by deformed or rough first-year ice exhibiting elevated σ^0 values.

For the present study, the available KNMI ASCAT (C-band, 2007 to 2018) and Oceansat-2 OSCAT (Ku-band, 2009 to 2014) sea-ice age proxy products were initially considered to provide coverage for the analysis interval. However, the temporal overlap to the round-robin exercise period was found to be insufficient. Instead, ASCAT backscatter data available within the RRDP were used, normalised to the same reference incidence angle (52.8°) as in the KNMI approach. The KNMI algorithm was then implemented on these ASCAT data, following the methodology of Belmonte et al. (2018), including winter-based threshold selection, temporal averaging, and geographical masking. Although the original algorithm was developed for the Northern Hemisphere, this study consistently applies the same processing framework to products in both hemispheres, since sea-ice age proxy data are available from KNMI for both hemispheres. KNMI (*Koninklijk Nederlands Meteorologisch Instituut*) is the Royal Netherlands Meteorological Institute.

2.4.3 Multi-sensor algorithms

2.4.3.1 The OSI SAF ice types

The OSI SAF sea-ice type product provides a daily near-real-time classification of sea ice into FYI, MYI, and ambiguous ice for both hemispheres. Unlike the C3S product, it is designed for operational monitoring rather than climate consistency. Frequent updates to sensors, tuning, and processing chains mean that the product should be used with caution in climate studies.

The retrieval follows the same Bayesian classification principles as those used in the C3S sea-ice type product. As in C3S, the primary passive microwave discriminator between FYI and MYI is the spectral gradient ratio GR3719V, with dynamic probability density functions derived from recent training data. A multi-sensor approach is applied, combining passive microwave observations from AMSR and SSMIS with C-band scatterometer backscatter from ASCAT. Single-sensor probabilities are computed in the swath domain and merged through a Bayesian multi-sensor analysis on the daily grid. AMSR acts as the primary passive microwave source, with SSMIS as backup, while ASCAT provides complementary information.

The current OSI SAF processing chain does not include all post-processing refinements implemented in the latest C3S sea-ice type version. In particular, the handling of warm-air intrusion events, marginal ice zone artefacts, and long-term climatological consistency is more limited, reflecting the operational focus of the product.

Sea-ice type is provided on a polar stereographic grid with 10 km spacing, together with pixel-level uncertainty estimates and detailed status flags. The classification is restricted to the cold season, with summer periods flagged as ambiguous due to melt-related loss of class separability.

2.4.3.2 The EC-ICE ice types

The Environment Canada Ice Concentration Extractor (EC-ICE) is an optimisation-based algorithm designed to retrieve total sea ice concentration (SIC) and partial concentrations of various ice types, primarily distinguishing between young ice (YI), first-year ice (FYI) and multiyear ice (MYI). The retrieval is based on a linear mixture model that assumes the total radiometric signature observed in a satellite pixel is the area-weighted sum of open water, YI, FYI, and MYI tie points. Instead of using single radiometric tie points, probability density distributions of the radiometric values per ice type (and open water) class are required. These are used to randomly select 1000 combinations of radiometric values and perform a Monte Carlo simulation.

The framework is designed to integrate data from passive microwave radiometers, such as AMSR or SSM/I, and active microwave scatterometers, such as ASCAT. The algorithm combines these sensors to utilise the different sensitivities of scatterometers and radiometers to surface and volume scattering characteristics of multiyear ice.

To maintain physical consistency during the Arctic winter, the original algorithm incorporates two automated correction modules. The first module uses atmospheric reanalysis data to identify periods when surface temperatures approach the melting point. This is necessary because liquid water in the snowpack can cause FYI to radiometrically mimic MYI. The second module uses ice drift vectors to verify that changes in the local multiyear ice (MYI) fraction are consistent with the physical advection of the ice pack. These post-processing corrections are not implemented in the intercomparison study in Chapter 3 where only the core of the algorithm is assessed. We will discuss post-processing later and apply it uniformly.

The processing workflow generates daily gridded fields of total SIC and MYI fraction, typically at a 12.5 km or 4.45 km spatial resolution. While the primary application is for the Arctic freezing season from October to May, the algorithm has been adapted for the Southern Ocean by modifying the a priori probability density distributions to reflect Antarctic ice properties. Each retrieval includes an estimate of uncertainty derived from the spread of Monte Carlo simulations, providing a metric for the statistical confidence of the ice-type classification.

2.4.4 Lagrangian algorithms

2.4.4.1 The NERSC sea-ice age

The NERSC sea-ice age algorithm derives ice age fractions from satellite sea-ice concentration and drift by tracking the evolution of multiyear ice (MYI) using a Lagrangian framework. Each year, a pan-Arctic MYI concentration field is initialised at the end of the melt season, when total ice extent is minimal. This MYI field is advected forward in time using daily satellite-observed ice drift, accounting for changes in concentration due to convergence, divergence, and melt. As advection proceeds, the difference between the observed total concentration and the advected MYI field defines the concentration of first-year ice, while subsequent annual differences between advected MYI fields yield second-year and older ice fractions. In this framework, all ice surviving the summer minimum is assigned a higher age class from 15 September onward.

Sea-ice age fractions are computed iteratively using a generic formulation in which ice of age N years is defined as the difference between MYI fields advected for N–2 and N–1 years. This approach produces a consistent set of age-resolved concentration fractions (first-year through multiyear ice) for each day. An average sea-ice age is then calculated as a concentration-weighted mean of the individual age classes. Compared to the earlier Eulerian implementation, which suffered from numerical diffusion and excessive smoothing, the new algorithm preserves sharper spatial gradients and a more physically consistent representation of ice history.

The key methodological advance is the use of a Lagrangian triangular mesh for advection. An initial triangular mesh is generated over the Arctic Ocean using the EASE2 projection, with node spacing of approximately 25 km and exclusion of inland regions. Mesh nodes are advected daily using interpolated ice drift vectors. To maintain mesh quality under strong deformation, the mesh is dynamically remeshed using well-defined geometric criteria, including edge length, element angles, area thresholds, and detection of flipped elements. Local mesh optimisation is then applied to ensure numerical stability while limiting unnecessary diffusion.

Sea-ice concentration is conservatively mapped between successive meshes using element-wise area preservation and weighted averaging for remeshed elements. Advected concentrations are conditioned by observed satellite concentrations to avoid unphysical values under strong convergence or observational inconsistencies. To avoid contamination of MYI by newly formed ice, the annual MYI field is defined as the minimum of several independently advected concentration fields around the seasonal minimum. The full workflow is applied to multi-decadal satellite drift and concentration records to generate the LM-SIAge climate data record, providing daily fields of ice age fractions and mean sea-ice age on a regular 25 km EASE2 grid suitable for climate analysis and downstream applications (Korosov et al., 2026).

The NERSC algorithm described in this chapter is referred to as the “LM” algorithm hereafter.

2.4.4.2 The NSIDC sea-ice age

The NSIDC sea-ice age algorithm is a Lagrangian tracking approach in which parcels of ice are initialised, advected by observed ice motion, and aged discretely in yearly steps. Each year on 15 September, when the seasonal ice minimum is reached, all grid cells with ice concentration above a fixed threshold are initialised as first-year ice (FYI). These ice parcels are represented as point clouds embedded in the NSIDC EASE-Grid projection. Daily ice motion fields (u, v) and ice concentration are ingested; missing values are filled using nearest-neighbour interpolation, and land points are masked to ensure that only ice over ocean is tracked.

Each day, ice parcels from previous years are advected forward by one time step using bilinear interpolation of the daily ice drift fields. After advection, parcels are filtered: points that move onto land or into grid cells with ice concentration below the threshold (15%) are discarded, representing melt or export. To limit numerical complexity and follow the NSIDC practice, only a fixed number of the youngest age classes (fractions) are retained. Newly formed ice is added only at the annual re-initialisation date, while surviving parcels retain their original "birth year," which implicitly increases their age by one year each September.

Sea-ice age is then rasterised onto a regular grid. All grid cells with ice concentration above the threshold are first assigned age 1 (FYI). Older age classes are overlaid by mapping the advected parcel locations to grid indices and assigning increasing integer ages (2, 3, …), corresponding to second-year ice and older multiyear ice. The resulting product is a daily gridded sea-ice age field, consistent with the NSIDC methodology, while the underlying Lagrangian parcel representation preserves the physical continuity of ice history across seasons.

In the round-robin exercise, we implemented the NSIDC ice age algorithm described above. It is referred to as "NSIDC2". In addition, we also evaluated the outputs of the original NSIDC algorithms, which are available at the NSIDC portal. These data are referred to as "NSIDC1". The main difference between NSIDC1 and NSIDC2 is the forcing files: as described below, the NSIDC2 product is forced by the OSI SAF ice drift and concentration, whereas the NSIDC1 product is forced by the NSIDC ice drift and concentration.

2.5 Harmonisation of the algorithms

2.5.1 Tie points for radiometric and backscatter-based algorithms

To enable fair comparisons between the algorithms, tie points were determined using a subset of the snapshot-RRDP dataset and applied uniformly to the algorithms that required them. Three winter seasons each, November through March 2016-17, 2018-19, and 2020-21 for the Arctic and March through August 2016, 2018 and 2020 for the Antarctic, are used for the tie point definition. The matched ice-type arrays, defined as the agreement of ice types across different providers, as described in Chapter 2.1.3, are used to identify areas of "pure" ice types with high certainty. To avoid disturbance by warm-air intrusions, a maximum temperature threshold of 273.15 K is applied. Additionally, a threshold of a minimum 80% sea-ice concentration and a buffer zone of two pixels (~50 km) at the boundaries between ice types are used before

the random choice of 100 pixels per day and ice type. From this data collection, the mean, median, standard deviation, and median absolute deviation (MAD) are calculated for the AMSR brightness temperatures, their channel combinations, and the ASCAT backscatter. These are stored in a .json file along with histograms that provide a typical distribution of observations per ice type (and open water).

The re-implemented algorithms that require tie points or distributions use these newly determined tie points. For the ECICE algorithm, the original tie points are also tested for comparison. The two versions are named ECICE_rtp and ECICE_otp, for the new RRDP tie points and the original tie points, respectively.

2.5.2 Harmonisation of forcing files for the LM and NSIDC algorithms

The same sea-ice drift and concentration datasets were used to drive both the NERSC and NSIDC algorithms.

For the period from 1991 through 2020, we use OSI-455, the EUMETSAT Ocean and Sea Ice Satellite Application Facilities (OSI SAF) global low-resolution sea-ice drift climate data record (CDR) (OSI SAF, 2022c; Lavergne and Down, 2023). For 2021–2024, we utilise the Global Low Resolution Sea Ice Drift product, OSI-405-c, which is a near-real time (NRT) product (OSI SAF, 2007; Lavergne et al., 2010). Both products utilise the same satellite sensors, SSM/I (>= F10) and SSMIS (CMSAF FCDR), AMSR-E (NSIDC) and AMSR2 (JAXA). The NRT product also utilises C-band scatterometer data from the ASCAT missions. Both products employ the same sea ice motion-tracking methodology, continuous maximum cross-correlation (CMCC, Lavergne et al. (2010)), which involves a fractional-pixel pattern matching of the brightness temperatures.

The global sea-ice concentration climate data record version 3 from SMMR/SSMI/SSMIS data (Lavergne et al., 2019; OSISAF, 2022b) was used for dates up to and including 2020, and the accompanying interim climate data record (ICDR) version 3 (OSI SAF, 2022a), based on SSMIS data, for dates 2021 – 2024. These products are both on a 25-km EASE-2 grid (Brodzik and Knowles, 2011). These products are generated using the SICCI3LF algorithm, based on 19 GHz and 37 GHz imagery. Information about the land/water mask was obtained from the status flag of the OSI SAF SIC product.

All OSI SAF products are available via the web portal at https://osi-saf.eumetsat.int/.

2.5.3 Harmonisation of the validation framework

The algorithms listed in Section 2.4 produce different types of outputs:

- Categorical classification: each pixel of the output product is assigned only one value - the sea-ice type index, for example, 2 for FYI, or 4 for MYI (see Section 2.1.3). Some algorithms differentiate only FYI and MYI and therefore produce binary classifications. Other algorithms can also derive SYI or other types of ice and produce multi-class classification.
- Probabilistic classification: each pixel of the output product is assigned several values: probabilities of belonging to different classes. In case of FYI/MYI classification, only two values are given: a

probability of FYI and a probability of MYI, e.g., [0.1, 0.9]. If more classes can be derived with an algorithm, more probabilities are provided.

- Concentration of ice of a particular age, for example, MYI concentration. The concentration can be expressed as a fraction of a grid cell or a fraction of the ice cover.

Moreover, the algorithms have different spatial extents: some cover only the Central Arctic, some extend farther south but exclude, for example, Baffin Bay, and some produce output for all sea-ice-covered pixels. The time coverage also varies, depending on the availability of input data.

For a harmonised intercomparison of all algorithms, the following approach is used:

Harmonisation of product presentations

All outputs are converted to categorical classification. Outputs of the categorical algorithms are grouped into two classes only: FYI and MYI. Outputs of the probabilistic and concentration algorithms are thresholded to produce only FYI and MYI classes. The categorical classification of the algorithms is evaluated against each individual ice chart to construct a confusion matrix and compute accuracy as specified in the PVP [RD-2].

The official C3S and OSISAF products are already provided as categorical classes. However, they include the ambiguous class, which corresponds to low classification probabilities and does not indicate whether the observation is more likely FYI or MYI. Simply excluding these pixels would remove the most uncertain regions and potentially bias the assessment. Therefore, instead of the final categorical output, intermediate probability fields from the C3S processing chain are used. As the binary FYI/MYI classes are derived from Bayesian probabilities with post-processing, the classification can be converted back into a probability field representing the likelihood of MYI. This allows the product to be treated consistently with the probabilistic classification approaches evaluated in the round-robin while retaining uncertain regions in the comparison. Hereafter, the common thresholds are used to convert into pure FYI/MYI classes.

The same procedure is applied to OSISAF and its probability-based representation (OSISAFp).

Harmonisation of spatial coverage

Two types of spatial masking are applied to the algorithm outputs. First, a common spatial mask, which is a temporally consistent intersection of coverages of all algorithms. The common spatial mask is limited to the smallest extent of all products. In the Southern Hemisphere, no such mask is applied. For the Northern Hemisphere, the mask is defined by the KNMI algorithm, which only applies in the central Arctic Ocean (see Section 2.4.2.1). Second, an individual mask for each algorithm. The individual mask is only limited by the extent of each product and the extent of available ice charts.

Harmonisation of temporal coverage

Similarly, two types of temporal masking are applied. First, a common temporal mask, which selects the dates (years and seasons) when outputs from all algorithms are available. Second, an individual temporal

mask, which selects the dates for each algorithm based on the availability of data from this algorithm and from the ice charts.

3 Round-robin results

3.1 Performance assessment of individual algorithms

All algorithms and simple channel combinations described above are compared to the ice-type classes from the ice charts in the snapshot-RRDP. As the difference between running algorithms on the morning and evening overpasses of AMSR is observed to be insignificant, the results shown and discussed are for the evening overpasses only, and the separation between the times will no longer be made.

For the first impression of algorithm performance, the balanced accuracy score (see PVP, Section 3.2) is computed only for dates when all algorithms per hemisphere are available, to avoid biases due to differing seasonal coverage or source data availability. In the Northern Hemisphere, only months from October to April are taken into account, for the Southern Hemisphere, months from March to September. This leads to a scope of 1313 days between January 2016 and December 2024 in the Northern Hemisphere, and 71 days between March 2016 and March 2022 in the Southern Hemisphere. Additionally, for the Northern Hemisphere, a spatial mask of minimum unified coverage is applied to prevent distortion of the results. The benefit of wider coverage, spatially as well as temporally, is discussed later.

The discussion and ranking are performed separately for both hemispheres, while including algorithms performing well in the Arctic are also included in the discussion for the Antarctic, as long as the algorithms are applied there.

3.2 Northern Hemisphere

Balanced accuracy score for all algorithms

Table 3.1 shows the balanced accuracy scores for all algorithms in the Northern Hemisphere, considering the joint time frame and area. The algorithms are sorted by performance and ranked in the first column with a number (to be used later to identify algorithms in the discussion). It is clear that, in most cases, the more complex and established algorithms perform better than the fundamental channel combinations.

Of the three Lagrangian algorithms, NSIDC1 (22), NSIDC2 (18) and LM(1), LM shows by far the highest score. The two NSIDC algorithm versions perform similarly well, with NSIDC2 performing slightly better than NSIDC1, and thus NSIDCS1 is excluded from further discussions.

Also, for the ECICE core algorithm (without post-processing), there is little difference between using the original tie points (7) or those developed from the snapshot-RRDP (5). Therefore, the discussion will focus on the version with the RRDP tie points only.

Table 3.1: Balanced accuracy score of all algorithms and channel combinations applied to the Northern Hemisphere using the spatial and temporal consistent mask. The algorithms that are the subject of further discussion are highlighted in bold.

	Algorithm	Balanced Accuracy
1	**LM**	**0.8867**
2	C3S	0.8842
3	**C3Sp**	**0.8697**
4	**ASIMIR**	**0.8624**
5	**ECICE_rtp**	**0.8580**
6	**BAYES_GR_18p7V-10p7V**	**0.8498**
7	ECICE_otp	0.8484
8	BAYES_GR_23p8V-10p7V	0.8460
9	OSISAF	0.8439
10	BAYES_GR_23p8V-6p9V	0.8407
11	BAYES_GR_36p5V-6p9V	0.8380
12	BAYES_GR_36p5V-10p7V	0.8370
13	**NasaTeam**	**0.8365**
14	**BAYES_DAL_36p5H-36p5V**	**0.8361**
15	BAYES_GR_18p7V-6p9V	0.8353
16	BAYES_DAL_18p7V-36p5H-36p5V	0.8329
17	BAYES_GR_23p8V-18p7V	0.8317
18	**NSIDC2**	**0.8287**
19	**OSISAFp**	**0.8269**
20	**BAYES_GR_36p5V-18p7V**	**0.8242**
21	BAYES_GR_36p5V-23p8V	0.8195
22	NSIDC1	0.8130
23	BAYES_GR_36p5H-6p9H	0.8105
24	BAYES_GR_23p8H-6p9H	0.8022
25	BAYES_GR_18p7H-6p9H	0.7937
26	BAYES_GR_36p5H-10p7H	0.7929
27	**ASCAT_KNMI**	**0.7894**
28	BAYES_GR_18p7H-10p7H	0.7887
29	BAYES_GR_10p7V-6p9V	0.7855
30	ASCAT_BAYES	0.7855
31	BAYES_GR_23p8H-10p7H	0.7842
32	BAYES_GR_36p5H-23p8H	0.7766
33	**GRice_23p8H-10p7H**	**0.7702**
34	GRice_10p7V-6p9V	0.7692
35	GRice_23p8H-6p9H	0.7659
36	BAYES_GR_36p5H-18p7H	0.7634
37	GR_36p5V-18p7V	0.7572
38	**GR_36p5V-10p7V**	**0.7564**
39	GR_36p5V-23p8V	0.7556
40	**DALice_36p5H-36p5V**	**0.7555**
41	GR_23p8V-18p7V	0.7549
42	GR_23p8V-10p7V	0.7475
43	GR_36p5V-6p9V	0.7422
44	BAYES_GR_89V-6p9V	0.7337
45	GR_18p7V-10p7V	0.7331
46	BAYES_GR_23p8H-18p7H	0.7330
47	GRice_36p5V-23p8V	0.7317
48	GR_36p5H-23p8H	0.7252
49	BAYES_GR_10p7H-6p9H	0.7237
50	GRice_23p8H-18p7H	0.7225
51	GRice_36p5V-18p7V	0.7194
52	DAL_36p5H-36p5V	0.7192
53	GR_36p5H-10p7H	0.7172
54	BAYES_GR_89V-10p7V	0.7164
55	GR_36p5H-18p7H	0.7150
56	GR_89V-6p9V	0.7104
57	GR_23p8V-6p9V	0.7074
58	GRice_36p5H-18p7H	0.7048
59	GR_89V-10p7V	0.7038
60	GR_18p7H-10p7H	0.7030
61	GR_23p8H-10p7H	0.7027
62	GRice_36p5V-6p9V	0.7021
63	GR_36p5H-6p9H	0.7002
64	BAYES_GR_89H-6p9H	0.6980
65	GRice_10p7H-6p9H	0.6957
66	GR_18p7V-6p9V	0.6879
67	GRice_36p5V-10p7V	0.6849
68	GR_23p8H-6p9H	0.6835
69	DALice_18p7V-36p5H-36p5V	0.6827
70	GRice_36p5H-6p9H	0.6802
71	GR_18p7H-6p9H	0.6771
72	GR_89H-6p9H	0.6767
73	GR_23p8H-18p7H	0.6742
74	GRice_18p7V-6p9V	0.6708
75	GRice_36p5H-10p7H	0.6638
76	GRice_23p8V-6p9V	0.6636
77	GR_89V-18p7V	0.6591
78	GRice_18p7H-6p9H	0.6586
79	GRice_23p8V-18p7V	0.6580
80	BAYES_GR_89H-10p7H	0.6567
81	GR_10p7V-6p9V	0.6562
82	GR_89H-10p7H	0.6557
83	BAYES_GR_89V-18p7V	0.6539
84	GRice_89V-6p9V	0.6512
85	GR_10p7H-6p9H	0.6456
86	GRice_89V-10p7V	0.6435
87	DAL_18p7V-36p5H-36p5V	0.6420
88	GRice_89H-6p9H	0.6324
89	GRice_89V-18p7V	0.6299
90	GRice_89H-10p7H	0.6121
91	GRice_18p7H-10p7H	0.6085
92	GRice_23p8V-10p7V	0.6077
93	BAYES_GR_89V-23p8V	0.6054
94	GR_89V-23p8V	0.6040
95	GR_89H-18p7H	0.6038
96	GRice_89V-23p8V	0.6029
97	BAYES_GR_89H-36p5H	0.5893
98	GRice_18p7V-10p7V	0.5887
99	GRice_89H-18p7H	0.5882
100	BAYES_GR_89H-18p7H	0.5876
101	GR_89H-36p5H	0.5864
102	GRice_89H-36p5H	0.5863
103	GRice_36p5H-23p8H	0.5809
104	GR_89V-36p5V	0.5668
105	BAYES_GR_89V-36p5V	0.5656
106	BAYES_GR_89H-23p8H	0.5633
107	GR_89H-23p8H	0.5630
108	GRice_89H-23p8H	0.5571
109	GRice_89V-36p5V	0.5543

For the C3S (2 and 3) and OSISAF (9 and 19) algorithms, the official categorical products show slightly higher scores than the probabilistic products (p). This is expected, as the pixels of an ambiguous class are excluded from C3S/OSISAF but remain in C3Sp/OSISAFp. The comparison of the two versions of the two categorical products is included for context; hereafter, only C3Sp and OSISAFp are considered in the round-robin analysis.

Currently, only limited pure scatterometer retrievals are available for the exercise, namely the KNMI (27) algorithm and the general Bayesian approach on backscatter (30). Both rely solely on ASCAT (C-band) data, although evidence suggests that Ku-band sensors (e.g. QuikSCAT and OSCAT) may provide improved performance. Despite its moderate performance, the ASCAT_KNMI ice type product (27) is included for further discussion.

The two algorithms, ASIMIR (4) and NASATeam (13), are also included for further discussion.

Comparison of channel combinations and different approaches

For the choice of channel combinations and different approaches to exploit them, a closer look is valuable to identify the channels and approach combinations that contain the most information for the ice-type algorithms. Figure 3.1 presents a compact comparison of all the different channel combinations. The figure shows a three-dimensional histogram of the balanced accuracy score based on the two (three for DAL) input channel frequencies and polarisations. The different approaches, like Bayes or MYI fraction, are separated by the markers and show the accuracy in the colour.

It is immediately clear that all approaches are less effective when 89 GHz (lower two rows in the figure) is included in the channel combination. That is because with higher frequencies and shorter wavelengths, the penetration depth in snow and sea ice becomes smaller due to strong scattering and absorption. Consequently, the observed brightness temperatures mainly reflect surface snow properties instead of the internal structure of the ice, limiting its usefulness for sea-ice type detection. There is a slight tendency towards higher accuracy when using V polarisation than H polarisation, probably because surface reflection affects H-pol more strongly. The Bayesian approach (crosses and stars) shows in general the highest accuracy scores, which are relatively stable for all frequency combinations as long as 89 GHz is excluded. The relative lower score with MYIf might be related to the choice of mean tie-points in its computation, which could be revisited in the future. For smaller differences between the frequencies the accuracy is slightly reduced.

It is important to note that the results in Figure 3.1 have applied the temporal and spatial masks and therefore reflect the conditions in the central Arctic Ocean wintertime. Whereas the more complex regions, like the marginal ice zones etc, are not considered so much here. In the rightmost part of the histogram, we see slightly better performance for DALice relative to DAL. A similar systematic difference in performance is not possible to find between the GR and GRice. For future work, it will be relevant to investigate selected channel combinations and approaches for their performance in more complex regions, such as the marginal ice zone and “shoulder” periods toward the summer months.

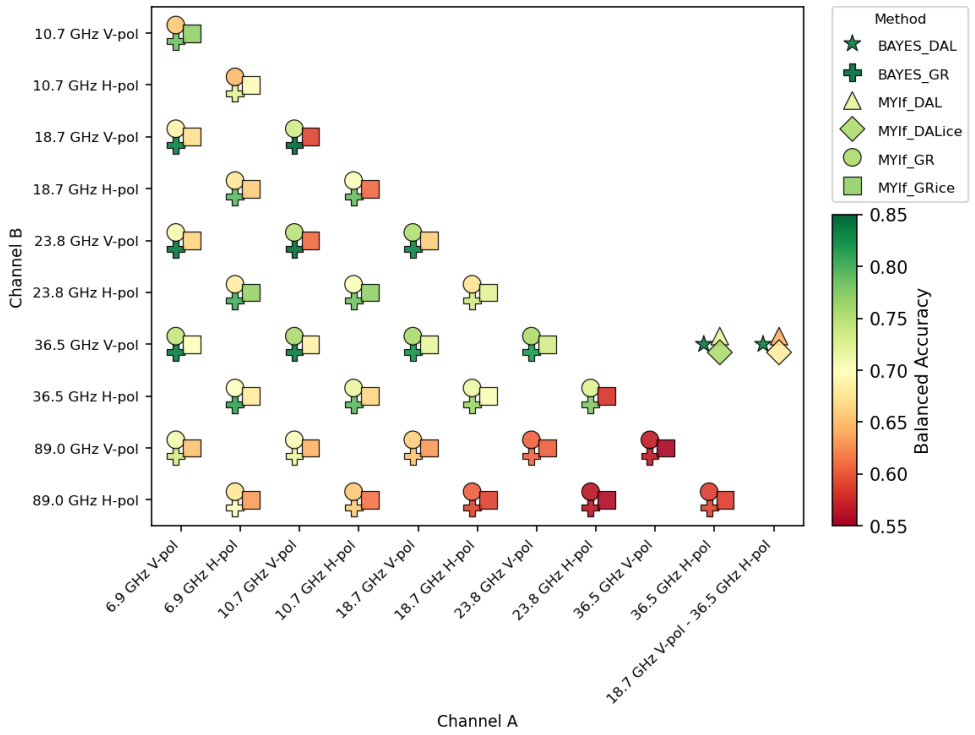


Figure 3.1: Balanced accuracy score of all channel combinations applied to the Northern Hemisphere using the spatial and temporal consistent mask.

The channel combination of 36.5 GHz and 18.7 GHz in V-pol, using the Bayesian approach (listed as number 20 in Table 3.1), is selected for further discussion, as it is fundamental to the C3Sp and OSISAFp algorithms. Based on Figure 3.1 and Table 3.1, and to include different approaches and frequencies, we have selected the following combinations for further comparison study:

(6) 18.7 GHz - 10.7 GHz V-pol with Bayesian GR approach

(14) 36.5 GHZ H-pol - V-pol with Bayesian DAL approach

(33) 23.8 GHz - 10.7 GHz H-pol with multiyear ice fraction on GRice approach

(38) 36.5 GHz - 10.7 GHz V-pol with multiyear ice fraction on GR approach

(40) 36.5 GHz H-pol - V-pol with multiyear ice fraction on DALice approach

All algorithms selected for the following discussion are highlighted in bold in Table 3.1. We recall that 18.7 GHz, 23.8 GHz, and 36.5 GHz are the frequency channels that are almost continuously available throughout the entire fleet of passive microwave missions. 36.5 GHz offers finer spatial resolution (than lower frequencies) and is thus interesting for building a CDR.

Comparisons of fourteen selected algorithms

Figure 3.2 shows the confusion matrices of the output of the 14 selected algorithms against the reference ice-type classification from the ice charts in the snapshot-RRDP. Below each matrix, the corresponding accuracy and balanced accuracy scores are provided. Overall, most algorithms capture a correct classification of the FYI class with high scores, seen by the higher values in the upper-left matrix cell. In general, they capture slightly less of the MYI (lower-right matrix cell) and also generally have a higher tendency to wrongly classify MYI as FYI (lower-left matrix cell) more often than the other way around. Only some channel-combination algorithms show reversed behaviour and strongly overestimate the presence of MYI.

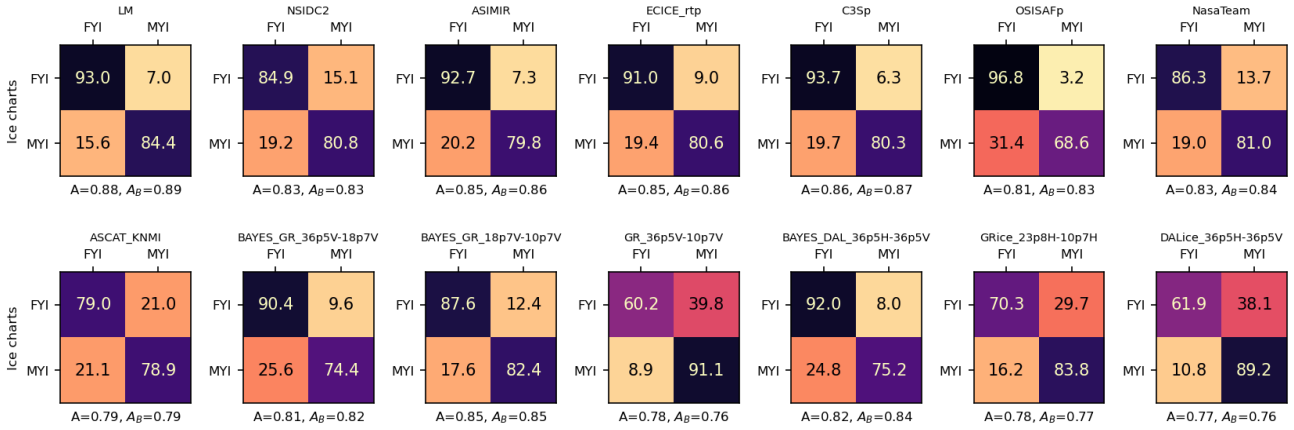


Figure 3.2: Confusion matrices of selected algorithms for the unified spatially covered area and temporal subset in the Northern Hemisphere. The name of the algorithm is shown above each matrix. The reference data is accumulated over all ice charts and reference data. Numbers in the matrix show the normalised number of predictions. The accuracy (A) and the balanced accuracy (AB) are given below the matrices.

The temporal evolution of the algorithms' accuracies is shown in Figure 3.3. For easier perceptibility, the selected algorithms are displayed in two groups of seven in different subplots of the figure. For each group, the daily balanced accuracy score (using a 30-day rolling mean) between 2015 and 2025 for the spatially unified coverage and temporal limitation of the majority of the algorithms is shown in the upper panel, and for the full temporal and spatial coverage provided for each algorithm is shown in the lower panel. For most of the established microwave-based algorithms, no summer classification is provided, and the main difference between the upper and lower panels is due to spatial masking. However, showing the full temporal coverage highlights a major benefit of the Lagrangian algorithms, which provide ice types through the entire summer season. The channel combinations are applied for the full season and emphasise this issue, as they show a significant drop in accuracy beyond any informative value in summer. Interestingly, the seasonal evolution of the algorithm performance within the freezing period from October to April features strong differences. While the scatterometer product and some microwave channel combinations perform worst at the start of winter and relatively better from January to April, the accuracy of most other algorithms decreases over the winter period, including both microwave-based and Lagrangian products.

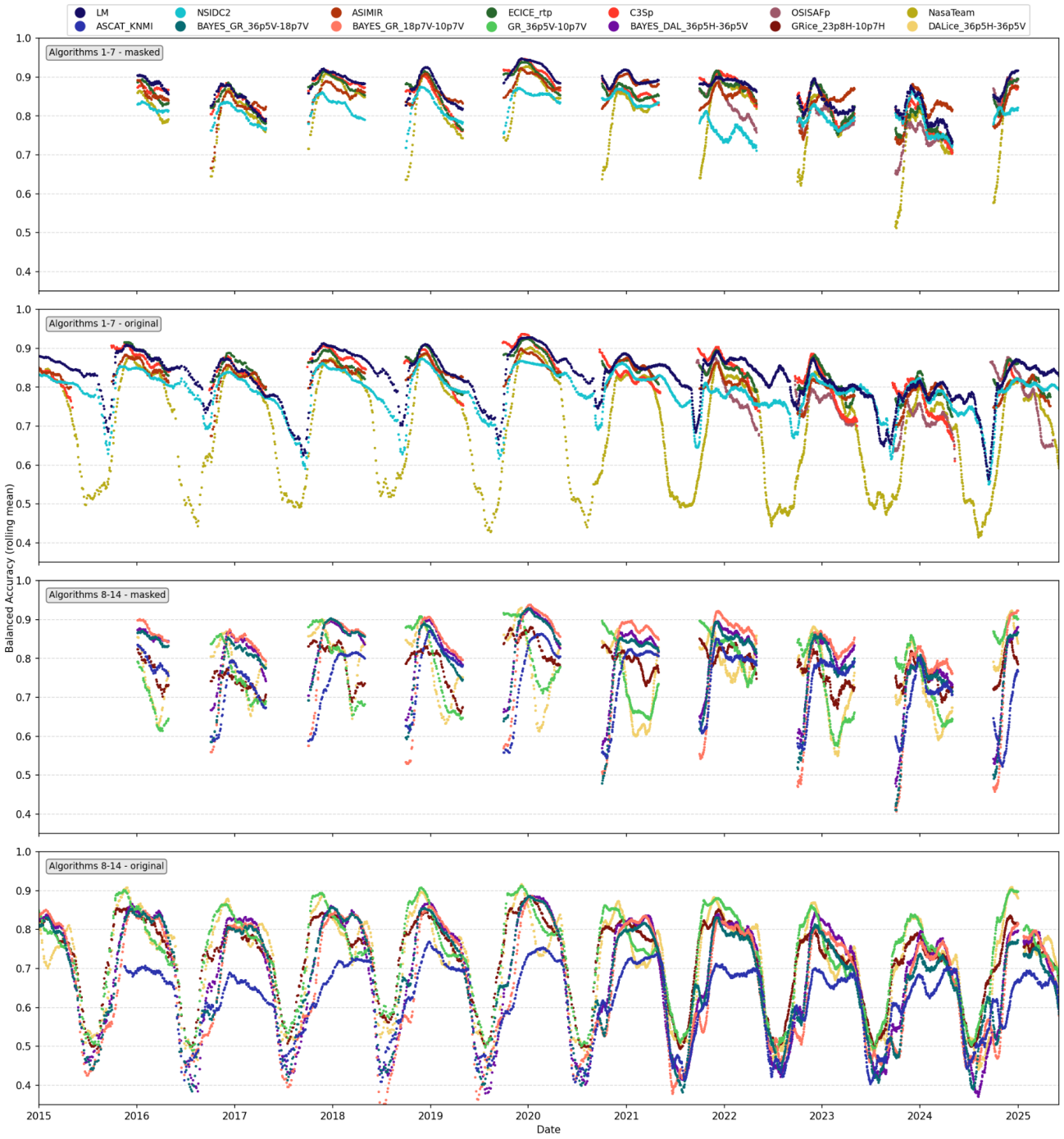


Figure 3.3: Temporal evolution of the balanced accuracy score for the selected algorithms in the Northern Hemisphere. Panel 1 and 3 show the results for the unified area mask and the overlapping days only, panels 2 and 4 for the original spatial area and time period of the algorithms, respectively. The algorithms are separated into two plots only for the sake of visibility.

The Lagrangian algorithms (LM and NSIDC2), in particular, show a notable decrease in accuracy just before the start of each freezing season, though they generally provide high accuracy. This is related to the different definitions of when the ice moves up in age class: in the LM algorithm, replenishment (i.e., transition from FYI to MYI) happens on 15 September every year by definition, whereas in the ice charts, the ice becomes multiyear ice only on 1st October. As a result, during the period from 15 Sep to 1 Oct, all MYI classifications by LM algorithms are mismatching the ice charts. In general, LM provides the highest accuracy. For further discussions the NSIDC2 is removed from the set of algorithms. So are three of the six channel combination approaches investigated so far: "BAYES_DAL_36p5H-36p5V" is preferred over "DALice_36p5H-36p5V", "BAYES_GR_36p5V-18p7V" is preferred over "BAYES_GR_18p7V-10p5V" because it is the fundamental combination used in "C3Sp" and "OSISAFp", and "GR_36p5V-10p7V" is kept as the only MYI-fraction approach channel combination representative.

The diverging seasonality of algorithm performance demonstrates the potential of combining retrieval approaches and products to obtain high-quality sea-ice age and type information. Not only the different spatial and seasonal coverage can be exploited by a merged product but also the benefits of different input data can be used.

Visual inspection and spatial comparison of maps

To investigate further which algorithms should be used and potentially combined, Figures 3.4 to 3.6 show spatial maps of the (normalised) MYI detection frequency within the first two weeks of December for the ten remaining algorithm products and the reference ice charts from the snapshot-RRDP. These figures illustrate the varying spatial coverage of the algorithms and reference datasets. Of the reference datasets, only the NIC ice charts provide full spatial coverage of the northern hemisphere. The AARI ice charts nearly do the same, but they lack data in the Canadian Arctic Archipelago because they often exclude landfast ice. The automated DMI ice charts have in times a more or less pronounced pole hole, and the CIS and DMI ice charts are only available for smaller regions. Regarding the algorithms, spatial coverage is mostly consistent, except for the Lagrangian algorithm LM, which does not cover the Canadian Arctic Archipelago (CAA) and Baffin Bay (BB).

Firstly, the maps reveal notable differences among the reference ice charts (maps at lowest row), which should be considered when interpreting the algorithm performance. The NIC charts show a larger extent and frequency of MYI in the Beaufort Sea (BS) in 2020 and 2022. Specifically, in 2022, both NIC and CIS show a circular pattern of MYI in BS that resembles the clockwise gyre in the region, which is not present in the other charts. Of interest is that, in 2022, the pattern showed an elongated MYI tongue extending from the North Pole towards the East Siberian Sea (ESS). NIC shows the highest frequency and extent of this pattern crossing all the way to the ESS coast, followed by the AARI chart with a slightly shorter and thinner tongue, while the autoDMI shows only a small pattern.

For the three selected periods, the reference charts show absence of MYI in the Baffin Bay. Similarly, the charts show MYI in parts of the CAA and Nares Strait, as well as a substantial amount of MYI following the East Greenland Current (EGC) with appearance of the northeast Greenland polonya in some of the charts.

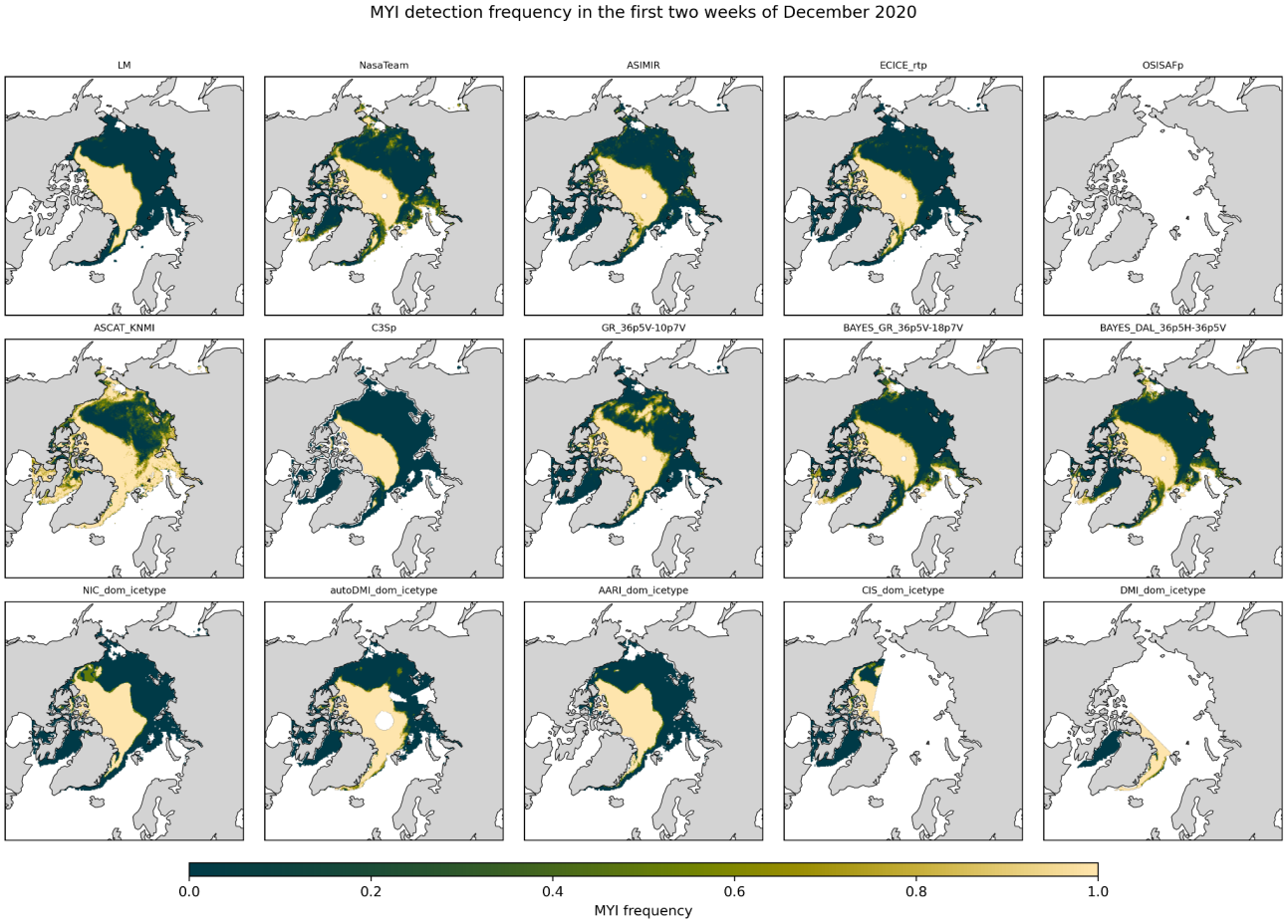


Figure 3.4: Spatial maps showing the normalised frequency of MYI detection by the selected subset of algorithms and the reference ice charts in the first two weeks of December, 2020, for the Northern Hemisphere. The name of the algorithm or ice chart source is shown above each map.

The algorithm outputs of the ice age and type products reveal systematic differences in their spatial behaviour. The microwave-based retrievals generally exhibit increased noise, with MYI classified in peripheral regions, the ice edge and narrow straits where such occurrences are unlikely. This effect is particularly pronounced for the ASCAT_KNMI, which shows extensive MYI detections along all marginal ice zones, even within its defined spatial mask. In contrast, the C3Sp product exhibits substantially less spurious MYI detection. This is a result of the inclusion of the post-processing steps that suppress physically implausible classifications. This reflects an inherent characteristic of direct interpretation approaches based on passive microwave and scatterometer data, where increased noise is expected and must be addressed through appropriate postprocessing and filtering. The Lagrangian algorithm (LM), on the other hand, does not show

the same type of artefacts, as MYI is only assigned based on tracked ice parcels originating from the seasonal minimum extent, and no MYI can suddenly appear in remote regions without a tracking history. Several microwave-based algorithms include masking based on drift information and the tracking history of ice parcels, e.g. C3S, ECICE, and OSISAF.

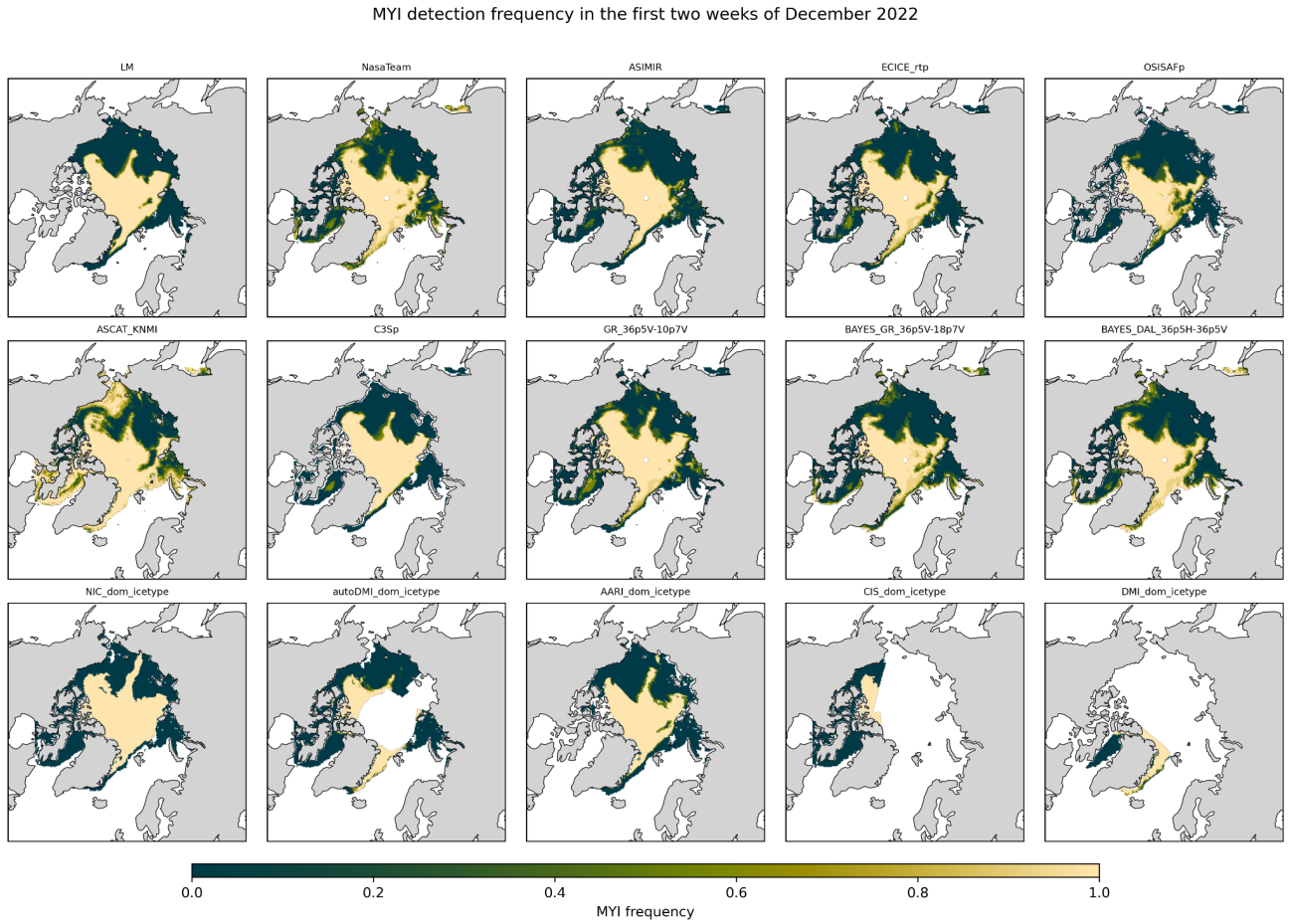


Figure 3.5: Spatial maps showing the normalised frequency of MYI detection by the selected subset of algorithms and the reference ice charts in the first two weeks of December, 2022, for the Northern Hemisphere. The name of the algorithm or ice chart source is shown above each map.

The characteristic elongated tongue of MYI in 2022 is well captured by several algorithms despite none showing the same extensive pattern as the NIC and AARI chart. In particular, LM, C3Sp and the GR_36.5V-10.7V combination reproduce this feature clearly and with a spatial extent comparable to the reference ice charts. This pattern is also indicated by most of the remaining algorithms, although the signal is less pronounced and less spatially coherent.

In several microwave-based algorithms, and for some years, MYI appears in the BB. This even happens in 2022 for both C3Sp and OSISAFp which include drift-correcting post-processing steps. For the three cases, the

reference ice charts do not support MYI in BB. However, it is not unlikely to happen and there has at times been observed MYI in both the Nares Strait and Baffin Bay (see e.g. Figure 2.4).

Capturing the outflow of MYI from the Arctic Ocean along the EGC is obviously challenging and varies widely in performance among algorithms and across the three case years. It is a known fact that the algorithms based on coarse-resolution SMMR-SSM/I-SSMIS data often lack MYI in EGC. This yields for the C3S product and is only partly corrected for in the post-processing steps. The multi-sensor algorithms, like OSISAF and ECICE, take advantage of the scatterometer influence, which has been documented to help in this regard (Breivik et al., 2012). The pure ASCAT_KNMI algorithm shows extensive MYI in the entire EGC for all three years and even overestimates it relative to the reference ice charts. The LM generally performs well at mapping MYI in EGC, despite some underestimation in 2024.

Overall, the LM demonstrates strong performance in comparison with the reference ice charts, combining spatial consistency with the ability to capture narrow and coastal MYI features without introducing spurious detections. Its reliance on ice tracking inherently constrains the classification and avoids the noise observed in direct interpretation methods. However, this dependency also introduces sensitivity to data gaps in the underlying drift or concentration products, which may affect tracking continuity unless re-initialisation strategies are applied.

Among the direct retrieval approaches, the established microwave-based algorithms (C3S, ASIMIR, ECICE, OSISAF) generally outperform the simple channel combinations, except the ASCAT_KNMI product, which is overall performing poorly with a lot of MYI detection, especially in the MIZ and coastal regions. This is likely because of a stronger deformed FYI that is misclassified as MYI by the radar scatterometer data, which is in particular sensitive to surface roughness.

A key aspect influencing the performance of these direct retrieval approaches is the application of post-processing. In general, all algorithms based on direct interpretation of passive microwave or scatterometer signals exhibit spurious MYI detections, particularly in marginal ice zones, coastal regions, and narrow straits, reflecting the sensitivity of the measurements to surface roughness, melt features, and floe edges. This behaviour is consistently observed across the different methods and highlights a fundamental limitation of the retrieval principle. The inclusion of post-processing steps, such as the use of drift information and temporal consistency constraints, significantly reduces this noise, as clearly demonstrated by the improved performance of products like C3Sp. Importantly, such post-processing is largely independent of the initial retrieval algorithm and can therefore be developed as a common framework applicable across methods, offering a promising pathway for improving overall product quality.

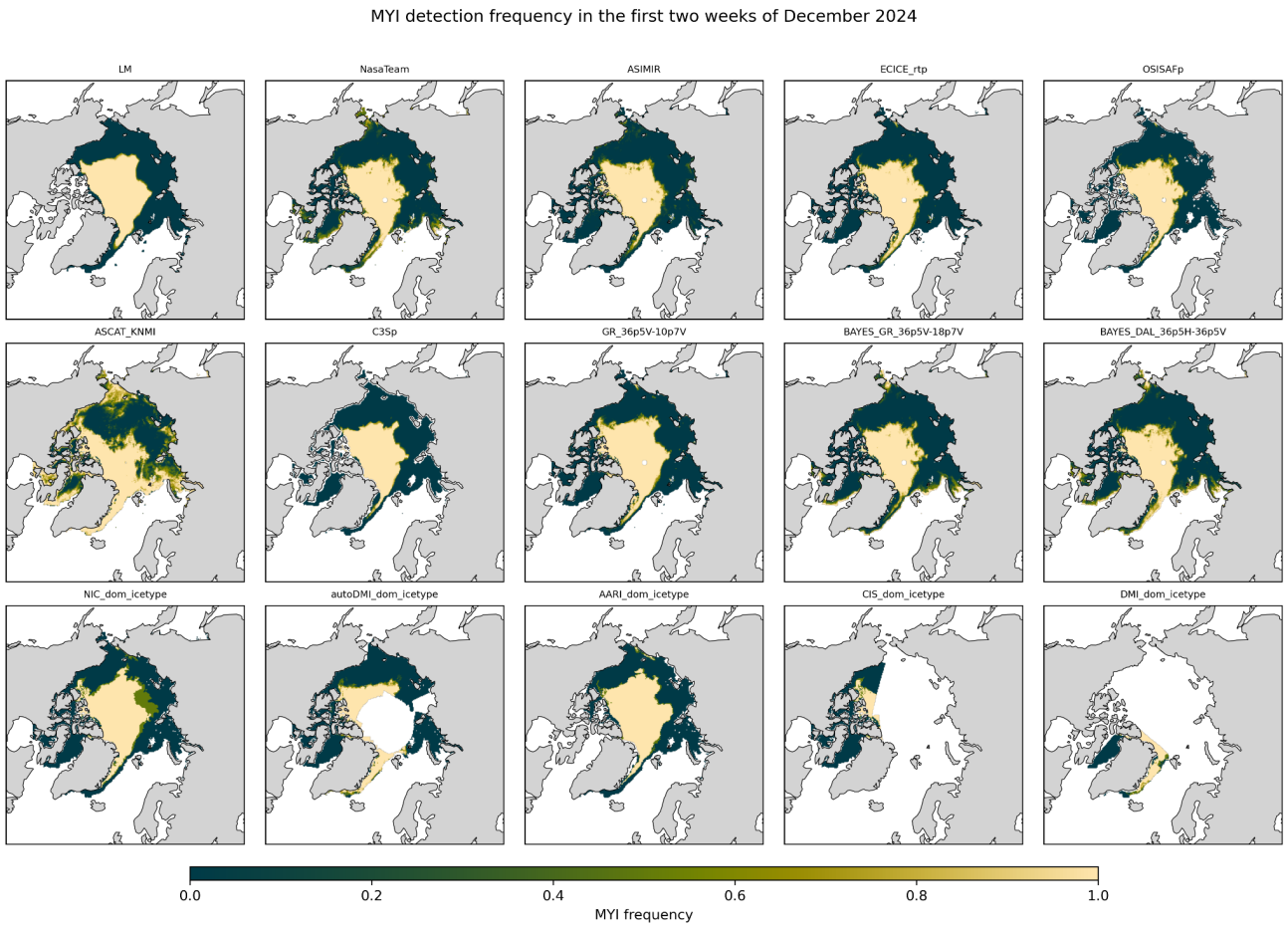


Figure 3.6: Spatial maps showing the normalized frequency of MYI detection by the selected subset of algorithms and the reference ice charts in the first two weeks of December, 2024, for the Northern Hemisphere. The name of the algorithm or ice chart source is shown above each map.

3.3 Southern Hemisphere

Balanced accuracy score for all algorithms

Table 3.2 shows the balanced accuracy scores for all algorithms in the Southern Hemisphere over the common time period, which is defined based on the availability of the other algorithms and does not account for the shorter record of OSISAFp (OSISAFp is included only for the latest years and would drastically reduce the data volume). The algorithms are ranked by performance.

	Algorithm	Balanced Accuracy
1	C3S	0.7797
2	**C3Sp**	**0.7691**
3	GR_18p7V-10p7V	0.7316
4	GR_23p8V-10p7V	0.7274
5	**GR_36p5V-10p7V**	**0.7221**
6	**ECICE_rtp**	**0.7134**
7	GRice_36p5V-10p7V	0.7126
8	**DALice_36p5H-36p5V**	**0.7106**
9	OSISAF	0.7100
10	GRice_36p5V-18p7V	0.7086
11	GRice_36p5V-23p8V	0.7075
12	GR_36p5V-18p7V	0.7065
13	GR_36p5V-6p9V	0.7053
14	GR_36p5V-23p8V	0.7041
15	GR_23p8V-18p7V	0.7035
16	GR_36p5H-10p7H	0.6890
17	**ASIMIR**	**0.6885**
18	GRice_23p8V-18p7V	0.6877
19	GRice_36p5V-6p9V	0.6852
20	GR_36p5H-23p8H	0.6843
21	GRice_23p8V-10p7V	0.6819
22	**NasaTeam**	**0.6800**
23	**OSISAFp**	**0.6772**
24	**GRice_36p5H-10p7H**	**0.6761**
25	GR_18p7H-10p7H	0.6755
26	GR_23p8V-6p9V	0.6754
27	**GRice_23p8H-10p7H**	**0.6704**
28	GR_23p8H-10p7H	0.6693
29	GRice_36p5H-18p7H	0.6688
30	GRice_18p7V-10p7V	0.6680
31	GR_36p5H-18p7H	0.6672
32	GRice_89V-6p9V	0.6629
33	ECICE_otp	0.6620
34	GRice_89V-10p7V	0.6599
35	GR_36p5H-6p9H	0.6580
36	GR_89V-6p9V	0.6575

	Algorithm	Balanced Accuracy
37	GR_89V-10p7V	0.6513
38	BAYES_GR_36p5V-10p7V	0.6498
39	BAYES_GR_36p5V-6p9V	0.6490
40	BAYES_DAL_18p7V-36p5H-36p5V	0.6472
41	BAYES_GR_23p8V-10p7V	0.6457
42	GR_18p7V-6p9V	0.6441
43	GRice_36p5H-23p8H	0.6431
44	DALice_18p7V-36p5H-36p5V	0.6430
45	BAYES_GR_36p5V-23p8V	0.6407
46	**BAYES_GR_36p5V-18p7V**	**0.6400**
47	DAL_36p5H-36p5V	0.6384
48	GRice_18p7H-10p7H	0.6381
49	GRice_23p8V-6p9V	0.6350
50	BAYES_GR_36p5H-6p9H	0.6329
51	BAYES_GR_23p8V-18p7V	0.6310
52	BAYES_GR_18p7V-10p7V	0.6309
53	**BAYES_DAL_36p5H-36p5V**	**0.6302**
54	BAYES_GR_36p5H-10p7H	0.6286
55	GRice_36p5H-6p9H	0.6284
56	GRice_89V-18p7V	0.6281
57	GRice_89H-6p9H	0.6257
58	BAYES_GR_23p8V-6p9V	0.6250
59	BAYES_GR_18p7H-10p7H	0.6247
60	BAYES_GR_89V-6p9V	0.6241
61	GR_89H-6p9H	0.6234
62	BAYES_GR_36p5H-23p8H	0.6231
63	GR_89V-18p7V	0.6208
64	GR_23p8H-6p9H	0.6202
65	BAYES_GR_89V-10p7V	0.6177
66	GRice_89H-10p7H	0.6147
67	BAYES_GR_23p8H-10p7H	0.6143
68	GR_23p8H-18p7H	0.6119
69	GRice_23p8H-6p9H	0.6111
70	GRice_18p7V-6p9V	0.6102
71	GR_89H-10p7H	0.6097
72	GRice_89V-23p8V	0.6087

	Algorithm	Balanced Accuracy
73	BAYES_GR_23p8H-6p9H	0.6058
74	BAYES_GR_89H-6p9H	0.6044
75	GR_89V-23p8V	0.6037
76	BAYES_GR_18p7V-6p9V	0.5967
77	BAYES_GR_89V-18p7V	0.5950
78	GR_18p7H-6p9H	0.5944
79	GR_89H-18p7H	0.5826
80	BAYES_GR_18p7H-6p9H	0.5815
81	BAYES_GR_89V-23p8V	0.5807
82	GRice_89H-18p7H	0.5778
83	BAYES_GR_36p5H-18p7H	0.5775
84	GRice_89H-23p8H	0.5772
85	GR_89H-23p8H	0.5746
86	GRice_18p7H-6p9H	0.5641
87	DAL_18p7V-36p5H-36p5V	0.5594
88	**ASCAT_KNMI**	**0.5571**
89	BAYES_GR_89H-10p7H	0.5527
90	ASCAT_BAYES	0.5527
91	GRice_23p8H-18p7H	0.5486
92	GR_10p7V-6p9V	0.5395
93	GRice_10p7V-6p9V	0.5380
94	GRice_10p7H-6p9H	0.5248
95	BAYES_GR_23p8H-18p7H	0.5247
96	GR_10p7H-6p9H	0.5241
97	BAYES_GR_10p7V-6p9V	0.5197
98	BAYES_GR_10p7H-6p9H	0.5081
99	BAYES_GR_89H-18p7H	0.5025
100	BAYES_GR_89H-23p8H	0.4936
101	GRice_89H-36p5H	0.4806
102	GR_89H-36p5H	0.4762
103	BAYES_GR_89H-36p5H	0.4742
104	GR_89V-36p5V	0.4601
105	GRice_89V-36p5V	0.4594
106	BAYES_GR_89V-36p5V	0.4539

Table 3.2: Balanced accuracy of all algorithms and channel combinations applied to the Southern Hemisphere using a temporal consistent mask. The algorithms that are the subject of further discussion are highlighted in bold.

Overall, four fewer algorithms are available compared to the Northern Hemisphere, as ASIMIR and the three Lagrangian data products are not provided in the Southern Hemisphere. Some of the algorithms selected in the Northern Hemisphere perform significantly worse in the Southern Hemisphere. In particular,

ASCAT_KNMI ice types and some of the channel combinations show much lower balanced accuracy, while other channel combinations perform better. Overall, the non-Bayesian approaches work much better, which becomes clearly visible in Figure 3.7.

Comparison of channel combinations and different approaches

Figure 3.7 again shows a three-dimensional histogram of the balanced accuracy score, based on the two (three for DAL) input-channel frequencies and polarisations. The different approaches, like Bayes or MYI fraction, are separated by the markers and show the accuracy in the colour.

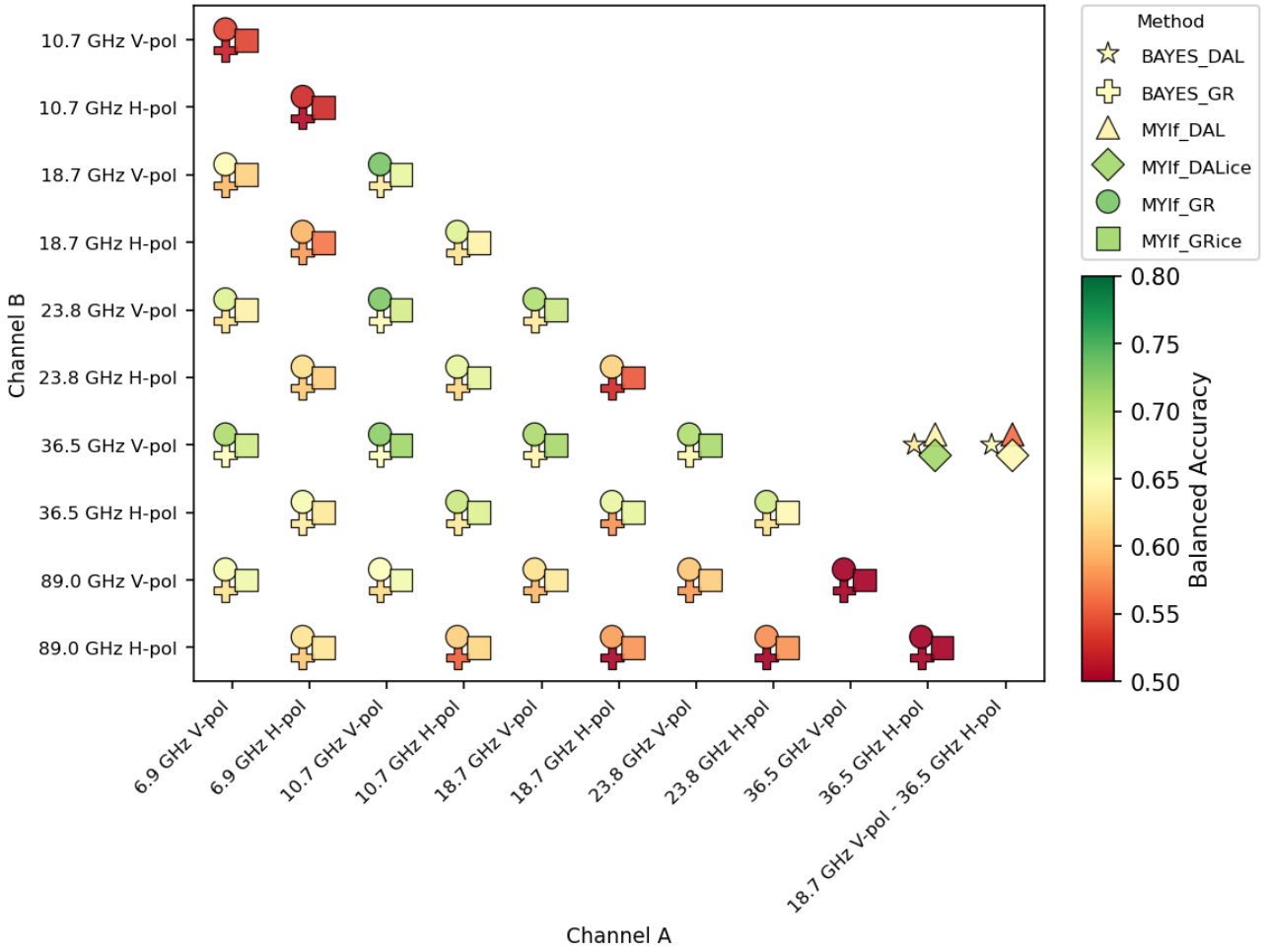


Figure 3.7: Balanced accuracy score of all channel combinations applied to the Southern Hemisphere using the spatial and temporal consistent mask.

The inverted behaviour of Bayesian approaches (crosses and stars), which now have consistently lower accuracies than non-Bayesian approaches, is likely due to the absence of a spatial mask, unlike in the Northern Hemisphere. In the Northern Hemisphere (NH), the marginal ice zone is mostly excluded from the comparison area by the spatially unified mask, whereas in the Southern Hemisphere (SH), no mask is applied, thus including the entire marginal ice zone, where the Bayesian approaches seem to be less reliable. The

channel combinations, including 89 GHz or combining frequencies with small differences, again show lower accuracy.

For further investigations, the same algorithms as for NH are used - if they are provided - but based on Figure 3.7 and Table 3.2, we have additionally selected GR_36p5V-10p7V, which is listed as number 5 in Table 3.2.

Comparisons of selected algorithms

Figure 3.8 shows the confusion matrices for the outputs of the 12 selected algorithms against the reference ice-type classification from the ice charts in the Snapshot-RRDP. These have been reduced to NIC and AARI ice charts for SH. Below each matrix, the corresponding accuracy and balanced accuracy scores are provided. With the exception of ASCAT_KNMI, all algorithms capture less of the MYI (lower-right matrix cell) and also generally have a tendency to wrongly classify MYI as FYI (lower-left matrix cell) more often than the other way around. This tendency is much more pronounced than in the Northern Hemisphere. ASCAT_KNMI, on the other hand, hardly detects FYI and is heavily overestimating the presence of MYI. This suggests that the backscatter threshold adapted from the NH is not appropriate for SH. All algorithms show poorer performance than in NH, though to different extents. This underlines the high complexity of detecting ice types and age in the Antarctic.

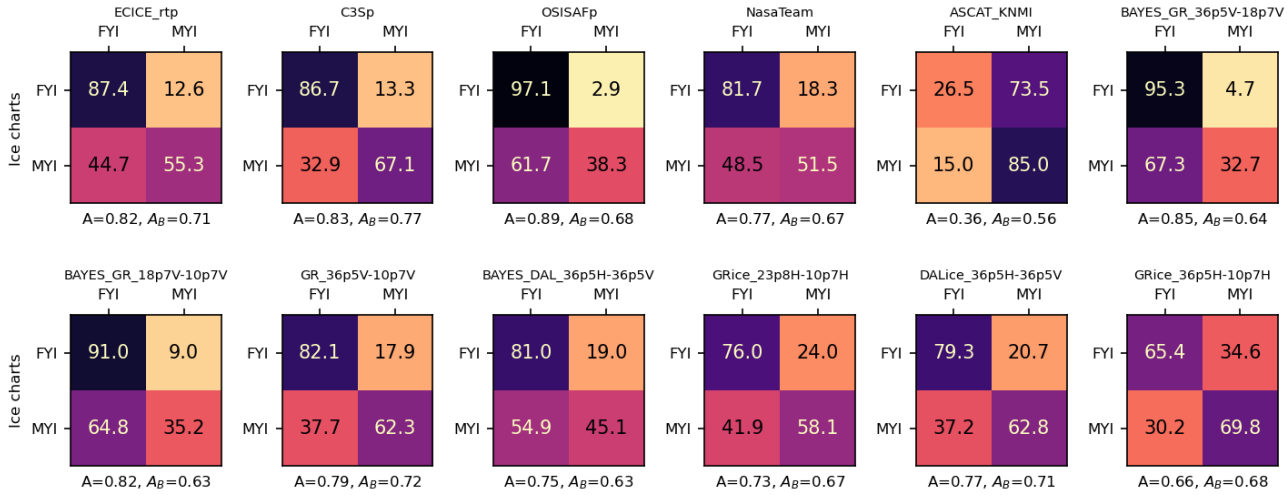


Figure 3.8: Confusion matrices of selected algorithms for the unified spatially covered area and temporal subset in the Southern Hemisphere. The name of the algorithm is shown above each matrix. The reference data is accumulated over all ice charts and reference data. Numbers in the matrix show the normalised number of predictions. The accuracy (A) and the balanced accuracy (AB) are given below the matrices.

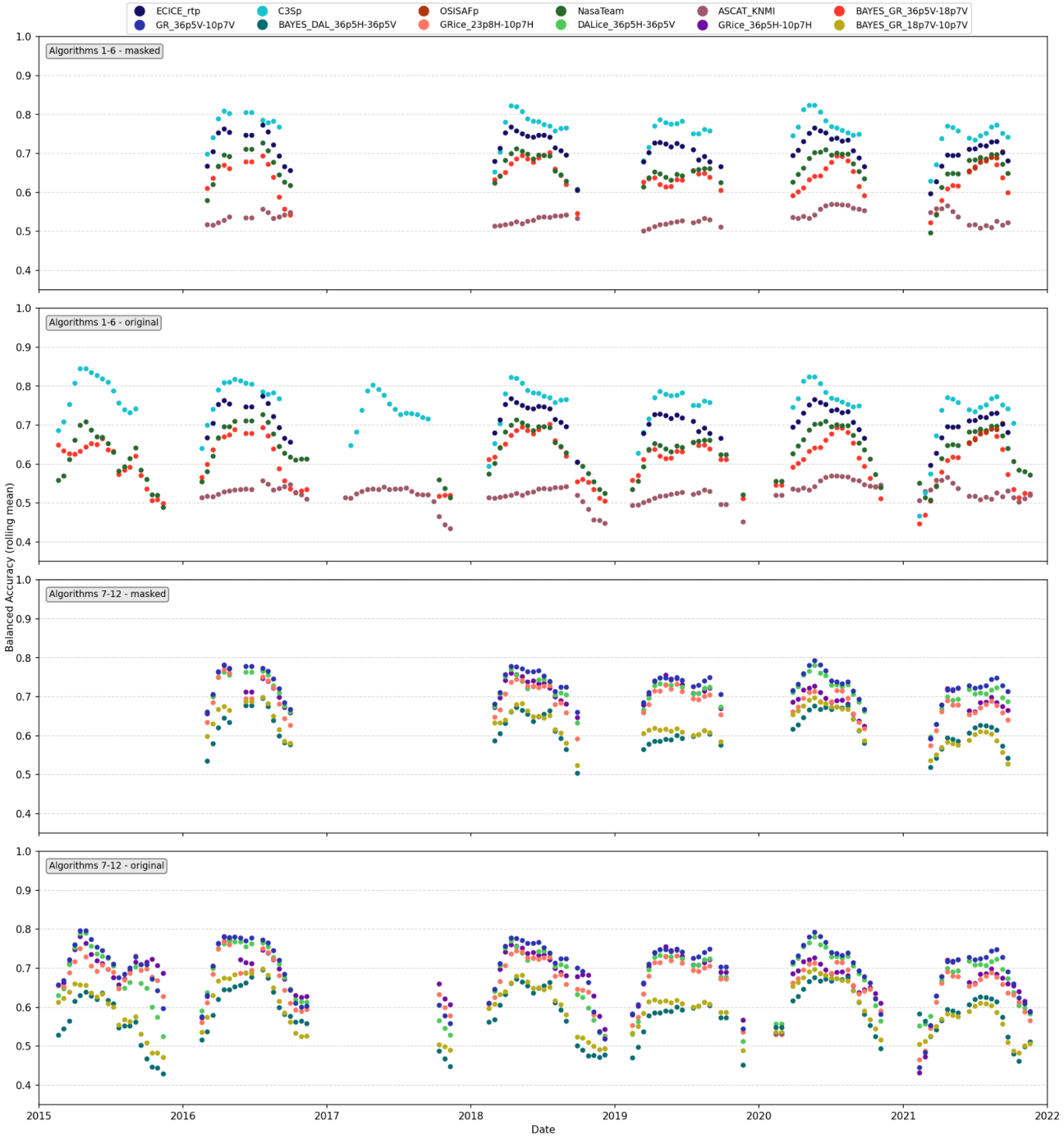


Figure 3.9: Temporal evolution of the balanced accuracy score for the selected algorithms in the Southern Hemisphere. Panel 1 and panel 3 show the results for the unified area mask and the overlapping days only, panels 2 and 4 for the original spatial area and time period of the algorithms, respectively. The algorithms are separated into two plots only for the sake of visibility.

Figure 3.9 shows the temporal evolution of the algorithms' balanced accuracies using a 30-day rolling mean, again separated into two subgroups of six algorithms each. As the Lagrangian algorithms are not provided for SH and the reference ice charts from the RRDP-snapshot are not provided through the peak of the melt season, the difference between the masked (upper panels) and the original coverage (lower panels) time series is not as pronounced as in the Northern Hemisphere. The seasonality of accuracy is also more similar across the algorithms, with a broad peak centred in the middle of the freezing season and lower accuracy at the beginning and end.

Without any significant exceptions, C3Sp demonstrates the highest accuracy, followed by ECICE_rtp. Comparing GR_36p5V-10p7V (blue dots in lower panel) and GRice_36p5V-10p7V (purple dots in lower panel) reveals slightly better performance when using the simple gradient ratio. Therefore, the latter is removed from further assessment. As the differences in performance between the channel combinations are not strong, the same collection of algorithms as in NH is selected for the spatial analysis of MYI detection.

Visual inspection and spatial comparison of maps

Figures 3.10 and 3.11 show the spatial maps of the (normalised) MYI detection frequency within the month of April (2018 and 2020) for the eight remaining algorithm products and the reference NIC and AARI ice charts from the snapshot-RRDP. Note that the maximum detection frequency here is two. The (white) data gaps close to the coastlines that are more or less pronounced in all products, as well as ice charts, are due to ice shelves.

In spatial terms, the C3Sp product shows the best agreement with the ice charts, with its strongest deviation occurring in the Amundsen Sea, where it detects too little ice, in particular in 2018. However, even the two ice charts do not fully agree, which underlines the complexity of ice-type detection in the Southern Hemisphere. ECICE_rtp also shows good agreement in 2018, despite some unreasonable MYI detection in the Ross Sea, which likely would be reduced when applying post-processing. In 2020, however, the deviation is enhanced. This is also the case for the NASATeam algorithm and the other channel combination approaches, which in general perform worse compared to the established algorithms.

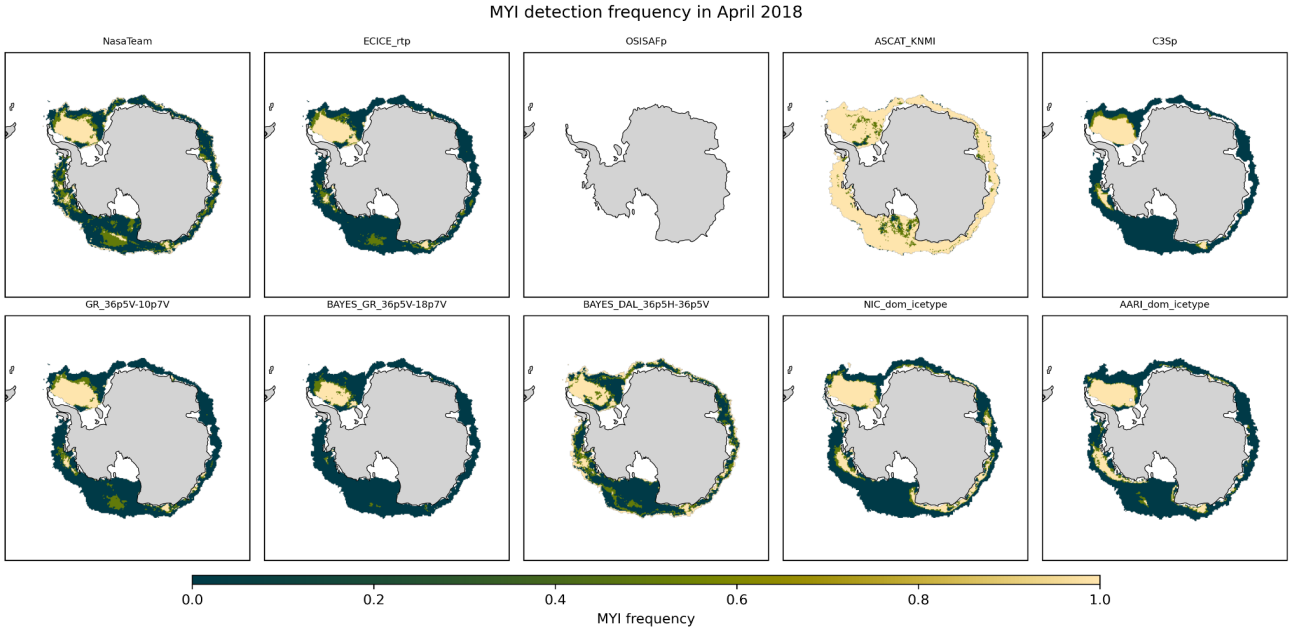


Figure 3.10: Spatial maps showing the normalised frequency of MYI detection by the selected subset of algorithms and the reference ice charts in April, 2018, in the Southern Hemisphere. The name of the algorithm or ice chart source is shown above each map.

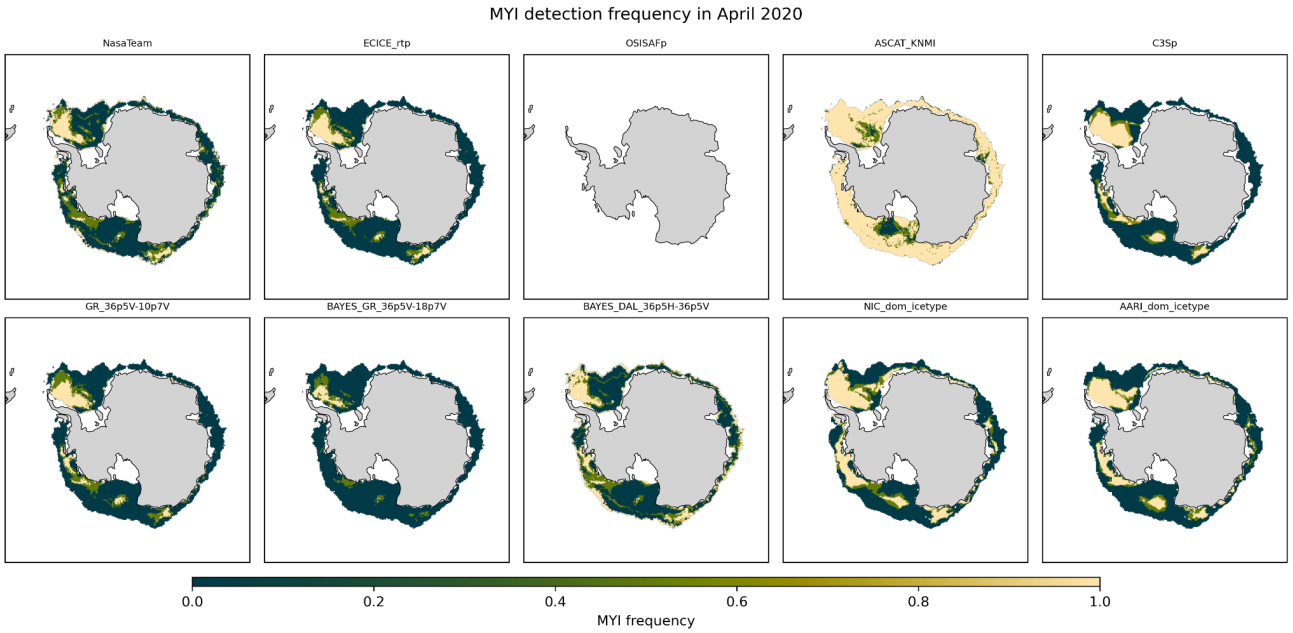


Figure 3.11: Spatial maps showing the normalised frequency of MYI detection by the selected subset of algorithms and the reference ice charts in April, 2020, in the Southern Hemisphere. The name of the algorithm or ice chart source is shown above each map.

4 Preliminary ranking of algorithms

4.1 Radiometric and backscatter-based algorithms

Among the radiometric and scatterometer retrievals considered in this exercise, C3Sp shows the highest accuracy in discriminating FYI and MYI, closely followed by ECICE_rtp and ASIMIR. These three algorithms therefore form the top-performing group in the intercomparison in this category. While ASIMIR is currently only implemented for the Northern Hemisphere, the method itself is not fundamentally restricted to this region and could in principle be extended to the Southern Hemisphere. It should also be noted that the C3Sp implementation includes dynamic tie-points (represented as probability distributions) together with the full set of post-processing steps applied in the operational product, whereas ECICE_rtp and ASIMIR were implemented directly within the RRDP framework using fixed tie-points and without additional post-processing. These differences should be kept in mind when interpreting the relative performance of the algorithms.

Although most algorithms in this category can, in theory, be applied during the summer melt period, the underlying physical conditions prevent reliable retrievals. These methods rely on the direct interpretation of microwave signatures, which lose their discriminatory power in summer. Consequently, seasonal coverage is eliminated as a selection criterion within this group of algorithms.

Table 4.1 Preliminary ranking of the radiometric and backscatter-based ice type algorithms.

Rank	Algorithm	Justification
1	C3Sp	C3Sp is among the algorithms with the highest accuracy in classifying FYI and MYI and is available for both hemispheres. This algorithm already includes drift correction and other post-processing steps as it was not re-implemented.
2	ECICE_rtp	ECICE_rtp is also amongst the highest-performing algorithms, with slightly lower accuracy than C3Sp. However, this algorithm has been reimplemented and, unlike C3Sp, does not include any post-processing corrections, which potentially could have increased its performance to a similar level as C3Sp. It is also provided in both hemispheres.
3	ASIMIR	ASIMIR shows accuracies similar to ECICE_rtp and does not rely on tie points. It is a stand-alone algorithm that does not contain post-processing steps but could benefit from future implementation of drift correction or other filters. However, ASIMIR is not available in SH and thus ranked lower than ECICE_rtp.
4	OSISAFp	Despite including its post-processing steps, OSISAFp performs notably worse than higher-ranked algorithms. In particular, it shows a strong bias toward FYI detection.

5	NASATeam	NASATeam performs well in NH but has stronger issues in SH than other established retrievals. It shows, however, that combining more (in this case, three) channels significantly improves performance compared to the simple two-channel combination approaches.
6	ASCAT_KNMI	Of the established algorithms, ASCAT_KNMI performs the worst with very low accuracy scores, especially in SH.

Overall, more complex and established algorithms outperform simple channel combinations. However, combining more microwave channels and combining radiometry and scatterometry to increase information content has shown potential. Purely scatterometer-driven ice-type classification algorithms perform much worse than microwave or combined algorithms, but currently they rely only on ASCAT data. Ku-band, conically scanning sensors like QuikSCAT potentially could improve the classification and might be added in a future version of the RRDP.

The C3Sp, ECICE and ASIMIR algorithms demonstrate the greatest potential. The intercomparison, however, is limited by the differences in input data and whether or not post-processing is applied. For this reason, the range of algorithms will initially be kept broad and will be narrowed down in the next version.

4.2 Lagrangian algorithms

The ranking of the Lagrangian algorithms, with justification according to the selection criteria, is provided in Table 4.2.

Table 4.2 Ranking of the Lagrangian sea ice age algorithms

Rank	Algorithm	Justification
1	LM	The algorithm has the highest accuracy in FYI / MYI classification. It provides a sea-ice age distribution in every pixel. It operates in winter and summer months, and whenever the upstream sea-ice drift and concentration products are available. The coverage is limited to the central part of the Arctic ocean and excludes Antarctica, the Canadian Archipelago, and the Bering Sea, however that can be improved in future and total coverage can be achieved.
2	NSIDC2	The algorithm has the second-highest accuracy in FYI / MYI classification among the Lagrangian algorithms. It provides only one ice age value per pixel and can operate in all seasons. The coverage is limited to the same region as for the LM algorithm.
3	NSIDC1	The algorithm has the lowest accuracy in FYI / MYI classification among the Lagrangian algorithms. It provides only one ice age value per pixel and can operate in all seasons. The coverage is better than for the LM or NSIDC1 algorithms as it

		includes also the Canadian archipelago.

Among the Lagrangian algorithms, only the LM algorithm will be selected for further development in the SAGE project because it performs significantly better than the two tested NSIDC versions among the Lagrangian algorithm category. The only advantage of the NSIDC1 algorithm over LM at the current state is its coverage of the Canadian Arctic Archipelago but the implementation of LM for this area is ongoing and will be included for the next version.

4.3 Lessons learned and potential for a merged algorithm

This document is the first version of the PVASR in the ESA CCI SAGE project. It documents some preliminary conclusions towards selecting an algorithm baseline for implementation in the production of a new sea-ice age/type climate data record. As of this version, the round-robin exercise does not conclude on a single algorithm for implementation and production; more work and analysis are required.

The rest of this section highlights some of the lessons learned so far from the algorithm intercomparison exercise.

Different product types and harmonisation. The different algorithms tested here have very different output types. Some are sea-ice age products tracking several age classes, others are sea-ice type products focused only on the FYI and MYI. The algorithms also return different quantities, including fractions, partial concentrations, probabilities, and binary classifications. We attempted to harmonise these different outputs to ensure a fair assessment. The harmonisation process will also be a central element of designing a merging algorithm to combine several approaches into one CDR.

Uncertainties in reference data. The reference ice charts in the snapshot-RRDP are often inconsistent in the extent and concentration of MYI, as well as in temporal availability. That requires using probabilistic approaches for evaluation of the ice age algorithms. The aforementioned harmonisation should include consistent computation of probabilities of ice types from ice charts and the derived products. It may contribute to better evaluation of the sea ice age product uncertainty.

Combination of microwave channels. We tested both well-established (GR) and somewhat newer (DAL) combinations of channels. While there are already clear indications that some channels (in particular 19, 23, 37 GHz) have the most potential for the retrieval of sea-ice type, our results indicate that more advanced ways to combine (more) channels can perform even better. These include Machine Learning approaches.

Importance of drift corrections and post-processing. The best performing algorithms so far (C3Sp and LM) rely heavily on sea-ice drift information. Sea-ice drift fields are at the core of the LM algorithm (that does not directly process microwave signals) and plays a key role in masking erroneous classifications in the C3S via

backtracking. This confirms that 1) the filtering and post-processing steps are probably as important as the choice of microwave channels and how they are combined in the “low level” algorithms, and 2) that there is a strong potential in SAGE to design and test improved merging methodologies between LM and microwave-based algorithms.

Scatterometry. For most of the CDR time span, data will come only from passive microwave missions, which is why this intercomparison has initially focused on this type of mission. We have included one algorithm/product based solely on scatterometry and two multi-sensor algorithms that combine radiometry and scatterometry. Both of these performed less than the radiometry-only algorithms. This might, however, be due to our only C-band scatterometers. Ku-band scatterometers (covering shorter periods) have been reported to offer better accuracy, at least in the Arctic (Belmonte et al., 2018, Ye et al., 2023).

Exploitation of the buoy-RRDP. The buoy-RRDP is biased toward FYI observations, and because the data points in a buoy time series are correlated, the dataset was not exploited for evaluating snapshot ice types. However, as the project progresses, this dataset could be used to derive significant value in evaluating the ice age in the LM algorithm and possibly validating drift corrections during post-processing. Additionally, the dataset might be useful to test the plausibility algorithms, retrieving more ice type classes than FYI and MYI

5 References

Belmonte Rivas, M., Otosaka, I., Stoffelen, A., and Verhoef, A.: A scatterometer record of sea ice extents and backscatter: 1992–2016, The Cryosphere, 12, 2941–2953, https://doi.org/10.5194/tc-12-2941-2018, 2018.

Breivik, L., Eastwood, S., Lavergne, T., and others, 2012: "Sea ice products for the EUMETSAT Ocean and Sea Ice SAF." Proceedings of the EUMETSAT Meteorological Satellite Conference.

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., & Savoie, M. H. (2012). EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets. ISPRS International Journal of Geo-Information, 1(1), 32-45. https://doi.org/10.3390/ijgi1010032

Cavalieri, D. J., P.Gloersen, and W. J.Campbell (1984), Determination of sea ice parameters with the NIMBUS 7 SMMR, J. Geophys. Res., 89(D4), 5355–5369, doi:10.1029/JD089iD04p05355.

GCOS report no. 244 (2022). The 2022 Implementation Plan for the Global Climate Observing System (GCOS-244). https://gcos.wmo.int

GCOS report no. 245 (2022). The 2022 Essential Climate Variables Requirements for the Global Climate Observing System (GCOS-245). https://gcos.wmo.int

Korosov, A. A., Rampal, P., Pedersen, L. T., Saldo, R., Ye, Y., Heygster, G., Lavergne, T., Aaboe, S., and Girard-Ardhuin, F.: A new tracking algorithm for sea ice age distribution estimation, The Cryosphere, 12, 2073–2085, https://doi.org/10.5194/tc-12-2073-2018, 2018.

Korosov, A., Edel, L., Regan, H., Lavergne, T., Aaboe, S., and Down, E. J.: A climate data record of sea ice age using Lagrangian advection of a triangular mesh, Earth Syst. Sci. Data, 18, 721–740, https://doi.org/10.5194/essd-18-721-2026, 2026.

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.

Melsheimer, C., Spreen, G., Ye, Y., and Shokr, M.: First results of Antarctic sea ice type retrieval from active and passive microwave remote sensing data, The Cryosphere, 17, 105–126, https://doi.org/10.5194/tc-17-105-2023, 2023.

Rückert, J. E., Huntemann, M., Tonboe, R. T., & Spreen, G. (2023). Modeling snow and ice microwave emissions in the Arctic for a multi-parameter retrieval of surface and atmospheric variables from microwave

radiometer satellite data. Earth and Space Science, 10, e2023EA003177. https://doi.org/10.1029/2023EA003177

Scarlat, R. C., Spreen, G., Heygster, G., Huntemann, M., Patilea, C., Pedersen, L. T., & Saldo, R. (2020). Sea ice and atmospheric parameter retrieval from satellite microwave radiometers: Synergy of AMSR2 and SMOS compared with the CIMR candidate mission. Journal of Geophysical Research: Oceans, 125, e2019JC015749. https://doi.org/10.1029/2019JC015749

Shokr M.,, A. Lambe and T. Agnew, "A New Algorithm (ECICE) to Estimate Ice Concentration From Remote Sensing Observations: An Application to 85-GHz Passive Microwave Data," in IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 12, pp. 4104-4121, Dec. 2008, doi: 10.1109/TGRS.2008.2000624.

Tharwat, A. (2020). Classification assessment methods. Applied Computing and Informatics, Emerald, 17(1), 168-192. doi:10.1016/j.aci.2018.08.003

Tschudi, M. A., Meier, W. N., and Stewart, J. S.: An enhancement to sea ice motion and age products at the National Snow and Ice Data Center (NSIDC), The Cryosphere, 14, 1519–1536, https://doi.org/10.5194/tc-14-1519-2020, 2020.

World Meteorological Organization & Intergovernmental Oceanographic Commission. (2010).
SIGRID-3: A vector archive format for sea ice charts (WMO/TD-No. 1214; JCOMM Technical Report No. 23, Rev. 2). WMO.
Accessed from MET Norway: https://cryo.met.no/sites/cryo/files/IceService_docs/wmo-Sigrid_3_2010.pdf

Y. Ye, G. Heygster and M. Shokr, "Improving Multiyear Ice Concentration Estimates With Reanalysis Air Temperatures," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 5, pp. 2602-2614, May 2016, doi: 10.1109/TGRS.2015.2503884.

Ye, Y.; Shokr, M.; Heygster, G.; Spreen, G. Improving Multiyear Sea Ice Concentration Estimates with Sea Ice Drift. Remote Sens. 2016, 8, 397. https://doi.org/10.3390/rs8050397

Ye, Y., Luo, Y., Sun, Y., Shokr, M., Aaboe, S., Girard-Ardhuin, F., Hui, F., Cheng, X., and Chen, Z.: Inter-comparison and evaluation of Arctic sea ice type products, The Cryosphere, 17, 279–308, https://doi.org/10.5194/tc-17-279-2023, 2023.