

Ocean & Sea Ice SAF

Global Sea Ice Concentration Climate Data Records

Algorithm Theoretical Basis Document

Document version: 3.0

Data set version: 3.0

OSI-450-a: [10.15770/EUM_SAF_OSI_0013](https://doi.org/10.15770/EUM_SAF_OSI_0013)

OSI-430-a: [10.15770/EUM_SAF_OSI_0014](https://doi.org/10.15770/EUM_SAF_OSI_0014)

OSI-458: [10.15770/EUM_SAF_OSI_0015](https://doi.org/10.15770/EUM_SAF_OSI_0015)

August 2022

*Thomas Lavergne¹, Atle Sørensen¹, Rasmus Tonboe^{2, now 3},
Roberto Saldo³, Leif Toudal Pedersen^{3, now 5}, Courtenay Strong^{4a},
Elena Cherkaev^{4b}, Kenneth M. Golden^{4b}, and Steinar Eastwood¹*

¹Norwegian Meteorological Inst., ²Danish Meteorological Insti., ³Technical Univ. of Denmark,
⁴University of Utah (^aDept of Atmos. Sciences, ^bDept of Maths), ⁵eolab.dk

The EUMETSAT
Network of
Satellite Application
Facilities



EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

Document Change Record

Version	Date	Change	Description	Responsible
1.0	23.06.2016		First version for review	John Lavelle
1.1	7.7.2016		Edited after PCR comments	John Lavelle
1.2	14.03.2019		Update for documenting ICDR OSI-430-b (including ORR comments)	Thomas Lavergne
2.0	21.04.2021		First version for OSI-450-a, OSI-430-a, and OSI-458, for review.	Thomas Lavergne
2.1	02.07.2021		Edited after PCR comments	Thomas Lavergne
3.0_draft	17.06.2022		Updated with final algorithms before ORR/DRR of OSI-450-a, OSI-430-a, and OSI-458.	Thomas Lavergne
3.0	24.08.2022		Revised after comments from ORR/DRR	Thomas Lavergne

Table of Contents

1	Introduction.....	5
1.1	The EUMETSAT Ocean and Sea Ice SAF.....	5
1.2	Disclaimer.....	5
1.3	Scope.....	6
1.4	Overview.....	6
1.5	Glossary.....	7
1.6	Reference and Applicable documents.....	8
1.6.1	Reference documents.....	8
1.6.2	Applicable documents.....	9
2	Input Data.....	10
2.1	The SMMR data.....	10
2.2	The SSM/I data.....	11
2.3	The SSMIS data.....	12
2.4	The AMSR-E data.....	12
2.5	The AMSR2 data.....	13
2.6	Numerical Weather Prediction data.....	13
2.7	Binary land mask.....	13
2.8	Sea Ice Extent Climatology.....	14
2.9	Beyond SSMIS.....	15
3	Overview of the SIC processing chain.....	16
4	Pre-processing on swath projection (L1P).....	17
4.1	Data volume reduction.....	17
4.2	Apply inter-platform calibration.....	17
4.3	Collocate channels.....	17
4.4	Land spill-over correction.....	18
5	Geophysical processing in swath projection (L2).....	20
5.1	A hybrid, self-tuning, self-optimizing sea-ice concentration algorithm.....	20
5.1.1	Merging equation for the hybrid algorithm.....	20
5.1.2	Applying the SIC algorithms.....	21
5.1.3	Tuning of the algorithm at 0% and 100% SIC.....	21
5.1.4	Dynamical selection of the SIC training data.....	22
5.2	Radiative transfer modelling for correcting atmospheric influence on T_b	23
5.3	Open-water filtering.....	26
5.4	Reducing systematic errors at high-concentration.....	29
5.5	Algorithm and tie-points uncertainties.....	31
6	Level 3 algorithms.....	32
6.1	Gridding and daily averaging.....	32
6.2	Gridding and smearing uncertainty.....	32
7	Level 4 algorithms.....	34
7.1	Interpolation of missing data.....	34
7.1.1	Temporal interpolation.....	34
7.1.2	Spatial interpolation.....	34
7.2	Total uncertainty.....	37

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

7.3 Flagging and final formatting.....	38
8 Higher-level Products.....	39
8.1 Monthly averaged SIC products.....	39
8.2 The sea-ice index OSI-420.....	39
9 Conclusions.....	40
10 References.....	41

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

1 Introduction

1.1 The EUMETSAT Ocean and Sea Ice SAF

The Satellite Application Facilities (SAFs) are dedicated centres of excellence for processing satellite data – hosted by a National Meteorological Service – which utilise specialist expertise from institutes based in Member States. EUMETSAT created Satellite Application Facilities (SAFs) to complement its Central Facilities capability in Darmstadt. The Ocean and Sea Ice Satellite Application Facility (OSI SAF) is one of eight EUMETSAT SAFs, which provide users with operational data and software products. More on SAFs can be read at www.eumetsat.int.

OSI SAF produces (on an operational basis) a range of air-sea interface products, namely: wind, sea ice characteristics, Sea Surface Temperatures (SST), Surface Solar Irradiance (SSI) and Downward Longwave Irradiance (DLI). The sea ice products include sea ice concentration, the sea ice emissivity at 50 GHz, sea ice edge, sea ice type and sea ice drift and sea ice surface temperature (from mid 2014).

The OSI SAF consortium is hosted by Météo-France. The sea ice processing is performed at the High Latitude processing facility (HL centre), operated jointly by the Norwegian and Danish Meteorological Institutes.

1.2 Disclaimer

All intellectual property rights of the OSI SAF products belong to EUMETSAT. The use of these products is granted to every interested user, free of charge. If you wish to use these products, EUMETSAT's copyright credit must be shown by displaying the words "Copyright © YYYY EUMETSAT" or the OSI SAF logo on each of the products used. Note : The comments that we get from our users is an important input when defining development activities and updates, and user feedback to the OSI SAF project team is highly valued.

Acknowledgement and citation

Use of the product(s) should be acknowledged with the following citations:

EUMETSAT Ocean and Sea Ice Satellite Application Facility, Global sea ice concentration climate data record 1978-2020 (v3.0, 2022), OSI-450-a, doi: 10.15770/EUM_SAF_OSI_0013, data (for [extracted period], [extracted domain],) extracted on [download date]

EUMETSAT Ocean and Sea Ice Satellite Application Facility, Global sea ice concentration interim climate data record (v3.0, 2022), OSI-430-a, doi: 10.15770/EUM_SAF_OSI_0014, data (for [extracted period], [extracted domain],) extracted on [download date]

EUMETSAT Ocean and Sea Ice Satellite Application Facility, Global medium resolution sea ice concentration climate data record 2002-2020 (v3.0, 2022), OSI-458, doi: 10.15770/EUM_SAF_OSI_0015, data (for [extracted period], [extracted domain],) extracted on [download date]

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

1.3 Scope

This document is targeted at OSI SAF product users and describes the scientific background, the source data and the processing steps used to create the OSI SAF Global Sea Ice Concentration Climate Data Record (OSI-450-a), the associated Interim Climate Data Record (OSI-430-a), and the OSI SAF AMSR Global Sea Ice Concentration Climate Data Record (OSI-458).

1.4 Overview

OSI-450-a, OSI-430-a and OSI-458 constitute the *third major release* of the OSI SAF Global Sea Ice Concentration Climate Data Records. In short, these three CDRs can be summarized as:

- OSI-450-a : The Global Sea Ice Concentration Climate Data Record, based on coarse resolution imagery from SMMR, SSM/I, and SSMIS and covering 1978-2020.
- OSI-430-a : The Global Interim CDR (ICDR) based on coarse resolution imagery from SSMIS and providing an extension of OSI-450-a starting January 2021.
- OSI-458 : The Global Sea Ice Concentration Climate Data Record based on medium resolution imagery from AMSR-E (2002-2011) and AMSR2 (2012-2020). OSI-458 is a Research to Operations transfer from ESA CCI.

OSI-450-a is a fixed-length climate data record (1978-2020) based on re-calibrated satellite data and the state-of-the-art ERA5 reanalysis. OSI-458 is also a fixed-length climate data record (2002 – 2020 with a data gap in 2011-2012) that achieves better spatial resolution than OSI-450-a by using the AMSR-E and AMSR2 satellites. OSI-430-a provides a timely extension of OSI-450-a starting in January 2021. OSI-430-a has two data streams: 1) the “nominal” ICDR that applies exactly the same algorithm as OSI-450-a and has a latency of 16 days, and 2) the “fast-track” ICDR that applies a slightly different algorithm (tie-point selection) and achieves a latency of 2 days. The fast-track ICDR was introduced for the third release after requests from operational climate users.

The *first major version* of the OSI SAF sea-ice concentration CDRs was called OSI-409 and was initiated in 2006 (Tonboe et al., 2016).

The *second major version* was OSI-450, complemented by the ICDR OSI-430-b (Lavergne et al., 2019). Some of the algorithms implemented in OSI-450 and OSI-430-b were contributed to by the ESA CCI Phase 1 and 2 projects. ESA CCI also released their own SIC CDR at medium resolution (based on AMSR-E and AMSR2 data), notably the SICCI-25km dataset.

For this *third major version*, new R&D contributions from ESA CCI (now CCI+) are acknowledged. On the one hand, new R&D contributions from ESA CCI contribute to the v3 SIC algorithms used in the OSI SAF CDR. In addition, OSI-458 is a newer and better version of SICCI-25km that was developed in the ESA CCI Phase 2 project (Research to Operation transfer from ESA CCI to EUMETSAT OSI SAF).

It is noteworthy that ESA CCI+ is also producing their own SIC datasets, that are complementary to those of the OSI SAF (see climate.esa.int/seaice for updates). The datasets from CCI+ Sea Ice will target higher spatial resolution than those of the OSI

SAF, at the cost of being shorter time series. The pros and cons of the sea-ice concentration datasets from OSI SAF and CCI+ are described in the Product User Manuals of both projects.



Figure 1: The ESA Climate Change Initiative Sea Ice project contributed to the OSI SAF climate data records through a number of algorithm developments.

1.5 Glossary

Acronym	Description
AMSR	Advanced Microwave Scanning Radiometer
ATBD	Algorithm Theoretical Basis Document
C3S	Copernicus Climate Change Service
CCI	Climate Change Initiative
CDOP	Continuous Developments and Operations Phase
CDR	Climate Data Record
CLASS	Comprehensive Large Array-data Stewardship System
DMI	Danish Meteorological Institute
DMSP	Defence Meteorological Satellite Program
ECMWF	European Centre for Medium-Range Weather Forecast
ESA	European Space Agency
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
FCDR	Fundamental Climate Data Record
FoV	Field Of View
FYI	First Year Ice
GR	Gradient Ratio
ICDR	Interim Climate Data Record
MET	Norwegian Meteorological Institute
NASA	National Aeronautics and Space Administration

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

NH	Northern Hemisphere
NSIDC	National Snow and Ice Data Center
NWP	Numerical Weather Prediction
OSI SAF	Ocean and Sea Ice Satellite Application Facility
OSTIA	Operational Sea Surface Temperature and Sea Ice Analysis
OWF	Open Water Filter
PCR	Product Consolidation Review
RTM	Radiative Transfer Model
SAR	Synthetic Aperture Radar
SH	Southern Hemisphere
SIC	Sea Ice Concentration
SICCI	ESA CCI Sea Ice project
SMMR	Scanning Multichannel Microwave Radiometer
SSM/I	Special Sensor Microwave/Imager
SSMIS	Special Sensor Microwave Imager Sounder
SST	Sea Surface Temperature
SDR	Sensor Data Record
Tb	Brightness Temperature
TBC	To Be Confirmed
TBD	To Be Determined
TBW	To Be Written
WF	Weather Filter

1.6 Reference and Applicable documents

User related aspects of the product (like file format and output specifications) are to be found in the Product User's Manual (PUM) [RD-1]. Results from validation against on-ice drifters are gathered in an associated Scientific Validation Report (SVR) [RD-2].

1.6.1 Reference documents

[RD-1] EUMETSAT OSI SAF Product User's Manual for the Global Sea Ice Concentration Climate Data Records v3 (OSI-450-a, OSI-430-a, OSI-458), SAF/OSI/CDOP3/MET/TEC/MA/288, version v3.0, August 2022

[RD-2] EUMETSAT OSI SAF Validation Report for the Global Sea Ice Concentration Climate Data Records v3 (OSI-450-a, OSI-430-a, OSI-458), SAF/OSI/CDOP3/DMI/SCI/RP/285, version v3.0, August 2022

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

1.6.2 Applicable documents

[AD-1] EUMETSAT OSI SAF Product Requirements Document,
SAF/OSI/CDOP3/MF/MGT/PL/2-001, version 1.90, 31/12/2021

2 Input Data

This chapter describes the SMMR, SSM/I, SSMIS AMSR-E and AMSR2 satellite data, as well as numerical weather prediction (NWP) data, used for this third version. Figure 2 shows a timeline of the satellite missions having relevant passive microwave sensors, many of which enter our SIC CDRs.

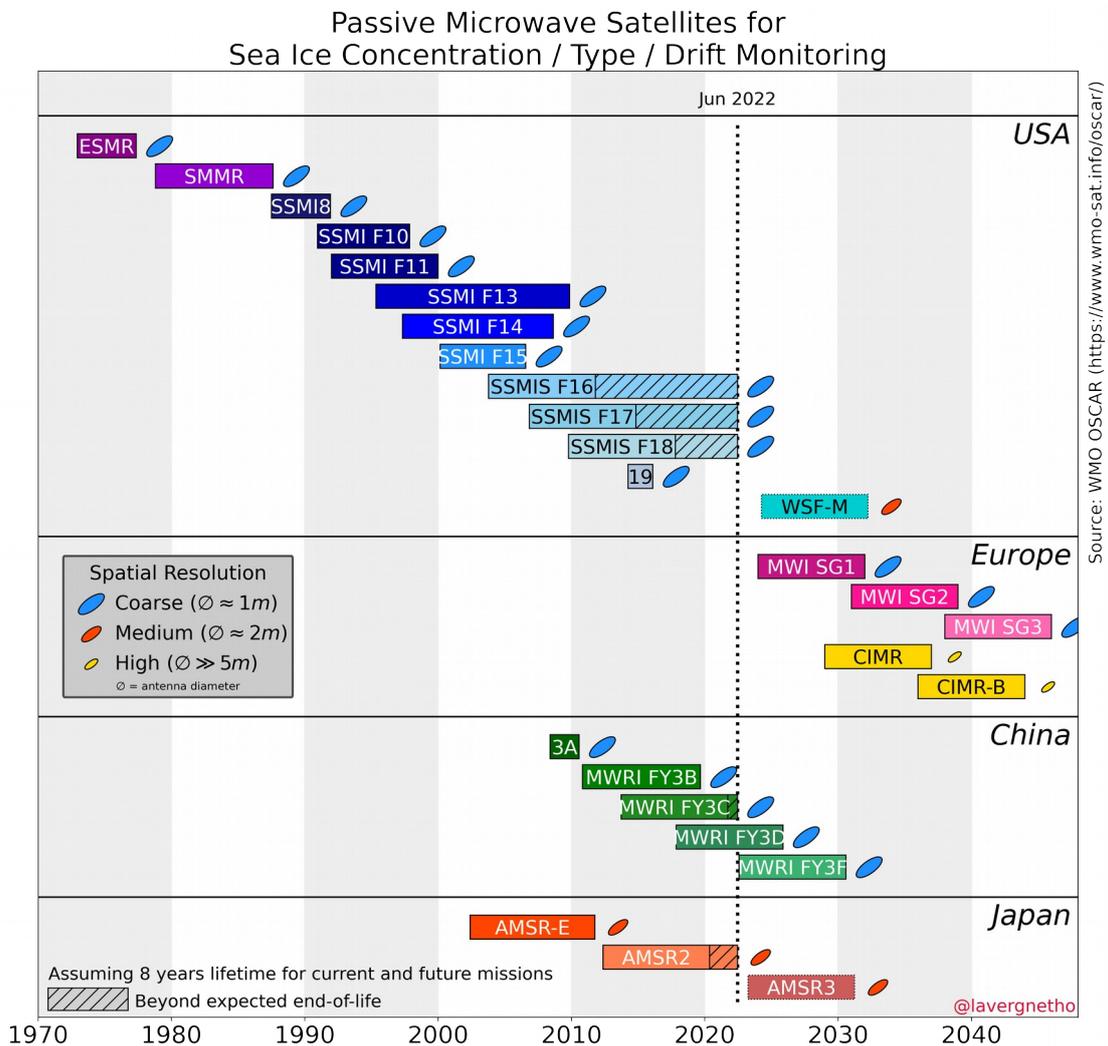


Figure 2: Timeline of the Passive Microwave satellite missions relevant for sea-ice concentration/extent/area monitoring with an indication of their spatial resolution capabilities. OSI-450-a and OSI-430-a use SMMR, all SSM/I, and SSMIS. OSI-458 uses AMSR-E and AMSR2. The horizontal bars represent satellite missions, which are colored by sensor family.

2.1 The SMMR data

The Scanning Multichannel Microwave Radiometer (SMMR) instrument on board the Nimbus 7 satellite operated from October 1978 to 20th August 1987 (Gloersen et al., 1992). For most of the period, the instrument was operated only every second day, due to power supply limitations. The instrument had 10 channels, from six Dicke

radiometers, at five frequencies (6.6, 10.7, 18.0, 21.0, 37.0 GHz) and vertical and horizontal polarization (Table 1). The scanning across track was ensured by tilting the reflector from side to side while maintaining constant incidence angle on the ground of about 50.2°. The scan track on the ground formed a 780 km wide arc in front of the satellite (Gloersen and Barath, 1977). Because of the satellite orbit inclination and swath width there is no coverage poleward of 84°.

Frequency (GHz)	Polarizations	Sampling (average)	Field of view	
			Along-track	Cross-track
6.6	H,V	25 km	148 km	95 km
10.7	H,V	25 km	91 km	59 km
18.0	H,V	25 km	55 km	41 km
21.0	H,V	25 km	46 km	30 km
37.0	H,V	25 km	27 km	18 km

Table 1: Characteristics of the Nimbus 7 SMMR channels (Gloersen and Barath, 1977).

Readers interested in the processing, calibration and quality check steps applied in the FCDR will find many more details in the CM-SAF documentation and Fennig et al. (2020).

2.2 The SSM/I data

The Special Sensor Microwave/Imager (SSM/I) sensors on board the Defence Meteorological Satellite Program (DMSP) started its record with the F08 satellite on 9th July 1987, shortly before the SMMR ceased to operate. The SSM/I is a total power radiometer, with a conical scan measuring the upwelling radiation from the Earth at a constant incidence angle of about 53.1° at four frequencies (19.3, 22.2, 37.0, 85.5 GHz). The swath width is about 1400 km and the polar observation hole extends to 87°.

The SSM/I data set used for this reprocessing was prepared by EUMETSAT CM SAF and covers the period of available DMSP satellites instruments from 1987 to 2008 (F08, F10, F11, F13, F14, F15) (see Table 2). Some SSM/I instruments continued their mission further than 2008, but these data are not included in the CM SAF FCDR.

The SSM/I instruments have five low frequency channels that are mostly similar to some of those on SMMR. In addition, two higher frequency channels at 85GHz, with twice the sampling rate and better spatial resolution, are available on the SSM/I starting with DMSP F10 (the 85 GHz channels had a malfunction on F08). Characteristics of the SSM/I channels are listed in Table 3.

Satellite	Period covered
F08	Jul 1987 – Dec 1991
F10	Jan 1991 - Nov 1997
F11	Jan 1992 – Dec 1999
F13	May 1995 – Dec 2008
F14	May 1997 – Aug 2008
F15	Feb 2000 – Jul 2006

Table 2: The different satellite missions carrying the SSM/I instrument and the periods they cover.

Frequency (GHz)	Polarizations	Sampling	Footprint size	
			Along-track	Cross-track
19.35	H,V	25 km	69 km	43 km
22.235	V	25 km	50 km	40 km
37.0	H,V	25 km	37 km	28 km
85.5	H,V	12.5 km	15 km	13 km

Table 3: Characteristics of the different SSM/I channels (from Wentz, 1991).

Readers interested in the processing, calibration and quality check steps applied in the FCDR will find many more details in the CM-SAF documentation and Fennig et al. (2020).

2.3 The SSMIS data

The SSMIS instruments are a slight evolution of the SSM/I concept, and most characteristics that drive the design of SIC CDRs are similar to SSM/I. Noticeable differences are the size of the polar observation hole (89°), and the center frequency of the high-frequency channels (91.1 GHz). The SSMIS instruments were also on board DMSP satellites, and we use F16, F17, and F18 missions (F19 was a short-lived mission, and F20 was never launched). DMSP F18 is thus the last available SSMIS instrument.

Data from three DMSP platforms are used in the CDR OSI-450-a: F16 (Nov 2005 - Dec 2013), F17 (Dec 2006 - Dec 2020), and F18 (Mar 2010 - Dec 2020). They are from CM SAF FCDR.

In addition, SSMIS F16, F17, and F18 are processed to extend the ICDR OSI-430-a from January 2021 onwards. This operational data stream is from NOAA CLASS.

2.4 The AMSR-E data

The AMSR-E instrument on board the Earth Observing System (EOS) satellite Aqua recorded passive microwave data from 1st June 2002 until 4th October 2011. This instrument measured vertically and horizontally polarized brightness temperatures at 7 frequencies (6.9, 7.2, 10.7, 18.7, 23.8, 36.5 and 89 GHz), thus 14 channels in all. Thanks to a larger antenna reflector, AMSR-E had significantly better spatial resolution than SSM/I or SSMIS. It also had a wider swath, and thus a smaller polar observation hole (89.5°).

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

OSI-458 uses AMSR-E data (June 2002 to October 2011) from NSIDC FCDR AE_L2A V003 by Ashcroft and Wentz (2013).

2.5 The AMSR2 data

The AMSR2 instrument on board the Global Change Observation Mission – Water (GCOM-W1) satellite provides similar data to the AMSR-E instrument (no 7.2 GHz channels), with slightly better resolution.

OSI-458 uses AMSR2 data (July 2012 to December 2020) from JAXA's archive of L1R data stream.

2.6 Numerical Weather Prediction data

The microwave radiation emitted by the ocean and sea ice travels through the Earth's atmosphere before being recorded by the satellite sensors. Scattering, reflection, and emission in the atmosphere add or subtract contributions to the radiated signal, and challenge our ability to accurately quantify sea-ice concentration.

An central step in our Level-2 processing is thus the explicit correction of the T_b for the atmospheric contribution to the top of the atmosphere radiation (see section 5.2). For this purpose, we access global hourly fields from C3S ERA5 reanalysis (produced by ECMWF, see Hersbach et al., 2020). Fields of 10m wind speed, 2m air temperature, total column water vapour and total column cloud liquid water are used.

The ERA5 reanalysis was recently extended to start in 1950, and is thus available throughout the time period of the CDRs OSI-450-a and OSI-458. The skill of ERA5 changes in Jan 1979 due to modifications in (1) the observing system, (2) the ocean boundary conditions (SIC data), and (3) the data assimilation system (background error covariances). Our investigations and validation results do not reveal a notable impact on the Oct-Dec 1978 SICs.

For the main stream of OSI-430-a, we access ERA5T data. ERA5T is the same data stream as ERA5 but is made available earlier (generally 5 days latency) but without the quality control going into ERA5.

The fast-track stream of OSI-430-a uses the operational IFS analysis and forecast from ECMWF.

2.7 Binary land mask

Land masks for the target 25×25 km grids (one for NH and one for SH) are prepared from two higher resolution sources. The ocean mask is from the ESA SST CCI OSTIA L4 product (version 2.1), at 0.05×0.05° resolution (~6×3 km in the polar regions). The lakes mask is from the ESA Lakes CCI (ESACCI-LAKES_mask_v1.nc) at 0.0083×0.0083° (~1×0.5 km in the polar regions). The OSTIA L4 land mask is selected because of the long tradition of preparing SST+SIC analyses. We have also investigated other land masks (such as that of the ESA CCI Land Cover project, but it did not make a difference at these spatial scales.

These high-resolution binary masks are first regridded to the EASE2 25×25 km NH and SH grids to prepare “density_of_land”, “density_of_ocean”, and “density_of_lakes” variables (all 3 in [0,1] and the sum equals 1). The Caspian Sea, which is water body both in the SST and Lakes mask, is considered a lake for our Sea Ice CDRs.

Then, a binary “smask” (surface mask) variable is prepared from the three density_of_* fields. Variable “smask” takes on values “0: ocean, 1: ocean coastline, 2: land, 4: lake coastline, 5: lake” (see Figure 3). In the final product, sea-ice concentration values are provided for all grid cells with smask 0 (ocean) and 5 (lakes).

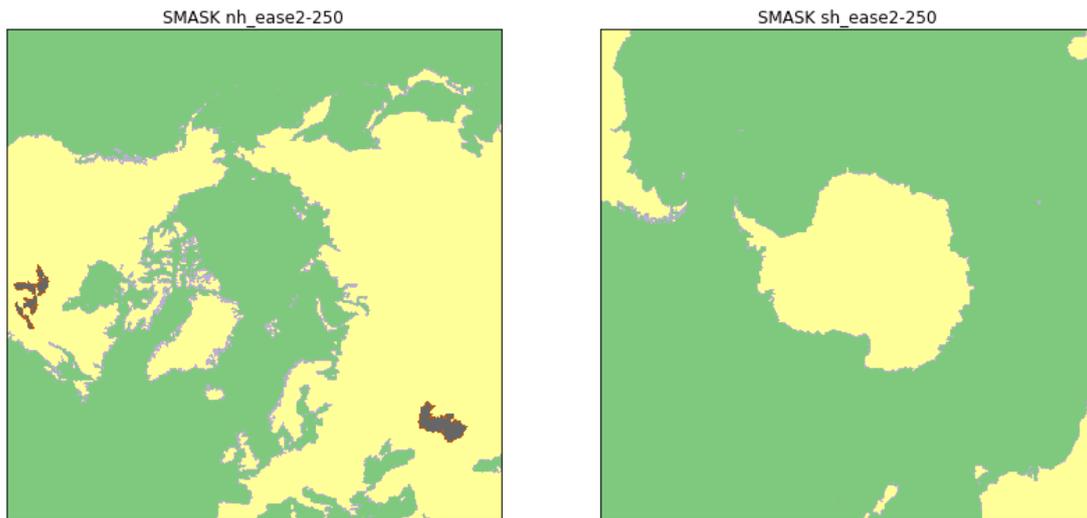


Figure 3: Variable “smask” used in the SIC CDRs on the EASE2 25.0 km grids (left: NH, right: SH). Ocean (smask=0): green, Ocean coastline (smask=1): light grey, Land (smask=2): yellow, Lake coastline (smask=4): red, Lake (smask=5): dark grey.

The binary smask is tuned to closely match that of the NSIDC SIC CDR (the NSIDC “SSM/I” 25 km Polar Stereographic mask) in the regions they have in common. On average, this corresponds to setting all 25×25 km grid cells with a fraction of land lower than 30 % to water (ocean or lake). There is no right or wrong binary land mask at such coarse resolution, and the decision to tune to the NSIDC SIC CDR land mask is to help an intercomparison of data records.

For the new “v3” CDRs, it was decided to only keep the largest lake systems (Caspian Sea and the US Great Lakes) and not consider smaller inland water (e.g. Great Bear Lake, Baikal, Ladoga). Experience from “v2” datasets indicates that our CDRs had limited value over smaller lakes due to the coarse resolution of the sensors.

2.8 Sea Ice Extent Climatology

We use a monthly varying maximum sea-ice extent climatology to filter out grossly erroneous sea-ice detection far from the polar regions, and along the coast lines at

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

mid to high latitudes (in section 7.3). A monthly varying climatology is required because of the large seasonal variability of the polar sea ice extent.

The monthly varying maximum sea-ice extent climatology implemented in the NSIDC SIC CDR v3 CDR (Peng et al. 2013 , Meier et al., 2017) was used as a basis for our climatology. The NSIDC climatology is described in their C-ATBD and covers the years up-until 2007.

We then did some modifications to the climatologies, mainly manual editing of some single pixels, based on US National Ice Center, Canadian Ice Service, and Norwegian and Finnish Ice Service ice charts (e.g. along the coast of northern Norway, for some summer months in the vicinity of Nova Scotia and in the Baltic Sea and Gulf of Finland). The climatology of peripheral seas and large freshwater bodies (e.g. Bohai and Northern Yellow Seas, Great Lakes, Caspian Sea, and Sea of Azov) was also revisited.

Compared to “v2”, the main change in the “v3” climatology is a focus on the coastal regions during summer, especially the Baltic Sea. These coastal regions are very challenging for our coarse resolution CDRs OSI-450-a and OSI-430-a (to a lesser extent OSI-458) because of land spill-over. In the preparation phase for “v3”, we liaised with the Finnish Ice Service and adapted the regional climatology using their input.

The cleaned climatologies are then expanded with a buffer zone of 150 km in the NH and 250 km in the SH. This expansion is not applied in the Baltic Sea during summer months. The larger expansion in SH is to cope with the slight positive trends in the SH sea-ice extent (Parkinson, 2019). Finally, the “v2” SIC CDR and ICDR (covering 1979-2020) was used to check that the new climatology does not cut away true sea-ice.

2.9 Beyond SSMIS

OSI-430-a relies on SSMIS for the daily extension of the data record. However, the three SSMIS are aging, and long past their design lifetime. SSMIS F19 failed early, and F20 was decommissioned before launch. The AMSR2 is also aging and is too dissimilar to SSMIS to ensure climate consistency at a daily SIC level. There is thus a non-zero risk that the SSMIS extensions of OSI-430-a would stop before the EPS-SG MicroWave Imager (MWI) is operational and processed in our chains (launch date currently 2023).

At the OSI SAF, a contingency plan has been to look into the processing of China’s MicroWave Radiometer Imager (MWRI) on board the Feng-Yun 3 satellites. Preliminary results are encouraging (FY3D). The chains will be finalized and brought to operation in the event when additional SSMIS fail before EPS-SG MWI is ready. This contingency plan is not only for OSI-430-a, but for several other sea-ice products.

3 Overview of the SIC processing chain

Figure 4 gives an overview of the processing chain for the SIC CDRs. The red boxes are data (stored in data files) and the blue boxes are processing elements that apply algorithms to the data. The whole process is structured into four chains, at Level 1P (left), Level 2, Level 3, and Level 4 (right). The input Level 1 (L1) data files hold the fields observed by the satellite sensors at the top of the atmosphere, in satellite projection: the brightness temperatures (Tb) are structured in swath files. The Level 1 Preprocessing (L1P) prepares the L1B swath files for SIC processing. The Level 2 (L2) chain transforms these into the environmental variables of interest, but still on swath projection: the SIC, its associated uncertainties, and flags. The L2 chain holds an iteration (marked by the “2nd iteration” grey box) similar to the workflow in Tonboe et al. (2016) and stemming from the developments of Andersen et al. (2006). This iteration implements two key correction schemes: the atmospheric correction algorithm at low-concentration range (section 5.2) and a correction for systematic errors at high-concentration range (section 5.4). The Level 3 (L3) chain collects the L2 data files and produces daily composited fields of SIC, uncertainties, and flags on regularly spaced polar grids. These fields can and will typically exhibit data gaps, e.g. in case of missing satellite data. The Level 4 (L4) chain fills the gaps, applies extra corrections, and formats the data files that make the CDR.

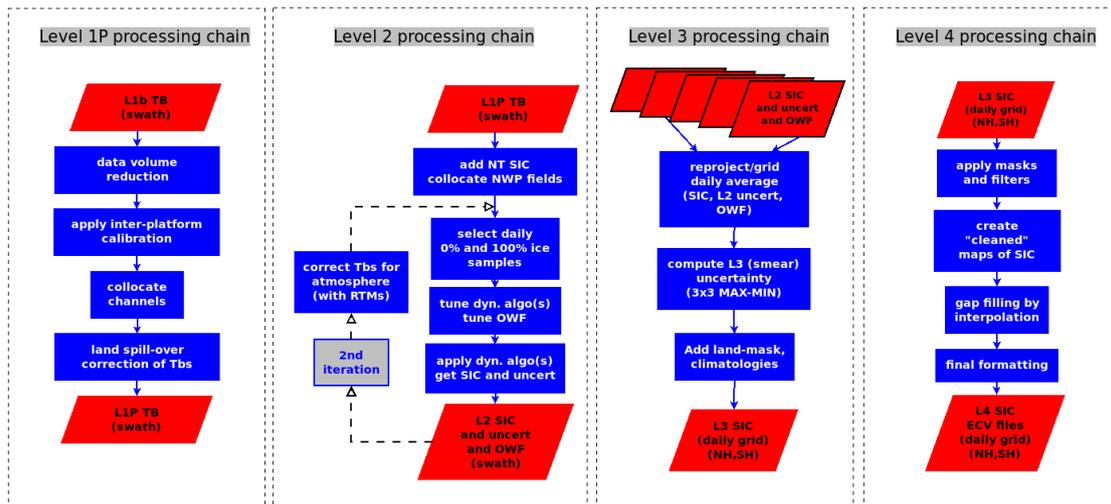


Figure 4: From left to right, the four main elements (Level 1P, Level 2, Level 3, and Level 4) in the sea-ice concentration (SIC) processing workflow. The red boxes depict data files, the blue boxes correspond to individual steps (a.k.a. algorithms) in the processing. The files that exit a processing chain (e.g. the “L2 SIC and uncert and OWF” at the bottom of the Level 2 processing chain) are the input for the next level of processing. Acronyms: NT is the Nasa Team algorithm, OWF is open-water filter, RTM is radiative transfer model, uncert stands for uncertainty.

The next four sections document the L1P, L2, L3, and L4 algorithms.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

4 Pre-processing on swath projection (L1P)

4.1 Data volume reduction

We focus our processing on the polar and sub-polar oceans. A first pre-processing step is thus to remove all instrument scans whose observations are all in the $[-35^{\circ}; +35^{\circ}]$ latitude band, as well as all scans well inside land and continents (200 km distance).

4.2 Apply inter-platform calibration

In the CM SAF Fundamental Climate Data Record (FCDR) of SMMR, SSM/I, and SSMIS data, the coefficients to inter-calibrate Tb data in the series of satellites are provided in the product files, not applied. We apply those coefficients as a processing step for OSI-450-a.

The brightness temperatures included in the SSMIS SDR data files accessed from NOAA CLASS cannot be used as-is in OSI-430-a. The reason is that the SDR brightness temperature algorithm (L1A->L1B) uses a set of scene-dependent coefficients that result in different calibration for pre-defined surface types (e.g. land, near coast, ice shelves/land ice, climatological ice, ice-free ocean, coast).

An overlap period of 12 months (year 2020) was used to tune scene- and sensor-dependent, re-calibration linear coefficients to transform the SDR Tbs to CMSAF-like Tbs up-front of the OSI-430-a processing. The coefficients and the quality of the re-calibration was found stable for all months and in both hemispheres for the overlap period, and thus applied for all later SSMIS data in OSI-430-a. The re-calibration of high-frequency (91 GHz) channels was less successful, but these are not used in OSI-450-a nor OSI-430-a.

In addition to this re-calibration, the SSMIS SDR files entering OSI-430-a have to go through a Quality Control (QC) step to detect and discard gross errors (e.g. bad scans).

For OSI-458, AMSR-E and AMSR2 data are used as-is, without re-calibration.

4.3 Collocate channels

Tb data are arranged as (scanline, scanpos) two-dimensional arrays in the L1B swath files. Due to the conical scan mechanism of the instrument, this indexing does not guarantee that observations from each of the different frequency channels at a given index in the 2D swath array are closest in space once projected onto the Earth's surface. This is especially true for the high-frequency channels (near-90 GHz) and at the edges of the swath.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

Since our SIC algorithms combine Tb data from different frequency channels, it is key that they are first collocated in terms of Earth location.

We perform this collocation with a nearest-neighbour method: we do not apply convolution/deconvolution methods to estimate Tbs at common locations and resolutions; we use the original Tb at their original location and resolution.

4.4 Land spill-over correction

Due to the coarse resolution of the sensors used, especially SMMR, SSM/I, and SSMIS, the Tb data are influenced by land emissivity several tens of kilometres offshore from the coastlines. The microwave emissivity of land is comparable to sea-ice emissivity and much higher than that of ocean water. This means that sea-ice concentration will be consistently overestimated in coastal regions with less than 100% sea ice concentration.

We adopt and adapt the swath-based correction algorithms of Maass and Kaleschke (2010). The basic principle is that a fine-resolution land mask is used together with the antenna viewing geometry to simulate (and correct for) the contribution of land emissivity to the observed Tb. The algorithm of Maass and Kaleschke (2010) was adopted with some modification and tuning, including (a) the computation of the fraction of land in each FoV in the view geometry of the antenna (not after projection to a map), and (b) the approximation of the antenna pattern functions as Gaussian (normal distribution) shapes indexed on the aperture angle from the central view direction, instead of distance on a projection plane. At the end of this step, Tb in FoV that overlap land and ocean (coastal FoVs) are corrected for the contribution by land, and can enter the Level 2 SIC algorithms.

The correction algorithm is described in details in Maas and Kaleschke (2010). The basic principle is that, for each FoV in the swath file, one separates Tb into two components, T_{sea} and T_{land} , as $T_b = (1 - \alpha)T_{sea} + \alpha T_{land}$ where α is a convolution of the antenna gain pattern and the land fraction in the footprint. Local T_{land} is calculated by using the neighbouring land-only FoVs ($\alpha > 0.95$). Then T_{sea} is computed from the formula above and can enter the sea-ice concentration algorithms.

For the land spill-over correction to work, the fraction of land within the FoV α must be accurate. This requires both an accurate land mask and a good representation of the antenna gain pattern. Both are revised for “v3”. First, we use the combined 0.05° landmasks of ESA CCI SST and Lakes (section 2.7) as the basis for our α . Second, we conducted a specific activity to tune the FoV aperture of the channels used in the SIC algorithms. The tuning of the aperture (expressed in radians from the view point of the satellite along its orbit) involve a satellite observation simulator (developed in-house in several projects) and a minimization of the mismatch between the simulated and actual Tb across land-ocean boundaries. Figure 5 illustrates the simulation of α with the satellite simulator, and the

comparison to measured Tb in the 19 GHz V-polarization channel of SSMIS F18 for three selected apertures.

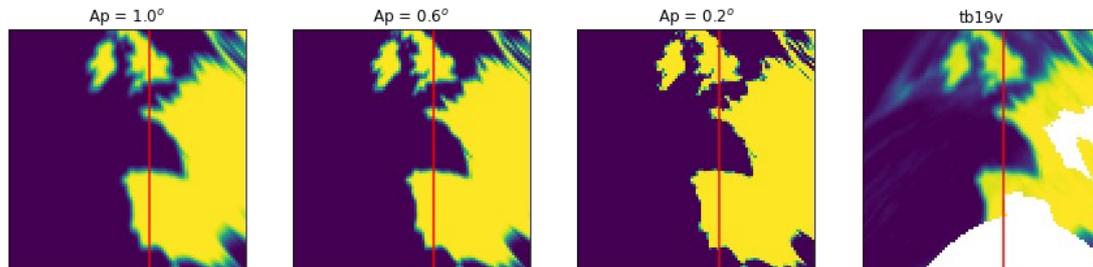


Figure 5: Simulated fraction of land α in each field of view for three apertures A_p : (1.0° , 0.6° , and 0.2°) using a satellite observation simulator. Far right: the brightness temperature field actually recorded by the 19 GHz V-polarization channel of the SSMIS F18 mission.

5 Geophysical processing in swath projection (L2)

5.1 A hybrid, self-tuning, self-optimizing sea-ice concentration algorithm

A new sea-ice concentration algorithm formulation was developed during the ESA CCI Sea Ice projects (2008-2017) and is used for the OSI SAF and CCI SIC CDRs. It is an evolution of the algorithms used in Tonboe et al. (2016). In this section, we describe both how the algorithm is trained to Tb training data sets, and how it is then applied to actual Tb measurements recorded by satellite sensors.

5.1.1 Merging equation for the hybrid algorithm

We call the SIC algorithm a hybrid algorithm because it combines two other SIC algorithms: one that is tuned to perform better over open-water and low-concentration conditions (named B_{OW} for “best open water”), and one that is tuned to perform better over closed-ice and high-concentration conditions (named B_{CI} for “best closed ice”). The combination equation is quite simply a linear weighted average of B_{OW} and B_{CI} results, where w_{ow} is the open-water weight and SIC is expressed as sea-ice fraction [0; 1]:

$$SIC_{hybrid}(\vec{T}) = w_{ow} * B_{OW}(\vec{T}) + (1 - w_{ow}) * B_{CI}(\vec{T})$$

$$\begin{cases} w_{ow} = 1 & \text{for } B_{OW} < 0.7 \\ w_{ow} = 0 & \text{for } B_{OW} > 0.9 \\ w_{ow} = \frac{B_{OW} - 0.7}{0.2} & \text{for } B_{OW} \in [0.7; 0.9] \end{cases}$$

where $\vec{T} = (T_{19v}, T_{37v}, T_{37h})$ is a triplet of brightness temperature. In the equations above and the rest of this document, the notations B_{OW} and B_{CI} thus correspond to SIC algorithms that are optimized for lowest retrieval uncertainty over Open Water and Closed Ice conditions, respectively.

The linear weighting equation with transition between 0.7 and 0.9 is the same as for the “v2” SIC CDRs, and stems from the R&D during the first phases of the CCI project. This form of transition was found better than the transition implemented in the “v1” SIC CDR (Tonboe et al., 2016) and better than the abrupt transition implemented in the Bootstrap algorithm between its Frequency and Polarization modes (Comiso, 1986).

5.1.2 Applying the SIC algorithms

Be \vec{T} a triplet of brightness temperatures measured from the satellite. The triplet holds the Ku V-pol channel (19v), Ka V-pol channel (37v) and Ka H-pol channel (37h). The SIC (noted C) is computed as:

$$C(\vec{T}) = \frac{\vec{v} \cdot (\vec{T} - \langle \vec{T}^W \rangle)}{\vec{v} \cdot (\langle \vec{T}^I \rangle - \langle \vec{T}^W \rangle)}$$

where $\langle \vec{T}^W \rangle$ (resp $\langle \vec{T}^I \rangle$) is the tie-point for 0% SIC (resp 100% SIC) and $\vec{v} = (v_{19v}, v_{37v}, v_{37h})$ is the vector of the algorithm coefficients.

The equation above is a generic form for computing SIC from algorithms B_{OW} and B_{CI} , once the tie-points and vectors \vec{v}_{OW} and \vec{v}_{CI} are known (see next sections).

5.1.3 Tuning of the algorithm at 0% and 100% SIC

The tuning of B_{OW} and B_{CI} is an evolution from the tuning of the Bristol (BRI, Smith and Barrett, 1994) algorithm. The Bristol algorithm is tuned to a training data set of purely 100% SIC and purely 0% SIC Tb samples by selecting a unit vector $v_{Bristol}$ that is orthogonal to the 100% SIC line (itself sustained by vector u). Vector $v_{Bristol}$ is selected so that the data plane ($u, v_{Bristol}$) holds the open water tie-point (the average point of all 0% SIC samples). This is illustrated in Figure 6.

When tuning B_{OW} (resp B_{CI}) we rather find the unit vector v_{OW} (resp v_{CI}) that minimizes the spread of the retrieved SIC at 0% SIC (resp 100% SIC) when the SIC algorithm is applied on the training data sets. The vectors v_{OW} (resp v_{CI}) are generally not aligned with $v_{Bristol}$ (Figure 6 left panel).

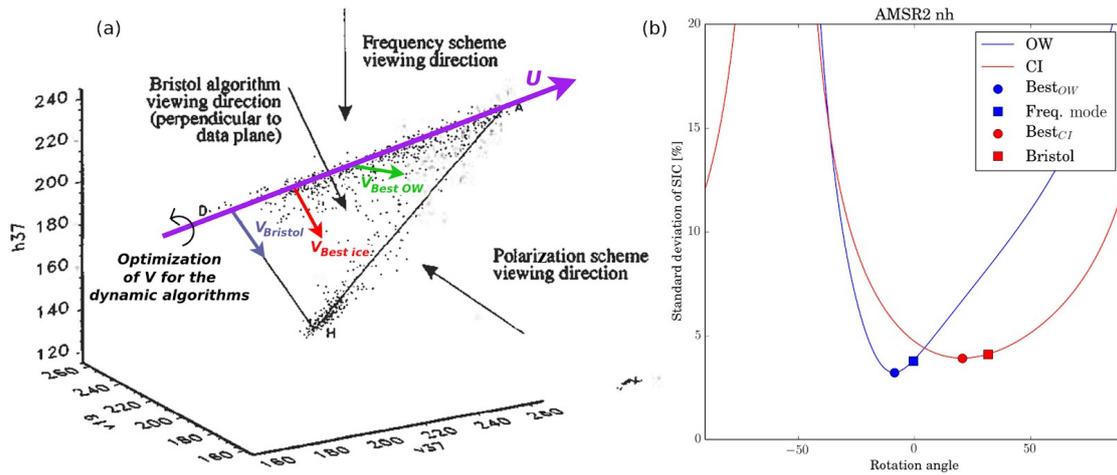


Figure 6: (a) Three-dimensional diagram of open-water (H) and closed-ice (ice line between D and A) brightness temperatures in a $19V$, $37V$, $37H$ space (black dots). The original figure is from Smith (1996). The direction U (purple) is shown, and vectors $v_{Bristol}$ (blue), $v_{Best-ice}$ (red), and $v_{Best-OW}$ (green) are added, as well as an illustration of the optimization of the direction of V for the dynamic (self-optimizing) algorithms. (b) Evolution of the SIC algorithm accuracy for open-water (blue) and closed-ice (red) training samples as a function of the rotation angle θ in the range $[-90^\circ; 90^\circ]$. Square symbols are used for the BFM (Bootstrap frequency mode, Comiso, 1986) and BRI (Bristol, Smith and Barrett, 1994) algorithms. Disc symbols locate the new, self-optimizing algorithms.

The optimization of vectors v_{ow} (resp v_{ci}) is implemented as a brute-force optimization over a 1-dimensional axis representative of the rotation angle vector \vec{v} has wrt. a reference direction. On the right panel of Figure 6, this optimization process is illustrated, with choosing 0° rotation angle as the Bootstrap Frequency Mode (BFM, Comiso, 1986) algorithm, in a case using AMSR2 data from the Northern Hemisphere. The solid lines plot the variation in the accuracy (measured as standard deviation of SIC, on the y axis) of the SIC algorithms defined by the rotation angle (x axis) against the 0% SIC (OW, blue) and 100% SIC (CI, red) training T_b data. The minimum of the blue (resp red) curves is reached at angle θ_{ow} (resp θ_{ci}) that defines v_{ow} (resp v_{ci}).

At the end of the tuning (this is repeated on a daily basis, see section 5.1.4), the tie-points $\langle \vec{T}^W \rangle$ and $\langle \vec{T}^I \rangle$ are stored, as well as vectors v_{ow} and v_{ci} . This is all that is needed to later apply the B_{ow} and B_{ci} algorithms on satellite swath data, and combine the results into the hybrid SIC value (section 5.1.1).

5.1.4 Dynamical selection of the SIC training data

As described in the previous section, tuning the algorithms requires two sets of training data: one from OW areas (SIC = 0 %) and one from areas we assume have fully CI cover (SIC = 100 %). The training of the algorithms is performed separately for each instrument and for each hemisphere. In addition, the training is updated for every day of the data record and is based on a $[-7; +7]$ days sliding window of T_b samples. The sliding window is relatively short (15 days) so that tie points react more rapidly to seasonal cycles, e.g. onset of melting.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

The dynamic training of the algorithms allows us to (a) adapt to interseasonal and interannual variations of the sea-ice and open-water emissivity, (b) cope with different calibration of different instruments in a series, or between different FCDRs, (c) cope with slightly different frequencies between different instruments (e.g. SMMR, SSM/I, and AMSR-E all have a different frequency around 19 and 37 GHz), (d) mitigate sensor drift (if not already mitigated in the FCDR), (e) compensate for trends potentially arising from the use of NWP reanalysed data to correct the T_b .

The CI training sample is based on the results of the NASA Team (NT) algorithm (Cavalieri et al., 1984): locations for which the NT value is greater than 95 % are used as a representation of 100 % ice (Kwok, 2002). Earlier investigations, e.g. during the ESA CCI Sea Ice projects, confirmed that NT was an acceptable choice for the purpose of selecting closed-ice samples. To ensure temporal consistency between the SMMR and later instruments, the closed-ice samples for NH are only used for algorithm tuning if their latitude is less than 84° N, which is the limit of the SMMR polar observation hole. The NT tie-points were revised in the context of the CCI+ Sea Ice project, and are one of the R&D input from CCI+ to the OSI SAF v3 CDRs.

The OW tie-point samples are selected in a 150 km wide belt that encompasses the extent of the sea-ice cover on a daily basis. In practice, NT sea-ice concentration are computed and gridded on a daily basis, and a 150 km wide belt starting 150 km away from the ice edge (defined as NT SIC = 15%) is defined. All observations from the swaths falling into this belt are selected for computing the OW tie-point. This ensures that the OW tie-point samples are representative of ocean and atmosphere conditions prevailing in the vicinity of the ice edge. This is new in “v3” since in “v2” the open water tie-point belt was around a monthly maximum ice extent climatology, thus further away from the actual sea-ice edge in recent years (in the Northern Hemisphere).

The [-7; +7 days] sliding window is used for the CDRs OSI-450-a and OSI-458 and for the main stream of the ICDR OSI-430-b. This window is the very reason why the latency of the ICDR is 16 days. This is because two sliding windows are required, one for the non-corrected tiepoints, and one for the corrected tie-points, see section 5.2. In order to achieve a latency of 2 days, the fast-track stream of OSI-430-a will instead use a [-7; -1 days] sliding window. This is the only difference between the main and fasttrack streams of the ICDR.

5.2 Radiative transfer modelling for correcting atmospheric influence on T_b

As described in Andersen et al. (2006) and confirmed in Ivanova et al. (2015), the accuracy of retrieved sea-ice concentration can be greatly improved when the brightness temperatures are corrected for atmospheric contribution by using a radiative transfer model (RTM) combined with surface and atmosphere fields from NWP reanalyses. For climate study purposes, the correction using NWP data is only

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

possible in combination with a dynamical tuning of the tie points, so that trends from the NWP fields are not introduced into the SIC data set. The correction scheme implemented in the CDRs is based on a double-difference scheme.

The scheme evaluates the correction offsets δT_b (one per channel), the difference between two runs of the RTM: $T_{b_{nwp}}$ uses estimates from NWP fields (in our case ERA5), while $T_{b_{ref}}$ uses a reference atmospheric state with the same air temperature as $T_{b_{nwp}}$, but zero wind, zero water vapour, and zero cloud liquid water. δT_b is thus an estimate of the atmospheric contribution at the time and location of the observation.

$$\begin{aligned}
 TB_{nwp} &= F(W_{nwp}, V_{nwp}, L_{nwp}; T_S, SIC_{ucorr}, \theta_0) \\
 TB_{ref} &= F(0, 0, 0; T_S, SIC_{ucorr}, \theta_{instr}) \\
 \delta_{TB} &= TB_{nwp} - TB_{ref} \\
 TB_{corr} &= TB_{obs} - \delta_{TB}
 \end{aligned}$$

For $T_{b_{nwp}}$, the RTM function F simulates the brightness temperature emitted at view angle θ_0 by a partially ice-covered scene with sea-ice concentration SIC_{ucorr} , and with surface and atmospheric states described by W_{nwp} (10 m wind speed, $m \cdot s^{-1}$), V_{nwp} (total column water vapour, mm), L_{nwp} (total column liquid water content, mm), and T_S (2 m air temperature). θ_{instr} is the nominal incidence angle of the instrument series (varies between satellites). The double-difference scheme is thus both a correction for the atmosphere influence on the T_b (as predicted by the NWP fields) and a correction to a nominal incidence angle. The typical values of δT_b range from about 10 K over open water to few tenths of a kelvin over consolidated sea-ice. The first-guess SIC (SIC_{ucorr}) is from the first pass through the SIC algorithm (before the 2nd iteration, see Figure 4). The corrected brightness temperature ($T_{b_{corr}}$) is then used in the 2nd pass of the SIC algorithm (2nd iteration).

For “v3”, we find that ERA5 has some skills at L_{nwp} , and that including L_{nwp} in the RTM-correction step has a positive impact on the uncertainty of the retrieved SICs at low SIC. The liquid water content from ERA-Interim (used in the CCI and OSI SAF “v2” CDRs) were not accurate enough (Lu et al., 2018). The T_b were thus not corrected for L ($L=0$ in both $T_{b_{nwp}}$ and $T_{b_{ref}}$) in “v2”.

In principle, any RTM will do for implementing F . Here we use the Remote Sensing Systems (RSSs) RTM, for which the tuning to different instruments is documented in Wentz (1983) for SMMR, Wentz (1997) for SSM/I and SSMIS, and Wentz and Meissner (2000) for AMSR-E and AMSR2.

Figure 7 and Figure 8 illustrate the positive impact of the RTM-based atmospheric correction in both hemispheres over the Open Water conditions. The correction with ERA5 data allows a reduction of the retrieval uncertainty from 3-5 % to 2 %. The smallest improvement is for AMSR-E and AMSR2 (OSI-458) because the 18.7 GHz frequency is further away from the water vapour absorption line (~ 22 GHz) than the 19.35 GHz of SSM/I an SSMIS. These plots are made from prototype “v3” SIC CDRs and demonstrate the expected retrieval uncertainties of the “v3” CDRs over open water and low concentration conditions.

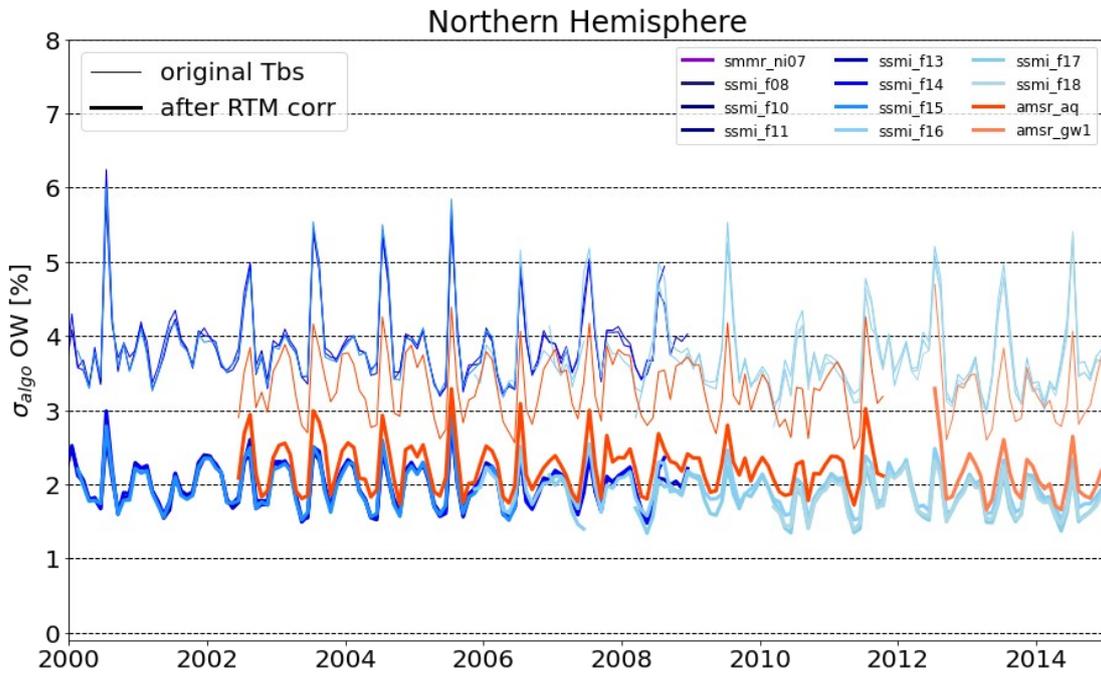


Figure 7: Timeseries of the SIC retrieval uncertainty σ_{algo} for OW conditions in the Northern Hemisphere over the period 2000-2015. Thin lines are for the SIC retrieval noise without atmospheric correction, and thick lines are with the atmospheric correction. Colors indicate the satellite missions (not all available missions were active during the period, see Figure 2).

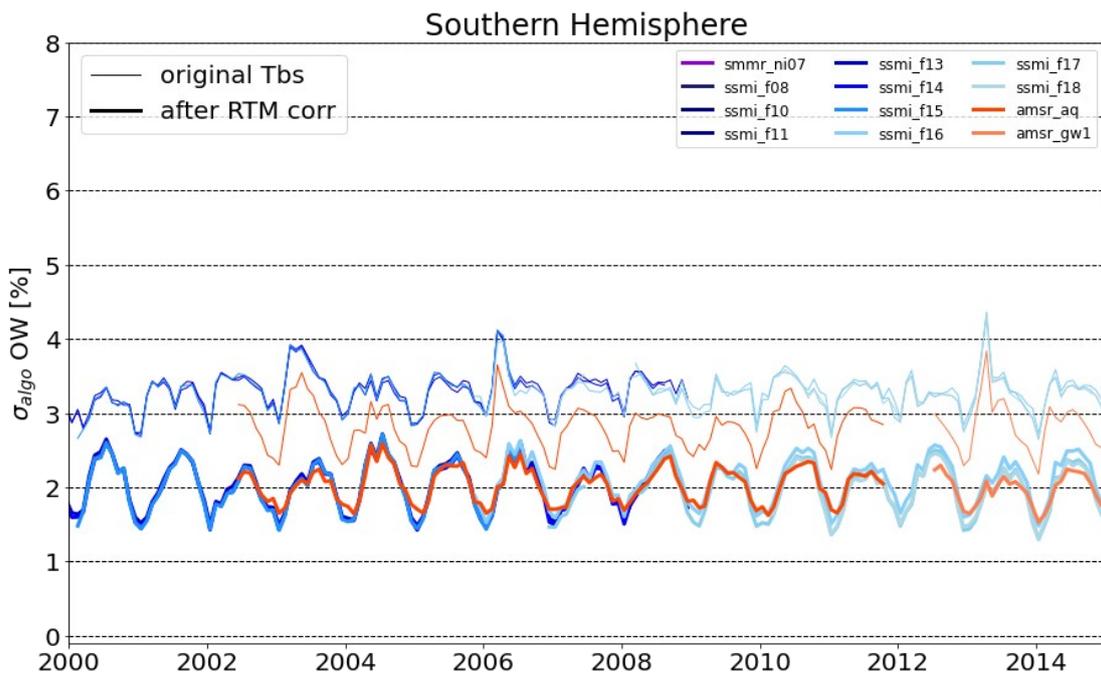


Figure 8: Same as Figure 7 but for the Southern Hemisphere.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

5.3 Open-water filtering

The weather filters (WFs) of Cavalieri et al. (1992) have been used in basically all available SIC CDRs (except the earlier EUMETSAT OSI SAF data sets, Tonboe et al., 2016). WFs are algorithms that combine Tb channels to detect when rather large retrieved SIC values (sometimes up to 50 % SIC) are in fact noise due to atmospheric influence (mainly wind, water vapour, cloud liquid water effects) and should be reported as open water (SIC = 0 %).

The WF by Cavalieri et al. (1992) detects (and consequently sets SIC to 0 % SIC) all observations with either $GR3719v > 0.050$ and/or $GR2219v > 0.045$. The GR notation stands for *gradient ratio* and is computed, e.g. as $GR3719v = (TB37v - TB19v) / (TB37v + TB19v)$. $T_{3719v} = 0.050$ and $T_{2219v} = 0.045$ are two thresholds that must be tuned in order to reach consistent behaviour across satellites.

Indeed, while WFs are effective at removing false sea ice in open-water regions, they are also “greedy”: they will always remove (detect as open water) some amount of low-concentration (and/or thin) sea ice, especially along the ice edge (Ivanova et al., 2015). The greediness of the filters is controlled by the thresholds T_{3719v} and T_{2219v} .

We note upfront that the name “weather filter” can be misleading, as the non-expert could understand that its effect is to filter-out weather effects (false sea ice) from calm open-water and low-ice-concentration conditions. However, this is not how the WF works, as it does remove true sea ice even in calm weather conditions. For this reason, we rather refer to these filters as an Open Water Filter (OWF).

For “v2” of the CDRs, following Lu et al. (2018), we used an OWF computed from Tb that has been corrected for atmospheric influence and features only the test for $GR3719v$ (because the atmospheric correction effectively handles the water vapour contamination that the $GR2219v$ controls). In addition, we dynamically tuned T_{3719v} on a daily basis using the dynamic tie-points samples. See also section 3.4.2 in Lavergne et al. (2019). The dynamic tuning of T_{3719v} was effective at ensuring temporal consistency across satellite missions (e.g. Fig. 1 and Fig 2. in Kern et al. 2019).

For “v3” of the CDRs, we continue the dynamic tuning of the OWF, but move away from $GR3719v$ (computed in the 2D (37V, 19V) space) and instead use a new quantity, noted d_{OWF} .

d_{OWF} is a normalized version of another scalar, noted d , the Distance Along the ice Line (DAL). The DAL was already introduced in the v2 CDRs but not for the purpose of the Open Water Filter. We first remind how d is computed, then how d_{OWF} is derived from d .

Once the unit vector u is known (section 5.1.3 : u is the unit vector sustaining the 100% sea-ice line), a scalar quantity d can be computed for each vector T (e.g. triplet) of Tb. We call d the “Distance Along the Line” (DAL).

$$d = u \cdot T$$

The *DAL* is then normalized using SIC, a tie-point for First-Year Ice <FYI> and a “low-weather” tie-point <LW>:

$$d_{OWF} = d - ((1 - SIC) \times u \cdot \langle LW \rangle + SIC \times u \cdot \langle FYI \rangle)$$

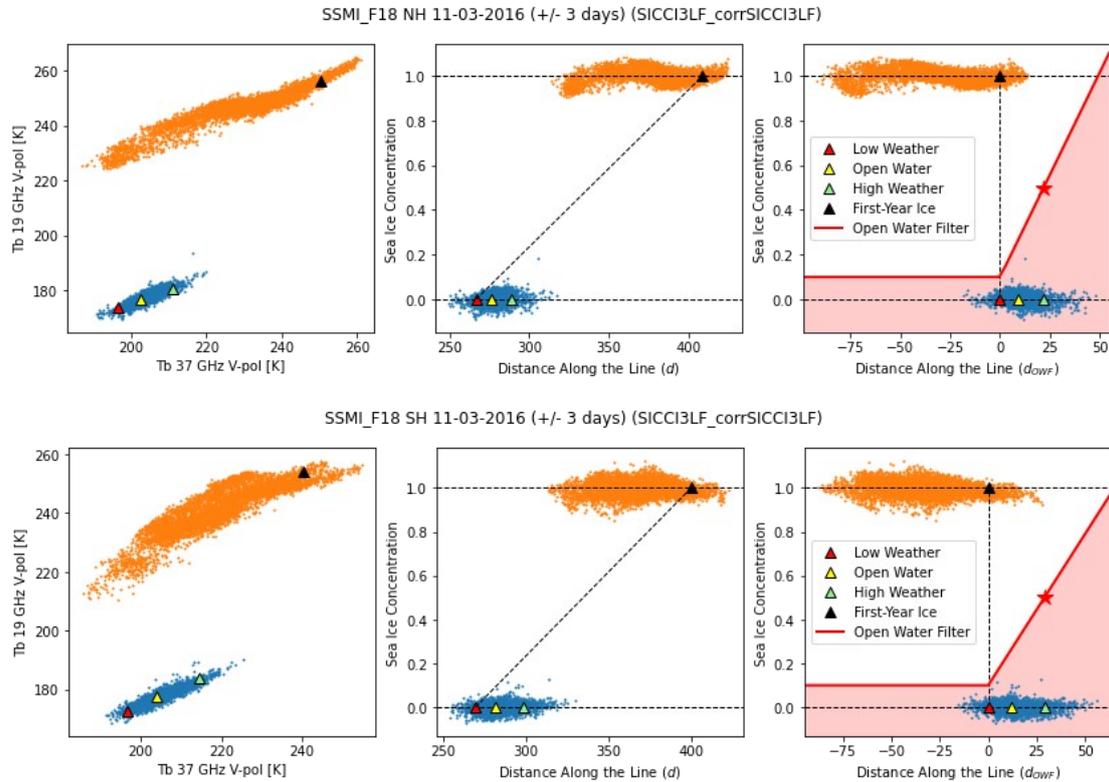


Figure 9: Example use of d and d_{OWF} for Open Water Filtering. Top row for the Arctic, bottom row for the Southern Ocean. Panels from left to right show the T_b samples (small dots) of 0% SIC (blue) and 100% SIC (orange) in a $(x:37v, y:19v)$ space (left panel), $(x:d, y:SIC)$ space (middle panel), and $(x:d_{OWF}, y:SIC)$ space (right panel). Tie-points (triangles) for Open Water (yellow), Low Weather (red), High Weather (green), and First-Year Ice (black) are also reported, as well as the domain detected as Open Water by the Open Water Filter (red area in right panel).

Figure 9 illustrates how the clusters of 0% and 100% SIC conditions are distributed in three 2D domain relevant for the Open Water Filter. The left-most panels are in the $(x:37v, y:19v)$ T_b space that sustained the OWF for the “v2” CDRs (through the GR3719v test). The middle panels show the same data in the $(x:d, y:SIC)$ diagram, that illustrates the intermediate step in the coordinate transform from the left-most to the right-most panels. In the right-most panels, the data points are plotted in the $(x:d_{OWF}, y:SIC)$ space that sustains the OWF in the “v3” CDRs. The $(x:d_{OWF}, y:SIC)$ space is built so that the intermediate SICs conditions will mostly fall onto and to the left of the (LW, FYI) line (vertical at $x=0$ in this 2D space).

In this space, the OWF is implemented with two binary tests (see red area in the right-most panels of Figure 9).

- Test1 = (SIC <= 0.1);

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

- $\text{Test2} = (\text{SIC} \leq (0.1 + (0.5 - 0.1) * d_{\text{OWF}} / d_{\text{HW}}));$

In Test2, d_{HW} is a point in the $(d_{\text{OWF}}, \text{SIC})$ space corresponding to “heavy weather contamination point” at the 95% percentile of the d_{OWF} values. Test2 detect as water all the T_b samples that fall below the line going through points $(x:0,y:0.1)$ and $(x:d_{\text{HW}},y:0.5)$ (the latter marked with a red star on Figure 9).

T_b observations that fulfill any of the two conditions above (Test1 OR Test2) are flagged ($\text{OWF}=\text{True}$) as “probably open water”, and their SIC is set to 0% in the final product. Figure 10 illustrates the concept and effect of the “v3” OWF in the $(d_{\text{OWF}}, \text{SIC})$ space.

In Figure 10, panel (c) shows the resulting OWF field for 11 March 2016. SICs in the red grid cells will be set to 0% SIC. The white grid cells at low latitudes (outskirt of the map) are not detected by the filter, possibly because the main noise source is Water Vapour. This is however not a problem since they will be set to 0% SIC by the maximum sea-ice extent climatology (section 2.8). We also note that the illustration in Figure 10 uses un-corrected T_b while the OWF in the processing chain uses RTM-corrected T_b s (see previous section) which translates into a further reduction of the amount of white grid cells outside the polar regions. The diagrams in Figure 9 were prepared using RTM-corrected T_b and are thus closed to the conditions in the CDR.

The OWFs are computed on swath projection (at L2), and not on daily averaged maps of brightness temperatures as is the case in many other climate data products (e.g. NSIDC CDR).

2016-03-11 NH ssmi_f18

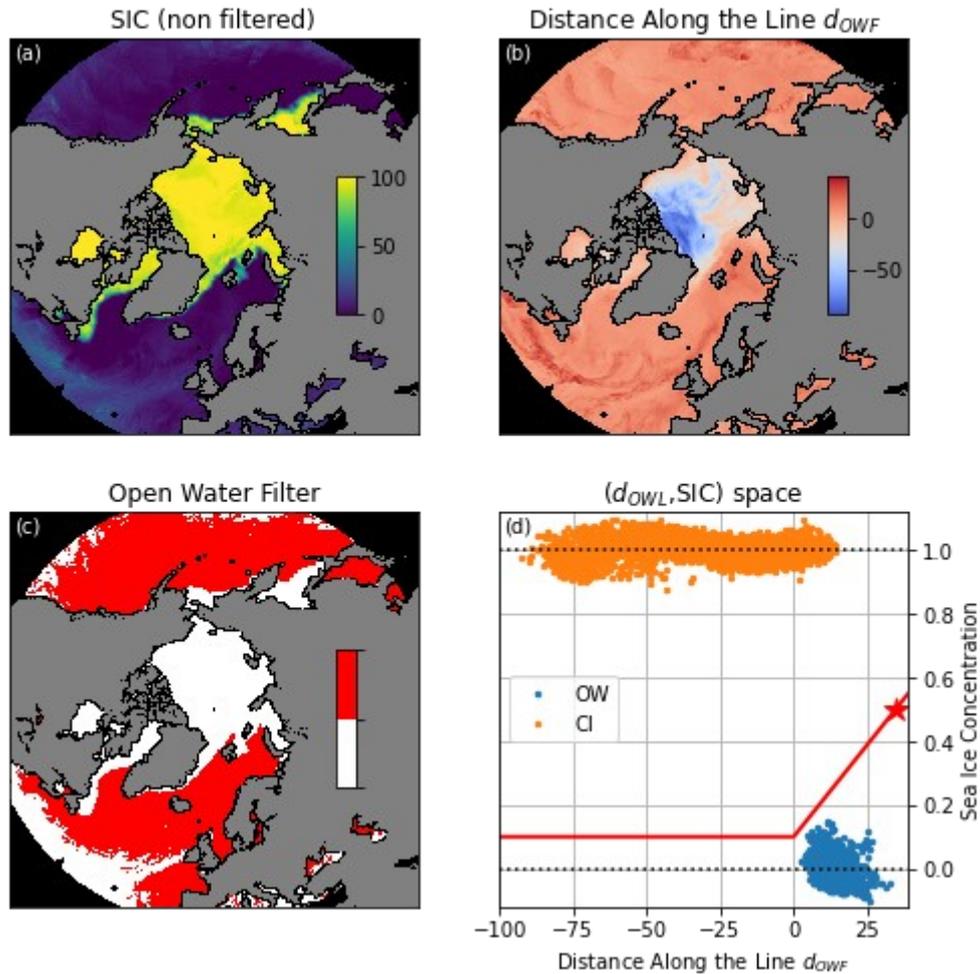


Figure 10: Illustration of the "v3" OWF concept with data from the SSMIS F18 mission, on 11 March 2016, in the Northern Hemisphere. (a) sea-ice concentration (non filtered), (b) the Distance Along the Line d_{OWF} , (c) the resulting Open Water Filter (red: "probably water") and (d) the filter (red line) visualized in the (d_{OWF}, SIC) space.

5.4 Reducing systematic errors at high-concentration.

By construction all SIC algorithms (Bootstrap, Bristol, our dynamic algorithms) consider that the SIC is exactly 100% when the input T_b falls on the consolidated, 100% SIC, ice line. The concept of an ice line has sustained the development of SIC algorithms for decades, since it allows algorithms to return SICs close to 100% for all consolidated ice conditions, whatever the type of sea ice (multi-year ice, first-year ice, mixture of types).

In Lavergne et al. (2019) we introduced a change from the concept of a sea-ice line to that of a sea-ice *curve*. The ice curve is also used in version 3 of the SIC CDRs, and relies on the Distance Along the Line (DAL) scalar d (see section 5.3).

Analysis of the spread of consolidated ice samples along the ice line reveals that systematic deviations occur, and that they are somewhat stable in time. These systematic deviations draw a sea-ice *curve*.

These deviations are computed in a coordinate system in which abscissae are $x: d=u.T$ (the distance along the line, DAL), and the ordinate is $B_{ci}(T)$ (the result of the best-ice SIC algorithm for a given T_b triplet). See Figure 9 middle panels (especially top-row, Arctic) for an illustration with SSMIS F18 data or the sea-ice curve: the consolidated ice tie-points (orange dots) draw a curve that varies around the ice line ($y=1$).

The proposed correction scheme moves the concept of an ice line to that of an ice curve, to follow the $B_{ci}(T)$ samples along the *DAL* axis more closely. An ice curve L is tabulated for each day in the record by binning the $B_{ci}(T)$ values by their $d=u.T$ values.

The correction algorithm for a given T_b observation triplet T goes then by:

- Compute $C = B_{ci}(T)$;
- Compute $d = u.T$
- Evaluate $L(d)$: value of L at index d (by 1D interpolation from tabulated values).
- The corrected SIC is $C / L(d)$
- Use the corrected SIC in the hybrid SIC formula (section 5.1.1) in place of B_{ci} .

SSMI_F18 NH 11-03-2016 (+/- 3 days) (SICCI3LF_corrSICCI3LF)

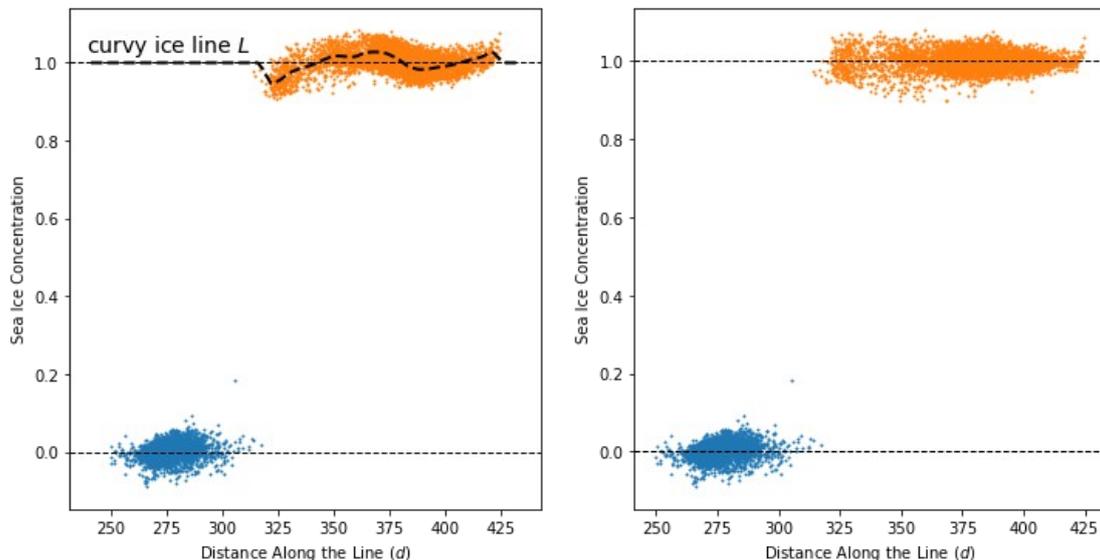


Figure 11: Illustration of the "curvy ice line" correction step: left panel: without correction, right panel: with correction. The dashed back curve in the left-most panel is used in the correction. Figure 11 illustrates the impact of this correction step: the left-most panel shows how the 100% SIC samples (orange dots) cluster along a curve, that is tabulated along

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

the $x:d$ axis (black dashed line). After the correction step, SICs align with the $y=1$ line, demonstrating the effectiveness of the correction step.

This consolidated ice curve defines the SIC 100 % isoline only during the 2nd iteration of the L2 chain, and has most effect on near-100% SIC regions. In the first iteration (before atmospheric correction), a “classic” (straight) ice line is used (see Figure 4).

5.5 Algorithm and tie-points uncertainties

Uncertainty estimates are needed when the SIC data are compared to other data sets or when they are assimilated into numerical models. The mean accuracy of some of the more common algorithms, used to compute ice concentration from SSM/I data, such as NASA Team and Bootstrap are reported to be 4-6 % in winter (Andersen et al., 2007; Ivanova et al. 2015) but the actual value varies with instrument, region, ice condition, etc. Time-varying maps of uncertainties are needed.

We express the total uncertainty as the sum of two contributions:

$$\sigma_{tot}^2 = \sigma_{algo}^2 + \sigma_{smear}^2$$

where σ_{algo} is the geo-physical uncertainty of the concentration algorithm and σ_{smear} is the uncertainty due to mismatch of the footprints at the two microwave frequencies (19 and 37 GHz) entering the SIC algorithm, and to the resampling to a grid where the sensor footprint covers more than one cell.

The term σ_{algo} is estimated at Level-2, while σ_{smear} is computed at Level-3. We first introduce how σ_{algo} is computed for a SIC algorithm, then how it is combined with σ_{smear} for producing the SIC uncertainty of an hybrid SIC algorithm.

Both the water surface and ice surface emissivity variabilities result in SIC uncertainties. Emission and scattering in the atmosphere also affect the T_b and the computed ice concentrations. Different algorithms have different sensitivities to these surface and atmospheric parameters (Andersen et al., 2007; Ivanova et al., 2015).

The algorithm uncertainty is the weighted average of the algorithm uncertainty obtained at 0% SIC and 100% SIC. Thus:

$$\sigma_{algo}^2 = (1 - C^2) \times \sigma_{CI}^2 + C^2 \times \sigma_{OW}^2$$

where σ_{ow} (resp σ_{ci}) is the uncertainty (1 standard deviation) obtained when applying the SIC algorithm on the 0% SIC (resp 100% SIC) training data samples.

The formula above is applied to both the B_{ow} and B_{ci} algorithms separately, and two uncertainty values are obtained: σ_{BOW} and σ_{BCL} .

These two uncertainty values are then combined to obtain the algorithm uncertainty of the hybrid algorithm: it is computed as a linear combination of the variances σ_{BOW}^2 and σ_{BCL}^2 . The same linear weights as used in the hybrid SIC algorithm (section 5.1.1) are used for mixing σ_{BOW}^2 and σ_{BCL}^2 , and the resulting standard deviation σ_{algo} is stored in the L2 product file.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

6 Level 3 algorithms

The Level 3 step contains the gridding and averaging of the swath data to daily fields, and the calculation of smearing uncertainties. Filters (e.g. OWF) are prepared as well, but not applied at this stage.

6.1 Gridding and daily averaging

The gridding and daily averaging loads all Level-2 observations within 24 hours, centered on 12:00 UTC, and grid these to the final output grids. There are two such grids, one for the northern and one for the southern hemispheres. Both have a grid spacing of 25 km.

The gridding is based on a KD-Tree search based on the distance between the cartesian coordinates (X_g, Y_g, Z_g) of the centers of the target grid cells and the cartesian coordinates (X_s, Y_s, Z_s) of the center of the satellite's FoVs. The search for N neighbours is constrained to a radius of influence of 12.5 km around each grid cell. Once the N closest FoV neighbours are known for each grid cell, they are combined in a daily averaged value, all with equal weight.

During the gridding, FoVs that overlap with land are discarded, to reduce the impact of land spill-over correction. In practice, we discard all FoVs for which the land area within the FoV of the 19 GHz frequency (α computed as in section 4.4) is above 0.1. The FoVs with α in $[0,0.1)$ will have limited contamination by land spill-over after having undergone the step described in section 4.4.

A gridded field is made for all the variables that might be of interest: the SIC estimates (both based on corrected and uncorrected Tbs), the algorithm uncertainties (combined as variances, not standard deviations), OWFs, etc.

For OSI-450-a and OSI-430-a, all SSM/I and SSMIS missions overlapping at any given time (Figure 2) are gridded together for each day of the data record. This is different from the strategy adopted e.g. for the NSIDC SIC CDR, and allow for less data gaps in the SIC fields, especially at the start of the SSM/I era. For OSI-458, AMSR-E and AMSR2 are gridded separately since they do not overlap.

6.2 Gridding and smearing uncertainty

The smearing uncertainty is the error due to the sensor footprint covering more than one pixel in the level 3 product grid. Footprint sizes for the channels used for ice concentration mapping range from over 50 km, for the 19 GHz channels, to about 30 km, for the 37 GHz channels. Further, these footprints, of uneven size, are combined in the algorithms when computing the ice concentration and this leads to an additional smearing effect. We call this the footprint mismatch error. The ice concentration data are represented on predefined and finer resolution grid (typically 10 or 25 km). The smearing and the footprint mismatch error cannot be estimated separately. However, the combined error can be simulated using high resolution ice concentration reference data and a model for the satellite measurement footprint patterns.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

The error is calculated taking cloud free 1 km MODIS images and assigning ice concentrations to all pixels based on the channel 1 imagery. For each pixel the corresponding brightness temperature is calculated for all relevant microwave channels based on standard tiepoints (Comiso et al. 1997). Using channel specific sensor footprints for weighting the ice concentration is calculated from the 1km brightness temperature image in the specified final resolution i.e. 10, 12, 25 and 50 km. This ice concentration is compared to the reference ice concentration (from MODIS), regridded to the same resolution. The standard deviation of the difference between these sets of ice concentration values is the standard deviation of the smeared points.

Using this approach, we parametrize the smearing uncertainty σ_{smear} with a proxy that measures the local variability of the ice concentration field. We found σ_{smear} to be proportional to the 3 x 3 pixel max - min sea ice concentration difference. This includes both the smearing and the footprint mismatch and it is thereby the total smearing error. To avoid computing the smearing below the sea ice concentration noise floor and not to exceed the range of values which were computed with the smearing simulator, the smearing error is:

```

if sigma_smear < sigma_SIC:
    sigma_smear = 0,
if sigma_smear >= sigma_SIC AND sigma_smear < 0.4:
    sigma_smear = 3 x 3 max - min sic difference,
if sigma_smear >= 0.4:
    sigma_smear = 0.4,

```

Where the sea ice concentration is a value between 0 and 1.

At the end of the Level-3 step, daily non-filtered fields of SIC and associated parameters are stored in netCDF files (internal). The fields might present data gaps (missing input data) and erroneous SIC values. Preparing clean and gap-free SIC is covered in the Level-4 step.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

7 Level 4 algorithms

This Level 4 step contains gap filling by interpolation of the areas with missing data, applying masks and coastal corrections, and final formatting of the ice concentration product.

7.1 Interpolation of missing data

For easing the use of the reprocessed data set, it was decided that some level of spatial interpolation should be performed for reducing the occurrence of gaps. Interpolated data points are clearly marked in the product file, so that users can choose to discard them and only ingest retrievals that rely on satellite signal.

Data gaps can occur in several forms, such as missing scan lines, missing orbits and the polar observation hole (NH only). While simple spatial interpolation might be efficient in filling small gaps (e.g. one or two missing scan lines), it blurs the sea ice concentration features. This effect becomes overwhelming when large areas are missing. To overcome this issue, yet implement a general approach for all cases, the ice concentration estimates from the previous and next daily products are used in the interpolation as well: it means that interpolation on a given date D uses pixels from 3 data files: $D-1$, D and $D+1$ for all instruments except for SMMR for which we use $D-2$, D , and $D+2$.

Gap-filling by interpolation is implemented in two steps: first a temporal interpolation, then a spatial interpolation.

7.1.1 Temporal interpolation

All gaps (ocean grid cells with missing SIC data) at day D are identified. For these gaps, the average of the SIC from $D-1$ and $D+1$:

$$SIC_{i,j,D} = 0.5 * (SIC_{i,j,D-1} + SIC_{i,j,D+1})$$

In cases where only one of the $D-1$ or $D+1$ maps have data at i,j coordinate, $SIC_{i,j,D}$ is set to this value. Many gaps will be filled by the temporal interpolation step, but some will remain, for example the polar observation hole in the NH.

7.1.2 Spatial interpolation

If there are still gaps in the SIC map at day D after the temporal interpolation step, these are filled by spatial interpolation using only the data from day D .

In v2 of the CDR, this was implemented by a Gaussian weighting function of the distance from the cell with missing data to the neighbouring cells with data. The approach was effective at filling the polar observation hole, but sometimes brought unrealistic spatial patterns and artifacts, especially at the beginning of the period with SMMR. Several users requested that better interpolation methods are investigated.

For the “v3” CDRs (specifically OSI-450-a), an alternative gap-filling method is being implemented in collaboration with University of Utah. The interpolation method is documented in Strong and Golden (2016). It expresses the SIC field in the polar observation gap as the sum of two contributions: a) a scalar field that is a solution of Laplace's equation with the boundary condition given by the observed ice concentration along the boundary of the polar observation hole, and b) a stochastic term that mimics the spatial autocorrelation lengths of the SIC field, and that is trained in the vicinity of the polar observation hole.

$$f(\lambda, \phi) = \psi(\lambda, \phi) + \Omega(\lambda, \phi).$$

In equation (1), ψ is the solution to Laplace's equation ($\nabla^2 \psi = 0$) using boundary conditions taken from the perimeter of the polar observation hole. Ω is a stochastic term that mimics the spatial autocorrelation lengths of the SIC field, and is trained in the vicinity of the polar observation hole. Steps are detailed below, and an example fill for the v3 data using these steps is shown in Figure 12.

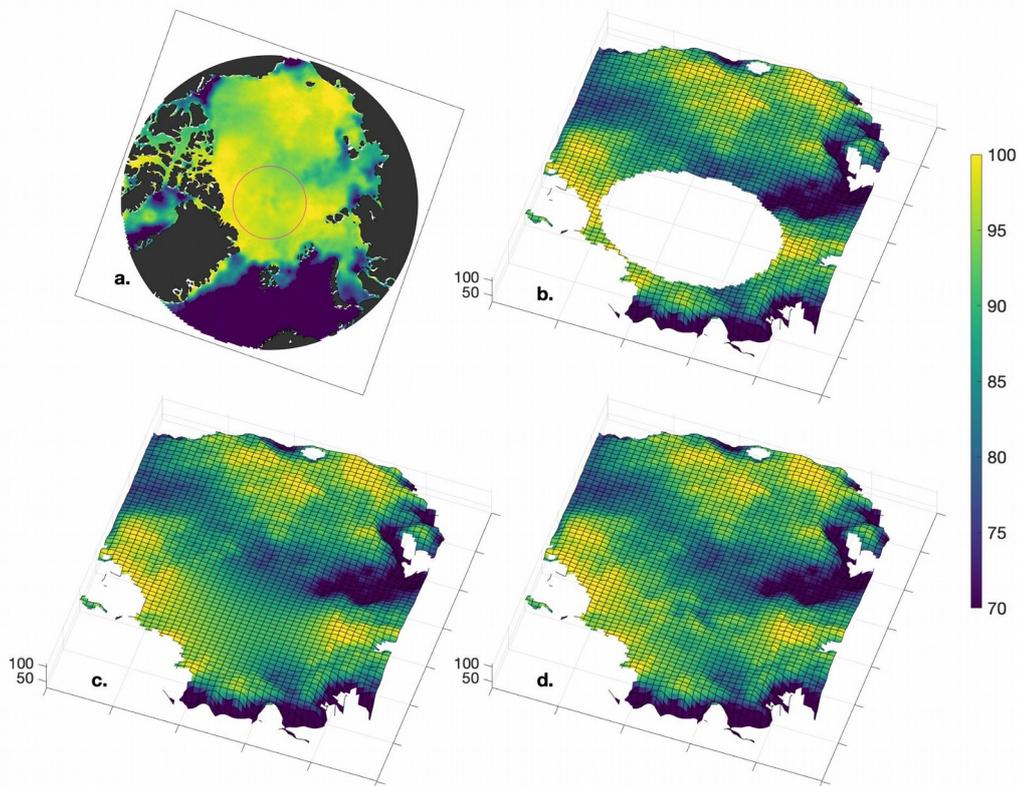


Figure 12: Illustration of the gap-filling method in SIC for 18 July 1983: a. magenta circle indicates location of polar data gap, b. three-dimensional surface plot of SIC around the gap (concentration indicated by vertical coordinate and shading), c. gap fill with only the SIC resulting from solving Laplace's equation, and d. gap fill with the added stochastic contribution. Note the colormap starts at 70% SIC.

As in Strong and Golden (2016), the stochastic term is formulated as the convolution

$$\Omega(x, y) = \frac{2h}{\eta\sqrt{\pi}}\Gamma(x, y) * \exp\left[-\frac{x^2 + y^2}{2\left(\frac{\eta}{2}\right)^2}\right]$$

where h is grid spacing, η is an autocorrelation length scale, and the Γ term is spatially uncorrelated Gaussian noise with time varying standard deviation

$$\sigma = \sigma_0 + \sum_{k=1}^{n_k} [\alpha_k \cos(2\pi k(t-1)/364) + \beta_k (2\pi k(t-1)/364)]. \quad (3)$$

Values for the parameters in the stochastic term $\{n_k, \sigma_0, \alpha_{1-n_k}, \beta_{1-n_k}, \eta\}$ are trained on the SIC field in the vicinity of the polar observation hole (Table 4). When performing the gap-filling, the stochastic term is generated in a reproducible manner by seeding the random number generator with an integer based on the SIC observation's serial number of days since January 0, 0000. Importantly, the gap-filling is done on SIC data that are not bounded to [0%;100%]: the gap-filling can (and does) return SIC values that are >100% SIC, that are set to 100% SIC in the last step of the analysis (see later in this chapter).

parameter	value
n_k	2
σ_0	2.20
α_1	-0.84
β_1	-1.01
α_2	-0.17
β_2	0.60
η	3.26

Table 4: Parameters for running the stochastic term equations, trained on the unbounded SIC data.

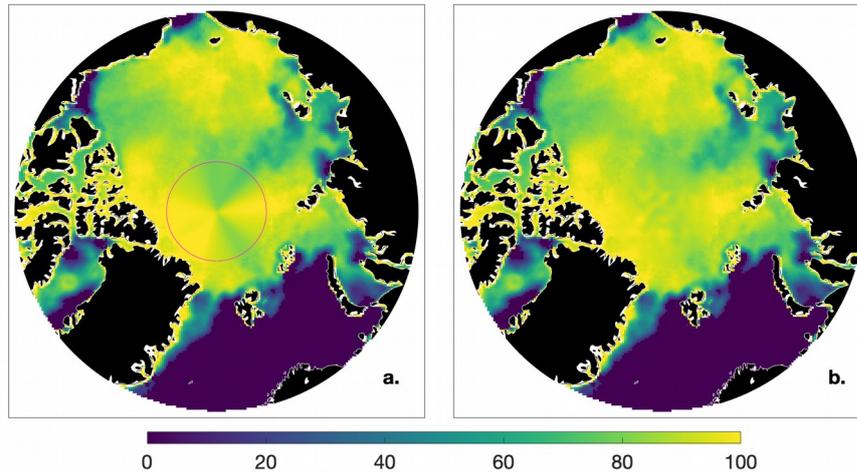


Figure 13: Two gap-filling method for 18 July 1983: a. isotropic interpolation as done in the v2 CDR, with magenta circle indicating the location of the SMMR polar data gap, b. new gap-filling for the v3 CDR following Strong and Golden (2016).

Figure 13 illustrates the result in SIC gap-filling for a date at the start of the OSI-450-a CDR. In the left panel, we show the gap-filling using the isotropic method that was used in the previous version of the CDR, in the right panel is the result of the new gap-filling.

It should be noted that such an advanced interpolation method is only needed for OSI-450-a in the 1978-2005 period since all other missions (SSMIS, AMSR-E and AMSR2) have a much smaller observation gap (limited to a small number of grid cells). It was not implemented for OSI-430-a nor OSI-458, for which the isotropic gaussian gap-filling algorithm is used.

7.2 Total uncertainty

The total uncertainty (as variance) is computed at this stage of the processing, as the sum of the variances of the (gridded and daily averaged) algorithm uncertainty (section 5.5), and of the smearing uncertainty (section 6.2):

$$\sigma_{tot}^2 = \sigma_{algo}^2 + \sigma_{smear}^2$$

Since some users have requested that the different uncertainty terms are provided, the final OSI SAF sea ice concentration product files provide the three level of uncertainties:

- total uncertainty,
- the algorithmic uncertainty,
- the smear uncertainty.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

7.3 Flagging and final formatting.

Details on the status and processing flags as well as the format of the product files are in the Product User Guide (PUG) [RD-1]. The file format is very similar (but not identical) to that used in “v2” CDRs.

Each file provide the SIC, its uncertainties (3 layers) and the status flag in addition to geo-location information. The main SIC variable contains the gap-filled and filtered SIC (i.e. with the effects of the OWF and constrained to the physical range [0-1]). An additional field contains the unfiltered and off-range SICs (aka “raw” values). Access to these off-range SICs was introduced in “v2” and is a pre-requisite for a meaningful evaluation and validation of SIC CDRs close to the end-points at 0% SIC and 100% SIC (Kern et al. 2019).

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

8 Higher-level Products

For this new version of the OSI SAF SIC CDRs, higher-level products are prepared on top of the Level-4 daily products, namely monthly averaged products and a sea-ice index (extent and area). This section briefly outlines the algorithms involved in the preparation of these higher-level products.

8.1 Monthly averaged SIC products

Monthly averaged SIC products are prepared from the daily Level-4 SIC files. The process is as follows:

1. For each day of the month, re-construct the “non-filtered” SIC field by combining the “filtered” and “raw” SIC maps. The “non-filtered” SIC field does not have the effect of the OWF, and has SICs both below 0% and above 100%.
2. Compute the monthly average of the non-filtered SIC fields.
3. Prepare a “filtered” and a “raw” monthly averaged fields. The “filtered” SIC field will apply a cut-off at 10% SIC (all values below are set to 0%) and at 100% (all values above are set to 100%). All SIC values that are modified by this cut-off are saved into the “raw” monthly averaged field.

This process is designed to take full advantage of the SIC range retrieved by the SIC algorithm (before true sea-ice is cut by the OWF) and to avoid introducing a low bias at 100% SIC. Indeed, the traditional way to compute a monthly mean SIC would be to average SICs bounded in [0%;100%] which, because of the non-zero retrieval noise would lead to pulling down the monthly average from 100%. To average the non-filtered SICs will allow the retrieval noise to pull the monthly average either ways and preserve the mean closer to 100% (for a full sea-ice cover).

For “v3” we did not compute uncertainties for the monthly averaged SIC products, because this had required a knowledge of the spatio-temporal correlation lengths of the daily uncertainty and because the monthly averaged product targets a user community that so far did not express the need for such uncertainty information.

8.2 The sea-ice index OSI-420

Sea ice extent and area per hemisphere are prepared from the daily Level-4 SIC product files. Well accepted definitions for sea-ice extent (total area of grid cells with SIC greater than 15% SIC) and sea-ice area (total area of all sea-ice, no threshold) are implemented. Lake ice is not included.

Daily sea-ice extent and area timeseries are prepared as netCDF/CF files. From these daily timeseries, monthly sea-ice extent and area will also be prepared. The baseline is to prepare the sea-ice index products for the Northern and Southern hemisphere, but regional indexes will also be considered.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

9 Conclusions

This Algorithm Theoretical Basis Document (ATBD) presents the algorithm baseline for the “version 3” Global Sea Ice Concentration Climate Data Records OSI-450-a and OSI-458 and the Interim Climate Data Record OSI-430-a.

Several improvements are implemented since the “version 2” (OSI-450, OSI-430-b, and SICCI-25km) baseline. The main improvements concern the land spill-over correction, the selection of open-water tie-points closer to the ice edge, a better tuning of the selection of sea-ice tie-points to limit the biases, the use of ERA5 fields, the use of new FCDR data, and the extension of the data record (both in 1978 and beyond 2015). OSI-458, a research-to-operation transfer from ESA CCI Phase 2, complements OSI-450-a and OSI-430-a with a finer spatial resolution.

Several of these enhancements listed above are contributed by the ESA Climate Change Initiative CCI Sea Ice project phases. This R&D input is acknowledged.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

10 References

- Andersen, S., L. Toudal Pedersen, G. Heygster, R. Tonboe, and L. Kaleschke, Intercomparison of passive microwave sea ice concentration retrievals over the high concentration Arctic sea ice. *Journal of Geophysical Research* 112, C08004, doi10.1029/2006JC003543, 2007.
- Andersen, S., R. T. Tonboe and L. Kaleschke. Satellite thermal microwave sea ice concentration algorithm comparison. *Arctic Sea Ice Thickness: Past, Present and Future*, edited by Wadhams and Amanatidis. Climate Change and Natural Hazards Series 10, EUR 22416, 2006A.
- Andersen, S., R. Tonboe, S. Kern, and H. Schyberg. Improved retrieval of sea ice total concentration from spaceborne passive microwave observations using Numerical Weather Prediction model fields: An intercomparison of nine algorithms. *Remote Sensing of Environment*, 104, 374-392, 2006B.
- Ashcroft, P. and Wentz, F. J.: AMSR-E/Aqua L2A Global Swath Spatially-Resampled Brightness Temperatures, Version 3 [2002–2010], NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA, https://doi.org/10.5067/AMSR-E/AE_L2A.003, 2013.
- Cavalieri, D. J., Gloersen, P., and Campbell, W. J.: Determination of sea ice parameters with the NIMBUS 7 SMMR, *J. Geophys. Res.*, 89, 5355–5369, 1984
- Cavalieri, D. J., J. Crawford, M. Drinkwater, W. J. Emery, D. T. Eppler, L. D. Farmer, M. Goodberlet, R. Jentz, A. Milman, C. Morris, R. Onstott, A. Schweiger, R. Shuchman, K. Steffen, C. T. Swift, C. Wackerman, and R. L. Weaver. 1992. *NASA sea ice validation program for the DMSP SSM/I: final report*. NASA Technical Memorandum 104559. National Aeronautics and Space Administration, Washington, D.C. 126 pages.
- Comiso J.C, D.J. Cavalieri, C.L. Parkinson, and P. Gloersen. Passive microwave algorithms for sea ice concentration: A comparison of two techniques. *Remote Sensing of Environment* 60, 357-384, 1997.
- Comiso J.C. Characteristics of arctic winter sea ice from satellite multispectral microwave observations. *Journal of Geophysical Research* 91(C1), 975-994, 1986.
- Fennig, K., Schröder, M., Andersson, A., and Hollmann, R.: A Fundamental Climate Data Record of SMMR, SSM/I, and SSMIS brightness temperatures, *Earth Syst. Sci. Data*, 12, 647–681, <https://doi.org/10.5194/essd-12-647-2020>, 2020.
- Gloersen, P., and F. T. Barath. A scanning multichannel microwave radiometer for Nimbus-G and SeaSat-A. *IEEE Journal of Oceanic Engineering OE-2(2)*, 172-178, 1977.
- Gloersen, P., W. J. Campbell, D. J. Cavalieri, J. C. Comiso, C. L. Parkinson, H. J. Zwally. Arctic and Antarctic sea ice, 1978-1987: satellite passive-microwave observations and analysis. *NASA SP-511*, Washington D. C., 1992.
- Gloersen, P., and D. J. Cavalieri (1986), Reduction of weather effects in the calculation of sea ice concentration from microwave radiances, *J. Geophys. Res.*, 91(C3), 3913–3919, doi:10.1029/JC091iC03p03913
- Hersbach, H, Bell, B, Berrisford, P, et al. The ERA5 global reanalysis. *Q J R Meteorol Soc.* 2020; 146: 1999– 2049. <https://doi.org/10.1002/qj.3803>

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

- Ivanova, N., Pedersen, L. T., Tonboe, R. T., Kern, S., Heygster, G., Lavergne, T., Sørensen, A., Saldo, R., Dybkjær, G., Brucker, L., and Shokr, M.: Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using passive microwave observations, *The Cryosphere*, 9, 1797-1817, doi:10.5194/tc-9-1797-2015, 2015.
- Kern, S., Rösel, A., Pedersen, L. T., Ivanova, N., Saldo, R., and Tonboe, R. T.: The impact of melt ponds on summertime microwave brightness temperatures and sea ice concentrations, *The Cryosphere Discuss.*, doi:10.5194/tc-2015-202, in review, 2016.
- Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., Tonboe, R. T., Saldo, R., and Sørensen, A. M.: Satellite passive microwave sea-ice concentration data set intercomparison: closed ice and ship-based observations, *The Cryosphere*, 13, 3261–3307, <https://doi.org/10.5194/tc-13-3261-2019>, 2019.
- Kunkee, D. B., G. A. Poe, D. J. Boucher, S. D. Swadley, Y. Hong, J. E. Wessel, and E. A. Uliana, 2008. Design and evaluation of the first special sensor microwave imager/sounder, *IEEE Trans. Geo. Rem. Sens.* 46(4), 863-883.
- Kwok, R.: Sea ice concentration estimates from satellite passive microwave radiometry and openings from SAR ice motion, *Geophys. Res. Lett.*, 29, 1311, doi:10.1029/2002GL014787, 2002.
- 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.
- MAAß, N. and KALESCHKE, L. (2010), Improving passive microwave sea ice concentration algorithms for coastal areas: applications to the Baltic Sea. *Tellus A*, 62: 393–410. doi: 10.1111/j.1600-0870.2010.00452.x
- Meier, W. N., F. Fetterer, M. Savoie, S. Mallory, R. Duerr, and J. Stroeve. 2017. NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 3. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: <https://doi.org/10.7265/N59P2ZTG>.
- Parkinson, C.: A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far exceeding the rates seen in the Arctic, *PNAS*, 116 (29) 14414-14423; DOI: 10.1073/pnas.1906556116, 2019.
- Pedersen, L. T., Saldo, R., Ivanova, N., Kern, S., Heygster, G., Tonboe, R., Huntemann, M., Ozsoy, B., Arduin, F., and Kaleschke, L.: Reference dataset for sea ice concentration, <https://doi.org/10.6084/m9.figshare.6626549.v3>, 2018.
- Peng, G., W. N. Meier, D. Scott, and M. Savoie. 2013. A long-term and reproducible passive microwave sea ice concentration data record for climate studies and monitoring, *Earth Syst. Sci. Data*. 5. 311-318. <https://doi.org/10.5194/essd-5-311-2013>
- Smith, D. M. Extraction of winter total sea ice concentration in the Greenland and Barents Seas from SSM/I data. *International Journal of Remote Sensing* 17(13), 2625-2646, 1996.
- Strong, C. and Golden, K. M.: Filling the Polar Data Gap in Sea Ice Concentration Fields Using Partial Differential Equations, *Remote Sens.*, 8, 442, <https://doi.org/10.3390/rs8060442>, 2016.

EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre	Algorithm Theoretical Basis Document for the OSI SAF Global Sea Ice Concentration CDR and ICDR	SAF/OSI/CDOP3/DMI_Met/SCI/ MA/270
---	--	--------------------------------------

Wentz, F. J. A model function for ocean microwave brightness temperatures. *Journal of Geophysical Research* 88(C3), 1892-1908, 1983.

Wentz, F. J. A well-calibrated ocean algorithm for SSM/I. *Journal of Geophysical Research* 102(C4), 8703-8718, 1997.