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

SAGE CCI

(Sea Ice Age and Drift)

End-to-End ECV Uncertainty Budget (E3UB)

Prime & Science Lead: Signe Aaboe
Met Norway, Norway

Technical Officer: Sarah Connors
ESA ECSAT, United Kingdom

Consortium: Norwegian Meteorological Institute (MET Norway)
S&T Norway 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	UB: Hannah Niehaus, Gunnar Spreen. NERSC: Anton Korosov. MET Norway: Signe Aaboe, Emily Down, Thomas Lavergene. S&T: Jacob Hay.	2026.04.17	V 1.0 consolidated	Released to ESA

Document Approval

Prepared by	Anton Korosov Lead Author, NERSC	
Issued by	Daniele Fantin, Project Manager, S&T	 Digitally signed by Daniele Fantin Date: 2026.05.11 15:44:47 CEST
Checked by	Signe Aaboe Science Leader, MET Norway	 2026.05.11
Approved by	Sarah Connors ESA Technical Officer	 Digitally signed by Sarah Connors Date: 2026.05.07 11:18:46 +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
ATBD	Algorithm Theoretical Basis Document
AWI	Alfred-Wegener-Institute for Marine and Polar Research
BGEP	Beaufort Gyre Exploration Project
C3S	Copernicus Climate Change Service
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
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
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
MEMLS	Microwave Emission Model of Layered Snowpacks
METNO or MET Norway	Norwegian Meteorological Institute
MIZ	Marginal Ice Zone
ML	Machine Learning
MPR	Monthly Progress Report
MS	MileStone
MYI	Multiyear Ice
NASA	National Aeronautics and Space Administration
NERSC	Nansen Environmental and Remote Sensing Center
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
PM	Progress Meeting, Project Manager
PMP	Project Management Plan
PMW	Passive Microwave
PSD	Product Specification Document
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
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	Science and Technology 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
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 End-to-End Uncertainty Budget (E3UB) for the SAGE project within CLIMATE-SPACE – THEME I – B. ADDITIONAL ESSENTIAL CLIMATE VARIABLES (ECVs) – NEW ECV PRODUCTS, in accordance with the contract [AD1], Statement of Work [AD2], and proposal [AD3–AD10].

The purpose of this document is to identify, describe, and, where possible, characterise the main sources of uncertainty affecting the SAGE products of sea-ice age and sea-ice drift. The document provides a brief overview of the algorithm, addressing the different algorithmic components of the SAGE system, including radiometry- and scatterometer-based classification, drift-based tracking approaches, and emerging machine learning methods. For a detailed description of the algorithms and the propagation of uncertainties into per-datum prognostic uncertainty estimates, see the Algorithm Theoretical Baseline Document (ATBD) [RD-XX].

This first version of the E3UB reflects the current development status of the SAGE system and provides an initial, primarily qualitative assessment of uncertainty.

1.2 Document Overview

This document is structured as follows:

- Chapter 1 - introduces this document.
- Chapter 2 - presents the overall structure of the SAGE system, introduces the main categories of uncertainties, and describes how uncertainties enter and propagate through the processing chain, including dependencies between productions.
- Chapter 3 - provides an overview of the uncertainties affecting radiometry- and scatterometer-based sea-ice type classification and their impact on the product.
- Chapter 4 - provides overview of the uncertainties in the Lagrangian sea ice age product
- Chapter 5 - describes uncertainties in the machine learning approach
- Chapter 6 -presents the uncertainty budget of the sea ice drift product
- Chapter 7 - provides guidance to users, validation activities, and ESA on the use of the product and priorities for future improvements in uncertainty characterisation.
- Chapter 8 -Summarises the main findings of the uncertainty budget and outlines key points for future development.

1.3 Applicable Documents

No	Doc. Id	Doc. Title	Date	Issue/ Revision/ Version
AD-1	4000147560/25/I-LR	ESA Contract No.	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-ATBD-001	Algorithm Theoretical Basis Document	16/04/2026	1.0
RD-2	METNO-ESA-SAGE-CCI-PVASR-001	Product Validation and Algorithm Selection Report	1/04/2026	1.0

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

2 Overview of the SAGE system and uncertainty framework

2.1 Overview

The SAGE system is designed to generate climate data records (CDRs) of sea-ice age (or type) and sea-ice drift (an extension of existing drift CDR backwards in time). The system consists of four main components:

- Sea-ice type
- Sea-ice age
- Machine learning component
- Sea-ice drift CDR extension

These components are not independent, as in particular sea-ice drift and concentration are shared across the system and propagate uncertainties between components.

2.2 General sources of uncertainty

The SAGE products are affected by uncertainties originating from measurements, auxiliary datasets, reference data, numerical processing, and intrinsic properties of the retrieval algorithms. These uncertainty sources can act independently or interact through the processing chain, and their influence may vary with season, region, sensor type, and ice conditions. Some components mainly introduce random noise, while others may generate systematic biases or temporally correlated errors that are particularly relevant for climate data records.

2.2.1 Satellite data

Satellite observations form the primary input to the SAGE system and therefore constitute a major source of uncertainty. Instrument noise in brightness temperatures and radar backscatter propagates directly into retrieval variables and can reduce the separability between ice classes or degrade the stability of drift tracking. Noise effects are usually strongest under low-contrast conditions, such as during melt or in heterogeneous ice cover.

Geolocation uncertainty is another important component. Small positioning errors may shift observed features between acquisitions and translate into displacement errors in sea-ice drift products. When daily composites are generated, geolocation errors can also introduce blurring, effectively reducing image sharpness and lowering tracking performance.

A further limitation is the relatively coarse spatial resolution of passive microwave and scatterometer sensors. Mixed surface types may occur within a single footprint, causing representativeness errors and ambiguity between open water, first-year ice, and multiyear ice. Coarse resolution also limits the ability to resolve small-scale deformation and narrow leads, and constrains the achievable accuracy of motion vectors.

2.2.2 Other upstream data

Several SAGE components rely on external geophysical datasets. Atmospheric reanalysis fields such as ERA5 air temperature, wind speed, and wind direction are used as auxiliary information or as direct forcing in

summer drift estimates. Errors in these fields, including biases in wind magnitude or timing, can propagate into model-based drift retrievals and seasonal classifications.

Sea-ice concentration products, such as those from OSI SAF, are also used in masking, conditioning, and age calculations. Uncertainties in concentration can affect the delineation of the ice edge, the identification of multiyear ice fractions, and the conditioning of advected fields. Since concentration products may themselves contain regional or seasonal biases, these effects can be inherited by downstream SAGE products.

Sea-ice drift products from SAGE or OSI SAF are a key upstream input for Lagrangian sea-ice age estimation and for consistency checks across products. Errors in drift magnitude or direction accumulate through repeated advection steps and may lead to misplaced trajectories, incorrect age assignment, and increased uncertainty in long-lived ice parcels.

2.2.3 Other data

Independent reference datasets are essential for training, calibration, and validation, but they also contain uncertainty. Operational ice charts are based on expert interpretation of multiple data sources and may differ between agencies or analysts in the placement of ice boundaries and the classification of ice types. Such discrepancies limit the precision with which satellite products can be validated and may introduce apparent disagreements that originate from the reference rather than the retrieval (see examples in the PVASR [RD-2]).

Reference buoy observations are widely used for drift validation. Although buoy positions are generally accurate, the motion of an individual buoy represents the trajectory of a single floe, whereas satellite drift products describe the average displacement over much larger spatial areas and longer temporal intervals. This scale mismatch creates representativeness uncertainty that cannot be fully separated from retrieval error. Additional quality-control issues, such as outliers or delayed transmissions, may also affect buoy datasets.

2.2.4 Numerical methods

Numerical processing introduces additional uncertainty beyond that contained in the original observations. Interpolation is required when transforming data between grids, collocating variables, or mapping swath observations to regular products. Interpolation can smooth gradients, shift extrema, and alter local variance, especially near coastlines or the ice edge.

Extrapolation is required when observations are missing or incomplete. This includes spatial gap filling, temporal continuation, and machine-learning prediction in regions where no direct observations are available. Because extrapolated values are not directly constrained by measurements, they are generally more uncertain than nominal retrievals and may fail under unusual or rapidly changing conditions.

Smoothing is commonly applied to suppress noise and improve product consistency. While beneficial for stability, smoothing can attenuate sharp spatial structures, damp short-term variability, and reduce the apparent magnitude of dynamic features such as leads, ridges, or deformation zones.

2.2.5 Inherent algorithm non-linearities

Some uncertainty sources arise from non-linear behaviour within the retrieval algorithms themselves. In sea-ice drift tracking, occasional rogue vectors may occur when the matching procedure locks onto an incorrect feature or when the signal is distorted by melt, atmospheric effects, or low contrast. These erroneous vectors can differ strongly from neighbouring values and, if not detected, may contaminate subsequent processing steps.

More generally, threshold-based decisions, Bayesian classification boundaries, and iterative optimisation schemes can amplify small input perturbations into disproportionately large output changes. Such non-linear effects are often intermittent and difficult to characterise with simple Gaussian error assumptions, but they can be important contributors to local uncertainty and outlier behaviour.

2.3 The processing chain and uncertainty pathways

The SAGE processing chain starts from satellite observations acquired by passive microwave radiometers and scatterometers (the blue box in Fig. 2.1). Passive radiometers provide brightness temperatures, while scatterometers provide radar backscatter, both of which contain information on sea-ice state and motion. These measurements are subject to instrument noise, calibration uncertainty, geolocation errors, and the limitations imposed by coarse spatial resolution. Surface melt, atmospheric moisture, changing snow conditions, and mixed pixels near the ice edge can further degrade the physical signal. Because these satellite observations form the common foundation of several downstream products, uncertainties introduced at this stage may propagate throughout the entire SAGE ecosystem.

The satellite observations are combined with additional upstream datasets, including sea-ice concentration products from OSI SAF and meteorological forcing from reanalysis or forecast systems such as ERA5 (the violet box in Fig. 2.1). Together, these inputs drive the sea-ice drift and sea-ice type algorithms and also serve as predictors for machine-learning components (the cyan boxes in Fig. 2.1). This creates strong dependencies between products: uncertainty in sea-ice concentration affects masks and ice-edge detection; uncertainty in winds or air temperature influences drift retrievals and summer motion estimates; and biases in training or predictor variables can be learned and transferred into machine-learning outputs. Since several algorithms rely on overlapping inputs, common upstream errors may appear coherently across multiple SAGE products rather than as isolated random noise.

The output products (the green boxes in the middle) are also linked to each other. Sea-ice drift fields, together with upstream sea-ice concentration, are ingested by the Lagrangian sea-ice age algorithm, where drift errors accumulate through repeated advection and can lead to misplaced trajectories or incorrect age classes. In this way, the quality of the sea-ice age product depends directly on the quality of the drift product and indirectly on the concentration data used to define the ice domain. Finally, sea-ice drift, sea-ice type, and sea-ice age are validated against independent reference datasets (the orange box in Fig. 1.3), including buoy trajectories and operational ice charts. These references are themselves uncertain because of scale mismatch, analyst interpretation differences, temporal sampling gaps, and occasional outliers. Validation therefore provides an

essential measure of performance, but part of the observed disagreement may originate from the reference data rather than the SAGE algorithms alone.

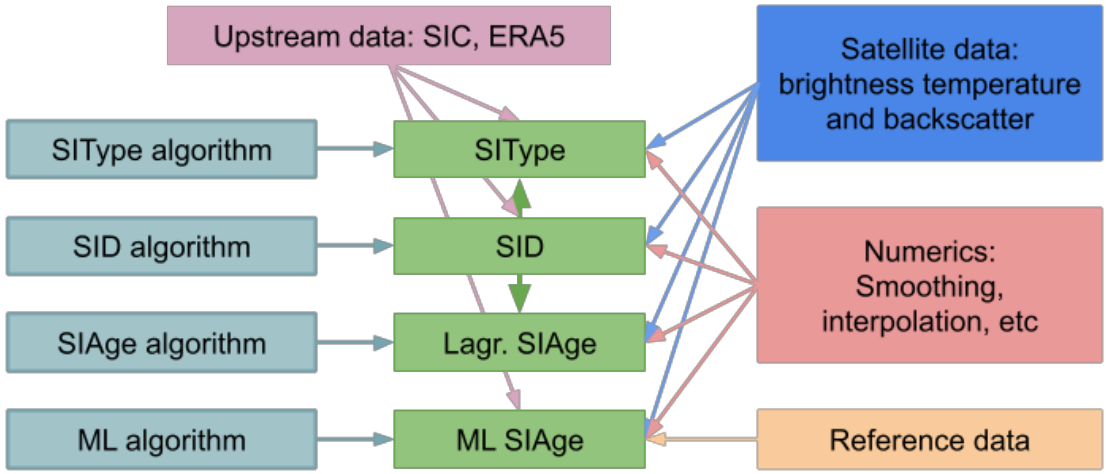


Figure 2.1 Pathways of uncertainties in the SAGE system from sources such as algorithms (cyan boxes), upstream data (violet box), satellite data (blue box), numerical methods (pink box), and reference data (orange box) into the SAGE products (green boxes) and between them (e.g., from SID into SIType or Lagrangian SIAge).

3 Sea-ice type - uncertainty budget

Sea-ice type is a diagnostic variable. It cannot be observed directly, but is inferred from satellite measurements using classification algorithms. In the context of SAGE, it is derived from passive microwave radiometry (PMW), and scatterometer data (SCAT), as well as auxiliary products such as sea-ice concentration, drift, and model data. Its uncertainty is therefore mainly driven by how the sensor signals are interpreted and by the assumptions used in the retrieval methods. Sea-ice type represents a classification of the ice cover into categories reflecting its thermodynamic history, typically reduced to the two main classes: first-year ice (FYI) and multiyear ice (MYI).

As outlined in Chapter 2 (Sections 2.2–2.3), uncertainties in the sea-ice type retrievals typically originate from satellite observations, auxiliary data, and processing steps, and propagate through the retrieval chain. For sea-

ice type, these uncertainties primarily affect the interpretation of the microwave and backscatter signals, the definition of class boundaries, and the representation of heterogeneous ice conditions within a grid cell.

Additional uncertainty arises from the selection of training data and prior assumptions in the classification, as well as from auxiliary inputs, such as sea-ice concentration and drift, which are often used in the processing.

Uncertainty varies with region and season. It is typically higher in areas with low ice concentration or mixed ice-water conditions, such as the marginal ice zone. In the summer months, the uncertainties become too large for a meaningful ice-type classification based solely on microwave observations. Therefore, no sea-ice type classification is provided in summer. Increased uncertainties are also expected during the transition periods into and out of the melt season, as well as during and following short-term warm-air events (see, e.g., section 4.6 in the ATBD [RD-1]).

The link between sea-ice type and sea-ice age introduces further complexity, as instantaneous classification may not always be consistent with the temporal evolution of the ice cover [RD-1, RD-2].

3.1 Main sources of uncertainty

In the context of this E3UB, the focus is not on individual algorithms, but on how uncertainties arise in microwave-based sea-ice type retrievals more generally, including examples from individual algorithms.

The candidate sea-ice type algorithms (selected from the PVASR exercise [RD-2]) include both pure PMW and PMW-SCAT multisensor approaches. These methods differ in their physical assumptions, input data, and classification frameworks, as described in the ATBD [RD-1]. Consequently, the individual components of the uncertainty budget do not affect each algorithm uniformly. Their relative impact and extent vary with the specific sensitivity of each algorithm to the assumptions or mechanisms, which are described in Table 3.1.

Table 3.1: Summary of all known error components contributing to the sea-ice type uncertainty budget.

Uncertainty component	Source / mechanism	Impact on product	Type (random vs systematic)	Data level?
Signal ambiguity	Overlap of FYI and MYI signatures in microwave radiometry and scatterometry; sensitivity to snow cover, melt, and deformation. (See Section 2.2.1)	Misclassification between FYI and MYI; reduced confidence in mixed or transitional conditions	Random + systematic	L3
Algorithm dependence	Different physical assumptions, input data, and classification schemes across PMW, scatterometer, and multisensor		Random + systematic	L2+L3

	methods. E.g. Bayesian statistical uncertainty and assumption of Gaussian distribution of training data			
Representativeness and sampling	Mixed ice conditions within sensor footprint; mismatch between footprint and grid cell; different product definitions (class vs concentration) (See Section 2.2.1)		Random + systematic	L3
Training data	Selection of representative training data	Biased classification if the training data are not representative	systematic	L3
Atmospheric correction	Mechanism/model selection to account for atmospheric effects on satellite measurements	Biases can propagate into the surface signature	systematic	L2
Auxiliary products	Errors in sea-ice concentration and drift products used for masking, training, correction, etc (See Section 2.2.2)	Propagation of errors into classification and post-processing	systematic	L3
Seasonal effects	Melt processes, wet snow, and changing surface properties	Strong degradation of FYI/MYI separability; unreliable classification in summer	systematic	L2

3.1.1 Algorithm dependence - C3S Bayesian approach

Classification confidence estimate

The C3S sea-ice type product includes an uncertainty estimate. The uncertainty is estimated within the Bayesian classification framework. The retrieval is based on probability density functions (PDFs) of the observed microwave signatures for each ice class, derived from training data. For a given observation, the algorithm computes the probability for each class (e.g., FYI or MYI) and assigns the class with the highest probability.

The associated Bayesian uncertainty is therefore expressed through these class probabilities:

$$uncertainty \propto \frac{1}{probability}$$

It represents the *degree of confidence in the classification*, given the overlap of the class-specific PDFs and the variability in the observed signal. Low classification confidence (i.e. similar probabilities for multiple classes) indicates ambiguous conditions, while high confidence reflects clearer separation in the observation space. This uncertainty measure primarily captures the effects of signature overlap, variability in training data, and assumptions in the statistical model, rather than full end-to-end uncertainty. It does not explicitly account for all sources of uncertainty, such as errors in auxiliary inputs or representativeness effects, and should therefore be interpreted as a measure of relative classification confidence within the Bayesian framework.

Gaussian assumption

Within the Bayesian classification framework, the distribution of the training data (i.e. observed microwave signatures for each ice type) is approximated by a Gaussian distribution. This provides a simple and efficient representation of the data and generally captures its central tendency and spread. However, the true distributions may deviate from a Gaussian shape, particularly under heterogeneous or transitional surface conditions, where skewed or multimodal behaviour may be more realistic.

Probability density functions

In addition, the class-specific PDFs are characterised by different variances, with MYI generally exhibiting a larger standard deviation than FYI. As a consequence, the MYI class-conditional probability density decays more slowly in the tails. For observations with high values, this can cause the Bayesian classifier to favour MYI over FYI, even when the observation is not typical of MYI. In practice, such high values may instead represent FYI or mixed conditions, for example, ice including areas of open water, rather than pure MYI. This behaviour biases the classification towards MYI, particularly in the marginal ice zone.

To mitigate this effect, the C3S algorithm applies an additional constraint by introducing an upper threshold based on the FYI distribution. Observations exceeding a defined limit (e.g. the mean plus one standard deviation of the FYI PDF) are not classified as MYI. This reduces the risk of false MYI classification in the high-value tail, although it also introduces an additional empirical assumption into the classification scheme.

An additional *source of variability* arises from the use of daily updated training data. The resulting PDFs are sensitive to variability in the observations, driven by short-term changes in atmospheric conditions and surface properties, as well as by sensor-related effects such as noise, viewing geometry, and inter-sensor differences.

These factors can influence the classification from one day to the next and may introduce temporal variability in the product, even where the underlying ice conditions are stable. The use of a 15-day data window provides some temporal stabilisation of the PDFs, but no other explicit smoothing is applied to the final product

3.1.2 Algorithm dependence - ASIMIR

The ASIMIR algorithm also uses Bayesian logic for uncertainty estimation. However, while the C3S method uses it for categorical classification, ASIMIR uses it for continuous inversion of a physical forward model. Within this Bayesian optimal estimation framework, the algorithm provides a consistent set of geophysical parameters and a pixel-wise estimate of retrieval uncertainty. The retrieval uncertainty is explicitly defined by the retrieved covariance matrix $\hat{S}$, where the square root of its diagonal elements provides the 1-sigma uncertainty for each variable. This calculation combines the sensitivities of the physical forward model with an effective measurement uncertainty matrix S_e. The model uncertainty was quantified through a sensitivity analysis involving 1000 MEMLS_ice simulations per frequency tuple.
This methodology accounts for instrumental noise and uncertainties arising from the simplified representation of microphysical surface parameters within the physical forward model. However, it does not incorporate structural model limitations or unsimulated physical processes, which remain additional sources of unquantified uncertainty.

3.1.3 Algorithm dependence - EC-ICE

For the EC-ICE algorithm, the retrieval uncertainty is quantified by calculating a confidence level CL for each pixel, which is derived from the statistical spread of the 1000 Monte Carlo realisations, which produce a distribution of possible concentration vectors for a single observation. This confidence level is mathematically defined as:

$$CL = 1 - \frac{MAD}{AD_{max}}$$

where MAD is the mean absolute deviation of the 1000 trial solutions from their median and AD_{max} is the maximum absolute deviation observed within those trials.

3.1.4 Training data, tiepoints and probability distributions

In categorical sea-ice type retrievals, the definition of tie-points or class-specific PDFs is typically based on training data representing relatively “pure” end-members of FYI and MYI. These are often derived from regions dominated by young FYI (e.g. shelf regions) and from the core areas of old, consolidated MYI. As a consequence, the intermediate range of observed signatures between these end-members is not explicitly represented in the training data.

Observations falling between the FYI and MYI tie-points may therefore have multiple interpretations:

1. Classification ambiguity: The observation lies in a region of overlapping signatures, leading to low confidence in the assignment (e.g. similar Bayesian probabilities for FYI and MYI).

2. Mixed ice conditions: The sensor footprint may contain a mixture of FYI and MYI, resulting in a combined signature that falls between the two end-members. This can then be interpreted as a fractional concentration of a certain ice class, e.g. EC-ICE.
3. Intermediate ice types: The observation may represent ice older than the “pure” young FYI but younger than the “pure” older MYI, and thus represents an intermediate ice age class.

This ambiguity is inherent to the FYI–MYI classification and constitutes a fundamental source of uncertainty in the retrieval. It also reflects a key difference between algorithm approaches, as different methods interpret observations in this intermediate range in different ways (see also discussion in the PVASR [RD-2]).

4 Lagrangian sea-ice age uncertainty budget

The uncertainty budget for sea ice age is derived through sequential propagation of uncertainties from the input Sea Ice Drift (SID) and Sea Ice Concentration (SIC) products through the full processing chain. First, uncertainty is propagated with advection of the concentration field, where the variance at time step (n+1) is obtained from the variance at step (n) scaled by the local area-change factor associated with divergence or convergence. The advected concentration is then conditioned by the observed SIC field, which constrains the solution by selecting the minimum of the advected and observed concentrations. Accordingly, the uncertainty of the conditioned field is taken from whichever of the two concentration estimates defines the minimum value. This provides the uncertainty in the advected MYI concentration field.

The conditioned concentration uncertainty is then combined with the uncertainty introduced by ice drift errors. To estimate this component, uncertainties from the gridded SID product are first spatially smoothed using neighbouring grid cells. These smoothed uncertainties are subsequently accumulated along the Lagrangian trajectories during repeated advection steps, yielding an integrated drift-position uncertainty for each mesh element. This integrated uncertainty defines a search radius around the advected location, representing the range of possible displaced positions. The resulting uncertainty in multi-year ice concentration is quantified as the standard deviation of concentrations of all mesh elements located within this radius. The total concentration uncertainty for each advected field is then obtained by combining the conditioned concentration uncertainty and the drift-induced uncertainty in quadrature.

In the final stage, uncertainties are propagated from concentration fields to age fractions and ultimately to the sea ice age variable itself. Individual age fractions are derived as differences between consecutive advected multi-year ice concentration fields; therefore, their variances are computed as the sum of the variances of the contributing fields. The sea ice age estimate is then calculated as a weighted mean of all age fractions, where each fraction is weighted by its corresponding age class. Uncertainty in the final sea ice age product is obtained using standard error propagation for a ratio of weighted sums, accounting for uncertainties in both the numerator (age-weighted concentrations) and denominator (total concentration of all age classes). This produces a complete end-to-end uncertainty budget that links uncertainties in the input ECV products directly to the uncertainty of the derived sea ice age Climate Data Record.

5 Machine learning classification - uncertainty budget

Machine learning (ML) is, fundamentally, a method for mathematical approximation (Hornik et al (1989)), and the utility of a model is empirically based on the sampled domain. Theoretically, an ML model is limited to interpolation over the learned space; conversely, the models may act unpredictably when tasked to extrapolate, which partially explains what we today refer to as hallucination. Therefore, a major source of potential future uncertainty will be in cases where we move beyond states that have been observed in the existing training data.

In principle, our models may contain probabilistic distributions. For deep neural networks, we use dropout as a Bayesian approximation (Gal, Y., & Ghahramani, Z. (2016)).

5.1 Accounting for the (co-) Domain

The simple summary is that our final model will probably only be a good (i.e. corresponding with reality) model for the data that it has seen. Probably does a lot of heavy lifting in the previous sentence, so to specify further; the pullback over the cospan of our channels and samples form a fibre product, which is a subset of the cartesian product of the active channels we use in our input.

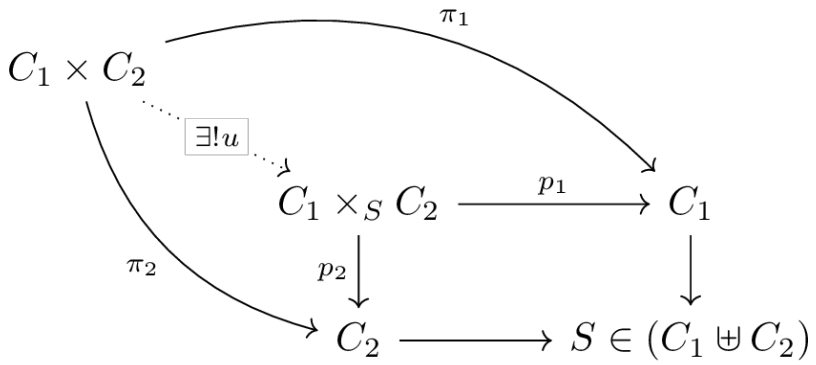


Figure 5.1: Commutative diagram of the fibre product ($C_1 \times_S C_2$ in the middle), from which training all data samples, S, are selected. This fibre product is a subspace of the cartesian product $C_1 \times C_2$.

Outlined in Figure 5.1 is a diagram showing the correspondence between the samples we have available, S, and the space that these samples outline. The "observed" space is limited to that which is outlined by the fibre product w.r.t. S, and is necessarily a subspace of the full space outlined by taking the product of our available channels, C. Samples exist within the disjoint union or coproduct of the sampled channels, specified here as sets where $\{s = (c_1, c_2) \mid c_1 \in C_1, c_2 \in C_2\}$ (allowing for some abuse of notation).

As mentioned, a model will generally do well with interpolation inside the sampled space, but is likely to struggle once it is forced to extrapolate. We therefore expect that our models are likely to perform well within the boundaries of conditions that have been seen in the training data, but underperform in novel scenarios where the input data deviates from what has been trained on.

This underlying map should also then account for why ablation or augmentation to the input data seems to work well, as it accounts for small variations around each of the input channels. Training on samples with an

introduced element of variation will slightly perturb the underlying product space, resulting in samples being drawn from a pseudo-distribution over the actual input space.
The result should be a model more robust to slight variations, that is better suited for the interpolation within the training domain.

5.2 (Pseudo-) Probability and predictive uncertainty

Machine learning models are subject to both aleatoric and epistemic uncertainty (Hüllermeier, E., & Waegeman, W. (2021)). In our case we are working with a form of supervised learning, where one of the most interesting aspects is the predictive uncertainty, i.e. an estimate of how uncertain the model is in the predicted outputs.
By default, the models we use will output pseudo-probabilistic distributions, where the values sum to, which are tempting to interpret as naive probabilities of an output. Oftentimes this is unproblematic, but it is important to keep in mind that this likelihood or probabilistic distribution is conditioned on the input.
However, in some cases, it may be pertinent to probe the model for a stochastic estimate. This can be done by running dropout at inference (e.g. Gal et al 2016), which introduces variance at the point of prediction. Adding dropout at inference time and repeatedly sampling the model will result in outputs that have some variance introduced, and if done correctly, can approximate predictive uncertainty.
When estimating the predictive uncertainty of our models, we will make use of this property to introduce noise, thus allowing us to report on the uncertainty of the model predictions.

5.3 Summary of uncertainties related to machine learning

Divided into primary and secondary sources of uncertainty where the primary sources are more likely to prevail through to the end of this project.

5.3.1 Primary sources

1) Aleatoric (irreducible) uncertainty. From statistical or stochastic error in the correlations which the neural network is tasked with, sometimes FIY may look like MYI in the training data.
    a) Overlaps in the input data w.r.t ground truth can be identified and mapped.
    b) Ground truth from buoys etc. may be used.
    c) Models may be tuned to account for proximity, e.g. using a form of random field smoothing to ensure that neighboring pixels will have a stronger tendency to agree.
2) Predictive uncertainty inherent to the models. Output will be a sample from a distribution across the different classes, which may be more or less subject to a degree of variance with minor perturbations of the inputs.
    a) Ensuring estimates of predictive uncertainty on the model outputs will help categorise and track potential issues in the final models.
    b) Tuning of the models can be focused on reducing predictive uncertainty insofar as that does not depreciate the target metrics. Added training after model convergence has also been shown to improve the stability of the latent categorisation.

5.3.2 Secondary sources

1. Epistemic (reducible) uncertainty. Models may lack the expressive capability to separate between nearby classes. Similarly an example of epistemic (or systemic) uncertainty may be due to lack of knowledge, or ignorance in how to approach the target classes.
    a. Accounting for potential mis-labeled data, and/or identifying cases of doubt for manual re-evaluation.
    b. Curating a high-quality subset of the data, and assigning a higher weight to performance on this may provide a forcing factor that removes a source of systemic uncertainty.
2. Data overrepresentation and bias. An overrepresented class may end up creating a bias in the model that results in unbalanced predictions.
    a. Appropriate weighting on both sampling and the loss-function (or criterion) should be able to compensate and result in the model being able to evaluate the input in a fair manner.
3. Geographic information being encoded into the model. If we do not endeavour to separate the model from the geographic location it is likely that the final model will retain some bias that is based on identifying which area the input originates from.
    a. This may or may not be an issue, depending on the intended use of the model. To the extent that we identify tendencies like this in the final model, it may be accounted for in a report and user guide.
    b. There may also be a benefit to using bias from geographic location as a factor in predicting the ice types. We may therefore decide at a later stage to explicitly provide this information to our network in order to allow a location-bound bias.
4. Missing data, or areas with partial coverage in either the inputs or labels. A concern here may be particularly that certain areas will be well supported, while others only have a fraction of the same coverage.
    a. Compensation for missing labels can be done by including pseudo-labels, or using consistency-based learning.
    b. When training we will be masking areas based on their data coverage. An accounting for the sum-coverage across the training period may be provided which may then indicate or inform, together with a quantified predictive uncertainty, areas that will benefit from more representation in future data collection.

6 Sea-ice Drift - uncertainty budget

To the best of our knowledge, there is no known methodology for uncertainty propagation for the sea-ice drift retrieved from motion tracking algorithms such as the Continuous Maximum Cross-Correlation method used in this project. Instead, the community computes statistics of validation errors (against a set of independent reference observations) and uses these statistics as the uncertainty. This approach to uncertainty is described in the rest of this section.

First, we briefly list the main sources of uncertainties that can contribute to the sea-ice drift tracking algorithm to return uncertain estimates. It can be useful to refer to the CCI SAGE ATBD [RD.1] for a general understanding of how the sea-ice motion tracking processing is performed.

- Lack of details of the input satellite imagery. For generating multi-decadal climate data records of sea-ice motion, we use rather coarse resolution satellite imagery from passive microwave (and sometimes scatterometer) missions. The spatial resolution of the FoVs is generally in the order of 10-15km. This coarse resolution automatically limits the sharpness of the imaged details, and thus the accuracy of the sea-ice motion vectors (slightly different motion vectors score relatively similarly). Sea-ice motion processed with similar algorithms from Synthetic Aperture Radar (SAR) imagery result in much more accurate motion vectors because of the much higher spatial resolution imagery.
- Geolocation accuracy. Offsets or noise in the geolocation will directly translate into errors in the sea-ice motion vectors. For the CDR processing, because we build daily averaged maps of brightness temperature as input to the motion tracking algorithm, the geolocation accuracy results in additional blurring and loss of details of the imagery (see previous point).
- Non-linear uncertainties / rogue vectors. The uncertainty sources listed above are mostly linear in effect. But some other sources (e.g. atmospheric opacity or surface melting) can lead to non-linear effects where the retrieved vector is far from the truth. Some steps in the algorithm are designed to detect and correct those rogue vectors. In a nominal production, there should thus be a limited amount of remaining non-linear uncertainties.
- Representativity uncertainty. The CDR sea-ice drift represents the averaged motion for large (~100x100 km) areas of the sea-ice surface, which is coarser resolution than many ocean/ice models, other satellite products, and buoy data. When comparing CDR vectors to other sources, there is thus always a component of the mismatch that can be attributed to the differences in spatial scales. The same applies to the temporal domain: the CDR sea-ice drift vectors represent the integrated Laplacian displacement which can be quite different from hourly velocities recorded by buoys. This is why we extract aggregated displacement vectors from the buoy trajectories to validate the CDR vectors.
- Summer sea-ice motion from wind-driven model. Because of surface melt and atmosphere opacity, CDR vectors cannot be reliably processed from satellite imagery, and a completely new technique is used during the summer melt season. This uses ERA5 reanalysis wind data and a free-drift model whose parameters are tuned against quality controlled satellite-based products. This model thus combines uncertainties due to the wind vectors, from the parameters, and from the physical limitation of the free-drift assumptions. This summer drift product is provided with specific quality flags and higher uncertainties.

- Spatial gap-filling. Gap-filling through spatial interpolation is required to fill areas where the main algorithms did not return useful estimates. These gap-filled grid cells are more uncertain than the nominal retrievals, and are clearly marked in the status flags of the product files.

The items above identify the main sources of uncertainties for the OSI SAF sea-ice drift CDR (1991-2020). The same sources apply for the back-extension processed by CCI SAGE. In particular, the pre-1991 period requires using coarser resolution imagery, potentially with larger geolocation errors. The ERA5 wind vectors might be more uncertain and the provision of daily complete fields will require relying more on spatial gap-filling.

6.1 Methodology

6.1.1 Access to buoy data

Buoy data must first be accessed and converted to a common format. Each per-buoy file should contain tuples of (time, latitude, longitude), plus metadata including the buoy name, source and overall time coverage of the file. There can be various different quality control issues with buoy data, such as delays in updating the GPS position causing sudden jumps in the track, erroneous single positions, or buoys that are stationary or on a ship. The following quality control steps were performed to detect and discard erroneous data:

1. Reorder the position records chronologically.
2. Remove duplicate position records (identical time/latitude/longitude).
3. Remove position records between which no displacement is observed (i.e. identical latitude/longitude).
4. Remove position records between which the velocity is larger than three times the standard deviation from the mean velocity, where the mean velocity is computed over the whole trajectory. This is to detect and remove rogue locations in the trajectory.
5. Visual inspection of each trajectory to remove any that could be identified as erroneous.

6.1.2 Collocation of product vectors and buoy vectors

In order to compare the product vectors with the buoy trajectories, they must be collocated so that they represent the same vector at the same time and geographical location. A single pair of collocated product and buoy vectors is called a “matchup”, and is fully specified by the values
($t_{0\ prod}$, $lat_{0\ prod}$, $lon_{0\ prod}$, $t_{1\ prod}$, $lat_{1\ prod}$, $lon_{1\ prod}$, $t_{0\ buoy}$, $lat_{0\ buoy}$, $lon_{0\ buoy}$, $t_{1\ buoy}$, $lat_{1\ buoy}$, $lon_{1\ buoy}$), where ($t_{0\ prod}$, $lat_{0\ prod}$, $lon_{0\ prod}$) refers to the start time ($t_{0\ prod}$) and the start latitude and longitude ($lat_{0\ prod}$, $lon_{0\ prod}$) of the product vector and ($t_{1\ prod}$, $lat_{1\ prod}$, $lon_{1\ prod}$) to the end time and position. In addition, for possible filtering later, the status flag field for the product vector is also retained.

The collocations are performed over a fixed period of 1 day, and collect the resultant matchups together in a single Net CDF file. Later, these files are concatenated over different periods for computing the statistics. A 3-d search of the time, latitude and longitude is performed between the buoy and the product data. For each record in each buoy trajectory, the following steps are conducted:

1. Converting the product vectors to the correct format for matching – the start and end times of the vectors are directly available in the product file, but the start and end positions require a conversion.

The collocation is also partially performed with x and y locations on the grid, so both (x, y) and (lat, lon) positions are retained during this process. The start ($lat_{0\ prod}$, $lon_{0\ prod}$) are defined by the grid cell coordinates, and these can be converted to ($x_{0\ prod}$, $y_{0\ prod}$) using the area definition of the EASE-2 grid. The drift components (dX_{prod}, dY_{prod}) of that grid cell are added to retrieve the end locations ($x_{1\ prod}$, $y_{1\ prod}$) of the product vectors:

$$x_{1\ prod} = x_{0\ prod} + dX_{prod}$$

$$y_{1\ prod} = y_{0\ prod} + dY_{prod}$$

These can then be converted back to ($lat_{1\ prod}$, $lon_{1\ prod}$) using the grid definition.

2. Mapping each in-situ record (t_{buoy}, lat_{buoy}, lon_{buoy}) to the product grid, i.e. converting (lat_{buoy}, lon_{buoy}) to grid coordinates (x_{buoy}, y_{buoy}) using the grid area definition. The result of this is a record of (t_{buoy}, x_{buoy}, y_{buoy}).
3. Spatially search for a possible matching buoy vector using a nearest neighbour search to find the closest product grid cell. This retrieves the closest product start point ($t_{0\ prod}$, $x_{0\ prod}$, $y_{0\ prod}$).

   Note that a nearest neighbour search is used rather than a bi-linear interpolation, since the latter may lead to artificially good validation statistics by averaging out part of the noise in the product.

4. Temporally assess whether the buoy record is a possible valid match for the product vector start time at that grid cell by checking if the following condition is satisfied:

   $$| t_{0\ prod} - t_{buoy} | < T_{max}$$

   where T_{max} is the maximum allowed time difference for the start of the drift vector (3 hours is used here).

5. Perform a check for multiple valid matches. If multiple buoy records meet the temporal condition for a single product grid cell, the one with the smallest spatial Euclidian distance is prioritised.
6. Ensure the possible match satisfies spatial constraints. The distance between (lat_{buoy}, lon_{buoy}) and ($lat_{0\ prod}$, $lon_{0\ prod}$) is checked to ensure it falls within a maximum specified radius (40 km here). If this record meets the criteria, the record can be defined as the startpoint of a possible matching buoy vector, ($t_{0\ buoy}$, $lat_{0\ buoy}$, $lon_{0\ buoy}$), but there are still some filtering steps.
7. Check that there is a buoy vector available from the trajectory records which matches the duration of the product vector. The buoy record with the closest duration to the duration of the product vector is found, such that $t_{diff\ prod} - t_{diff\ buoy}$ is minimised, where $t_{diff\ prod} = t_{1\ prod} - t_{0\ prod}$ and equivalent for the buoy. The buoy vector is retained as a possible match if:

   $$| t_{diff\ prod} - t_{diff\ buoy} | < D_{max}$$

where D_{max} is the maximum duration difference (here 1 hour). This buoy record ($t_{1\ buoy}$, $lat_{1\ buoy}$, $lon_{1\ buoy}$) then defines the endpoint of the possible matching buoy vector.

8. Check that the buoy vector is surrounded by four valid product vectors. This means that each of the corners of the bounding box in which the buoy vector starts, has a valid product vector which starts from that grid point. Although only the nearest of these four is considered in the statistics, this constraint is introduced to avoid validation data in the outer edges of the vector field, such as in the marginal ice zone or in coastal regions (fast ice). Since some of the ice-tethered buoys are designed to continue floating when sea ice melts, this constraint is also an effective way to avoid collocating ocean current measurements with sea ice vectors.
9. Finally, for the matched vectors, the dX_{buoy} and dY_{buoy} can be found from the ($x_{0\ buoy}$, $y_{0\ buoy}$) and ($x_{1\ buoy}$, $y_{1\ buoy}$), since the statistical analysis is performed by comparing the vector displacements.

Even with carefully selected and quality controlled buoy data, and careful collocation between the product and in-situ data, there is a possibly large and mostly uncontrolled source of error from the representativity mismatch between the scales sampled by the buoy and the satellite product. A buoy samples the motion of the ice floe it was deployed on, which are often large floes chosen by the field scientists in order to limit the risk of losing the buoy prematurely. The CDR satellite drift product on the other hand samples the motion of a much larger area of the sea ice surface, according to the feature tracking window size.

The mismatch between the two scales of motion contributes to part of the error budget and it is not possible to separate this representativity error from the measurement error of the satellite product with a simple two-way statistical analysis (see Hwang and Lavergne, 2010 for further discussion).

6.1.3 Analysis of collocated vectors

Individual matchup files are concatenated per hemisphere to cover the required time period of the statistics. The validation statistics can then be calculated directly on the matchup pairs of drift vectors (dX_{prod}, dY_{prod}) and (dX_{buoy}, dY_{buoy}).

The difference (or mismatch) between the product and buoy displacements in x and y is given by:

$$\Delta dX = dX_{prod} - dX_{buoy}$$
$$\Delta dY = dY_{prod} - dY_{buoy}$$

The biases in the displacements are the mean of the differences:

$$\mu_x = \overline{\Delta dX} = (1 / N) \Sigma_{i=1}^{N} (dX_{prod,i} - dX_{buoy,i})$$
$$\mu_y = \overline{\Delta dY} = (1 / N) \Sigma_{i=1}^{N} (dY_{prod,i} - dY_{buoy,i})$$

The standard deviations of the differences are derived from the diagonal elements of the covariance matrix, which is defined as

$$covar_M = \begin{bmatrix} cov(\Delta dX, \Delta dX) & cov(\Delta dX, \Delta dY) \\ cov(\Delta dY, \Delta dX) & cov(\Delta dY, \Delta dY) \end{bmatrix}$$

where each element of the matrix is given as (for example)

$$cov(\Delta dX, \Delta dY) = (1 / N\text{-}1) \Sigma_{i=1}^{N} (\Delta dX_i - \overline{\Delta dX})(\Delta dY_i - \overline{\Delta dY})$$

The standard deviations for the *x* and *y* components are then
$\sigma_x = \sqrt{(covar_M[0, 0])}$
$\sigma_y = \sqrt{(covar_M[1, 1])}$

An overall standard deviation can be calculated with a simple mean of the *x* and *y* components
$\sigma = 0.5 \times (\sigma_x + \sigma_y)$

And finally the correlation coefficient between ΔdX and ΔdY is given by
$\rho_{x,y} = covar_M[0, 1] / (\sigma_x \cdot \sigma_y)$

Note that when assigning the uncertainties, different uncertainties can be determined for different status flags by sub-dividing the matched product-buoy vectors according to the product status flag before retrieving the statistics for each subset.

6.1.4 Adapting uncertainties for time mis-registration

The previously calculated uncertainty values are valid when using the single-senor ice drift vectors with the space-varying start t_0 and stop t_1 time information provided in the product files. These can vary between 8 UTC and 16 UTC across the product grid, depending on the orbit and instrument characteristics, rather than all being from exactly 12 UTC to 12 UTC. When t_0 and t_1 are not accounted for, and the vectors are used as if from 12 UTC to 12 UTC, the uncertainties must be raised. Here a 2nd order polynomial formula is used to calculate the raised uncertainty $\sigma_k^{12}(\delta t)$.

$\delta t = | t - t_0 |$
$\sigma_k^{12}(\delta t) = 0.015 \times \delta t^2 - 0.005 \times \delta t + \sigma_k$

where δt has units of hours, and σ_k is that from the previous section.

It is easily seen that the uncertainties for vectors with t_0 close to 12 UTC will not be raised significantly from σ_k.

The numerical coefficients in this equation were re-used from those of the NRT product OSI-405. In brief, the polynomial shape was obtained by collocating buoys and the OSI-405 satellite products with wrong time information, deliberately misregistered from -10 hours to +10 hours. The mismatch was fitted to determine the polynominal coefficients.

6.2 Early Results

Tables 6.1 and 6.2 show the bias and standard deviation validation statistics for the integrated winter periods (summer for gap-filling wind drift) for the various frequency and polarisation combinations of each instrument. For SMMR-NI07 in the Northern Hemisphere, the standard deviations are generally within around 6km, and for the SSMI instruments, around 3-5 km. Southern Hemisphere has much higher standard deviations, up to nearly 10 km for SMMR-NI07. Both hemispheres show quite variable moderate bias. Most periods have a

relatively low number of matched vectors in these validation statistics, which obviously lowers the quality of these statistics.

Table 6.1: Northern Hemisphere bias and standard deviation statistics (x- and y-components) for the frequency-polarisation combinations of each instrument. All are integrated vectors over all winter periods, except for wind drift, which is the summer season.

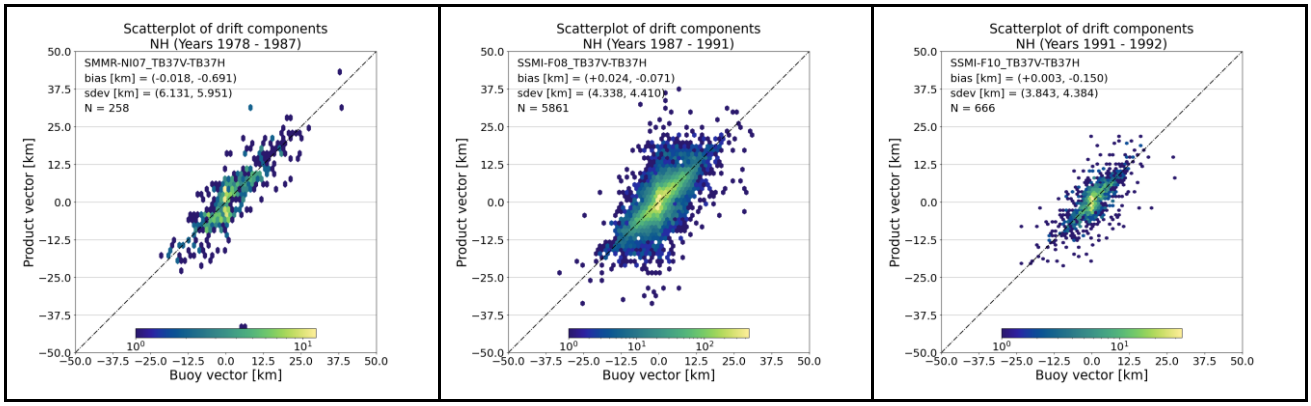
Inst	Freq (GHz)	Pol	Seas	Bias (x)(km)	Bias (y)(km)	Std (x) (km)	Std (y) (km)	Num vecs
SMMR-NI07	37	V & H	W	+0.170	-0.568	4.312	5.185	152
SMMR-NI07	37	V	W	+0.130	+0.060	4.893	6.099	152
SMMR-NI07	37	H	W	+0.322	-1.085	5.462	5.289	152
SSMI-F08	37	V & H	W	+0.002	-0.132	2.681	2.792	2983
SSMI-F08	37	V	W	+0.015	-0.127	2.667	2.741	2981
SSMI-F08	37	H	W	-0.022	-0.174	3.190	3.295	2982
SSMI-F08	85	V & H	W	-0.087	-0.133	3.823	3.363	1577
SSMI-F08	85	V	W	+0.375	-0.490	4.124	4.294	876
SSMI-F08	85	H	W	-0.114	-0.113	3.126	3.232	2274
SSMI-F10	37	V & H	W	+0.028	-0.435	3.450	3.848	277
SSMI-F10	37	V	W	+0.157	-0.310	3.511	3.960	278
SSMI-F10	37	H	W	-0.066	-0.511	3.621	4.248	276
SSMI-F10	85	V & H	W	+0.207	-0.524	3.658	3.537	276
SSMI-F10	85	V	W	+0.285	-0.465	3.801	3.476	276
SSMI-F10	85	H	W	+0.144	-0.670	3.689	3.627	273
Wind			S	-0.070	+0.314	2.628	2.366	11052

Table 6.2: Same as Table 6.1 but for the Southern Hemisphere.

Inst	Freq (GHz)	Pol	Seas	Bias (x)(km)	Bias (y)(km)	Std (x) (km)	Std (y) (km)	Num vecs
SMMR-NI07	37	V & H	W	+2.796	+0.828	7.744	8.589	51
SMMR-NI07	37	V	W	+2.854	+1.987	9.654	9.059	46

SMMR-NI07	37	H	W	+0.266	-0.211	4.954	5.525	499
SSMI-F08	37	V & H	W	+2.125	+1.159	7.396	8.059	61
SSMI-F08	37	V	W	+0.161	-0.159	4.951	5.254	499
SSMI-F08	37	H	W	+0.388	-0.332	5.593	6.310	499
SSMI-F08	85	V & H	W	+0.319	-0.451	4.946	4.365	249
SSMI-F08	85	V	W	-0.046	-0.665	4.932	3.721	237
SSMI-F08	85	H	W	+0.348	+0.292	5.296	4.932	293
SSMI-F10	37	V & H	W	+0.218	+0.112	5.087	4.841	524
SSMI-F10	37	V	W	+0.335	+0.138	4.832	5.378	508
SSMI-F10	37	H	W	+0.289	-0.172	5.955	4.881	517
SSMI-F10	85	V & H	W	-0.002	-0.061	4.186	3.700	535
SSMI-F10	85	V	W	-0.053	+0.030	4.293	3.671	559
SSMI-F10	85	H	W	+0.047	-0.183	4.492	4.053	504
Wind			S	+0.641	-1.110	4.420	5.824	655

Figures 6.1 (37 GHz combined V and H polarisations) and 6.2 (85 GHz combined V and H polarisations) show some examples of validation plots for the different sensors in the Northern and Southern Hemispheres, plotting product vector x- and y- displacements against matched buoy displacements. It can be seen by comparison of the standard deviations and biases, that the 37 GHz and 85 GHz results are broadly comparable in terms of quality of the vectors in validation. It is not necessarily true that the sources of scatter in both channels arise from the same mechanisms however, instead it may be expected that the 37 GHz channels may have more scatter due to lower resolution, whereas the 85 GHz channels may have more scatter due to rogue vectors.



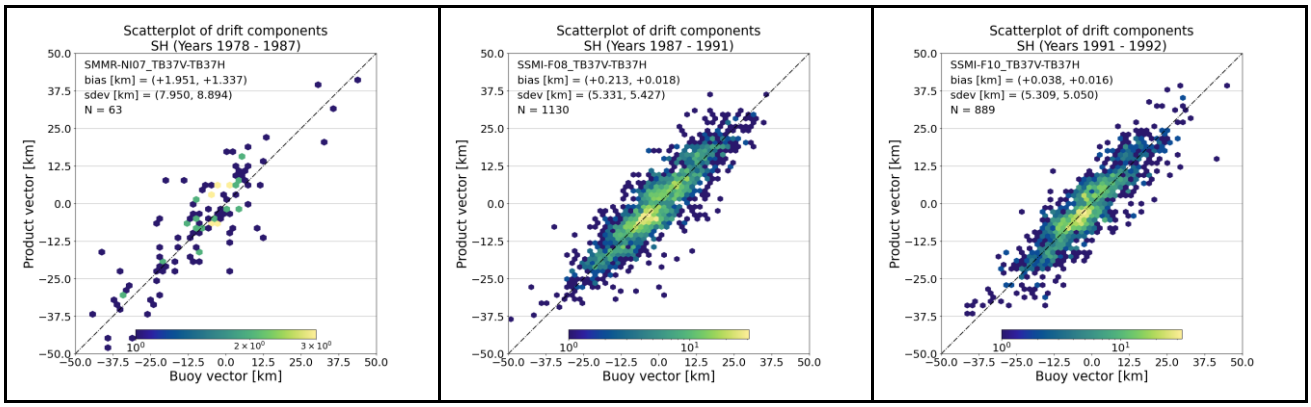


Figure 6.1: Scatter plots of 37 GHz combined V and H polarisations for SMMR-NI07 (left), SSMI-F08 (middle) and SSMI-F10 (right), for Northern Hemisphere (top row) and Southern Hemisphere (bottom row).

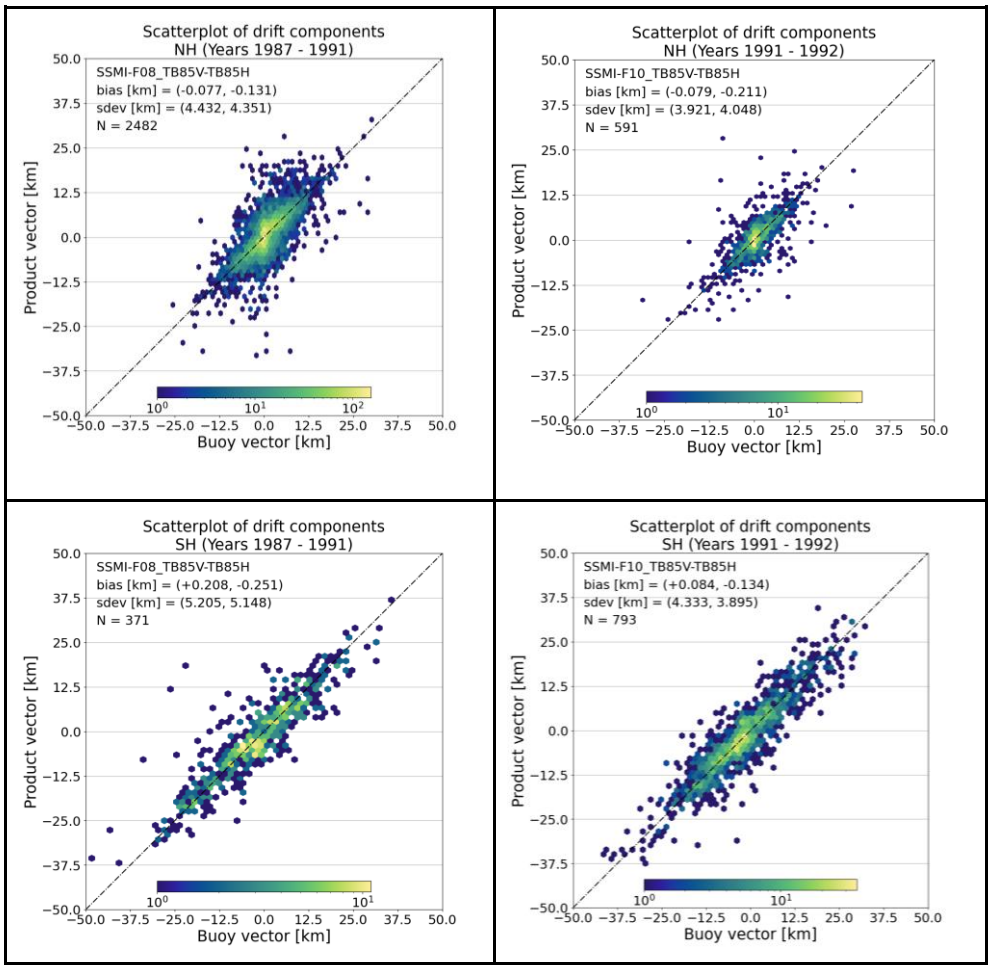
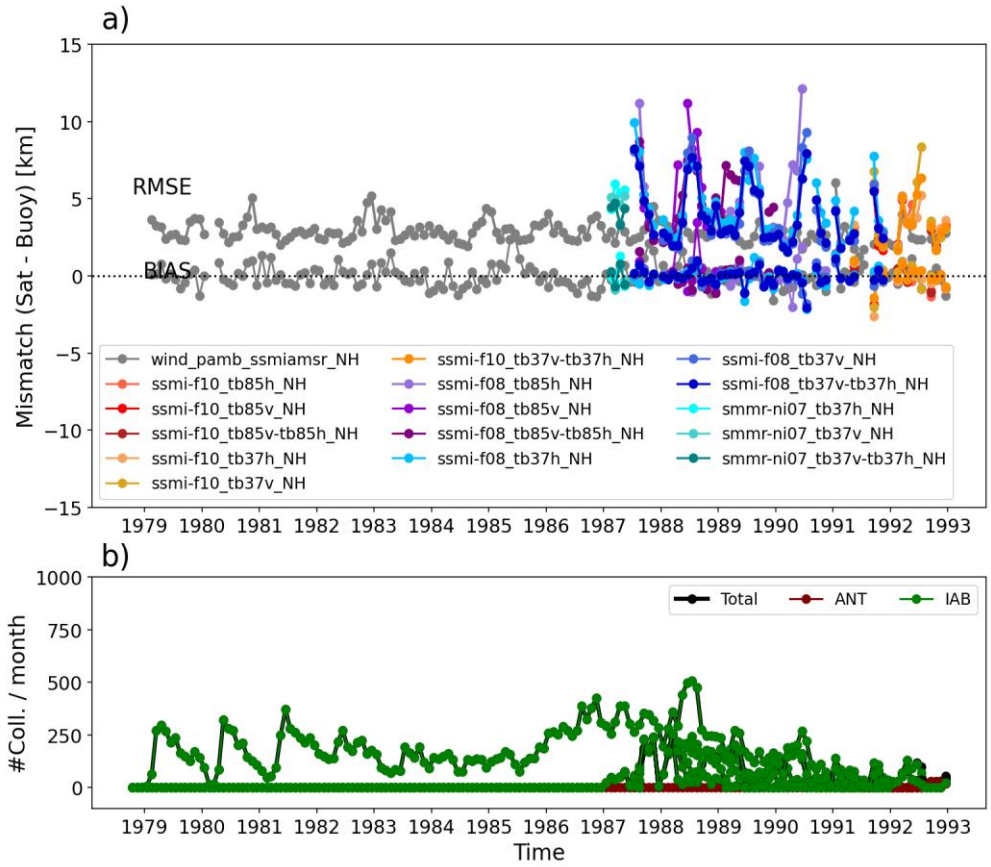


Figure 6.2: Scatter plots of 85 GHz combined V and H polarisations for SSMI-F08 (left) and SSMI-F10 (right), for Northern Hemisphere (top row) and Southern Hemisphere (bottom row).

Figure 6.2 shows timeseries of the standard deviation and bias from validation against buoys for the various instruments with the different frequency and polarisation combinations. In addition, the number of buoy collocations per month is shown in order to demonstrate which periods were possible to validate in Figure 6.3. The RMSE and bias show quite high variation generally but are not very different between the instruments. There are not very strong differences between the quality of the vectors between the 37 GHz and 85 GHz frequencies for SSMI-F08. Vectors calculated from the Free Drift Model based on the ERA5 wind reanalysis are shown in green, and it can be seen that these are suitably matched to be able to be used for gap filling in summers and when data is missing. It should however be noted that the wind-derived drift vectors do not capture long-term trends and therefore should not be used for that kind of analysis.

Despite the relatively larger uncertainties in this first version of the sea-ice drift back-extension, it is very encouraging that processing these early, coarse resolution imagery does not seem to result in significant biases. This will be further investigated and confirmed in future versions of this back-extension work.


a)
Mismatch (Sat - Buoy) [km]
RMSE
BIAS
wind_pamb_ssmiamsr_NH
ssmi-f10_tb85h_NH
ssmi-f10_tb85v_NH
ssmi-f10_tb85v-tb85h_NH
ssmi-f10_tb37h_NH
ssmi-f10_tb37v_NH
ssmi-f10_tb37v-tb37h_NH
ssmi-f08_tb85h_NH
ssmi-f08_tb85v_NH
ssmi-f08_tb85v-tb85h_NH
ssmi-f08_tb37h_NH
ssmi-f08_tb37v_NH
ssmi-f08_tb37v-tb37h_NH
smmr-ni07_tb37h_NH
smmr-ni07_tb37v_NH
smmr-ni07_tb37v-tb37h_NH
b)
#Coll. / month
Total
ANT
IAB
Time

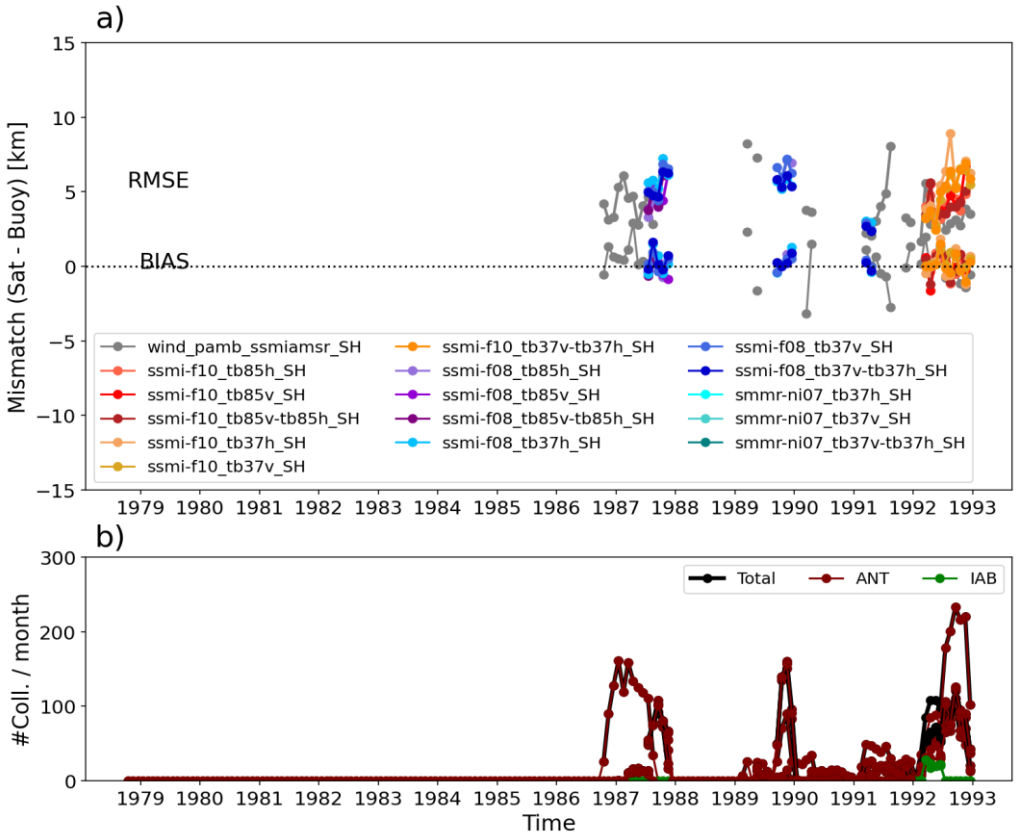


Figure 6.3: Timeseries of the RMSE and bias (top frames) and number of buoy collocations (bottom frame) for the Northern Hemisphere (top plots) and Southern Hemisphere (bottom plots). No validation was possible until 1986 in the Southern Hemisphere due to lack of buoys.

7 Recommendations

Given the current development status of the SAGE system and the preliminary nature of this first version of the E3UB, the following recommendations should be regarded as initial suggestions and are expected to be further developed in the next version of the E3UB as the blended algorithm is designed.

Recommendations to Users

- Beware of ambiguities in variable definitions, including differences in how sea-ice type and sea-ice age are defined, represented (e.g. categorical vs continuous)
- Exercise caution in high-uncertainty conditions, in particular:
    - the marginal ice zone,
    - transition zones or regions with heterogeneous ice conditions,
    - the melt season,
    - where the separability of ice types is reduced, e.g. the ambiguous sea-ice class in the C3S product
    - transition from one satellite mission to the next, especially relevant for the sea-ice drift back-extension
- Beware of the biases in the uncertainty of the data coming from different satellites as it may introduce trends or sudden transitions in the uncertainty time series

Recommendations to the Validation Team

- Beware of ambiguities in variable definitions, including differences in how sea-ice type and sea-ice age are defined, represented (e.g. categorical vs continuous), and associated with uncertainty estimates across products
- Beware of inconsistencies in the reference sea ice charts. For the same day, sea ice charts may provide quite different maps of multi-year ice, which may differ substantially on weekly time scales.

Recommendations to ESA

- Variability of the reference ice charts should be quantified and estimated in terms of uncertainty. This source needs to be added to the overall uncertainty budget.
- Impact of the quantified prognostic and diagnostic uncertainties on the biases in the sea ice age product should be better evaluated
- More attention should be given to assess the limitations of the developed uncertainty framework and identifying unquantified uncertainties

8 Summary and conclusions

The SAGE uncertainty framework recognises that product quality is determined by the full processing chain rather than by any single algorithm component. Uncertainty originates in the primary satellite observations, including noise, geolocation errors, and spatial resolution constraints, and is further influenced by auxiliary upstream datasets such as sea-ice concentration, atmospheric forcing, and external drift products. Additional contributions arise from reference datasets used for training and validation, numerical processing steps such as interpolation, extrapolation, and smoothing, and from intrinsic non-linear behaviour within the retrieval algorithms. Because the SAGE products are strongly interconnected, uncertainties may propagate from one component to another, accumulate through sequential processing, or appear consistently across several products that rely on common inputs.

An initial uncertainty framework has been developed to combine product-specific assessments and capture dependencies between sea-ice drift, sea-ice type, sea-ice age, and machine-learning components. Validation against buoys, ice charts, and other independent observations remains essential, but should always be interpreted in light of the uncertainty and representativeness limits of the reference data themselves.

In this report we do not only quantify the errors through comparison with reference data (i.e., evaluate the diagnostic uncertainties), but also aoutlines approaches towards providing transparent and traceable uncertainty information that supports scientific interpretation, operational use, and long-term climate applications of the SAGE data record (through prognostic uncertainties).

9 References

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning (arXiv:1506.02142). arXiv. https://doi.org/10.48550/arXiv.1506.02142

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Machine Learning, 110(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8