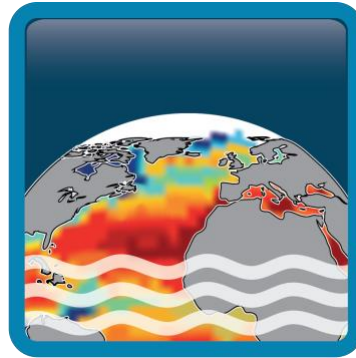


Climate Change Initiative+ (CCI+) Phase 2

Sea Surface Salinity



[D2.1] Product Validation and Algorithm Selection Report (PVASR)

Customer: ESA

Ref.: ESA-EOP-SC-AMT-2021-26

Version: v4.0

Ref. internal: 4000123663/18/I-NB

Revision Date: 20/06/2023

Filename: SSS_cci-PHASE#2-D2.1-PVASR_v4r0.docx

Deliverable code: D2.1



Signatures

Author	Fabrice Bonjean Jacqueline Boutin Dimitry Khvorostyanov		
Reviewed By	Jean-Luc Vergely		
	Rafael Catany		
	Roberto Sabia		
Approved By	Jacqueline Boutin (Science Leader)	LOCEAN	
	Nicolas Reul (Science Leader)	IFREMER	
	Rafael Catany (Project Manager)	ARGANS	
Accepted By	Roberto Sabia (Technical Officer)	ESA	
	Susanne Mecklenburg (Technical Officer)	ESA	

Diffusion List
Sea Surface Salinity Team Members
ESA (Susanne Mecklenburg, Roberto Sabia)

Amendment Record Sheet

Document Change Record		
Date / Issue	Description	Section / Page
15/07/2019 / v1.0	Update template and review formatting (v1.0) – First issue to ESA	Whole document
03/12/2019 / v1.1	Add reference documents	Section 1.3.2 / page 1
03/12/2019 / v1.1	Add Definitions (new section 2) terms relevant to this document	Section 2 / page 3
03/12/2019 / v1.1	Added Figure 1	Section 2 / page 4
03/12/2019 / v1.1	Improved description of the Description of the algorithm & ancillary data tested during the round robin	Section 5
28/05/2020/ v2.0	Update all comparisons with v2.3 and 2.4 CCI+SSS	Sections 4 to 8
28/05/2020/ v2.0	Add comparisons with ship binned tracks in the northern Atlantic Ocean and with Tara circle experiment in the Arctic Ocean	Section 4
28/05/2020/ v2.0	Add some comparisons with ISAS maps	Section 6.2
24/06/2020 / v2.1	Added ESA's feedback and suggestions (minor changes to previous version)	N/A
26/07/2021/ v3.0	Update all comparisons with V3.2 and 2.3 CCI+SSS	Sections 4 to 8

Table of Contents

Signatures	iii
Amendment Record Sheet	v
List of figures	vii
List of tables	x
Acronyms.....	xi
1 Introduction	1
1.1 Executive Summary	1
1.2 Scope	1
1.3 References	2
1.3.1 Applicable Documents	2
1.3.2 Reference Documents.....	2
1.4 Structure of the document	3
2 Definitions	4
3 Overview	7
4 Round robin methodology	9
4.1 In situ data	9
4.1.1 Monthly binned ship tracks in North Atlantic.....	9
4.1.2 Research vessel TSG in the Southern Ocean	10
4.1.3 Gridded in-situ datasets.....	12
4.2 Colocation methodology	12
4.2.1 Monthly binned ship tracks in northern Atlantic.....	12
4.2.2 Monthly binned ship tracks in Antarctica	12
4.3 Metrics.....	13
5 Description of the algorithms & ancillary data tested during the round robin exercise	15
5.1 Global products	16
5.2 Polar research products	16
6 Algorithm/Product evaluation.....	17
6.1 Verification at global scale	17
6.2 Results of the round robin exercise	29
6.2.1 North Atlantic comparisons	29
6.2.2 Southern Ocean comparisons.....	44
6.3 Products evaluation summary	51
6.4 Open issues and discussion	52
7 Conclusion and future work	53

List of figures

Figure 1: Scale portraying the typical depth at which near-surface salinity is measured by various sensors/platforms. The small squares show the average measurement depth and the capped lines show the range for that average. For profiling platforms (ASIP, Bow Bridle, STS-Argo, Argo) the range represents the variability of the top-most point in the profile. For platforms with standardized configurations that measure at fixed depths (Salinity Snake, SSP, Wave Glider) the mean and range of each sensor at a particular depth are shown. For platforms where there are multiple sensor configurations (drifters, mooring, shipborne TSG) or that sample at different depths depending on the specifics of the platform, the range of measurement depths across all platforms is shown. Radiometric penetration depths were calculated using the Stogryn (1997) relationship and show penetration depths at 1.43 GHz over the salinity range of 20 pss to 38 pss and temperature range of $-2\text{ }^{\circ}\text{C}$ to $35\text{ }^{\circ}\text{C}$ (where the “mean” value shown in the figure is for $20\text{ }^{\circ}\text{C}$ and 35 pss). (Figure taken from [RD 06]). -----6

Figure 2: Map of boat transects and their division into regular boxes. An example of a ship track is shown as a solid black line along B-AX02 (Reverdin et al., 2018). In this work, only the B-AX01 and B-AX02 transects are used. ----- 10

Figure 3: TSG tracks spanning from 2010 to 2022, with each type of TSG cruise being represented by a different color. ----- 11

Figure 4: Number of monthly-averaged TSG data within about $25\times 25\text{ km}^2$ boxes along recurring ship tracks around Antarctica. ----- 11

Figure 5: Median difference between CCI L4 and ISAS 17+NRT SSS in three latitudinal bands (70S-40S, 40S-40N, 40N-70N) obtained with a) V3.2 and V4.1 further than 1000km from coast, b) V3.2 and V4.1 at less than 1000km from coast, with c) V3.2 and V4.2 further than 1000km from coast, d) V3.2 and V4.2 at less than 1000km from coast. ----- 19

Figure 6: Robust std difference between CCI L4 and ISAS 17+NRT SSS in three latitudinal bands (70S-40S, 40S-40N, 40N-70N) obtained with a) V3.2 and V4.1 further than 1000km from coast, b) V3.2 and V4.1 at less than 1000km from coast, with c) V3.2 and V4.2 further than 1000km from coast, d) V3.2 and V4.2 at less than 1000km from coast ----- 20

Figure 7: Correlation coefficient between CCI L4 and ISAS 17+NRT SSS in three latitudinal bands (70S-40S, 40S-40N, 40N-70N) obtained with a) V3.2 and V4.1 further than 1000km from coast, b) V3.2 and V4.1 at less than 1000km from coast, with c) V3.2 and V4.2 further than 1000km from coast, d) V3.2 and V4.2 at less than 1000km from coast ----- 21

Figure 8: Robust standard deviation of the centered reduced difference between CCI L4 and ISAS 17+NRT SSS in three latitudinal bands (70S-40S, 40S-40N, 40N-70N) obtained with a) V3.2 and V4.1 further than 1000km from coast, b) V3.2 and V4.1 at less than 1000km from coast, with c) V3.2 and V4.2 further than 1000km from coast, d) V3.2 and V4.2 at less than 1000km from coast ----- 22

Figure 9: Mean differences between CCI L4 and ISAS 17+NRT SSS averaged over all longitudes as a function of latitude and time, (Left) with CCI V3.2 and (Right) with CCI V4.2. ----- 23

Figure 10: Std differences between CCI L4 and ISAS 17+NRT SSS averaged over all longitudes as a function of latitude and time, (Left) with CCI V3.2 and (Right) with CCI V4.2. ----- 24

Figure 11: Rms difference between CCI L4 SSS and ISAS 17+NRT SSS for (Left) V3.2 and (Right) V4.2. -----	25
Figure 12: Comparison of monthly number of CCI+ SSS estimates between version 3.2 (Left) and version 4.2 (Right) over the period from January 2010 to September 2020. -----	26
Figure 13: Mean differences between CCI L4 and ISAS 17+NRT SSS averaged over all longitudes as a function of latitude and time, (Left) with CCI V4.3 and (Right) with CCI V4.2. -----	27
Figure 14: Std differences between CCI L4 and ISAS 17+NRT SSS averaged over all longitudes as a function of latitude and time, (Left) with CCI V4.3 and (Right) with CCI V4.2. -----	28
Figure 15: Interpretation keys for the box plots in the next figures-----	29
Figure 16: Results of box plots for each bin along the North Atlantic B-AX01 TSG tracks, numbered from West to East (Figure 2). The plotted datasets for each bin result are arranged from left to right, namely: ISAS-20, CCI+SSS V3.2, CCI+SSS 4.2, CCI polar research res0, res1, res1_noWS, res3, and res4. The box for each product, as shown in Figure 15, represents the median and interquartile range (IQR) for the entire 2010-2022 period, and includes a standard outlier threshold (bars) set to 1.5*IQR. -----	30
Figure 17: Same as Figure 16, but for B-AX02, and TSG tracks are numbered from South to North. -----	31
Figure 18: Time series of TSG data and products within different North Atlantic B-AX01 boxes.	33
Figure 19: Time series of TSG data and products within different North Atlantic B-AX02 boxes	34
Figure 20: (Color) Mean absolute difference (mad) between each product and TSG SSS within each box for B-AX01. Number couples indicate the 95% confidence interval of mad using a bootstrap procedure. -----	35
Figure 21: (Color) Mean absolute difference (mad) between each product and TSG SSS within each box for B-AX02. Number couples indicate the 95% confidence interval of mad using a bootstrap procedure. -----	36
Figure 22: (Color) Correlation between each product and TSG SSS within each box for B-AX01. Number couples indicate the 95% confidence interval of correlation using a bootstrap procedure. -----	37
Figure 23: (Color) Correlation between each product and TSG SSS within each box for B-AX02. Number couples indicate the 95% confidence interval of the correlation using a bootstrap procedure. -----	38
Figure 24: (Color) Robust standard deviation of difference between each product and TSG SSS within each box for B-AX01. Number couples indicate the 95% confidence interval of the robust std using a bootstrap procedure. -----	39
Figure 25: (Color) Robust standard deviation of difference between each product and TSG SSS within each box for B-AX02. Number couples indicate the 95% confidence interval of the robust std using a bootstrap procedure. -----	40

Figure 26: Hovmoller plots of mean seasonal difference between products and TSG data for each TSG box (x-axis), for (Top) B-AX01 and (Bottom) B-AX02----- 41

Figure 27: Hovmoller plots of anomaly difference with respect to mean seasonal cycle between products and TSG data for each TSG box (x-axis), for B-AX01. ----- 42

Figure 28: Hovmoller plots of anomaly difference with respect to mean seasonal cycle between products and TSG data for each TSG box (x-axis), for B-AX02. ----- 43

Figure 29: Joint distribution of each product with the TSG data within the entire domain around Antarctica. ----- 44

Figure 30: Same as Figure 29, but restricted to the north of 60°S----- 44

Figure 31: (Color and Numbers) Correlation, mean absolute difference and robust std of difference between each product and the TSG data over the entire domain.----- 45

Figure 32: Correlation, mean absolute difference and robust std of difference between each product and the TSG data over the domain restricted to latitudes north of 60°S.----- 46

Figure 33: Time series of the averaged products and TSG data over the entire Antarctica domain where there are collocations. ----- 47

Figure 34: Collocation data number and comparison metrics between each product and the TSG data along ship tracks from Tasmania to Antarctica. ----- 49

Figure 35: Collocation data number and comparison metrics between each product and the TSG data along ship tracks from South Africa to Antarctica. ----- 50

List of tables

Table 1: Newly processed CCI+SSS products compared to V3.2 in the RR tests. ----- 15

Acronyms

AD	Applicable Document
Aquarius	Aquarius NASA/SAC-D sea surface salinity mission
ATBD	Algorithm theoretical basis documents
CCI	The ESA Climate Change Initiative (CCI) is formally known as the Global Monitoring for Essential Climate Variables (GMECV) element of the European Earth Watch programme
CCI+	Climate Change Initiative Extension (CCI+), is an extension of the CCI over the period 2017–2024
E3UB	End-to-end ECV Uncertainty Budget
EASE	Equal-Area Scalable Earth (EASE) Grid
ECMWF	European Centre for Medium Range Weather Forecasts
ESA	European Space Agency
ISAS	In Situ Analysis System
LEGOS	Laboratoire d'Etudes en Géophysique et Océanographie Spatiales
LOCEAN	Laboratoire d'Océanographie et du Climat, Expérimentations et Approches Numériques
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OTT	Ocean Target Transform
PIRATA	Prediction and Research Moored Array in the Tropical Atlantic
PMEL	Pacific Marine Environmental Laboratory
PSD	Product Specification Document
PUG	Product User Guide
PVASR	Product Validation and Algorithm Selection Report
RD	Reference document
RFI	Radio Frequency Interference
RMSD	Root Mean Square Differences
RR	Round Robin
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Ocean Salinity
SoW	Statement of work
SSS	Sea Surface Salinity
SSSOS	Sea Surface Salinity Observation Service
SST	Sea Surface Temperature
URD	User Requirement Document
WP	Work package
WS	Wind Speed



1 Introduction


1.1 Executive Summary

This document holds the Product Validation and Algorithm Selection Report (PVSAR) prepared by the CCI+SSS team as part of the activities included in the [WP220] of the Proposal (Task 2 from SoW ref. ESA-EOP-SC-AMT-2021-26).

Both similarities and differences can be observed when comparing satellite-derived SSS products with the ISAS dataset. V4.1 shows a slight reduction in seasonal variation biases at high latitudes compared to V3.2, while V4.2 further improves this reduction, highlighting the need for addressing remaining biases in SMAP and Aquarius SSS. Comparing V4.2 to V4.1, it can be observed that correlation improvements are present in areas far from the coast. Detailed comparisons between V4.2 and V3.2 demonstrates reduced differences in latitudinal profiles. In regions at very high latitudes, the CCI+SSS phase 1 has identified V3.2 as overly filtered for ice contamination. However, V4.2 has shown significant improvements in this aspect, allowing for satellite derived SSS to be recovered closer to areas with sea ice. Nevertheless, challenges persist in these regions due to increased uncertainties in satellite SSS measurements in icy waters, potential remaining ice contamination, uncertainties in ISAS SSS data, and the likelihood of increased representativity errors near ice. V4.2 shows significant improvement globally compared to V4.1 and V3.2. V4.2 displays reduced differences along the North Atlantic TSG tracks in high latitude regions and polar research products, except in regions with increased spatial coverage. The time variations of SSS indicate visible improvement in V4.2 over V3.2, particularly along the North Atlantic tracks. The Southern Ocean exhibits enhancements in V4.2 compared to V3.2, especially north of 60°S. Additionally, V4.2 demonstrates increased spatial coverage south of 60°S. However, due to reduced ice filtering in colder regions, larger uncertainties are present compared to the north of 60°S. This observation highlights the persisting challenges associated with sea-ice contamination despite the improvements made. V4.2 improves bias reduction and correlation, with some challenges observed near coastlines and in extremely high latitudes. On a global scale, the RFI-corrected V4.3 demonstrates results that are highly comparable to V4.2. However, since the RFI correction in V4.3 is still in its preliminary stage, further detailed analysis is needed at the regional level to assess its effectiveness in the treated regions.

1.2 Scope

The report summarizes the results of the first round-robin algorithm comparisons for the CCI+SSS phase 2 project. It evaluates the performance of CCI+SSS 4.1, 4.2, and 4.3 global products and five polar research products. The round-robin exercise is global, with focuses on the North Atlantic and Antarctic areas. The primary objective of the exercise is to assess the progress made since the last phase 3.2 final version. In addition, the report provides insights into the strengths and limitations of the products, helping to identify areas for improvement and further development.

	<p align="center">Climate Change Initiative+ (CCI+) Phase 2 Product Validation and Algorithm Selection Report</p>	<p>Ref.: ESA-EOP-SC-AMT-2021-26 Date: 20/06/2023 Version: v4.0 Page: 2 of 65</p>
--	--	--

1.3 References

1.3.1 Applicable Documents

ID	Document	Reference
AD01	CCI+ Statement of Work	SoW
AD02	Product User Guide (PUG)	PUG
AD03	User Requirement Document (URD)	SSS_cci-D1.1-URD-i1r0
AD04	Product Specification Document (PSD)	SSS_cci-D1.2-PSD-v1r4
AD05	Algorithm Theoretical Baseline Document	SSS_cci-D2.3-ATBD_L3_L4-i1r0_v1.1

1.3.2 Reference Documents

ID	Document	Reference
RD01	Alory G., T. Delcroix, P. T��chin��, D. Diverr��s, D. Varillon, S. Cravatte, Y. Gouriou, J. Grelet, S. Jacquin, E. Kestenare, C. Maes, R. Morrow, J. Perrier, G. Reverdin and F. Roubaud, 2015. The French contribution to the Voluntary Observing Ships network of Sea Surface Salinity. <i>Deep Sea Res.</i> , 105, 1-18, doi:10.1016/j.DSR.2015.08.005.	
RD02	Reverdin, G. and Alory, G. (2018) "Monthly binned sea surface salinity, temperature, and density in the North Atlantic subpolar gyre." The French Sea Surface Salinity Observation Service (SSS OS). doi: 10.6096/sss-bin-nasg.	
RD03	Robert R. Sokal & Rohlf, F. James, 1936- joint author (1981). <i>Biometry the principles and practice of statistics in biological research</i> (2d ed). San Francisco W. H. Freeman	
RD04	X. Yin, J. Boutin, P. Spurgeon, Analysis of biases between measured and simulated SMOS brightness temperature over ocean, <i>IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing</i> , doi: 10.1109/JSTARS.2013.2252602, 2013.	
RD05	Bureau International des Poids et Mesures, Guide to the Expression of Uncertainty in Measurement (GUM), JCGM 100:2008, 2008. Available online at http://www.bipm.org/en/publications/guides/gum.html	
RD06	Boutin, J., Y. Chao, W.E. Asher, T. Delcroix, R. Drucker, K. Drushka, N. Kolodziejczyk, T. Lee, N. Reul, G. Reverdin, J. Schanze, A. Soloviev, L. Yu, J. Anderson, L. Brucker, E. Dinnat, A.S. Garcia, W.L. Jones, C. Maes, T. Meissner, W. Tang, N. Vinogradova, B. Ward (2016b), Satellite and In Situ Salinity: Understanding Near-surface Stratification and Sub-footprint Variability, <i>Bulletin of American Meteorological Society</i> , 97(10), doi: 10.1175/BAMS-D-15-00032.1.	
RD07	Vinogradova, N., Lee, T., Boutin, J., Drushka, K., Fournier, S., Sabia, R., Stammer, D., Bayler, E., Reul, N., Gordon, A., Melnichenko, O., Li, L., Hackert, E., Martin, M., Kolodziejczyk, N., Hasson, A., Brown, S., Misra, S., & Lindstrom, E. (2019). Satellite Salinity Observing System: Recent Discoveries and the Way Forward. <i>Frontiers in Marine Science</i> , 6(243), 23p. Publisher's official version : https://doi.org/10.3389/fmars.2019.00243 .	



1.4 Structure of the document

The PVASR is structured as follows:

- Section 1: Introduction.
- Section 2: Definition of key terms used throughout the document.
- Section 3: Overview of the tasks performed, and the comparison results obtained.
- Section 4 describes the Round Robin (RR) methodology. Specifically, Section 4.1 covers the in-situ data (ship tracks) used in the RR tests, while Section 4.2 explains the collocation methodology between in-situ and satellite data for ship tracks. Finally, Section 4.3 describes the metrics utilised in the RR tests.
- Section 5 presents the CCI+SSS satellite products evaluated and compared to in situ data.
- Section 6 presents the results of the evaluation: independent comparisons using the ISAS dataset (Section 6.1), RR exercise in North Atlantic and around Antarctica (Section 6.2), a summary of the most significant comparison features between the products (Section 6.3), and open issues (Section 6.4).
- Additionally, Section 7 outlines potential future perspectives for CCI+SSS PVASR.



2 Definitions

This document includes relevant definitions and considerations from [RD 05] for the SSS product algorithm assessment:

Measurand: quantity subject to measurement, in our case, the salinity, defined as the relative amount of salt dissolved in seawater (corresponding to gram of salt per kilogram of seawater) at the sea surface.

Error: result of a measurement minus a true value of the measurand. Since the 'true' value of the measurand is unknown, the error's 'true' value is unreachable.

Uncertainty: parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand. Uncertainty of measurement comprises, in general, many components. In the case of RR, since comparisons with measurements in the fields validate measurements, 'experimental standard deviations' classically evaluated from the statistical distribution of the results of a series of measurements achieved in the same conditions cannot be estimated. Hence, in the case of RR, the uncertainty is evaluated from assumed probability distributions of the measurand derived, with some uncertainty, from in situ measurements.

In [RD 05], 'it is understood that the result of the measurement is the best estimate of the value of the measurand and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion'. In the case of satellite radiometric measurements, the absolute calibration of the SSS needs to be better known and essential differences between the various satellite SSS come from the different systematic corrections that are applied. Therefore, we will distinguish between 'uncertainties associated with systematic effects' (a bias can quantify that - see below), from the 'uncertainties associated with random errors' coming from the noise of the measurements (linked to the radiometric resolution), from errors that are not well characterized given the present knowledge of the sources of errors.

Discrepancy: The difference between the data product and the validation value.

(Relative) Bias: The mean value of the discrepancy.

Validation: The process of independent assessment means the quality of the data products derived from the system outputs.

Precision: The difference between one result and the mean of several results obtained by the same method, i.e., reproducibility (includes non-systematic errors only).

Observational errors: Observational errors are the ones corresponding to the precision of the instruments, plus, when available, the ones due to inaccurate absolute calibration. The precision of in situ SSS is generally less than 0.01 for an individual measurement. However, the absolute calibration of merchant ships' TSG can be as large as 0.1 for a given transect. For satellite SSS, the



absolute calibration error is usually unknown; the precision is on the order of 0.4 - 0.6 for individual SSS in warm regions as retrieved from Aquarius or SMOS and SMAP, respectively. These observational errors are reduced at level 3 and level 4 according to the number of satellite passes occurring in the same pixel over one week, by roughly a factor $\sqrt{2}$ for Aquarius and a factor 2 to 3 for SMOS and SMAP. Since an absolute reference is usually not available, what is provided in the products is an **observational uncertainty** (see E3UB report).

Sampling errors: According to [RD 07], sampling errors arise when one data type does not represent a process (or scale) that the other does, e.g., due to the differences in their spatial and/or temporal samplings. The “expected” differences, i.e., the low bound at which two estimates are allowed to differ, are in the following called **sampling uncertainties**.

Satellite SSS: Sea Surface Salinity within the first centimetre of the sea surface, by nature integrated over a surface that depends on the radiometer characteristics and the data processing.

In-situ SSS: Near Surface Salinity measured at several cm to several meter depth (see Figure 1).

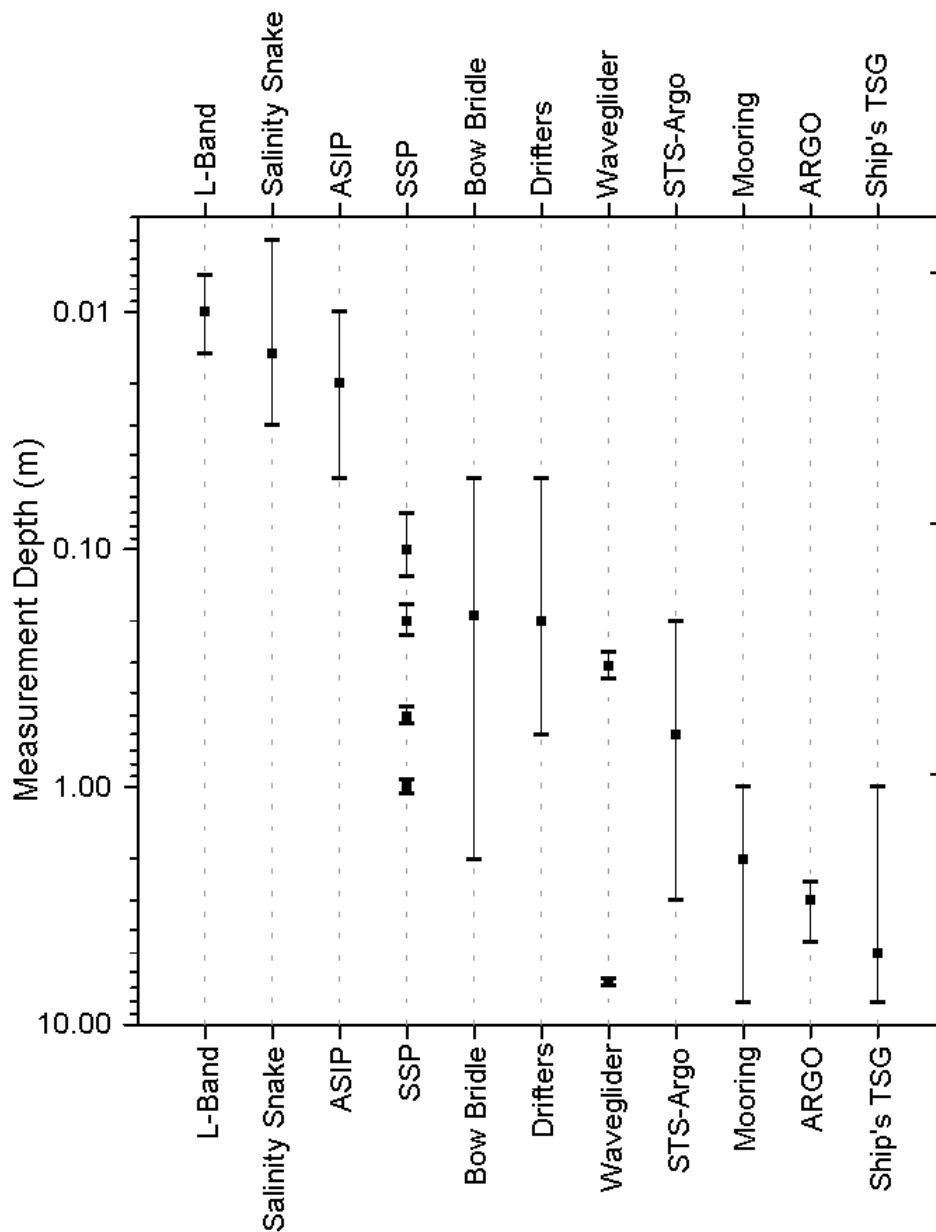


Figure 1: Scale portraying the typical depth at which near-surface salinity is measured by various sensors/platforms. The small squares show the average measurement depth and the capped lines show the range for that average. For profiling platforms (ASIP, Bow Bridle, STS-Argo, Argo) the range represents the variability of the top-most point in the profile. For platforms with standardized configurations that measure at fixed depths (Salinity Snake, SSP, Wave Glider) the mean and range of each sensor at a particular depth are shown. For platforms where there are multiple sensor configurations (drifters, mooring, shipborne TSG) or that sample at different depths depending on the specifics of the platform, the range of measurement depths across all platforms is shown. Radiometric penetration depths were calculated using the Stogryn (1997) relationship and show penetration depths at 1.43 GHz over the salinity range of 20 pss to 38 pss and temperature range of -2°C to 35°C (where the “mean” value shown in the figure is for 20°C and 35 pss). (Figure taken from [RD 06]).



3 Overview

Three metrics are utilized to evaluate the performance of the algorithms/products. These metrics aim to describe uncertainties that arise from distinct types of errors that are managed differently during satellite processing:

- **M1** measures the **robust standard deviation between satellite and in situ SSS**. This metric characterizes random errors that result from measurement noise, poorly characterized error sources, and other factors. Using median calculation instead of mean makes the statistical robustness stronger, as extreme values and outliers less influence the median.
- **M2** calculates the **difference between the satellite and in situ SSS** to identify any bias. This helps to identify any systematic errors caused by factors such as radiometer calibration issues, land-sea contamination, sun contamination.
- A similar metric, the **absolute bias between satellite SSS and in situ SSS**, is used throughout the report to consider only absolute differences.
- **M3** measures the **coefficient of determination between satellite SSS and in situ SSS**. Therefore, it is sensitive to overly stringent filtering or smoothing of extremum values, such as low SSS in river plumes.
- **M4** evaluates the statistical distribution properties of the **centred reduced variable** and provides insights on the appropriateness of CCI L4 SSS uncertainties. Therefore, M4 will not be assessed in this report.

We evaluate the significance level of each metric M1, M2, and M3 for each case. To provide a reliable estimate, we accompany each calculation with a bootstrap procedure that provides a 95% confidence interval. It is important to note that the bootstrap method only estimates the metrics' sampling distribution and does not consider observational and sampling uncertainties currently. Nonetheless, we can use the comparisons between confidence intervals for different products to evaluate them against each other. This approach enables us to assess the products and provide valuable insights into the satellite-derived SSS.

In this report, we have utilized multiple datasets to ensure an evaluation of the satellite-derived SSS products that have been available so far. Specifically, for this first implementation of the PVASR, we have focused on the polar regions and products using repetitive ship tracks across the Northern Atlantic Ocean, and ship tracks around the Antarctic. In addition, for a more general evaluation of the global SSS products, we have incorporated the ISAS 17+NRT datasets. Finally, we also use ISAS 20 as an in-situ gridded product for intercomparison with the satellite datasets.

We co-locate satellite and in-situ data. The results of the product validation exercise demonstrate that the new version 4.2 of CCI+SSS provides superior global results compared to the previous product 3.2. Additionally, the results indicate that V4.2 is more dependable than the top-performing product among the polar research products in the tested regions. Furthermore, at the global level, the RFI-corrected V4.3 strongly agrees with V4.2. However, as the RFI correction in V4.3 is currently in its preliminary phase, a more thorough analysis is required at the regional level to evaluate its effectiveness, specifically in the treated areas.



Climate Change Initiative+ (CCI+) Phase 2
Product Validation and Algorithm
Selection Report

Ref.: ESA-EOP-SC-AMT-2021-26

Date: 20/06/2023

Version: v4.0

Page: 8 of 65

In this report, we do not focus on the very high latitudes of the Arctic Ocean, such as the Beaufort and Chukchi Seas, Nordic Seas, and Barents Seas. Instead, specific validations will be performed there by the validation team involved in the CCI+SSS option.



4 Round robin methodology

This section outlines the RR test methodology used to compare and validate various satellite products. First, assessments are done at a global scale using ISAS products. Then, we focus on the Atlantic Ocean, specifically the northern part, due to contrasting SSS regimes, RFIs in certain areas, and independence from the SMOS Ocean Target Transformation region. This region is constantly monitored through in-situ measurements along repetitive ship tracks, which include maximum and minimum SSS regions. Consequently, the data quality can be monitored throughout the satellite period.

The RR method is a validation technique used to compare different algorithms (and related products), where we apply the different algorithms to the same set of in-situ data, and the results obtained with various CCI+SSS versions are compared. In general, this method is used to assess the different algorithms' performance and determine which is most suitable for a particular application.

In the context of PVASR work, the round-robin method is applied by selecting a set of validation data and systematically using it for each algorithm evaluation. Next, each algorithm is compared to the data, and the results are compared using standard validation metrics. The PVASR report then presents a comprehensive analysis of the performance of each algorithm, including statistical measures such as bias, standard deviation, and correlation coefficients. The report also provides algorithm selection recommendations based on the round-robin analysis results.

4.1 In situ data

4.1.1 Monthly binned ship tracks in North Atlantic

These data are derived from measurements made by merchant ships using thermosalinographs (TSG) along two transects: B-AX01 between southern Greenland and Denmark, and B-AX02 between Newfoundland and Iceland (Figure 2), monthly averaged in geographic boxes of a typical size of 150 km on a side. These monthly averages are then temporally smoothed using a three-month sliding average with coefficients 1-2-1. The measurement depth can vary from 5m to about 10m. The boxes will be numbered from west to east along the two transects. [RD 02] provides a detailed description of these data. These data now cover the period 1993-2018. The dataset used is available on the Laboratory for Studies in Geophysics and Spatial Oceanography (LEGOS) website at the following address: <https://doi.org/10.6096/SSS-BIN-NASG>.

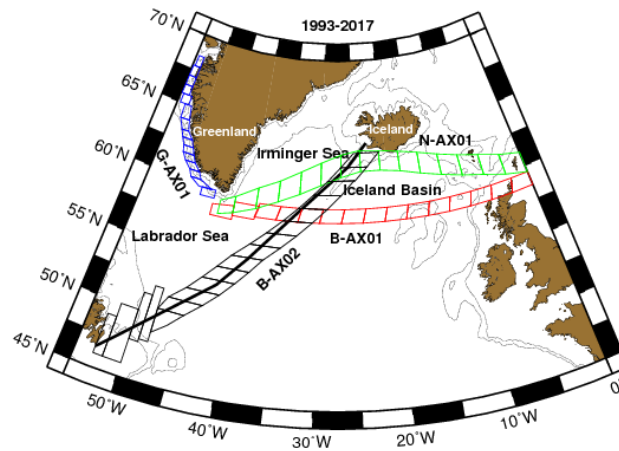


Figure 2: Map of boat transects and their division into regular boxes. An example of a ship track is shown as a solid black line along B-AX02 (Reverdin et al., 2018). In this work, only the B-AX01 and B-AX02 transects are used.

4.1.2 Research vessel TSG in the Southern Ocean

The Southern Ocean has several in-situ datasets, including recurrent measurements along specific paths over the years (Figure 3). The temporal coverage of seasons and year-to-year variations are essential for the RR task, and the recurrence over the years provides this.

For this report, we have considered two ensembles of paths that are relatively well-sampled and only measure at depths above 11m. The initial path group is focused on a straight line that starts from the southernmost point of Africa and goes towards the 0° meridian, which then divides and heads towards Antarctica (Figure 4). The second path is centred on a straight line near the 145°E meridian from Tasmania to Antarctica. These data ensembles were collected from the Pi-MEP facility, which is available at <https://www.salinity-pimep.org/>:

- The TSG-NCEI-0170743 dataset contains sea surface temperature and salinity data collected between 2010 and 2017 in the South Atlantic Ocean and Southern Ocean from SA Agulhas and Agulhas-II research vessels as part of scientific activities by the South African National Antarctic Programme (SANAP), South African Department of Environmental Affairs (DEA), and Italian National Antarctic Research Programme (PNRA). The measurements were obtained through a thermosalinograph (TSG) during several cruises to Antarctica and sub-Antarctic islands, and the TSG devices were regularly calibrated and monitored.
- The TSG-POLARSTERN dataset has been gathered through the <https://www.pangaea.de/> data warehouse utility from 2010/01/01 to the present.
- The TSG-LEGOS-Survostral (<https://www.legos.omp.eu/survostral/>) dataset is a collection of delayed mode regional data from the TSG installed on the Astrolabe vessel during round trips between Hobart and the French Antarctic base at Dumont d'Urville. The Survostral project provides it and is available via FTP. The dataset contains modified data and only TSG data marked with quality flags 1 and 2 (i.e., including only good-quality data).
- The TSG-LEGOS-DM dataset (<https://www.legos.omp.eu/sss/>) collects sea surface salinity data from voluntary observing ships. The French Sea Surface Salinity Observation Service validates

and archives the data, ensuring that only high-quality data flagged as 1 or 2 are included in the dataset. When possible, we utilize adjusted values.

Platform Names within ~25x25km² Boxes

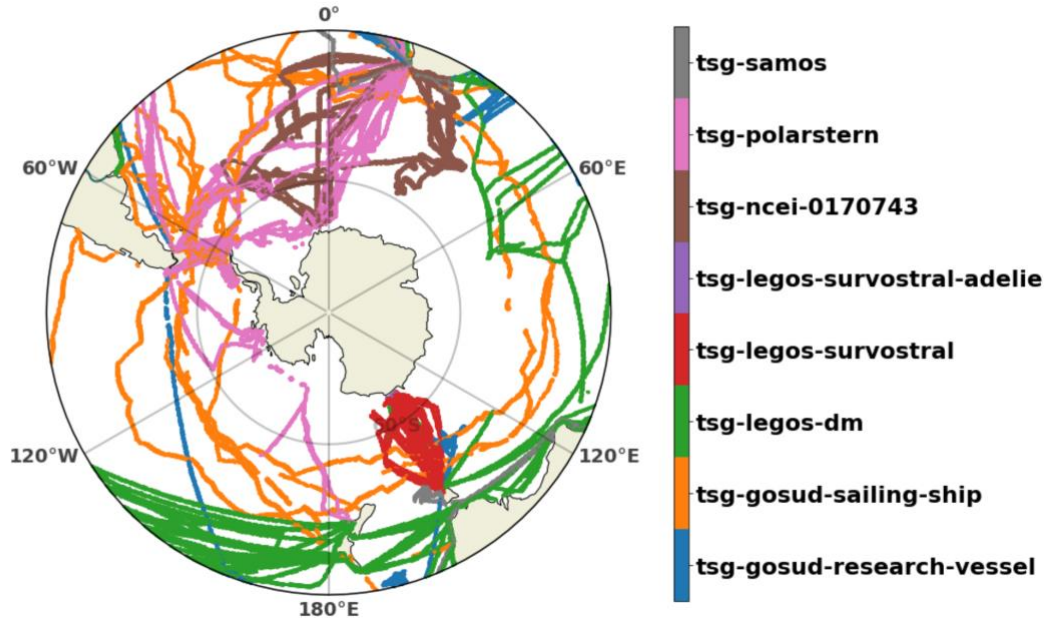


Figure 3: TSG tracks spanning from 2010 to 2022, with each type of TSG cruise being represented by a different color.

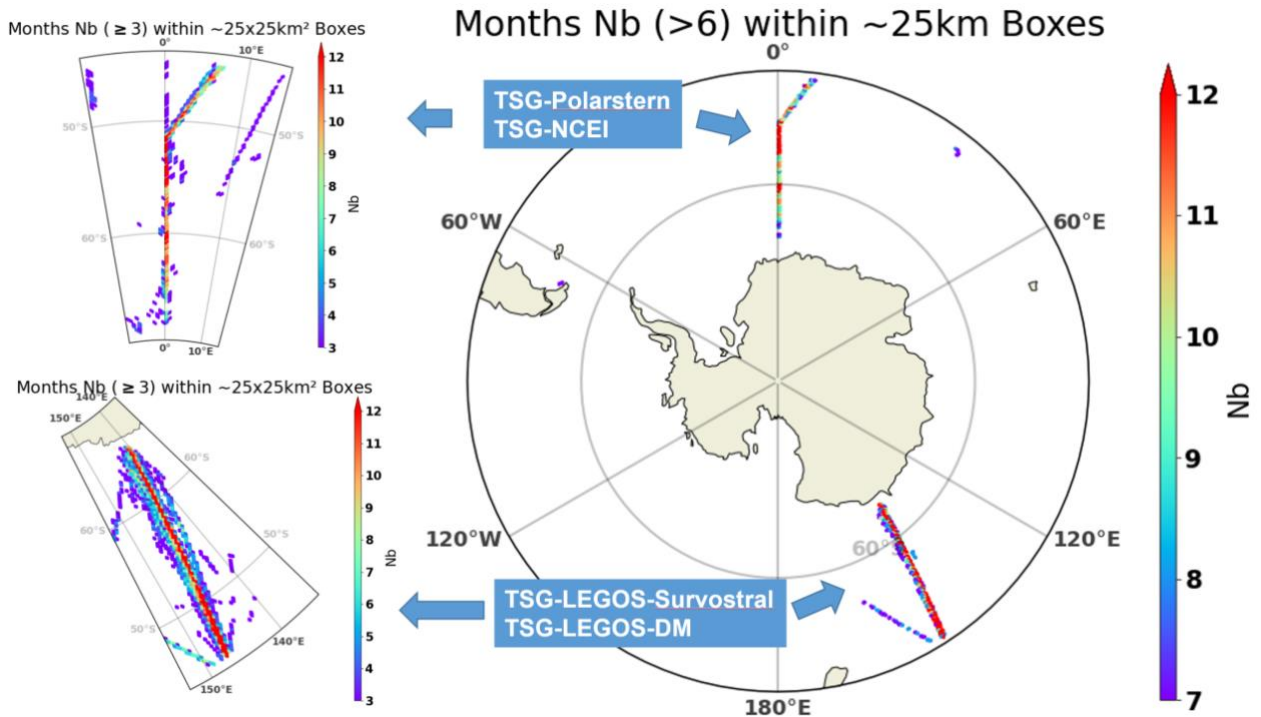



Figure 4: Number of monthly-averaged TSG data within about 25x25 km² boxes along recurring ship tracks around Antarctica.

	<p>Climate Change Initiative+ (CCI+) Phase 2 Product Validation and Algorithm Selection Report</p>	<p>Ref.: ESA-EOP-SC-AMT-2021-26 Date: 20/06/2023 Version: v4.0 Page: 12 of 65</p>
--	---	---

4.1.3 Gridded in-situ datasets

Gridded in-situ datasets based on Argo float surface data and additional in-situ platform data provide non-satellite gridded realizations of the SSS field and temporal variations.

We use the following two in-situ products for intercomparisons:

- ISAS-17

ISAS-17 provides a monthly global analysis of sea surface salinity and temperature, with a horizontal resolution of $0.5^{\circ} \times 0.5^{\circ}$ and coverage of the 5m depth used for comparison with satellite observations. This analysis employs the most recent statistics and ISAS version 8 to interpolate data from various in-situ measurements, such as Argo and Deep-Argo profiles, to supplement areas where Argo sampling is limited, especially at higher latitudes. The dataset spans from 2002-2017. In addition, this ISAS 17 dataset is extended to 2022 via a delayed time and near real-time processing of updated observations (ISAS delayed mode, ISAS NRT).

- ISAS-20

ISAS-20 (only Argo) is a dataset that provides interpolated temperature and salinity measurements at different depths and locations in the global ocean. It was created using the ISAS version 8 and updated statistics, covering 2002 and 2020. The dataset is based on Argo and Deep-Argo data and is gridded on a $0.5^{\circ} \times 0.5^{\circ}$ horizontal grid with coverage of 5m depth level (among others). We also provide monthly climate data and annual standard deviation values.

Note that the ISAS-17 and ISAS-20 datasets available on SEANOE were of degraded quality in the Arctic Ocean due to issues with the SSS climatology used prior to the ISAS optimal analysis. Therefore, in this report, we use ISAS-17 and ISAS-20 versions reprocessed by N. Kolodziejczyk employing an updated climatology as prior (N. Kolodziejczyk, pers. comm.).

4.2 Colocation methodology

Our approach to colocation focuses on satellite measurements. This entails adapting the sampling of the on-site measurements to ensure that it is representative of the data from the satellite product being analysed.

4.2.1 Monthly binned ship tracks in northern Atlantic

We averaged the satellite data over a month to compare satellite products with the monthly binned ship tracks dataset. Figure 2 describes the boxes utilised during the averaging process.

4.2.2 Monthly binned ship tracks in Antarctica

When evaluating products in the Antarctic region, we begin by focusing on predefined 25×25 km² grid bins that have frequent ship track measurements. These measurements are then averaged over a month within each bin. We use the averaged location of TSG data as a reference for collocating satellite data instead of the center of the box. This generates collocation pairs



relative to the 25 x 25 km² pixelization of the domain and monthly satellite products are matched to their respective averaged TSG locations. Once collocations are completed, metrics are computed within 150km radius discs. The Earth is divided into an approximate ~25 x 25 km² grid using the Healpix pixelisation method which can be accessed at <https://healpix.jpl.nasa.gov/>.

4.3 Metrics

For selecting the algorithm, the metrics introduced in section 3 are computed as described below (*horizontal bars indicate the mean over a set of measurements*).

- Standard deviation of the differences (*std diff*):

$$std_diff = \sqrt{\overline{(SSS_{satellite} - SSS_{in-situ})^2} - \overline{(SSS_{satellite} - SSS_{in-situ})}^2}$$

- Robust standard deviation (*std diff rob*, or M1 in section 3):

$$\frac{\overline{median(|(SSS_{satellite} - SSS_{in-situ}) - median(SSS_{satellite} - SSS_{in-situ})|)}}{0.6745}$$

- Bias (or M2 in section 3):

$$bias = \overline{SSS_{satellite} - SSS_{in-situ}}$$

- Mean absolute difference (*mad*):

$$mad = \overline{|SSS_{satellite} - SSS_{in-situ}|}$$

- Coefficient of determination (r^2) of the linear regression y between in-situ SSS and satellite SSS (or M3 in section 3):

$$r^2 = \max \left[0, 1 - \frac{\sum(Y - y)^2}{\sum(Y - \bar{Y})^2} \right], \text{ with } Y = SSS_{satellite} - SSS_{in-situ}$$

We indifferently call r^2 the coefficient of determination and c the correlation coefficient

- Standard deviation of the reduced centered difference (*std diff cr*).

The SSS difference is divided by the time and space varying uncertainty magnitude as given below:

$$\Delta SSS_{cr} = \frac{\Delta SSS}{sat_{uncertainty}}$$



for each measurement point, $sat_{uncertainty}$ is the satellite uncertainty at that point (random uncertainty estimated from the L4 generation).

The std is calculated using this scaled and non-dimensional formulation instead of $SSS_{satellite} - SSS_{in-situ}$. In addition, variable centring is applied as part of the std diff formulation above.

- Robust std of the reduced centered difference (*std diff cr rob*):

As mentioned above, using the median calculation.

In these equations, $SSS_{in-situ}$ corresponds to the salinity of the in-situ measurements, after the colocation processing described in section 4.2.

$SSS_{satellite}$ corresponds to the salinity sensed by the satellite.

In addition, we compute the $rmsd$:

$$rmsd = \sqrt{(SSS_{satellite} - SSS_{in-situ})^2}$$

We do not use $rmsd$ for evaluating the algorithm as this coefficient is related to the bias and std diff that are already considered separately:

$$rmsd = \sqrt{std_diff^2 + bias^2}$$

5 Description of the algorithms & ancillary data tested during the round robin exercise

During this RR test, we evaluated nine versions of the CCI+SSS products (Table 1), including three global product versions and five polar research product versions. These versions were newly processed during CCI phase 2, and the final version of phase 1 was considered a previous reference. Therefore, our report focuses on the comparisons between CCI+SSS V3.2 (phase 1) and global V4.1, V4.2, V4.3 (phase 2), as well as polar research products Res0, Res1, Res1noWS, Res3, and Res4. The comparisons were made using CCI monthly fields.

Table 1: Newly processed CCI+SSS products compared to V3.2 in the RR tests.

Version	Changes with respect to previous version
V4.1	<p>With respect to 3.2</p> <p>Change in input L2 data:</p> <ul style="list-style-type: none"> • Official SMOS L2 SSS v7 products instead of CCI v671 reprocessing. • SMAP RSS v5 instead of SMAP RSS v4. <p>Added/modified corrections:</p> <ul style="list-style-type: none"> • Addition of a SSS bias depending on wind speed and ocean state. • Rain rate correction now depends on wind speed. • Correction for dielectric constant: BVZ model instead of BV. • Seasonal latitudinal correction adjusted on mean latitudinal profiles (instead of median latitudinal profiles). • Use ISAS SSS mean latitudinal profile (instead of SMOS best dwell line median profile) for performing latitudinal seasonal correction. • Added corrections depending on SST and WS for SMAP and Aquarius. • Change in ice flag: use of Acard < 40. • Regular 0.25° grid instead of EASE 2 global grid.
V4.2	<p>Same as V4.1 plus:</p> <ul style="list-style-type: none"> • Seasonal latitudinal correction for SMAP and Aquarius.
V4.3	<p>Same as V4.2 plus:</p> <ul style="list-style-type: none"> • RFI corrections around three strongly contaminated areas (Samoa, Barbados and the Gulf of Guinea). • A slight adaptation of the bias correction parameters (sigbias) to optimise RFI correction. • Correction of ice flagging from Jan 2010 to May 2010 (CATDS data)
Res0 Arctic only	<p>With respect to 3.2</p> <p>Identical, except processing is switched to EASE2 polar grid.</p>
Res1 Arctic only	<p>SMAP RSS v5 is the reference product for calibrating the other products (SMOS and Aquarius). However, L4 objective analysis procedure remains the same.</p> <p>Rationale: The L4 construction method involves identifying an internal SSS reference among the three products - SMOS, Aquarius, and SMAP. Once the reference product is identified, a seasonal correction is calculated during its overlapping period with the other two products</p>



	and then propagated to the other products. Based on seasonal biases, SMAP has been determined as the most suitable reference product compared to Aquarius and SMOS.
Res1noWS Arctic + Antarctic	Same as Res1 plus: the used SMOS L2 version is that resulting from the L2OS inversion without wind speed retrieval
Res3 Arctic only	Same as Res1 plus: <ul style="list-style-type: none"> the time collocation between SMOS and SMAP is set to 3 days; the absolute product calibration is carried out using the ISAS quantiles. <p>Rationale: For Res0, Res1, and Res1_noWS, collocations between SMOS and SMAP are conducted using monthly-averaged products. However, the ice edge can exhibit significant variability within a month. Therefore, collocations should be performed within a window of one week at most.</p> <ul style="list-style-type: none"> The absolute calibration using ISAS is a test for the general improvement of the product. This is done only for that specific Res3 product.
Res4 Arctic only	Same as Res3 plus: No absolute calibration using ISAS.

5.1 Global products

Three versions of the global CCI+SSS L4 product have been generated and tested in the first part of the CCI+SSS phase 2 project. We summarize in Table 1 the main changes in each version.

5.2 Polar research products

As we progress through phase 2 of CCI+SSS, we have generated five polar research L4 products in the northern hemisphere (from 45°N northward) and one polar research L4 product in the southern hemisphere (from 45°S southward). These products have been built using the following source data:

- CCI SMOS L2 data, specifically the v671 data obtained from the ERA5 auxiliary data over the polar grids.
- RSS SMAP L3 v5 data.
- Aquarius v5 data.

The construction of the five polar research L4 products involves similar processing techniques, with some variations aimed at enhancing the initial product quality (Table 1).

The report utilises the 30-day time-averaged configuration of the polar products instead of the 7-day version yet to be tested.



6 Algorithm/Product evaluation

6.1 Verification at global scale

Below, we present the significant similarities or differences that we noticed when comparing to ISAS 17 + delayed mode + nrt (we will refer to it as ISAS 17+NRT for simplicity). The difference magnitudes helped us identify enhancements or deteriorations in the tested versions.

Note: ISAS SSS comparisons are preferred here due to their ease of handling and ability to provide global and regular spatiotemporal samplings. Despite being utilised in the CCI+SSS processing, the influence of ISAS SSS is limited to very large-scale and protracted temporal statistics. The verifications presented below are based on comparing CCI L4 maps with ISAS 17+NRT. Only pixels with PCTVAR lower than 80% are considered to ensure a significant signal-to-noise ratio.

In V4.1, we observed a slight reduction in the seasonal variation of biases at high latitudes compared to V3.2 (Figure 5a and b). However, in V4.2, this reduction became even more significant, highlighting the importance of addressing remaining seasonal latitudinal biases in SMAP and Aquarius SSS (Figure 5c and d). Additionally, Figure 6 shows a decrease in the robust standard deviation of difference. In areas far from the coast, correlation has mostly increased from V4.1 to V4.2, particularly over the global ocean (Figure 7c). However, closer to the coast, the improvement in correlation is less evident (Figure 7d), possibly due to V4 not going as close to the coast where the maximum variability occurs. This is because SMOS L2 SSS v7 was more filtered close to the coast than SMOS L2 CCI v671. The CCI L4-ISAS SSS differences normalized by the CCI L4 SSS uncertainties behave similarly in V4 as in V3 (Figure 8). This is worth noting that the above comparisons were obtained using monthly CCI products, and we also observed improvements in weekly products.

In light of the significant improvement achieved with V4.2 compared to V4.1, we conduct more detailed comparisons between V4.2 and V3.2 here. The latitudinal profiles demonstrate a marked reduction in both the mean difference (as shown in Figure 9) and the std differences (as shown in Figure 10) with V4.2, particularly within the latitudinal range of 60°N to 60°S. V4.2 also exhibits a notable decrease in seasonal biases in the high latitudes. However, in very high latitudes, the extent of CCI L4 is increased (due to the alteration in ice flagging), and the mean difference and standard difference remain high. This could be attributed, in part, to uncertainties in ISAS 17+NRT SSS, which may be smoothed over large scales in these regions relative to their variability scales.

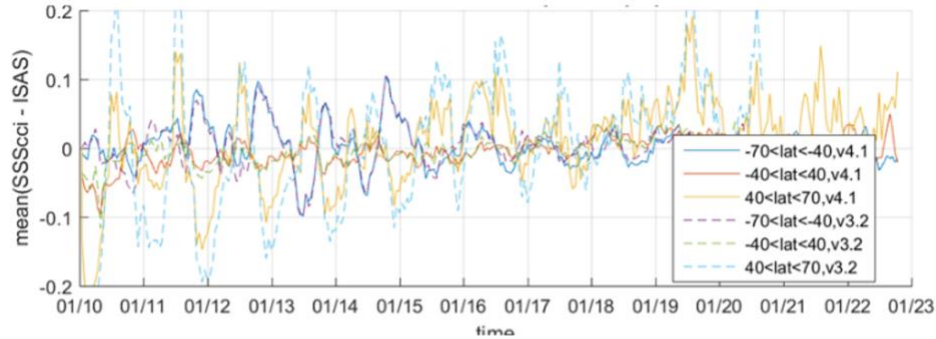
The RMS difference maps of CCI L4-ISAS SSS during the overlapping period of V3.2 and V4.1 (shown in Figure 11) exhibit similarities, except for areas near the coast or islands, where the stronger RFI filtering in SMOS L2 v7 eliminates several points. Additionally, SMOS L2 v7 has a more extensive land mask than SMOS L2 V671.



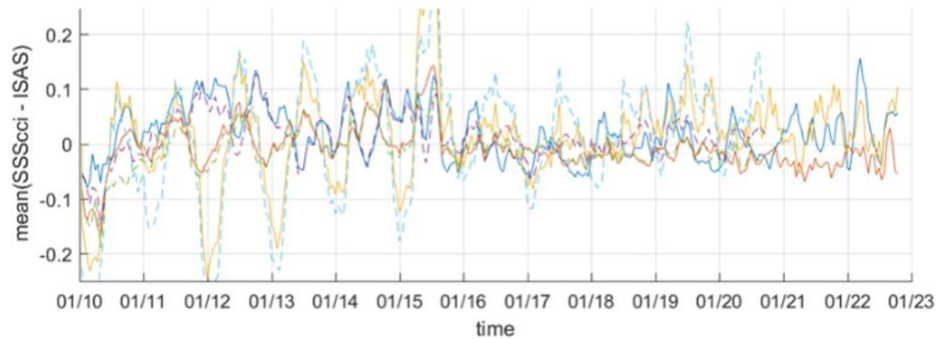
Furthermore, comparing a monthly number of CCI+ SSS estimates between version 3.2 and V4.2 reveals significant differences (Figure 12). In V4.2, there is a higher number of SSS estimates observed at high latitudes, while V3.2 incorrectly removes a substantial number of data points.

The CCI+SSS version 4.3 is similar to version 4.2, except for implementing RFI corrections in three specific areas measuring approximately 20° latitude by 20° longitude. These regions are heavily affected by RFI sources. Global comparisons between V4.3 and V4.2 indicate a high level of similarity, as demonstrated in Figure 13 and Figure 14. This outcome is expected due to the limited extent of the treated regions for RFI. Preliminary evaluations conducted in these specific areas revealed improvements in comparison to the ISAS reference dataset, but some slight degradation was also observed at few points (not shown). Further analysis is necessary in these regions, particularly by incorporating additional available in-situ references such as mooring data.

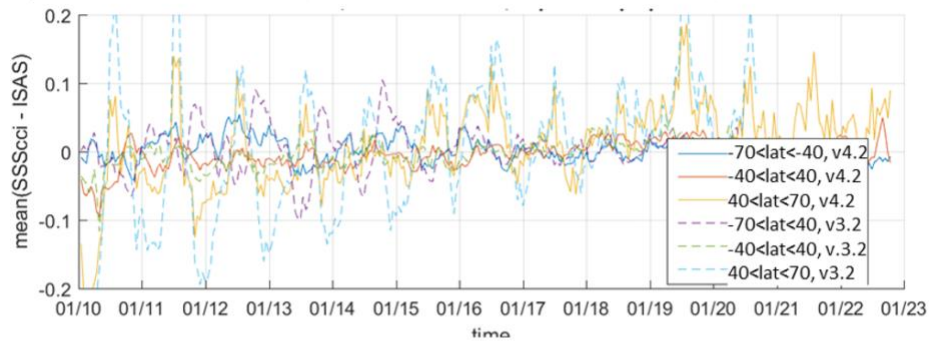
a) Median difference; distance to coast >1000km; v3.2 and v4.1



b) Median difference; distance to coast <1000km; v3.2 and v4.1



c) Median difference; distance to coast >1000km; v3.2 and v4.2



d) Median difference; distance to coast <1000km; v3.2 and v4.2

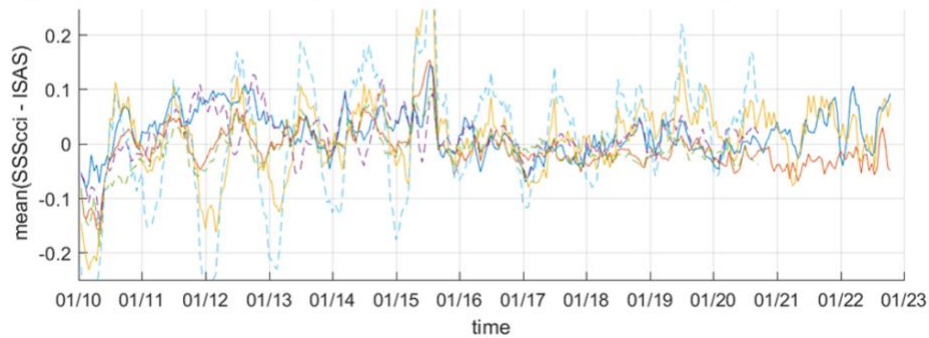


Figure 5: Median difference between CCI L4 and ISAS 17+NRT SSS in three latitudinal bands (70S-40S, 40S-40N, 40N-70N) obtained with a) V3.2 and V4.1 further than 1000km from coast, b) V3.2 and V4.1 at less than 1000km from coast, with c) V3.2 and V4.2 further than 1000km from coast, d) V3.2 and V4.2 at less than 1000km from coast.

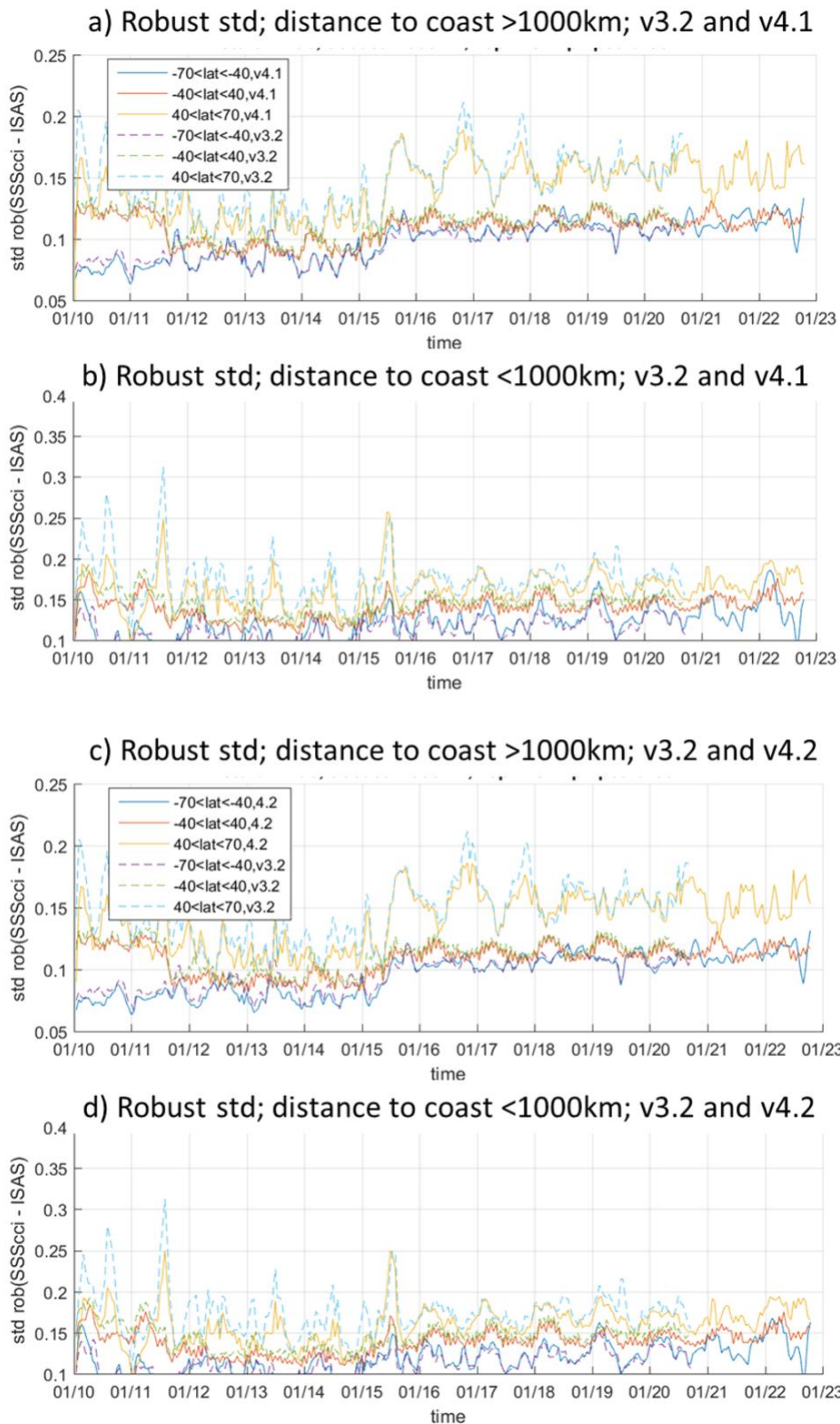


Figure 6: Robust std difference between CCI L4 and ISAS 17+NRT SSS in three latitudinal bands (70S-40S, 40S-40N, 40N-70N) obtained with a) V3.2 and V4.1 further than 1000km from coast, b) V3.2 and V4.1 at less than 1000km from coast, with c) V3.2 and V4.2 further than 1000km from coast, d) V3.2 and V4.2 at less than 1000km from coast

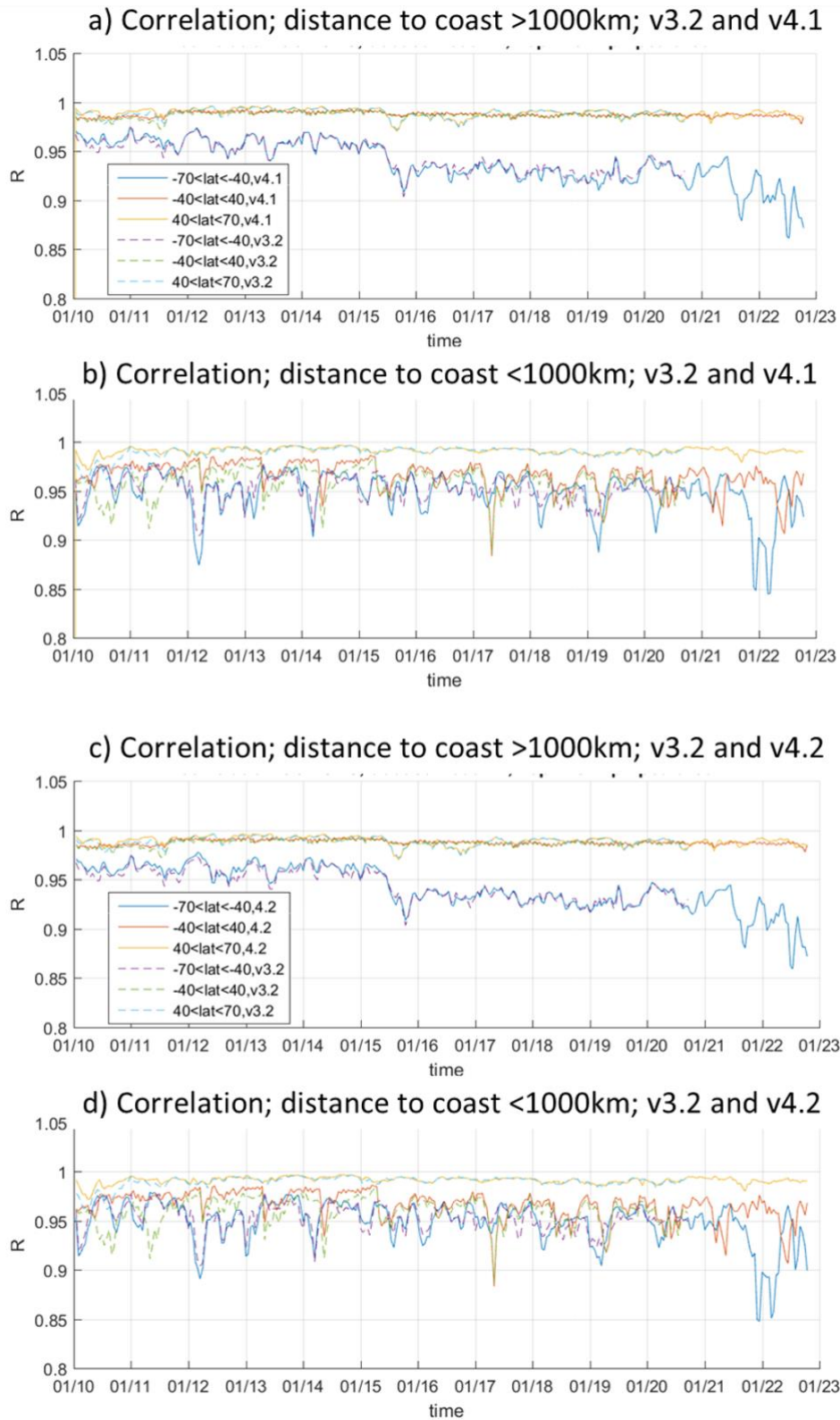


Figure 7: Correlation coefficient between CCI L4 and ISAS 17+NRT SSS in three latitudinal bands (70S-40S, 40S-40N, 40N-70N) obtained with a) V3.2 and V4.1 further than 1000km from coast, b) V3.2 and V4.1 at less than 1000km from coast, with c) V3.2 and V4.2 further than 1000km from coast, d) V3.2 and V4.2 at less than 1000km from coast

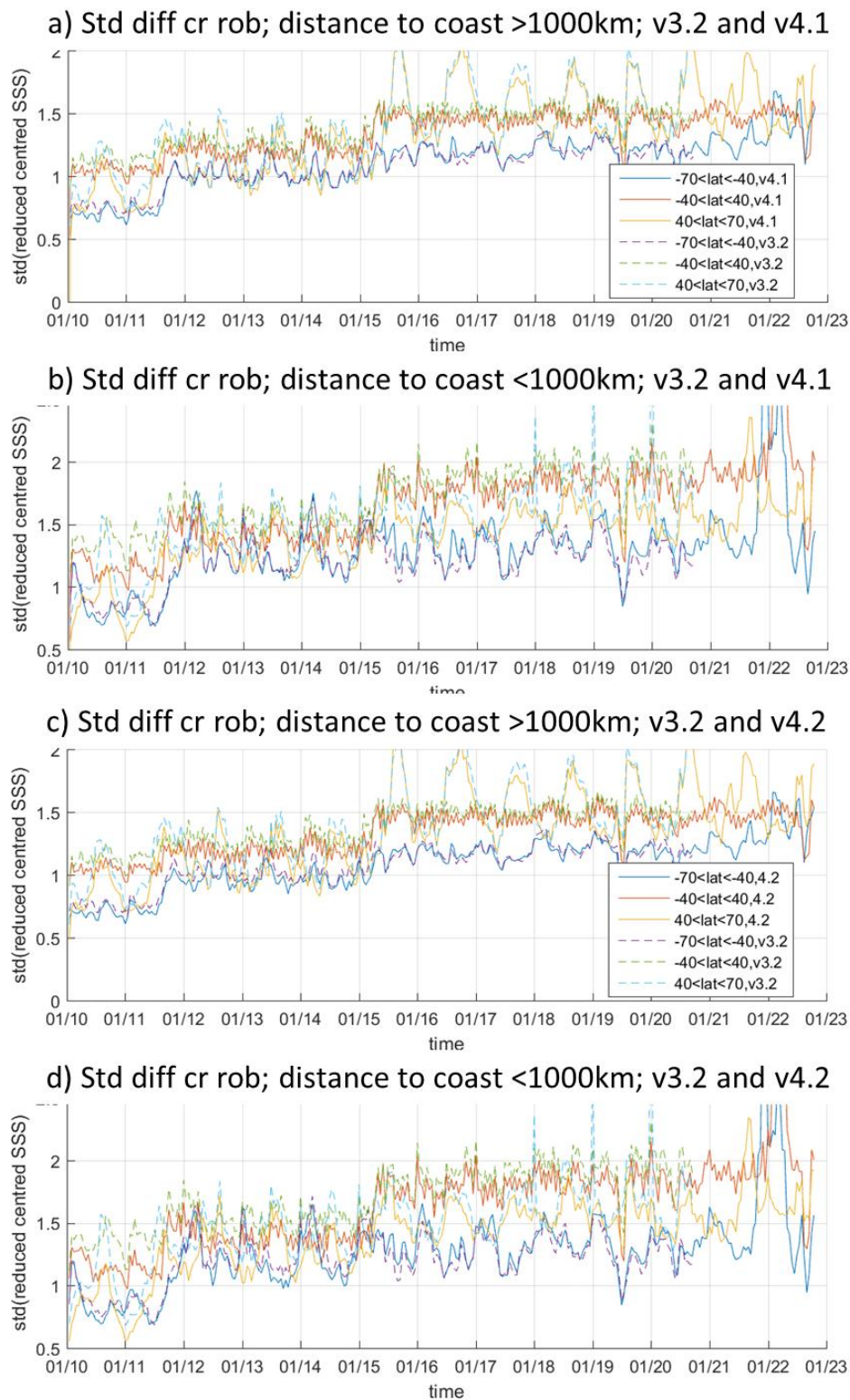


Figure 8: Robust standard deviation of the centered reduced difference between CCI L4 and ISAS 17+NRT SSS in three latitudinal bands (70S-40S, 40S-40N, 40N-70N) obtained with a) V3.2 and V4.1 further than 1000km from coast, b) V3.2 and V4.1 at less than 1000km from coast, with c) V3.2 and V4.2 further than 1000km from coast, d) V3.2 and V4.2 at less than 1000km from coast



CCI v3.2

CCI v4.2

Mean for CCI_3.2-E25-G30d-R0.25-AD-S0.5-L0.5-I0.5- - ISAS-I0.50-S30d-R0.25-AD-P80- [Global]

Mean for CCI_4.2-E25-G30d-R0.25-AD-S0.5-L0.5-I0.5- - ISAS-I0.50-S30d-R0.25-AD-P80- [Global]

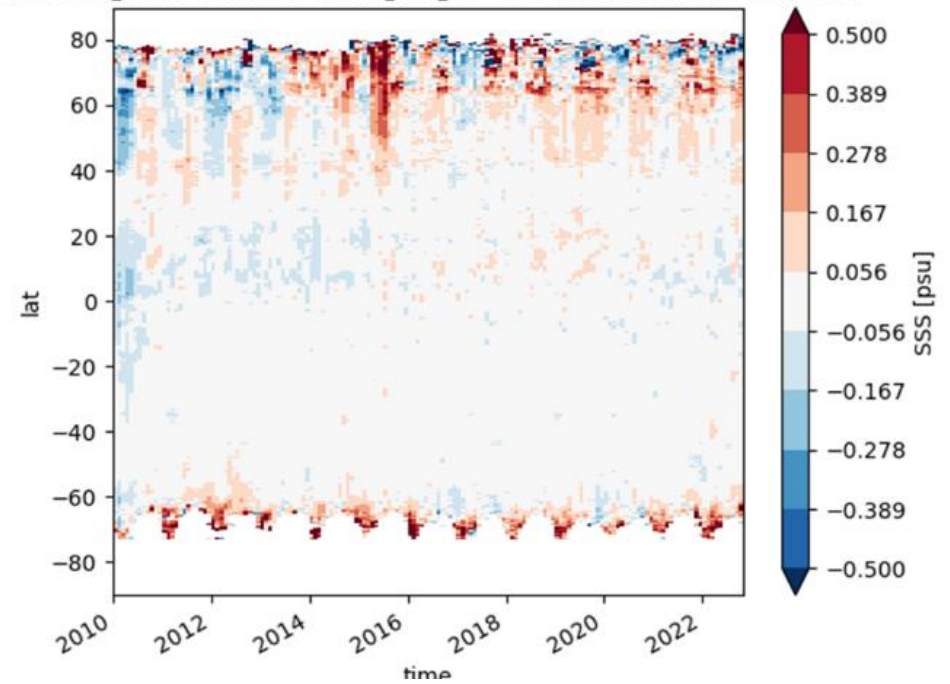
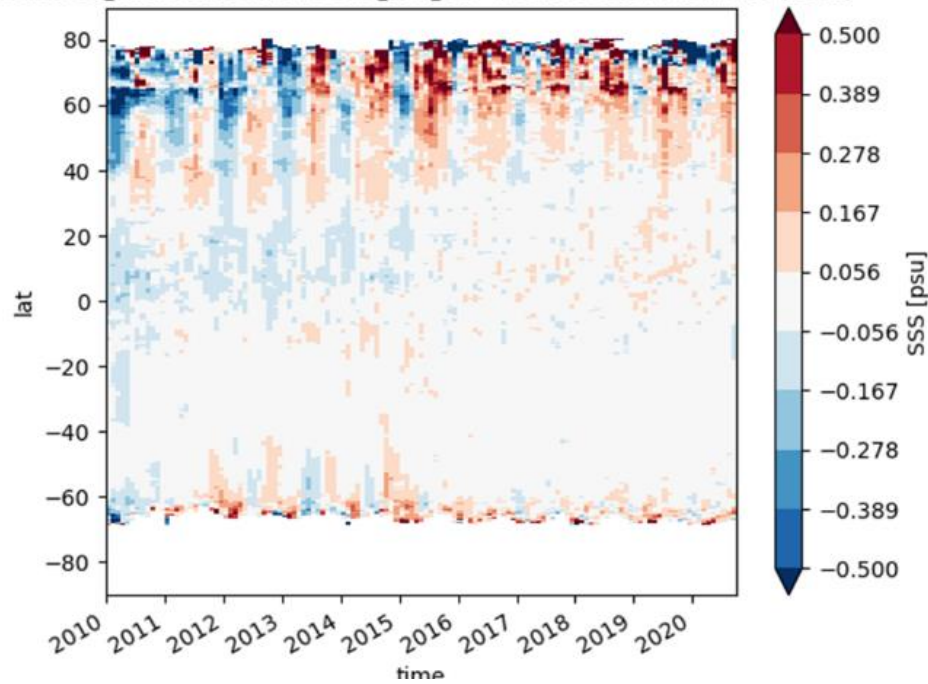


Figure 9: Mean differences between CCI L4 and ISAS 17+NRT SSS averaged over all longitudes as a function of latitude and time, (Left) with CCI V3.2 and (Right) with CCI V4.2.



CCI v3.2

CCI v4.2

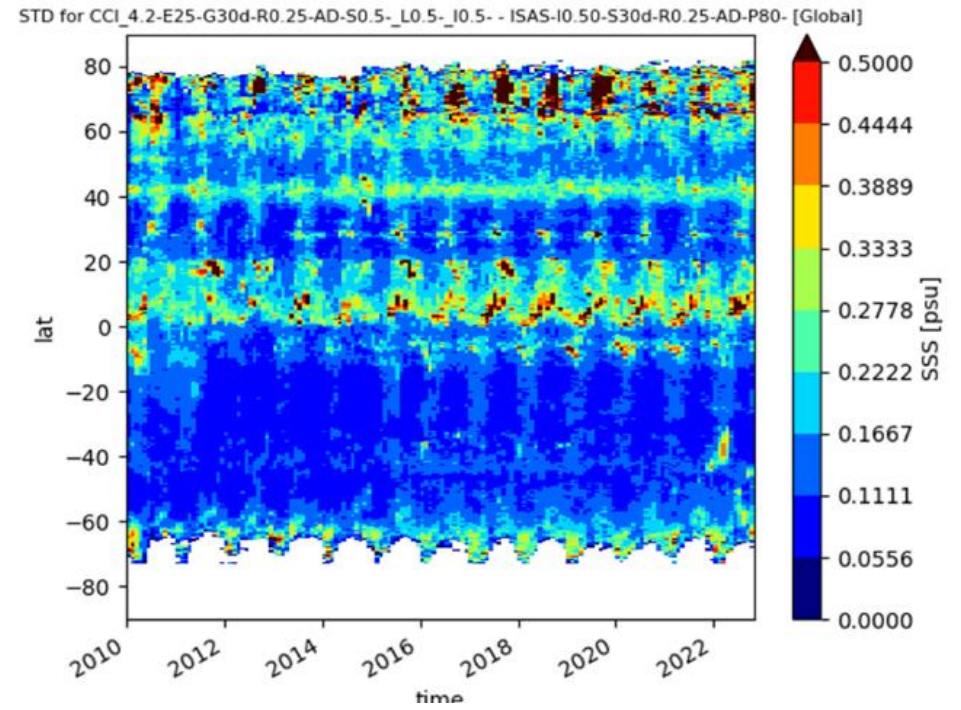
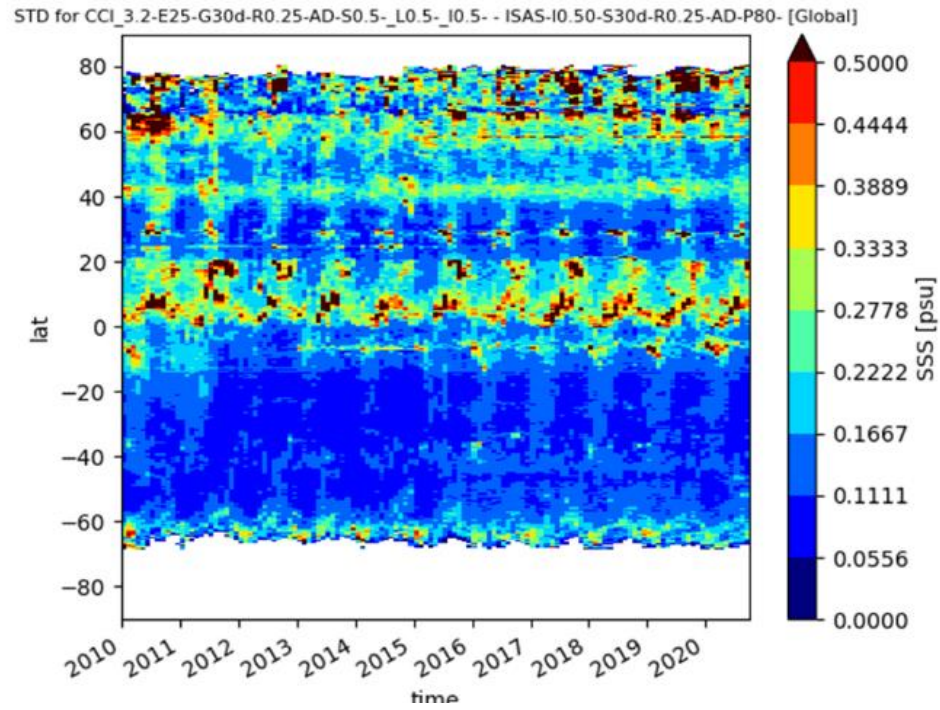
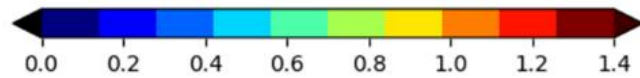
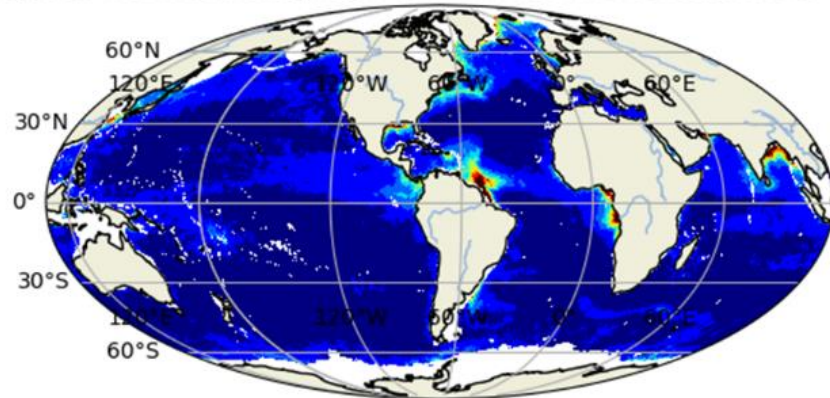


Figure 10: Std differences between CCI L4 and ISAS 17+NRT SSS averaged over all longitudes as a function of latitude and time, (Left) with CCI V3.2 and (Right) with CCI V4.2.



CCI v3.2

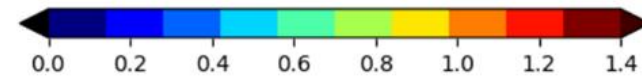
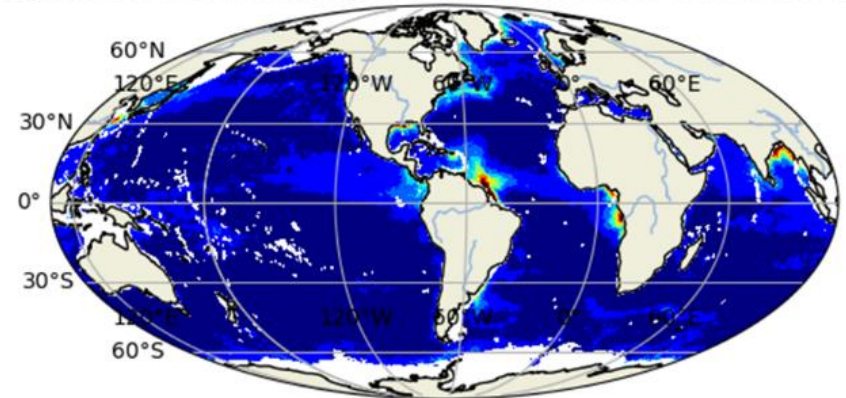
STD, CCI_3.2-E25-G30d-R0.25-AD-S0.5-_L0.5-_I0.5- vs ISAS-I0.50-S30d-R0.25-AD-P80-, 2010-01-15 - 2020-09-15



(C) LOCEAN, 2023-04-26

CCI v4.2

STD, CCI_4.2-E25-G30d-R0.25-AD-S0.5-_L0.5-_I0.5- vs ISAS-I0.50-S30d-R0.25-AD-P80-, 2010-01-15 - 2020-09-15



(C) LOCEAN, 2023-04-26

Figure 11: Rms difference between CCI L4 SSS and ISAS 17+NRT SSS for (Left) V3.2 and (Right) V4.2.

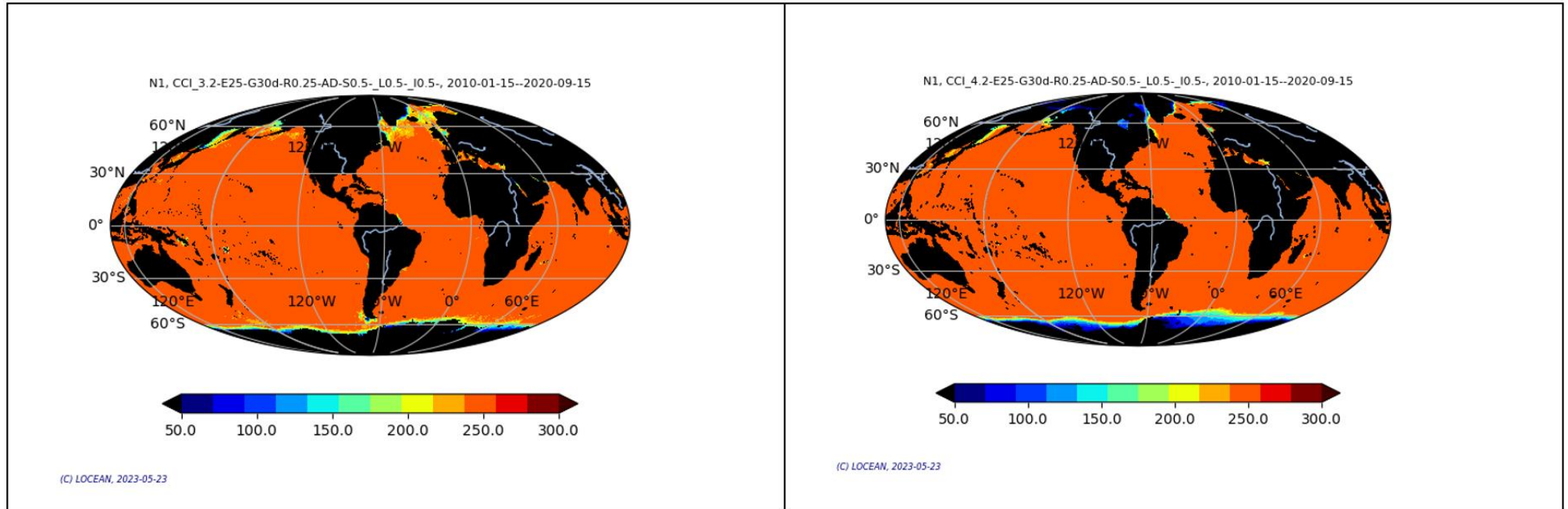
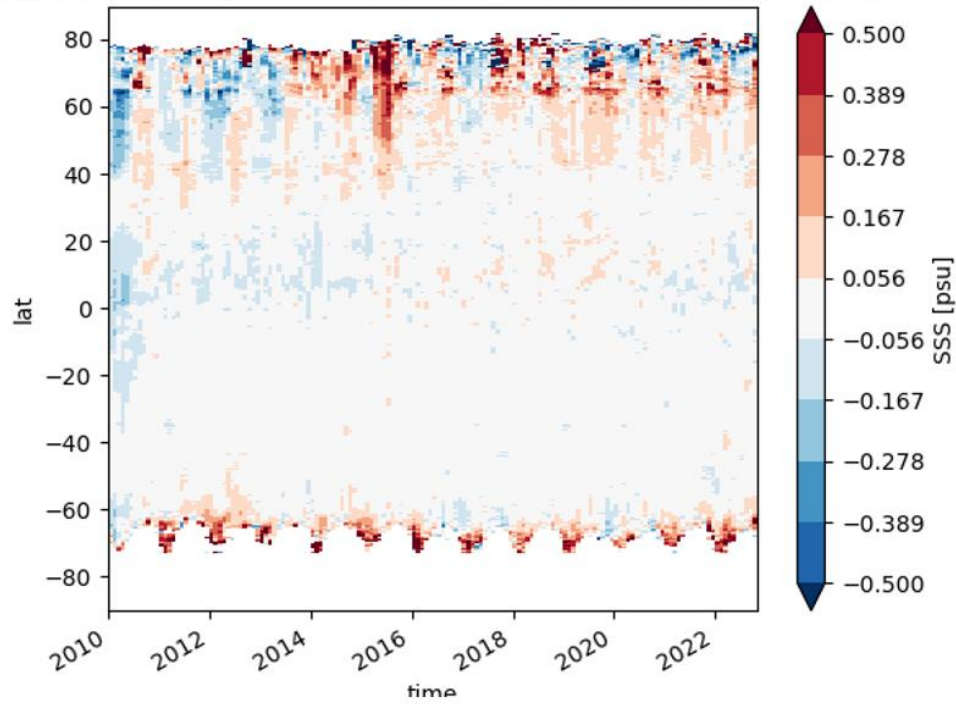


Figure 12: Comparison of monthly number of CCI+ SSS estimates between version 3.2 (Left) and version 4.2 (Right) over the period from January 2010 to September 2020.



CCI v4.3

Mean for CCI 4.3-E25-G30d-R0.25-AD-S0.5-L0.5-I0.5- - ISAS-I0.50-S30d-R0.25-AD-P80- [Global]



CCI v4.2

Mean for CCI 4.2-E25-G30d-R0.25-AD-S0.5-L0.5-I0.5- - ISAS-I0.50-S30d-R0.25-AD-P80- [Global]

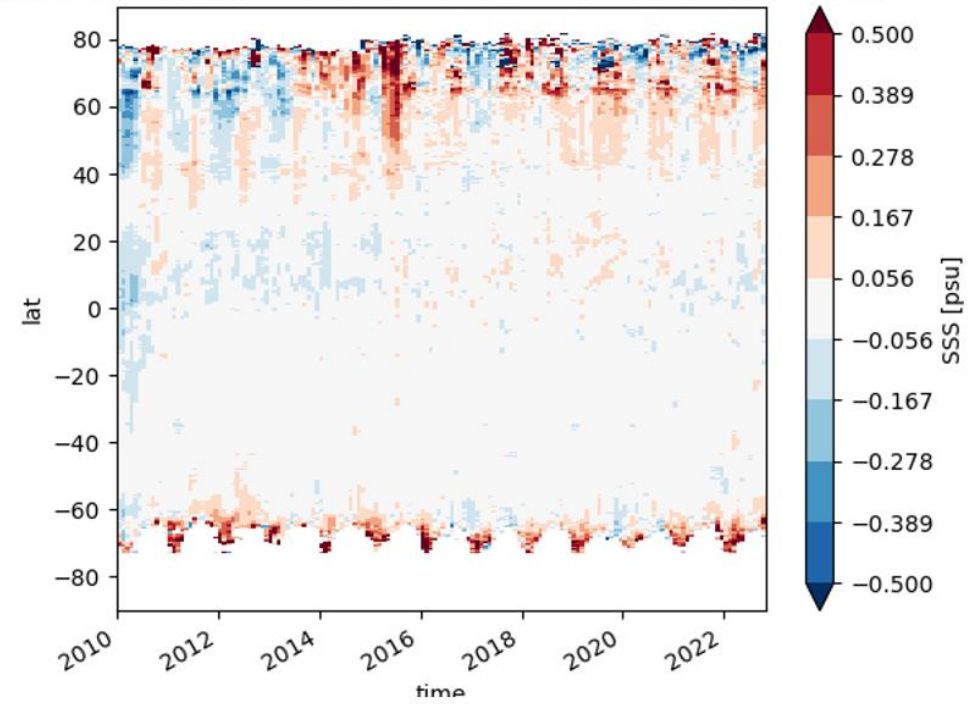
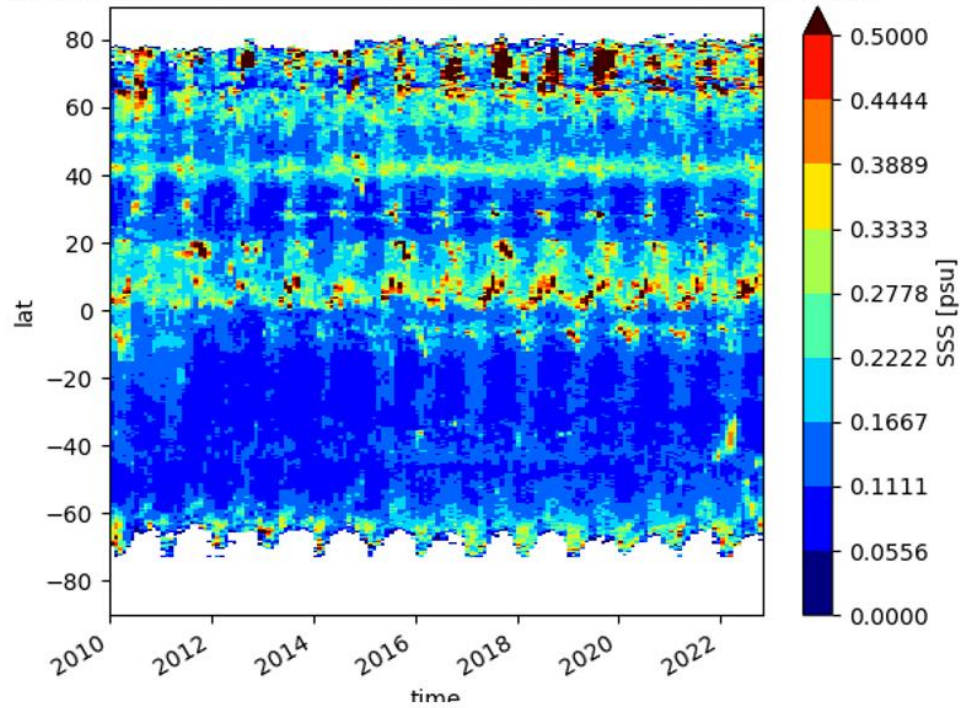


Figure 13: Mean differences between CCI L4 and ISAS 17+NRT SSS averaged over all longitudes as a function of latitude and time, (Left) with CCI V4.3 and (Right) with CCI V4.2.



CCI v4.3

STD for CCI_4.3-E25-G30d-R0.25-AD-S0.5-_L0.5-_I0.5- - ISAS-I0.50-S30d-R0.25-AD-P80- [Global]



CCI v4.2

STD for CCI_4.2-E25-G30d-R0.25-AD-S0.5-_L0.5-_I0.5- - ISAS-I0.50-S30d-R0.25-AD-P80- [Global]

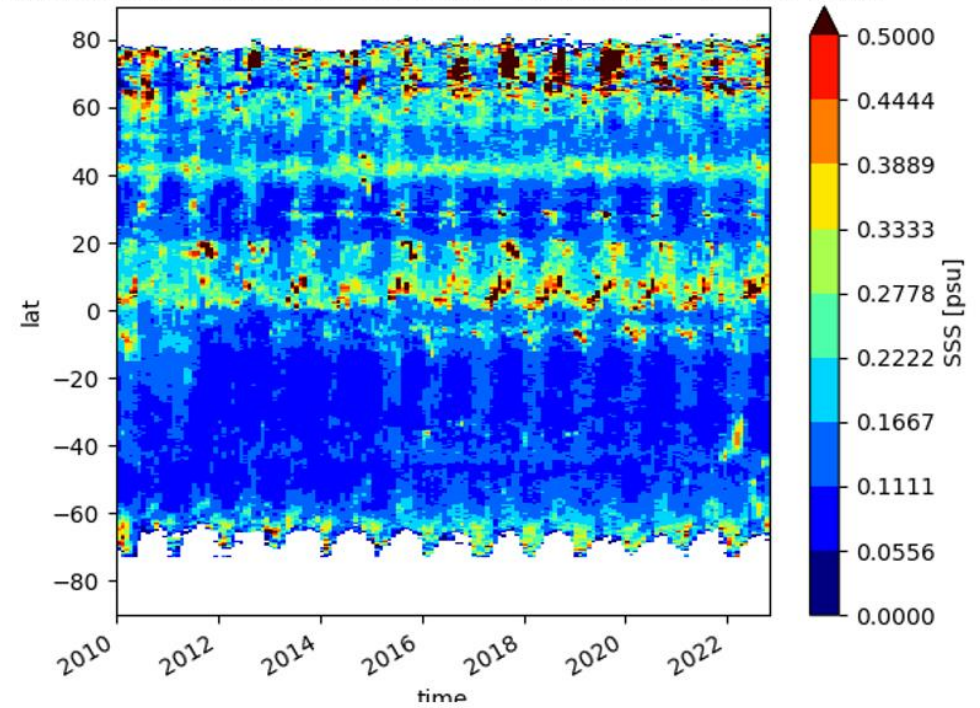


Figure 14: Std differences between CCI L4 and ISAS 17+NRT SSS averaged over all longitudes as a function of latitude and time, (Left) with CCI V4.3 and (Right) with CCI V4.2.

6.2 Results of the round robin exercise

In Section 6.1, results showed that the CCI+SSS global product V4.2 outperformed V4.1 by a significant margin. Therefore, the RR exercise did not include V4.1. Another notable finding from Section 6.1 was that, globally, V4.2 improved over V3.2. Therefore, to evaluate the progress in phase 1, the RR exercise will mainly compare V3.2 with V4.2, along with polar research products, in areas with high latitudes. Additionally, the newly produced datasets will also be compared among themselves.

6.2.1 North Atlantic comparisons

In the North Atlantic at high latitudes along the TSG tracks, the V4.2 product displays reduced differences compared to other satellite products, as shown in Figure 16 and Figure 17. While the mean bias for V4.2 can sometimes be slightly smaller or larger than V3.2 and other products, the variance of the difference is generally smaller, which is evident from the IQR. On the other hand, the polar research products generally show poorer performance, although Res1noWS is the most reliable among the polar research products.

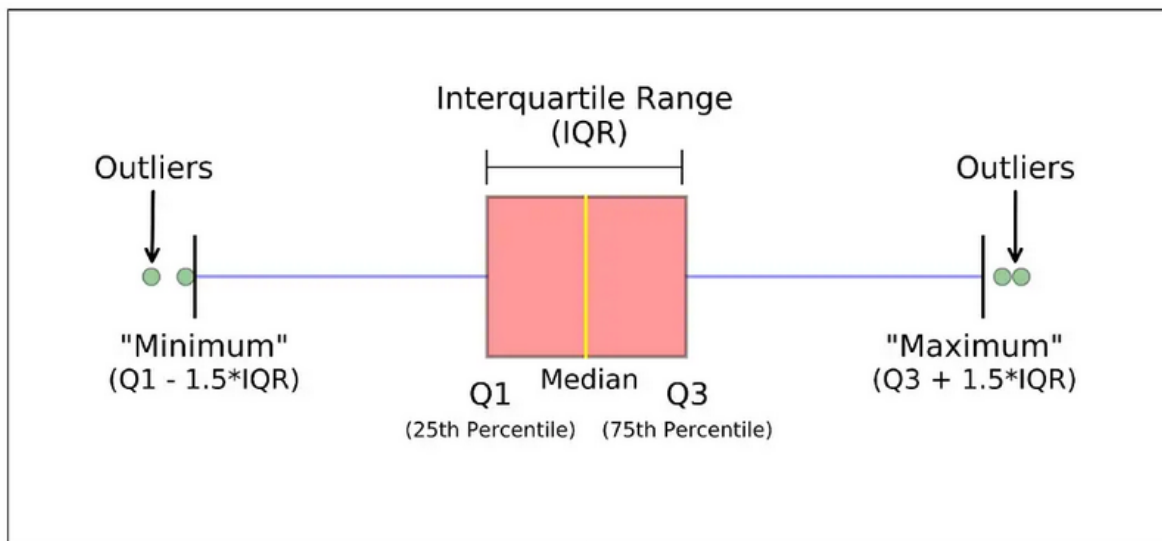


Figure 15: Interpretation keys for the box plots in the next figures

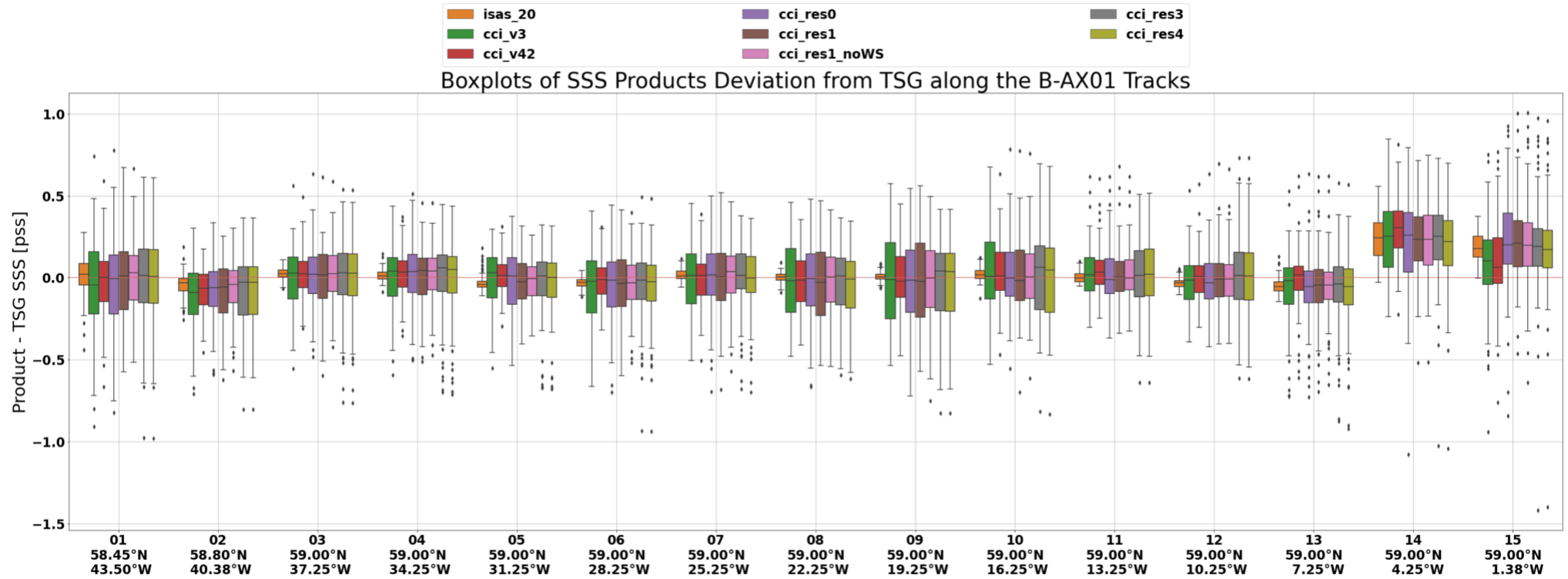


Figure 16: Results of box plots for each bin along the North Atlantic B-AX01 TSG tracks, numbered from West to East (Figure 2). The plotted datasets for each bin result are arranged from left to right, namely: ISAS-20, CCI+SSS V3.2, CCI+SSS 4.2, CCI polar research res0, res1, res1_noWS, res3, and res4. The box for each product, as shown in Figure 15, represents the median and interquartile range (IQR) for the entire 2010-2022 period, and includes a standard outlier threshold (bars) set to 1.5*IQR.

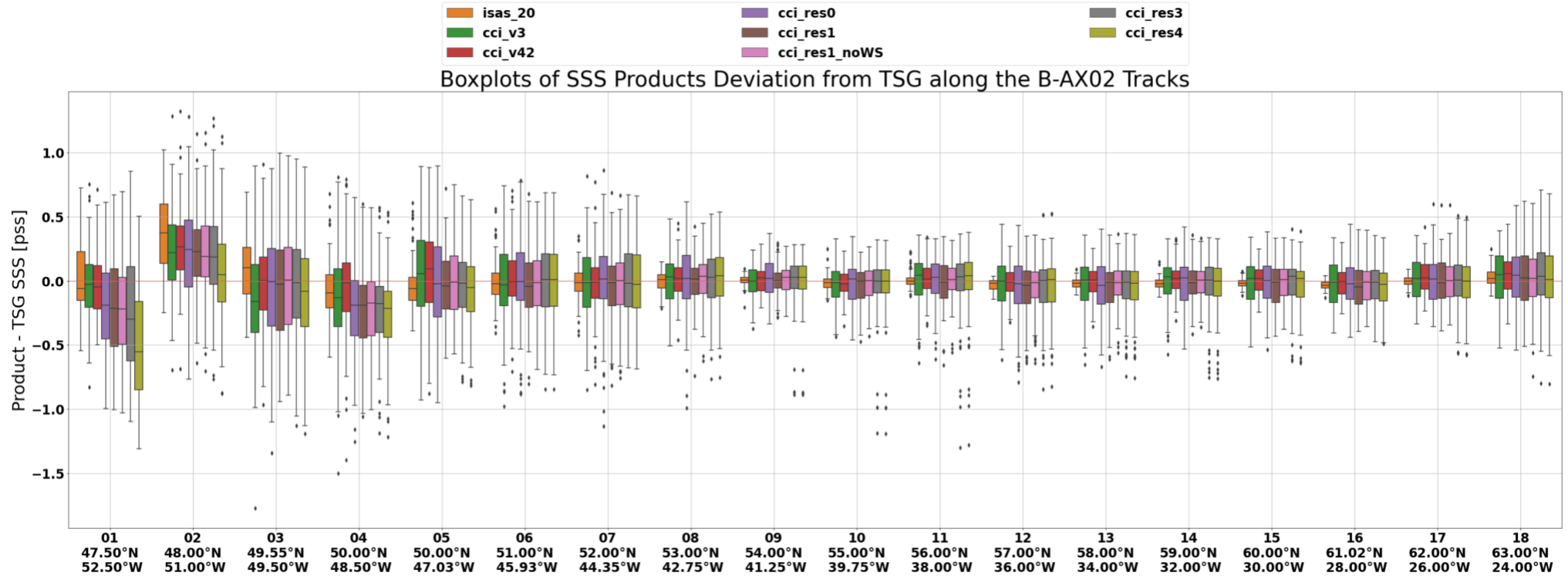



Figure 17: Same as Figure 16, but for B-AX02, and TSG tracks are numbered from South to North.

	<p>Climate Change Initiative+ (CCI+) Phase 2 Product Validation and Algorithm Selection Report</p>	Ref.: ESA-EOP-SC-AMT-2021-26 Date: 20/06/2023 Version: v4.0 Page: 32 of 65
--	---	---

The time variations of SSS displayed by satellite products are generally less reliable than those shown by the ISAS 20 product, as evidenced by Figure 18 and Figure 19. However, there are some instances at specific times where the improvement of V4.2 over V3.2 is clearly visible, such as in B-AX01 07.

The comparison metrics between the satellite-derived SSS products and TSG data reveal significant improvements with the V4.2 version relative to the V3.2 version along the North Atlantic TSG tracks, as demonstrated in Figure 20, Figure 21, Figure 22, Figure 23, Figure 24 and Figure 25. Particularly noteworthy is the substantial reduction in robust standard deviation of difference along both B-AX01 and B-AX02 tracks, as illustrated in Figure 24 and Figure 25. While Res1noWS is confirmed as the best polar research product, it is not as accurate as V4.2.

The superiority of V4.2 compared to V3.2 is evident in the mean seasonal analysis, as shown in Figure 26, particularly during the December to June period, as well as in July and August. During December to June, V3.2 and to a lesser extent, the polar research products exhibit negative differences relative to the TSG data. However, these discrepancies are significantly reduced or absent in V4.2.

The interannual variability analysis of the difference between the satellite-derived SSS and TSG data also indicates some progress of V4.2 over V3.2 (Figure 27, Figure 28). Nevertheless, there is still a noticeable regime shift along B-AX01 before and after April/May 2015, which corresponds to the beginning of SMAP. Specifically, the SSS anomaly differences show a predominantly negative pattern before 2015, but they become predominantly positive after that time. This pattern is also discernible along B-AX02, albeit to a lesser extent.

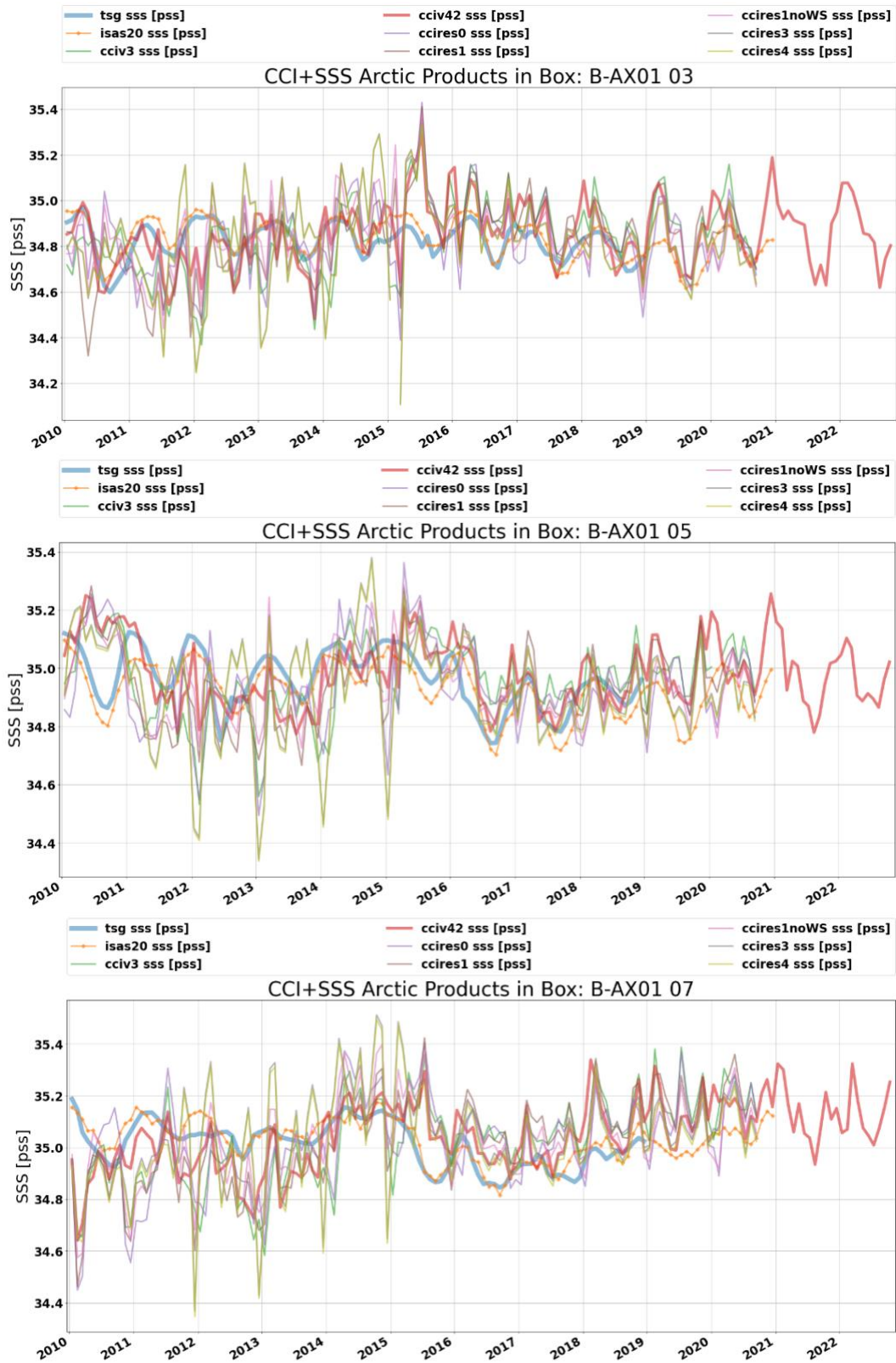


Figure 18: Time series of TSG data and products within different North Atlantic B-AX01 boxes.

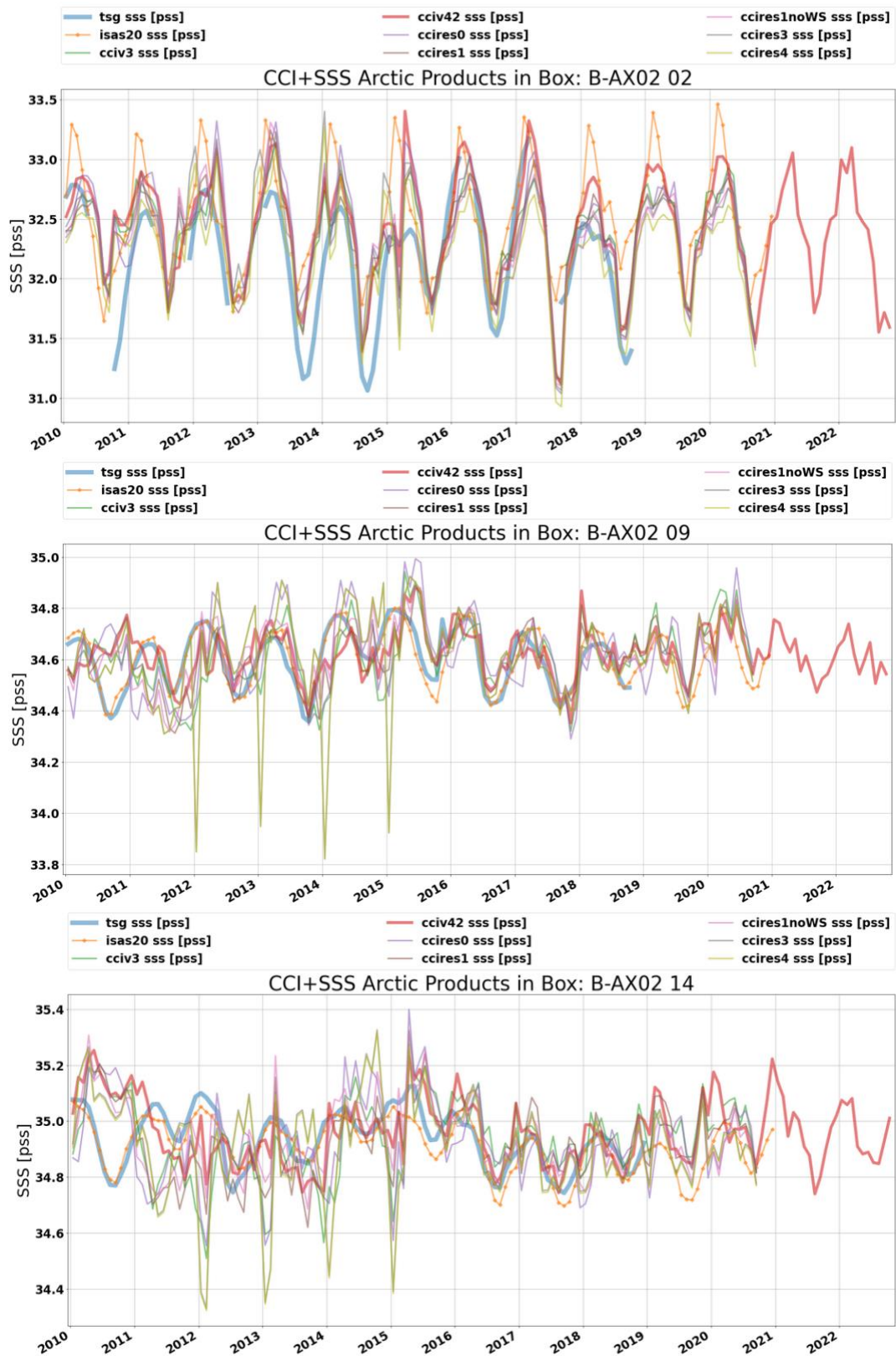


Figure 19: Time series of TSG data and products within different North Atlantic B-AX02 boxes



95% Confidence Interval of the Temporal Mean of Absolute Difference
 between Each Product and TSG SSS Within Each Box
 Along B-AX01

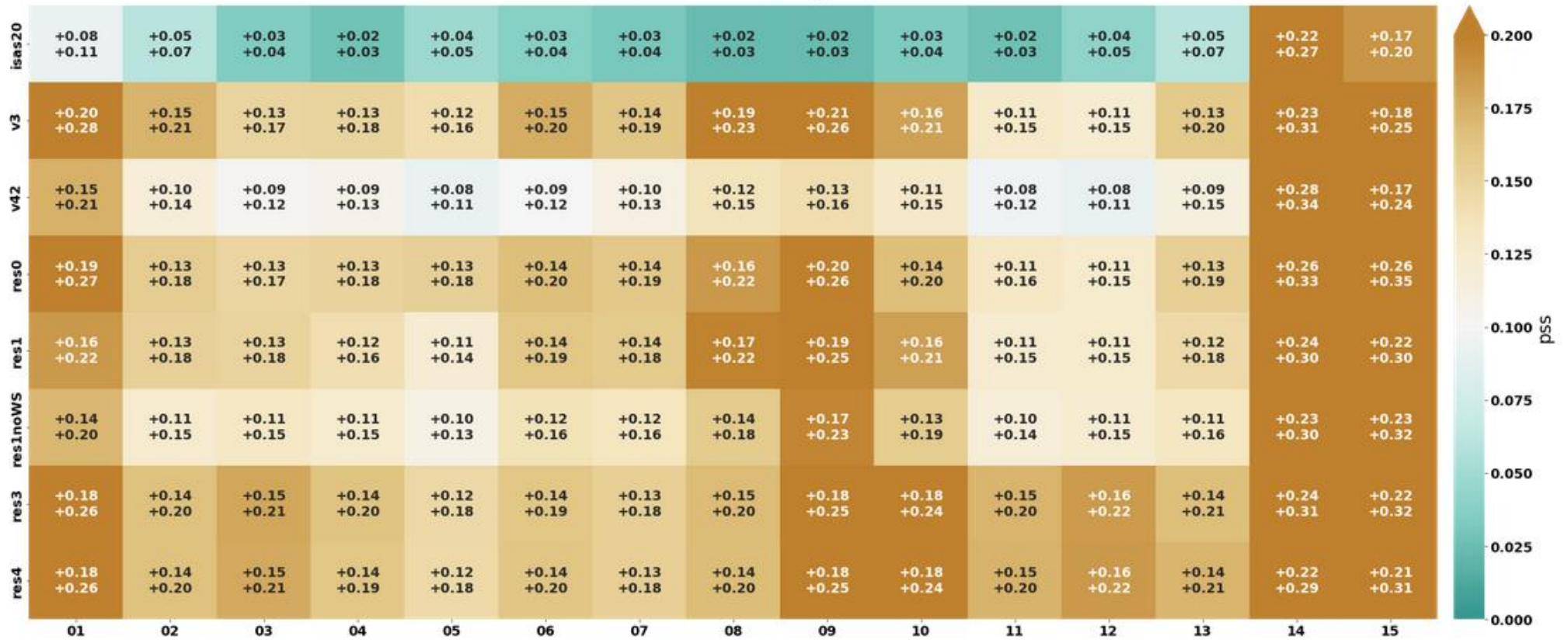


Figure 20: (Color) Mean absolute difference (mad) between each product and TSG SSS within each box for B-AX01. Number couples indicate the 95% confidence interval of mad using a bootstrap procedure.



95% Confidence Interval of the Temporal Mean of Absolute Difference
 between Each Product and TSG SSS Within Each Box
 Along B-AX02

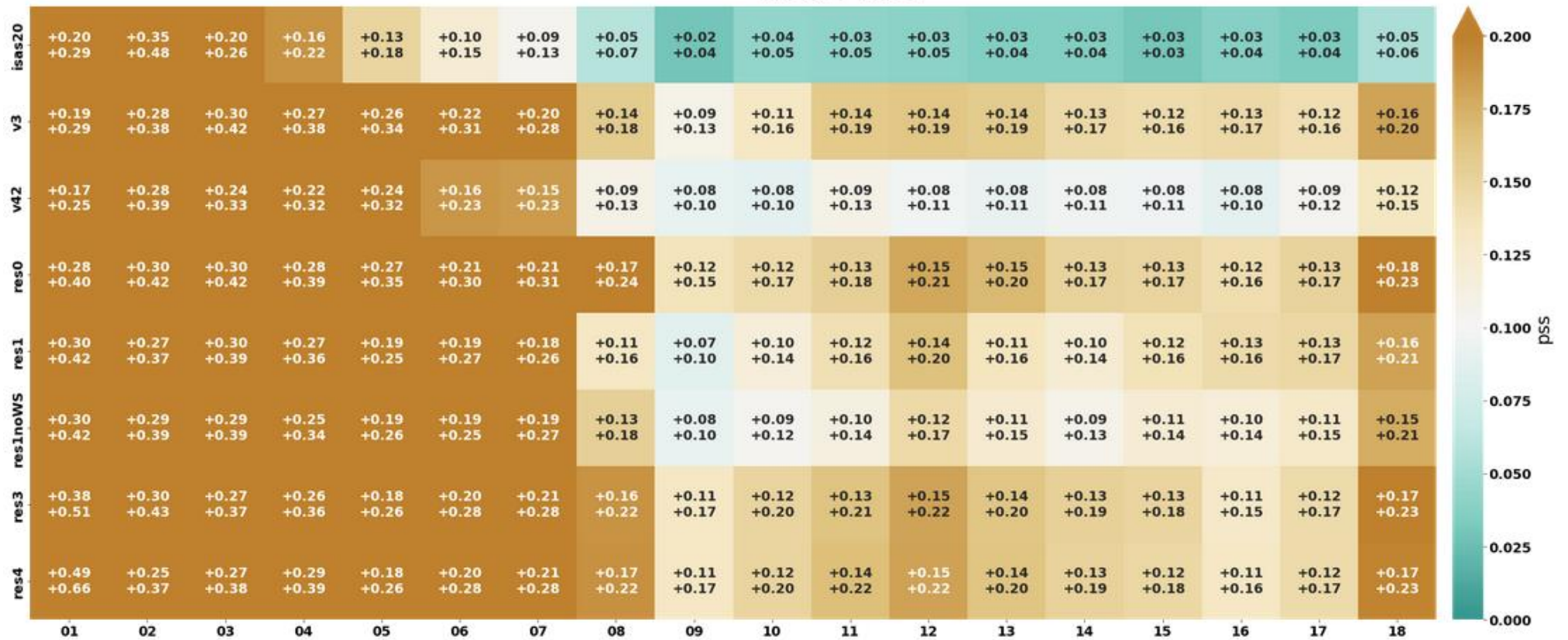


Figure 21: (Color) Mean absolute difference (mad) between each product and TSG SSS within each box for B-AX02. Number couples indicate the 95% confidence interval of mad using a bootstrap procedure.



95% Confidence Interval of the Temporal Correlation
 between Each Product and TSG SSS Within Each Box
 Along B-AX01

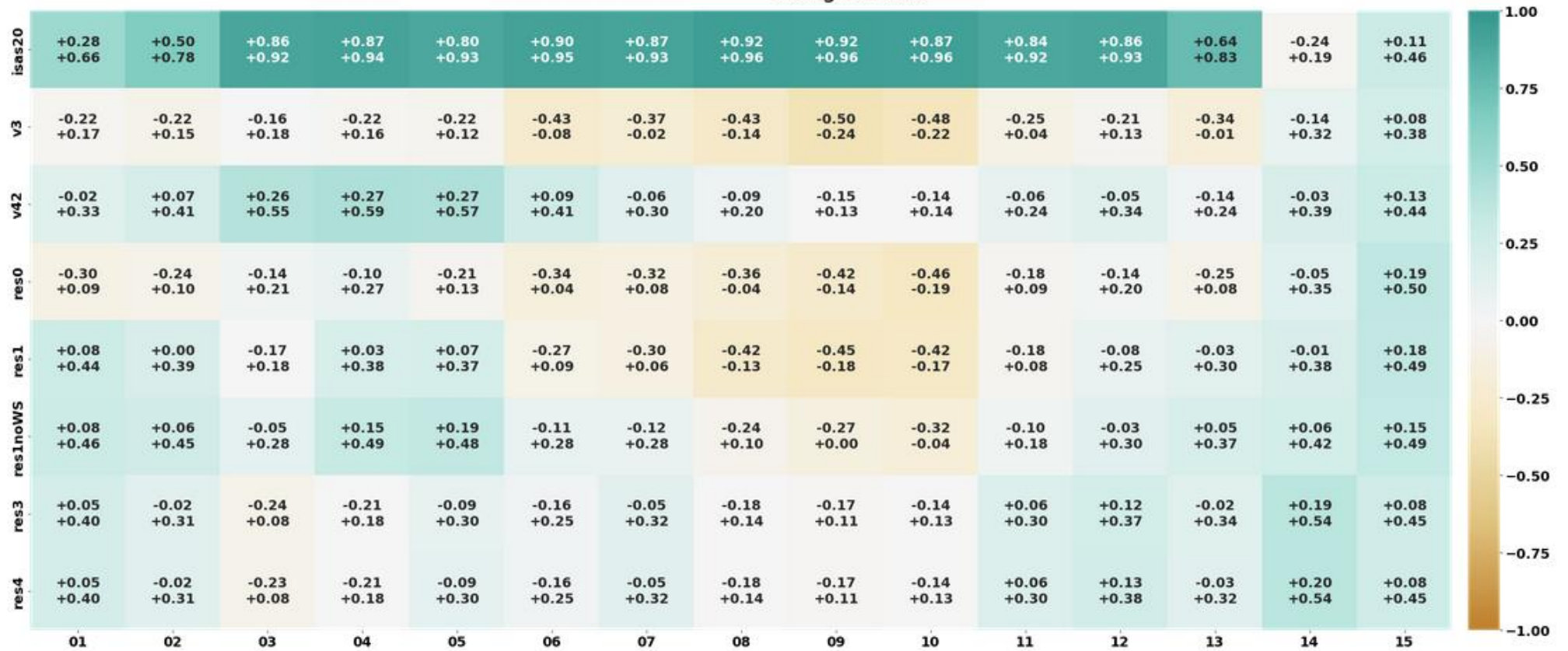


Figure 22: (Color) Correlation between each product and TSG SSS within each box for B-AX01. Number couples indicate the 95% confidence interval of correlation using a bootstrap procedure.



95% Confidence Interval of the Temporal Correlation
 between Each Product and TSG SSS Within Each Box
 Along B-AX02

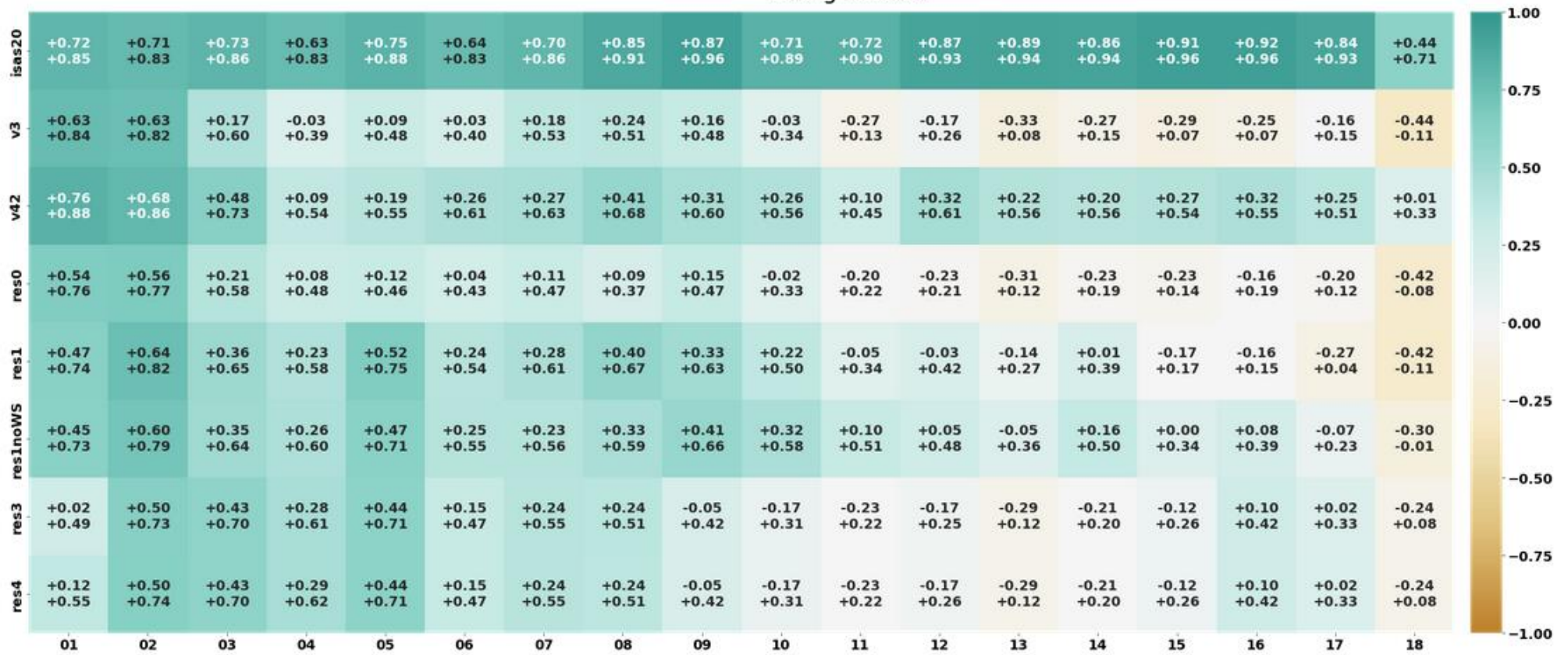


Figure 23: (Color) Correlation between each product and TSG SSS within each box for B-AX02. Number couples indicate the 95% confidence interval of the correlation using a bootstrap procedure.



95% Confidence Interval of the Robust Std of Difference
 between Each Product and TSG SSS Within Each Box
 Along B-AX01

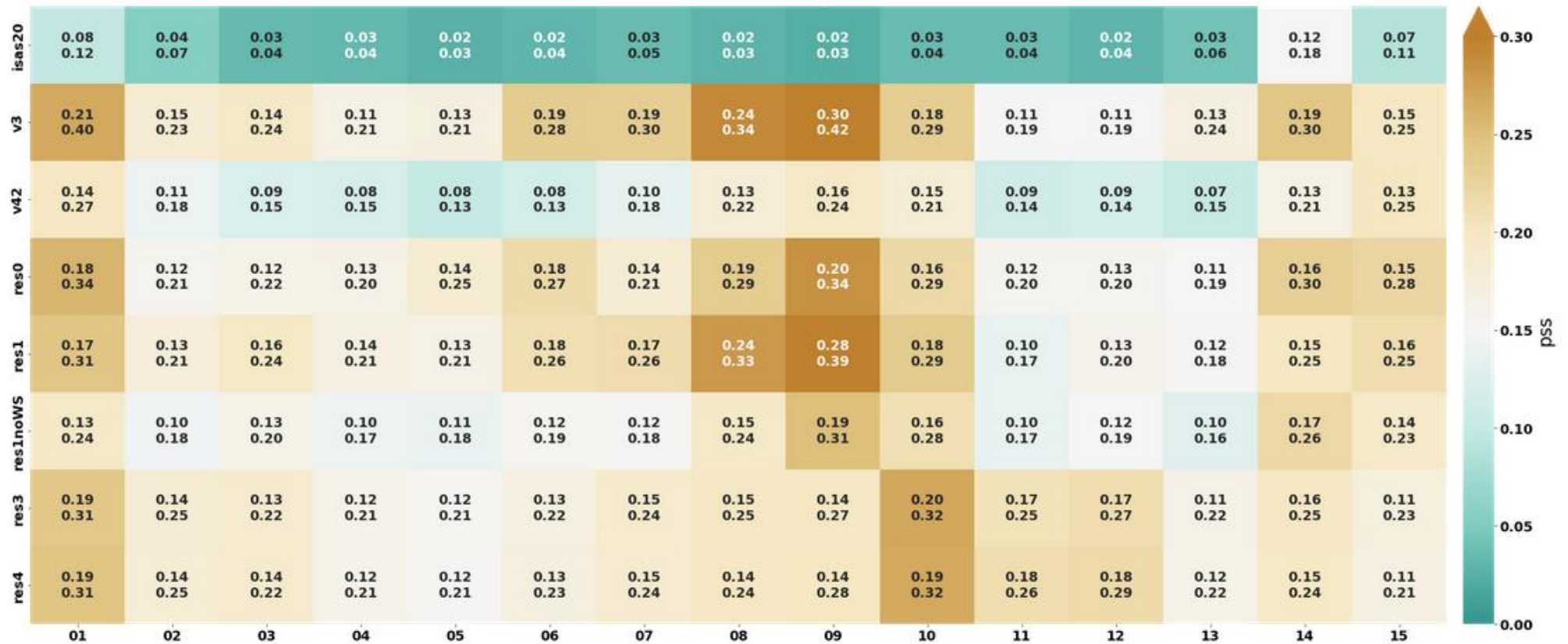


Figure 24: (Color) Robust standard deviation of difference between each product and TSG SSS within each box for B-AX01. Number couples indicate the 95% confidence interval of the robust std using a bootstrap procedure.



95% Confidence Interval of the Robust Std of Difference
 between Each Product and TSG SSS Within Each Box
 Along B-AX02

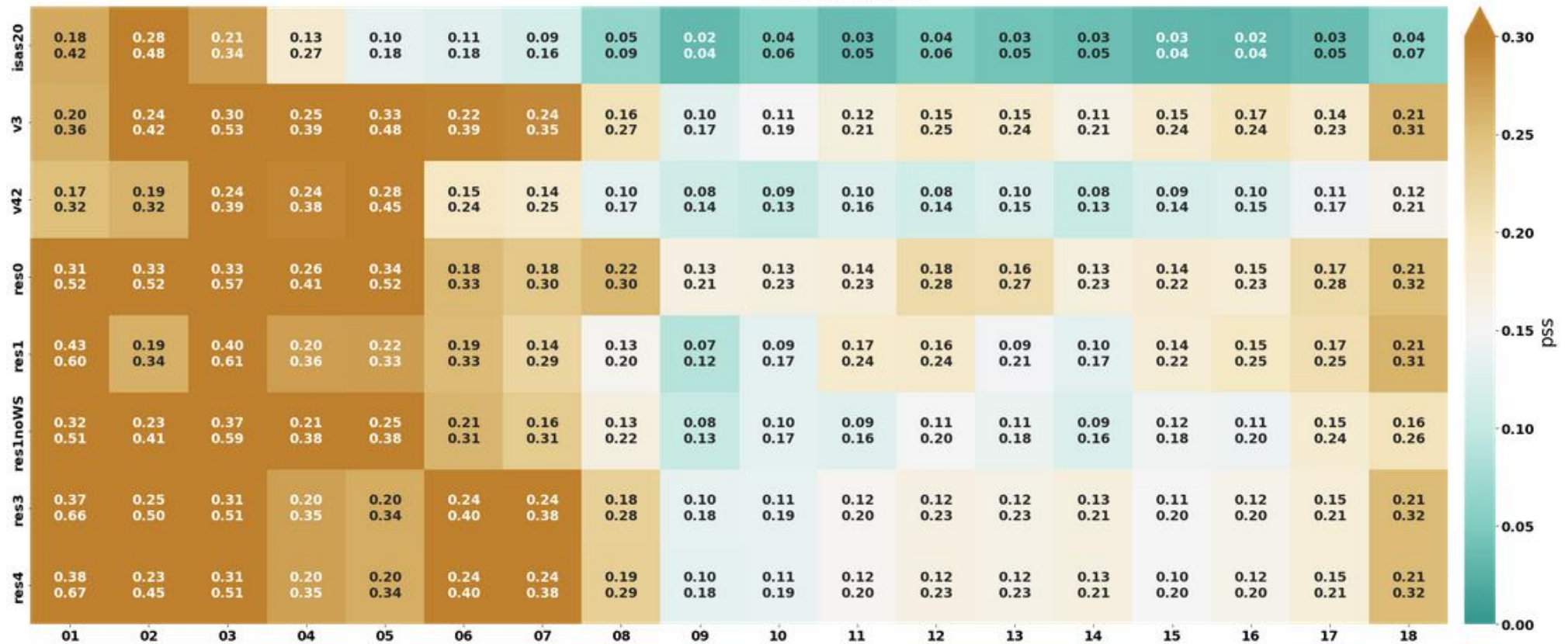


Figure 25: (Color) Robust standard deviation of difference between each product and TSG SSS within each box for B-AX02. Number couples indicate the 95% confidence interval of the robust std using a bootstrap procedure.

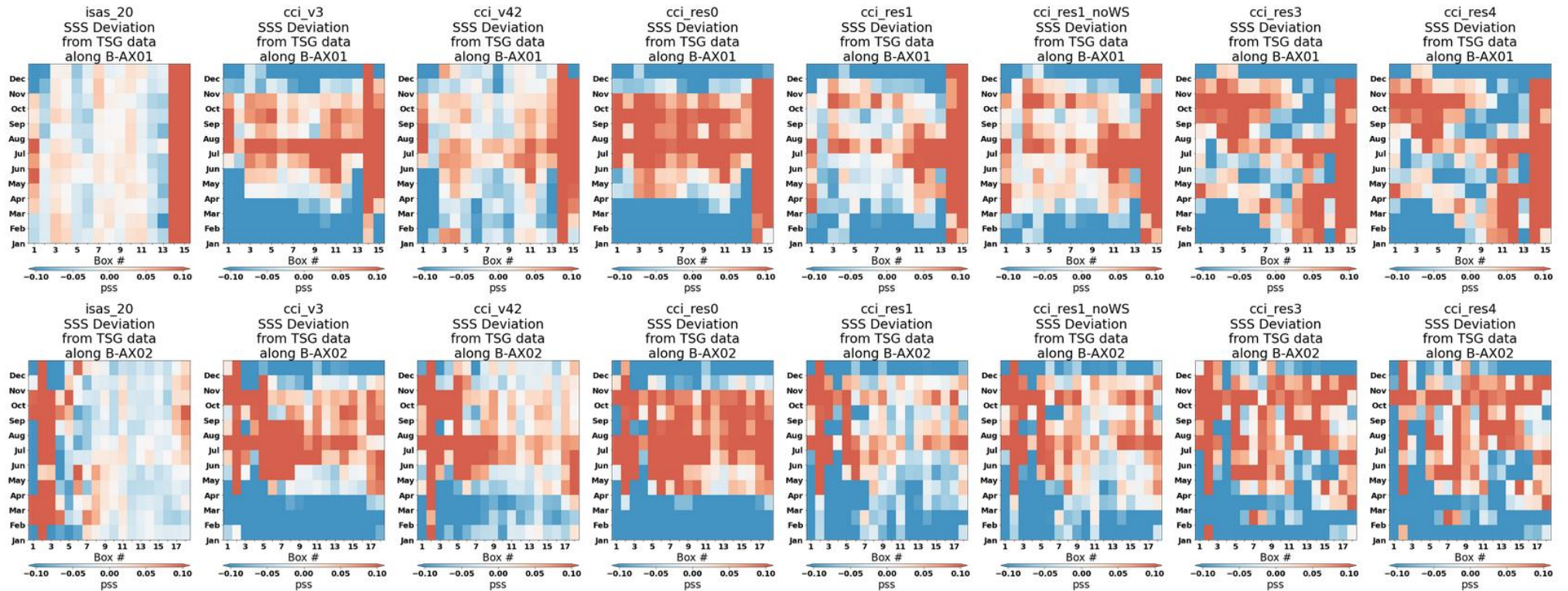


Figure 26: Hovmöller plots of mean seasonal difference between products and TSG data for each TSG box (x-axis), for (Top) B-AX01 and (Bottom) B-AX02

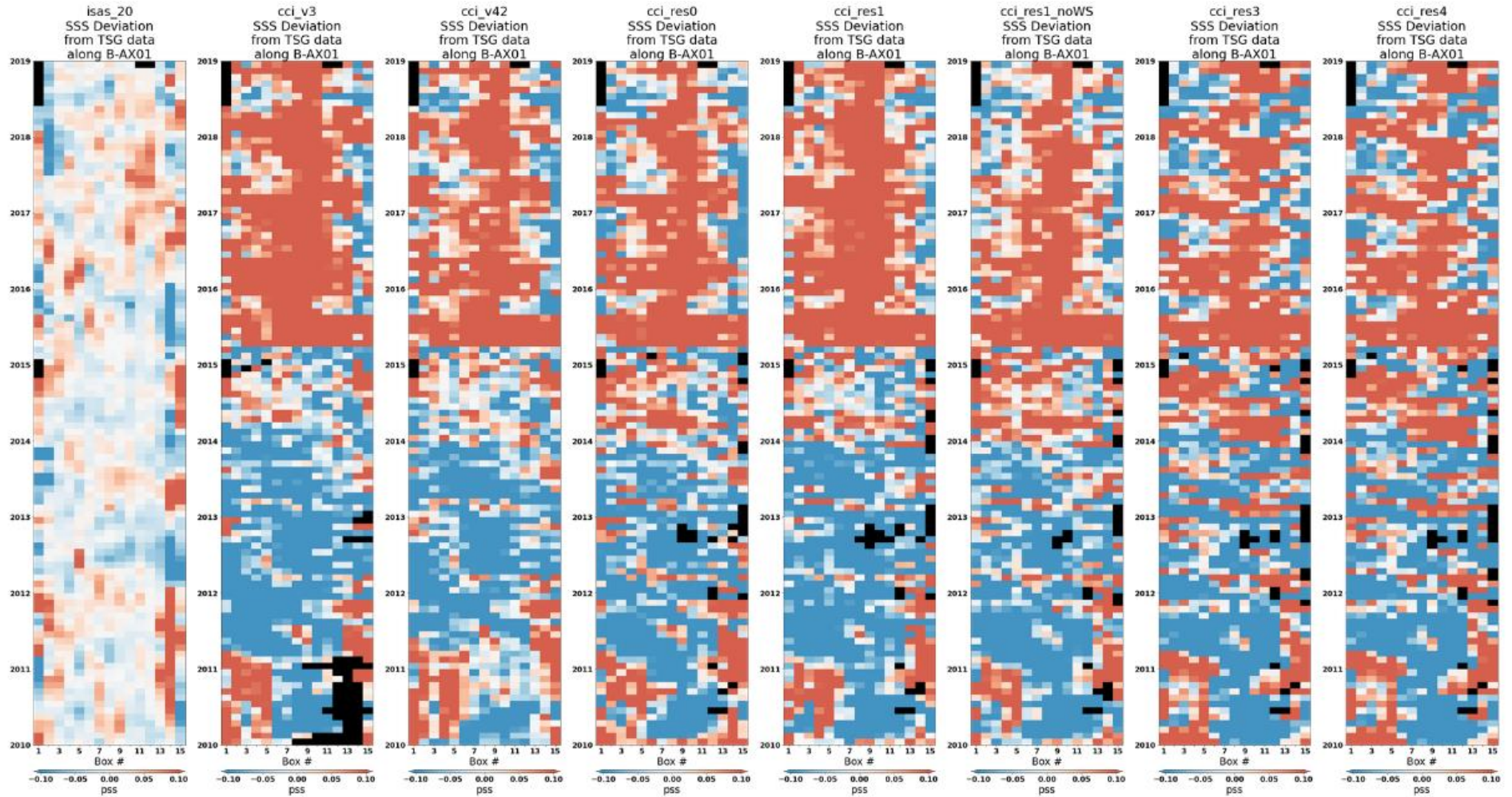


Figure 27: Hovmoller plots of anomaly difference with respect to mean seasonal cycle between products and TSG data for each TSG box (x-axis), for B-AX01.

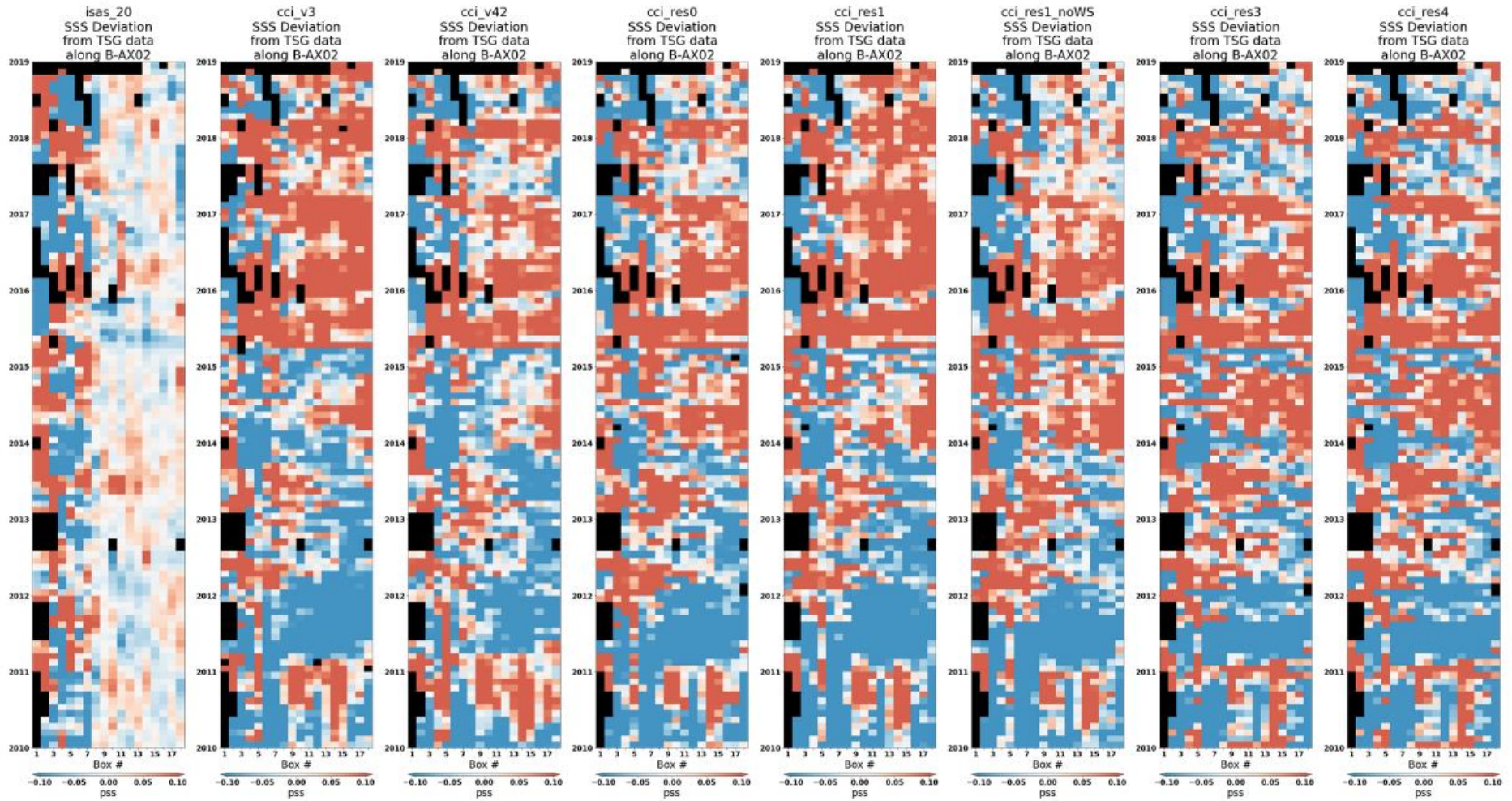


Figure 28: Hovmoller plots of anomaly difference with respect to mean seasonal cycle between products and TSG data for each TSG box (x-axis), for B-AX02.

6.2.2 Southern Ocean comparisons

Note: hereafter in the figures the term "CCIbeta" refers to the Res1noWS version of the polar research product, and "CCIv3" to the V3.2 version.

Considering all collocated data (Section 4.1.2) in the Southern Ocean region, the global comparison between V4.2 and V3.2 does not, at first sight, exhibit significant improvement. The joint distribution analysis highlights that V4.2 has a slightly larger spread than V3.2 when considering the TSG data across the entire region (Figure 29). Furthermore, when first examining the comparison metrics across the entire domain, V4.2 shows relatively poorer performance than V3.2 (Figure 31). This can be attributed to including a higher number of pixels with V4.2 compared to V3.2. However, it should be noted that in regions where the number of pixels is the same as those north of 60°S (as indicated below), V4.2 demonstrates improved performance. Like the findings along the North Atlantic tracks, an improvement is still observed compared to Res1noWS.

Conversely, when the collocation data is restricted to latitudes north of 60°S, farther away from Antarctica, V4.2 demonstrates better scores than V3.2 (Figure 30, Figure 32). This can be attributed to the fact that V4.2 has extended data coverage toward the ice edge, resulting in increased challenges related to sea-ice contamination compared to V3.2.

An exciting observation arises from analysing the mean values across the domain, as illustrated in Figure 33. Notably, there are distinct instances where V4.2 exhibits noticeable improvement compared to V3.2, particularly evident in the years 2010 and 2013.

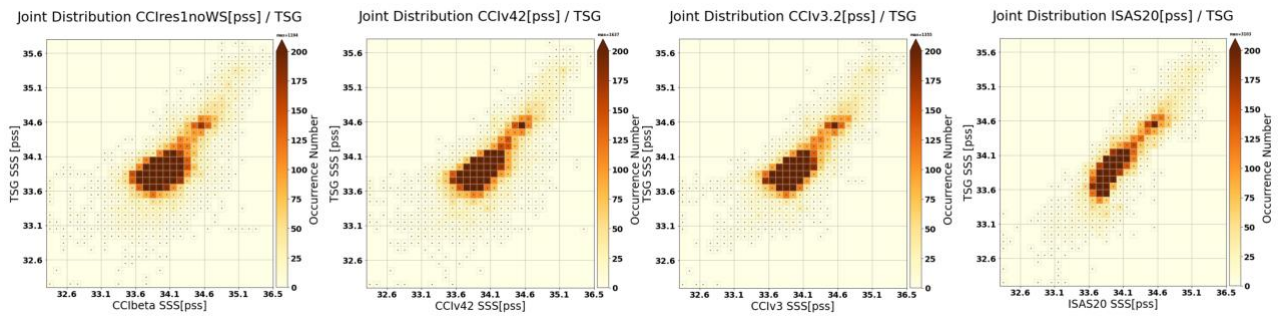


Figure 29: Joint distribution of each product with the TSG data within the entire domain around Antarctica.

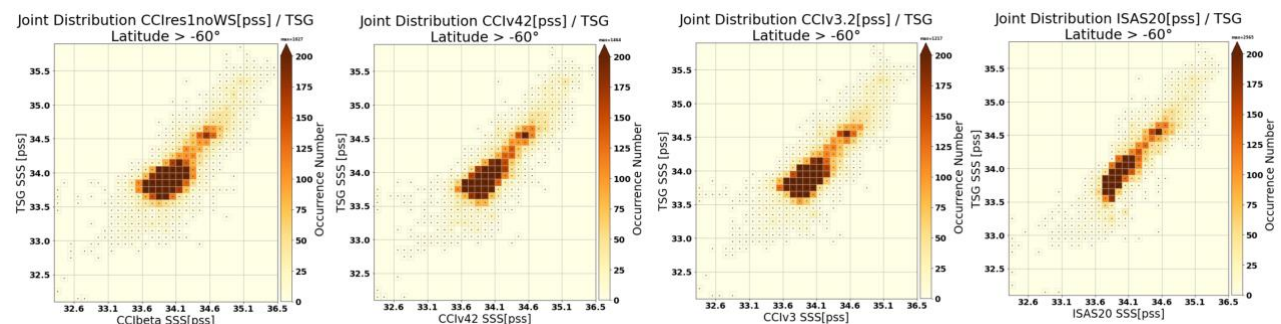


Figure 30: Same as Figure 29, but restricted to the north of 60°S

Comparison Metrics over the entire Antarctica Domain

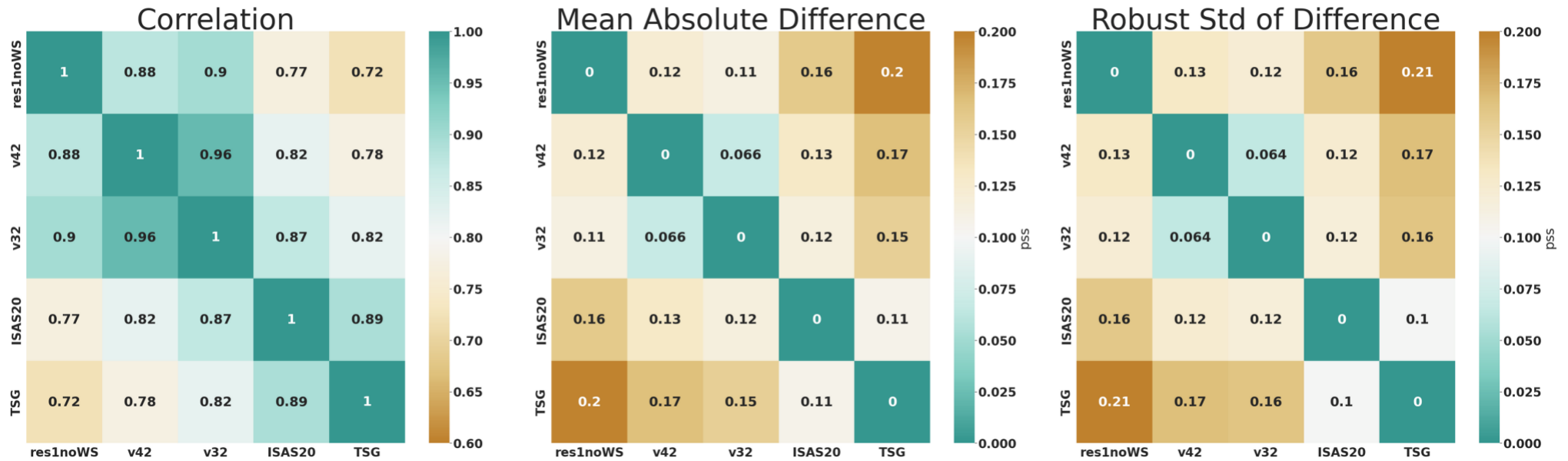


Figure 31: (Color and Numbers) Correlation, mean absolute difference and robust std of difference between each product and the TSG data over the entire domain.

Comparison Metrics over the Antarctica Domain Latitude > -60°

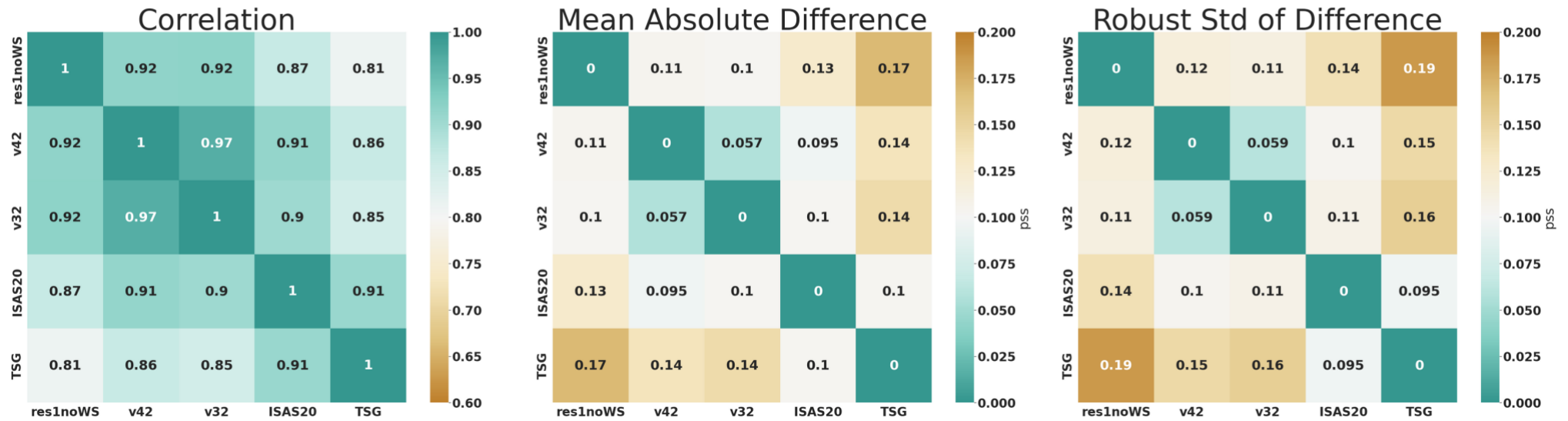


Figure 32: Correlation, mean absolute difference and robust std of difference between each product and the TSG data over the domain restricted to latitudes north of 60°S.

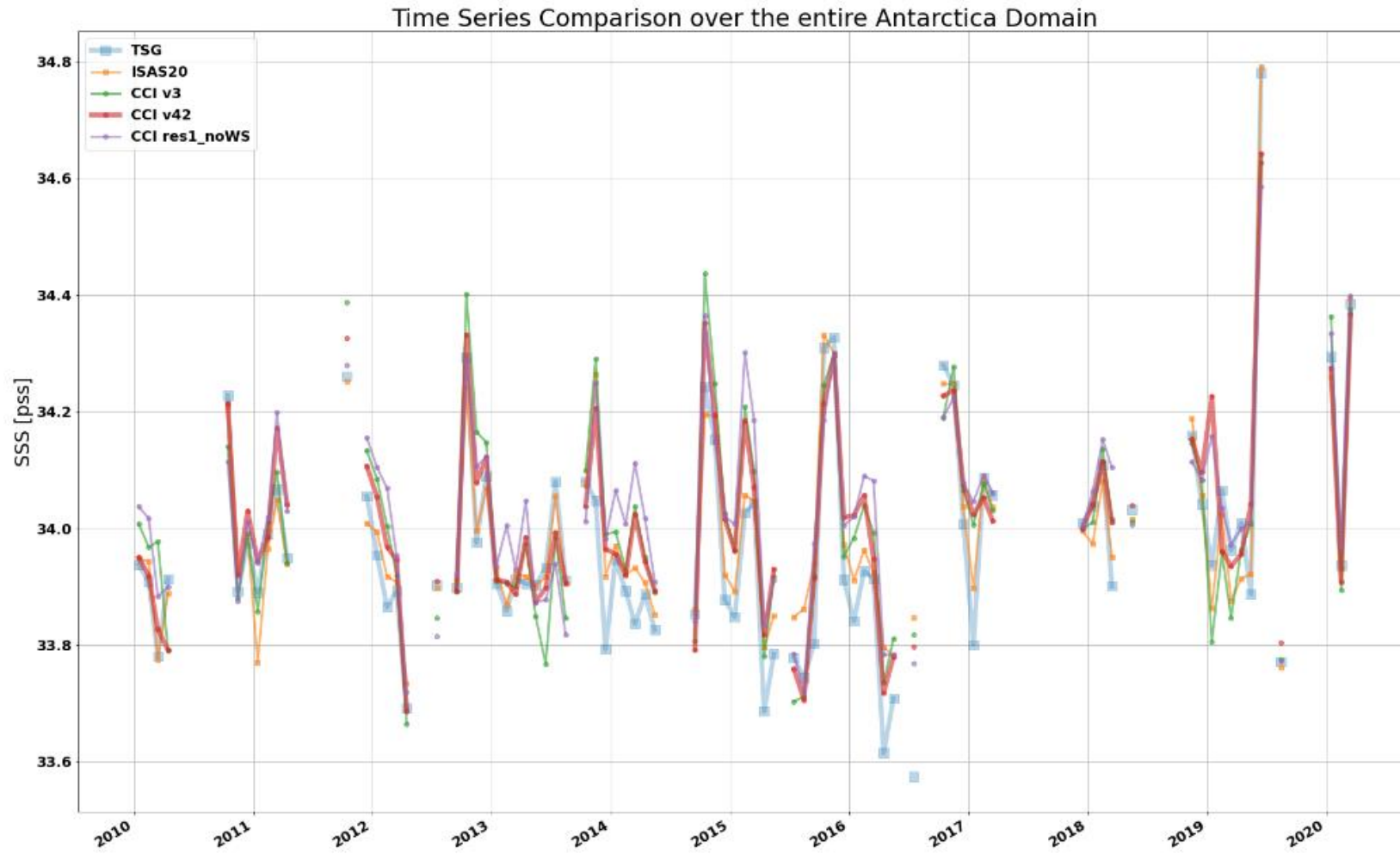


Figure 33: Time series of the averaged products and TSG data over the entire Antarctica domain where there are colocations.



The results along the recurring TSG tracks from Tasmania to Antarctica provide a positive outlook for V4.2 compared to V3.2 (Figure 34). The number of collocated data in each disc is large (at least 20). As a result, the correlation between V4.2 and TSG data is higher than that of V3.2 (and Res1noWS), particularly for boxes 2, 3, 4 and 5, except near Antarctica (box 7). Moreover, the mean difference and standard deviation of the difference are smaller for V4.2, particularly in boxes 2, 3, 4, and 5, although not in box 7.

When assessing the comparison metrics along the South Africa-Antarctica TSG tracks, the performance of V4.2 appears less favourable than the Tasmania-Antarctica comparison, as depicted in Figure 35. This is because the number of available collocated data in the discs is always smaller than 20. In boxes 1, 3, and 6, there is a slightly higher correlation between V4.2 and TSG data, along with a slightly smaller mean difference and robust standard deviation compared to V3.2. Conversely, in boxes 2, 4, and 5, V3.2 demonstrates a slight advantage over V4.2. This is important to consider that these findings should be nuanced because the number of collocated data is lower in these cases compared to the Tasmania-Antarctica case. Additionally, no collocated data is available for V3.2 in boxes 7 and 8, which are closest to Antarctica.

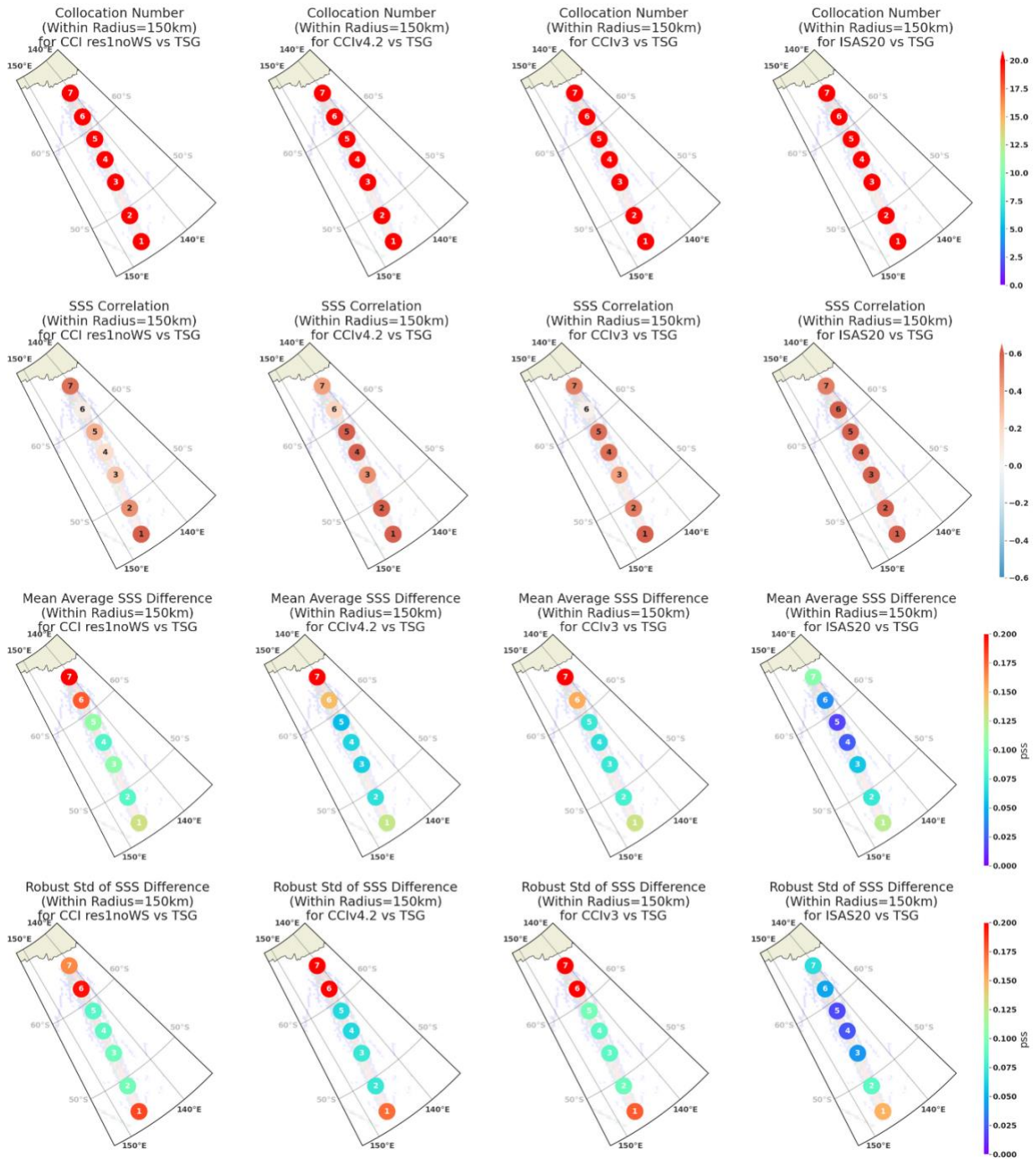


Figure 34: Collocation data number and comparison metrics between each product and the TSG data along ship tracks from Tasmania to Antarctica.

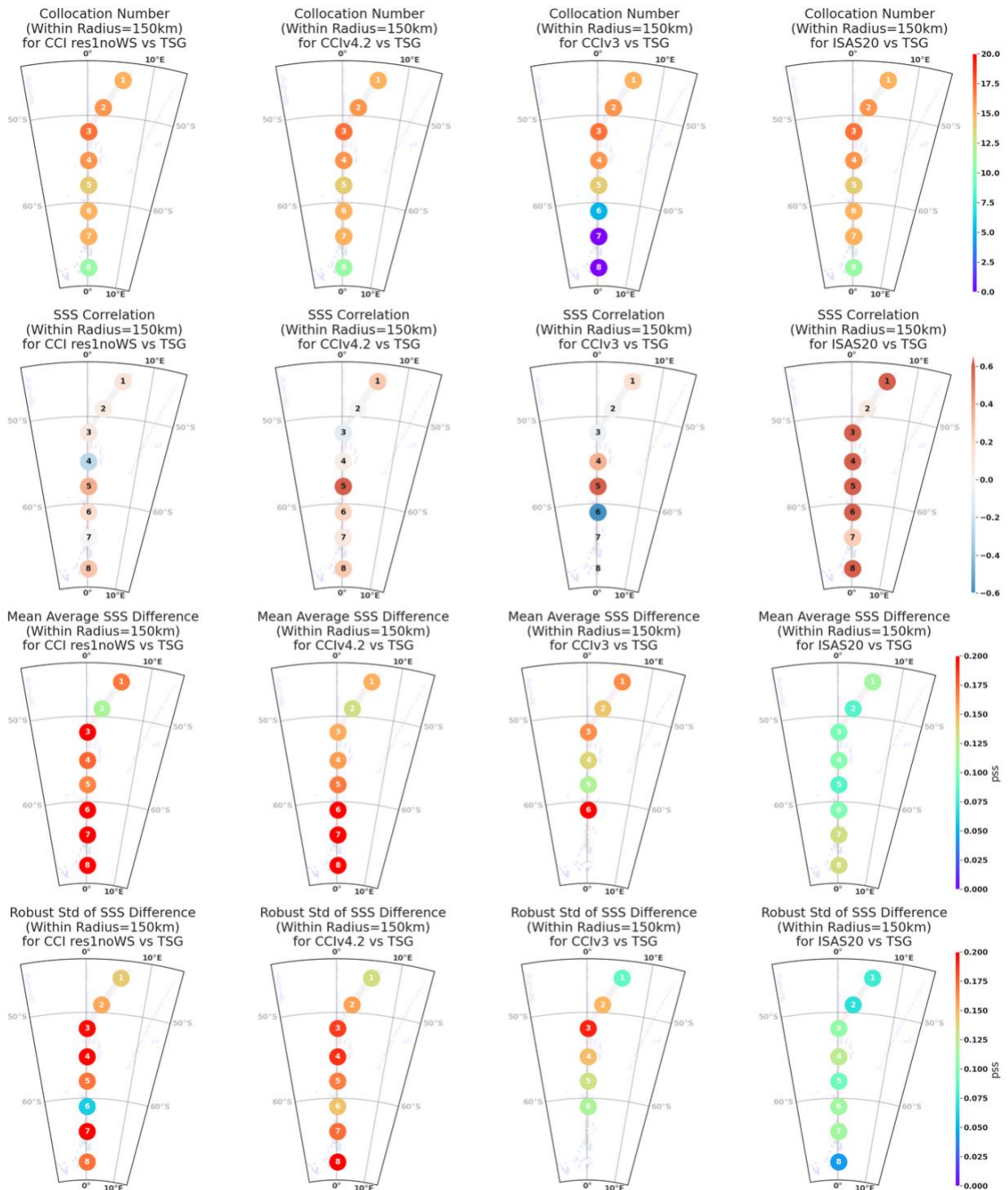


Figure 35: Collocation data number and comparison metrics between each product and the TSG data along ship tracks from South Africa to Antarctica.



6.3 Products evaluation summary


The comparison between the satellite-derived Sea Surface Salinity (SSS) products and ISAS dataset revealed significant similarities and differences. The differences in magnitudes helped identify enhancements or deteriorations in the tested versions. The analysis focused on comparing CCI L4 maps with ISAS, considering only pixels with PCTVAR lower than 80%.

In V4.1, there was a slight reduction in seasonal variation biases at high latitudes compared to V3.2. V4.2 further improved this reduction, emphasising the need to address remaining seasonal latitudinal biases in SMAP and Aquarius SSS. In addition, the robust standard deviation of the difference decreased, and correlation mainly increased in areas far from the coast in V4.2 compared to V4.1, particularly over the global ocean. However, improvements in correlation were less evident closer to the coast, possibly due to differences in coastal coverage between SMOS L2 SSS v7 and SMOS L2 CCI v671. The comparisons were based on monthly CCI products, with observed improvements noted in weekly products.

Further detailed comparisons between V4.2 and V3.2 showed a reduction in mean and standard differences in latitudinal profiles within the range of 60°N to 60°S. However, in extremely high latitudes, the mean and standard differences remained high due to some remaining issues near the ice and likely uncertainties in ISAS SSS. The RMS difference maps between V3.2 and V4.1 exhibited similarities, with more significant differences near the coast or islands due to RFI filtering in SMOS L2 v7. The PVASR highlighted the significant improvement achieved with V4.2 over V4.1 and V3.2 globally.

Detailed comparisons were conducted between V3.2 and V4.2, focusing on high-latitude regions and polar research products. V4.2 demonstrated reduced differences along the North Atlantic TSG tracks compared to other satellite products, while the polar research products generally showed poorer performance. The time variations of SSS indicated visible improvement in V4.2 over V3.2 at specific times. Comparison metrics with TSG data confirmed the superiority of V4.2 over V3.2 along the North Atlantic tracks, particularly regarding the reduced robust standard deviation of the difference. In the mean seasonal analysis, V4.2 showed significant improvement, especially from December to June and July to August, with reduced negative differences compared to TSG data. The interannual variability analysis indicated some progress of V4.2 over V3.2, but a regime shift was observed in April/May 2015, corresponding to the beginning of SMAP.

In the Southern Ocean region, V4.2 outperforms V3.2 in terms of spatial coverage, as V3.2 was heavily filtered near ice. North of 60°S, V4.2 demonstrates superior performance compared to V3.2. However, metrics south of 60°S seem comparatively worse, which can be partially attributed to the different spatial coverage in frigid waters. A detailed analysis of recurring Temperature-Salinity-Depth (TSG) tracks from Tasmania to Antarctica reveals higher correlation and smaller differences between V4.2 and TSG data, except near Antarctica. Along the South Africa-Antarctica TSG tracks, V4.2 shows a slightly higher correlation and smaller mean differences in specific areas than V3.2, but V3.2 performs slightly better in other areas. These

	<p>Climate Change Initiative+ (CCI+) Phase 2 Product Validation and Algorithm Selection Report</p>	<p>Ref.: ESA-EOP-SC-AMT-2021-26 Date: 20/06/2023 Version: v4.0 Page: 52 of 65</p>
--	---	---

findings need to be considered in the context of the availability of collocated data and the proximity to Antarctica.

Overall, V4.2 demonstrated notable advancements over V3.2, specifically in mitigating biases and enhancing correlations across diverse regions. These findings remained consistent when evaluating the products with different in-situ datasets. However, specific challenges and disparities were noted in coastal areas and extremely high latitudes.

6.4 Open issues and discussion

A significant uncertainty factor in product evaluations is the difference in spatial coverage, particularly near coastlines and the ice edge. Therefore, to enhance the PVASR, it is recommended in the future to systematically provide metrics calculated using two approaches: 1) considering all available collocated data for each product and 2) considering collocated data at locations/times that are common across all products, as demonstrated in the Antarctic study.

It is crucial to incorporate diverse in situ platforms, not limited to Argo float products, to ensure comprehensive evaluations. However, it should be acknowledged that using different platforms can introduce variations in measurement depths, leading to increased vertical sampling uncertainty and decreased reliability in product rankings. Therefore, addressing the vertical sampling uncertainty of the in situ data should be a focus for the next iteration of the PVASR.

While recurring measurements such as regular ship routes or fixed mooring measurements over the years are valuable in RR tests, it is worth noting that such measurements are scarce at very high latitudes. Nonetheless, the PVASR must be conducted with the available data to ensure comprehensive evaluations.



7 Conclusion and future work

In conclusion, the comparison between the satellite-derived Sea Surface Salinity (SSS) products and the ISAS 17+NRT dataset revealed significant improvements in V4.2 compared to V3.2. V4.2 showed a reduction in seasonal biases at high latitudes and a decrease in the robust standard deviation of differences. Correlation mainly increased over the global ocean but was less evident closer to the coast. The comparisons were based on monthly CCI products, with observed improvements noted in weekly products.

Detailed comparisons between V4.2 and V3.2 demonstrated a marked reduction in mean and standard differences within the latitudinal range of 60°N to 60°S. However, in extremely high latitudes, differences remained high due to issues near the ice and uncertainties in ISAS 17+NRT SSS. The comparison metrics with TSG data showed superior performance of V4.2 along the North Atlantic TSG tracks compared to other satellite products. In addition, the time variations of SSS indicated visible improvement in V4.2 over V3.2 at specific times.

The global comparison between the RFI-corrected V4.3 and V4.2 reveals high similarity in the results. However, due to the preliminary nature of the RFI correction in V4.3, a more comprehensive analysis at the regional level is required to evaluate its effectiveness in the treated regions.

In the Antarctic region, when restricted to latitudes north of 60°S, V4.2 demonstrated better scores, despite challenges related to sea-ice contamination. The recurring TSG tracks from Tasmania to Antarctica showed higher correlation and smaller differences between V4.2 and TSG data compared to V3.2, except near Antarctica. Along the South Africa-Antarctica TSG tracks, V4.2 had a slightly higher correlation and lower mean difference in some boxes than V3.2, but V3.2 performed slightly better in other boxes. These findings should consider the availability of collocated data and proximity to Antarctica.

Overall, V4.2 showcased notable advancements over V3.2 in mitigating biases and enhancing correlation across diverse regions. However, challenges and disparities were observed in coastal areas and extremely high latitudes. Incorporating diverse in-situ platforms in the PVASR is crucial to ensure comprehensive evaluations and address vertical sampling uncertainty in in-situ data. Additionally, providing metrics calculated based on all available collocated data and common locations/times across all products can enhance the PVASR's reliability. Finally, despite limitations in recurring measurements at very high latitudes, the PVASR should be conducted with the available data to ensure comprehensive assessments.

End of document