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: v5.0

Ref. internal: 4000123663/18/I-NB

Revision Date: 22/11/2024

Filename: SSS_cci-D2.1-PVASR-v5r0

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	
	Roberto Sabia Digitally signed by Roberto Sabia Date: 2025.02.24 16:22:36 +01'00'		

Diffusion List
Sea Surface Salinity Team Members
ESA (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

Table of Contents

List of figures

List of tables

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 Executive Summary

The V5.x versions of the CCI+SSS L4 product aim to provide improved-quality SSS estimates over version V4.4 by addressing both random and systematic uncertainties. Random uncertainties from the L2 products are propagated through to the L3 and L4 levels, while systematic uncertainties are reduced through sensor-specific corrections (e.g., wind speed, rain rate, SST, and RFI corrections) and empirical adjustments such as seasonal and latitudinal corrections. Additionally, outlier filtering is applied to further improve data quality.

For V5, the following updates were applied:

- Reprocessed SMOS data with L1v700 to generate L2 SMOS CCI v731, covering January 2010 to December 2023.
- Filtered SMAP v5.3 data per RSS ATBD guidelines, including rain rate (RR) filtering, with a 40 km resolution.

Updates to SMOS L2OS processing included:

- Using ISAS for OTT instead of WOA.
- Adopting ERA5 reanalysis for auxiliary parameters (wind speed, SST, wave height) instead of ECMWF forecasts.
- Resolving sea state flag and sun glint issues.
- Incorporating the BVZ dielectric constant model for improved accuracy.
- Applying RFI correction globally.
- Extending data coverage to December 2023.

Key Differences between V5.x Versions

Each V5.x version introduces specific calibration and filtering techniques, primarily focused on polar regions and calibration adjustments.

Table I: Versions CCI+SSS 5.x and their characteristics.

Version	Key Characteristics
V5.1	Calibration after ice filtering based on the current ISAS climatology.
V5.2	Correction of V5.1 by adjusting for the difference between the median of the current ISAS climatology and the median of the new climatology by Nicolas Kolodziejczyk.
V5.3	Calibration based on the median of L4 climatological fields (after ice filtering) and the new climatology by Kolodziejczyk.
V5.4	Hybrid version combining V5.1 and V5.3, with a hyperbolic tangent (tanh) transition function applied between 65°N and 70°N.

V5.5	Another hybrid version combining V5.1 and V5.3, with a cosine transition function applied between 65°N and 70°N.

Important Note

In the following evaluations, V5.x algorithms are assessed across various regions. For regions with a northern boundary south of 65°N, versions 5.1, 5.4, and 5.5 are identical, so either version 5.1 or versions 5.4 and 5.5 are excluded from the graphics, as their results are equivalent. However, all V5.x versions are distinct in regions that include areas north of 65°N.

Reference Data

The V5.x algorithm validation continues to use the reference datasets from V4.x, excluding ISAS SSS gridded fields (details follow). ISAS-20 SSS results, based on comparisons with in-situ measurements, still serve as indicators.

As in V4.x, reference datasets for CCI+SSS algorithm validation include individual in-situ SSS measurements, such as TSG data from the North Atlantic and Southern Ocean.

For V5.x, however, in-situ SSS data have been expanded to cover the Atlantic (40°N–40°S, 80°W–20°E), Arctic (45°N–90°N), and Antarctic (45°S–90°S), integrating all available sources, including Argo, TSGs, drifters, moorings, sail-drones, mammals and others, updated through the CCI v5.x timeline ending in 2023 (see Figure 1). This expanded dataset is prioritized for quantitative assessments of V5.x both relative to V4.4 and in comparisons within V5.x versions.

Hereafter, "in-situ" refers to all SSS data accessible from Pi-MEP within these three regions. The collocation method here also differs from Pi-MEP's standard approach (details follow). While "Arctic" and "Antarctic" traditionally denote CCI products above 45° latitude, covering mid- and high latitudes, all CCI V5.x products remain global in coverage.

Major Results

Unless otherwise noted, all evaluated products are CCI+SSS monthly (30-day) datasets.

The V5.x versions cover an extended period compared to V4.4, which ended in October 2022, now reaching through December 2023 (see illustration with V5.5 in Figure 2).

Significant differences between versions V5.x and V4.4 are observed in the open ocean, primarily north of 60°N and south of 60°S, consistently over time (Figure 2, Bottom). Differences are particularly notable during specific periods, including 2015 north of 45°N and 2010–2011 across all latitudes.

1) Performance comparison of CCI+SSS V5.x with reference data in open ocean regions

Comparison with collocated in-situ data located beyond coastal areas and sea-ice edges, as depicted in the box plots (Figure 3) and the corresponding numerical values (Table II -Table IV), indicates that differences between CCI+SSS V5.x versions and reference datasets are generally minor. Percentiles (1st, Q1, median, 75th, and 99th), along with the standard deviation of differences (stddiff), mean absolute difference (mad), and correlation (r), remain close across versions, suggesting no single V5.x version consistently outperforms others.

Potential improvements of V5.x over V4.4 are indicated by green cells in Table II - Table IV. In Atlantic, version 5.1 (identical to 5.4 and 5.5, as noted earlier) generally performs better, though not consistently (Table II). In Arctic, version 5.5 shows similarly good performance (Table III). However, in Antarctic, V4.4 slightly outperforms V5.x in metrics such as stddiff, mad, and specific percentiles (Table IV), although V5.2 demonstrates moderate improvements in some instances.

Overall, V5.x versions exhibit relatively strong agreement with reference data in Atlantic and Arctic, collectively providing most incremental improvements over V4.4, though without substantial breakthroughs. In Antarctic, V5.x versions do not consistently improve upon V4.4.

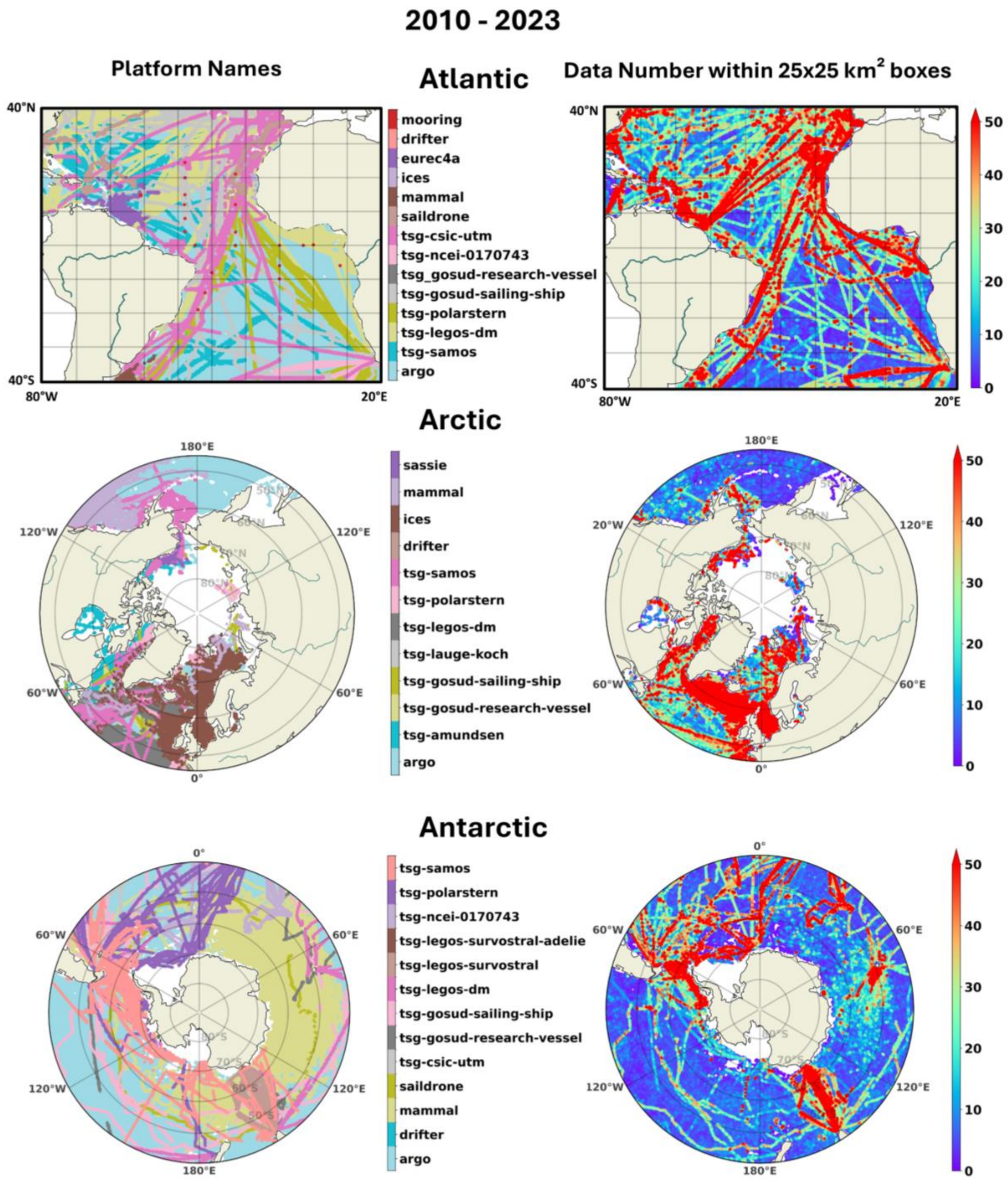

Figure 1: All available in-situ SSS data from Pi-MEP for 2010–2023 in Atlantic (Top, 80°W-20°E, 40°S-40°N), Arctic (Middle, Latitude ≥ 45°), and Antarctic (Bottom, Latitude ≤ -45°) regions, displayed using Healpix gridding with 25x25 km² boxes (see text for explanation). Left: Platform names (with Argo at the bottom in light blue). Right: Total data count over the entire period (no time binning is applied yet).

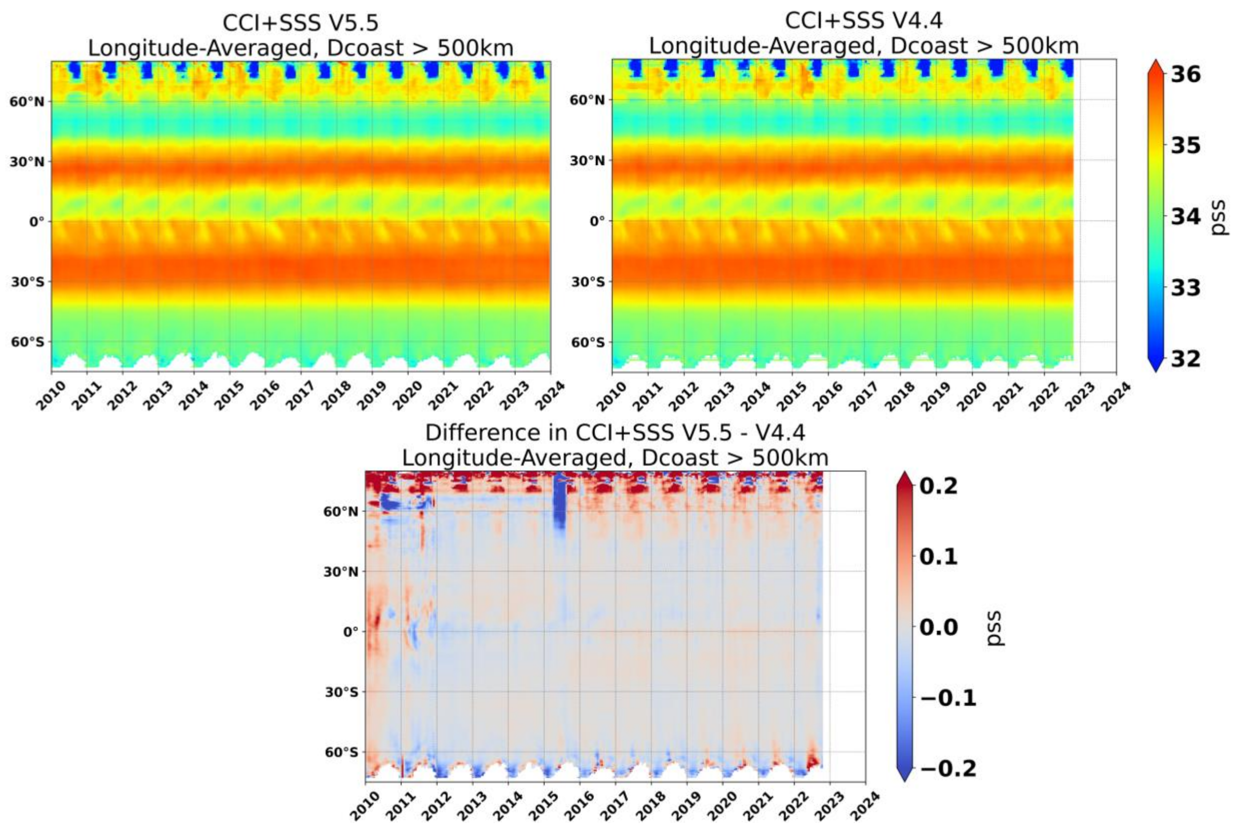

Figure 2: Hovmöller diagrams of CCI SSS datasets and their difference over time (x-axis) and latitude (y-axis), globally averaged across all longitudes for locations more than 500 km from the coast. Top Left: CCI+SSS V5.5, extended to December 2023. Top Right: CCI+SSS V4.4. Bottom: V5.5 – V4.4 SSS difference.

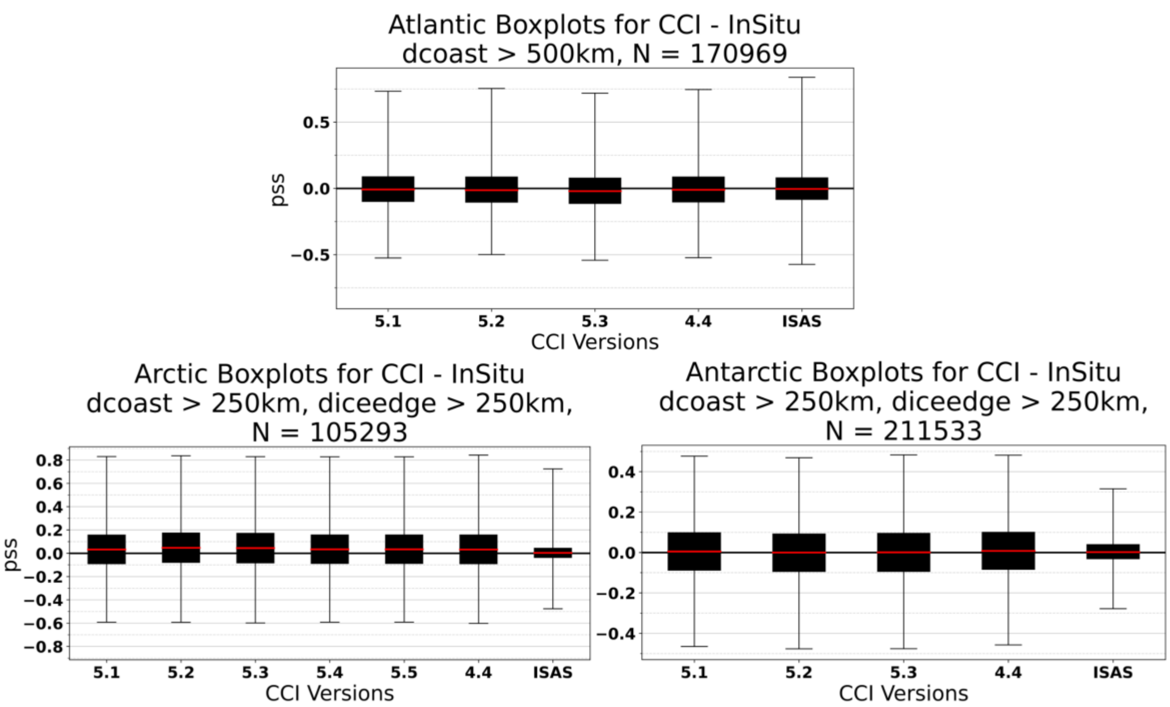

Figure 3: Boxplot diagrams showing the 1st, 25th, 50th (median), 75th, and 99th percentiles of differences between each CCI+SSS version and reference data (in-situ). For in-situ comparisons, ISAS20 (Argo-based) is also included as a reference

product. Regional details: Top: Atlantic, 80°W to 20°E, 40°S to 40°N, >500 km from coasts; Bottom-Left: Arctic, 45°N to 90°N, >250 km from coasts and sea-ice edge; Bottom-Right: Antarctic, south of 45°S, >250 km from coasts and sea-ice edge. Versions 5.1 to 5.5 are compared to version 4.4. For Atlantic and Antarctic, V5.4 and V5.5 are omitted (see note in text). Numerical values, including std of difference, mean absolute difference and correlation are given in Table II.

Table II: Numerical values in pss of the 1st, 25th, 50th (median), 75th, and 99th percentiles for differences between each CCI+SSS version and reference in-situ data across Atlantic, Arctic, and Antarctic regions, as visualized in Figure 3. The table includes standard deviation, mean absolute difference and correlation results. For Atlantic, data are beyond 500 km from coasts; for Arctic and Antarctic regions, distances exceed 250 km from both coasts and sea ice edges. Green cells indicate improvements in version 5.x over 4.4, with darker green for greater improvements, while orange marks better performance by version 4.4, showing degradation in 5.x (there is no such occurrence here in Atlantic). Blue cells indicate ISAS20 outperformance over CCI, noting that ISAS20 exclusively uses Argo data.

Region	Product	1st Percentile [pss]	Q1 (25%) [pss]	Median [pss]	Q3 (75%) [pss]	99th Percentile [pss]	StdDiff [pss]	MeanAbsDiff [pss]	r
Atlantic 40°S-40°N 80°W-20°E Dcoast>500km N = 170,969 Note: V5.1=V5.4=V5.5	**CCI5.1**	-0.525	-0.099	-0.008	0.087	0.732	0.244	0.142	0.953
	CCI5.2	-0.499	-0.104	-0.014	0.086	0.753	0.246	0.143	0.952
	CCI5.3	-0.542	-0.114	-0.020	0.078	0.718	0.246	0.145	0.952
	CCI4.4	-0.523	-0.102	-0.011	0.086	0.746	0.252	0.144	0.949
	ISAS20	-0.573	-0.083	-0.005	0.079	0.838	0.264	0.141	0.945

Table III: Same as Table II but for Arctic region.

Region	Product	1st Percentile [pss]	Q1 (25%) [pss]	Median [pss]	Q3 (75%) [pss]	99th Percentile [pss]	StdDiff [pss]	MeanAbsDiff [pss]	r
Arctic 45°N-90°N +/-180° Dcoast>250km Dicedge>250km N = 105,293	**CCI5.1**	-0.593	-0.090	0.032	0.156	0.831	0.270	0.175	0.971
	CCI5.2	-0.593	-0.078	0.047	0.173	0.838	0.269	0.180	0.971
	CCI5.3	-0.598	-0.082	0.044	0.171	0.829	0.263	0.180	0.973
	CCI5.4	-0.592	-0.088	0.034	0.157	0.828	0.261	0.174	0.973
	CCI5.5	-0.591	-0.089	0.034	0.157	0.828	0.261	0.174	0.973
	CCI4.4	-0.602	-0.089	0.032	0.157	0.843	0.260	0.175	0.973

	ISAS20	-0.476	-0.036	0.004	0.042	0.724	0.191	0.086	0.986

Table IV: Same as Table II but for Antarctic region.

Region	Product	1st Percentile [pss]	Q1 (25%) [pss]	Median [pss]	Q3 (75%) [pss]	99th Percentile [pss]	StdDiff [pss]	MeanAbsDiff [pss]	r
Antarctic 80°S-45°S +/-180° Dcoast>250km Dicedge>250km N = 211,533 Note: V5.1=V5.4=V5.5	CCI5.1	-0.465	-0.087	0.005	0.098	0.476	0.177	0.126	0.807
	CCI5.2	-0.477	-0.094	-0.000	0.092	0.469	0.178	0.127	0.807
	CCI5.3	-0.476	-0.093	0.001	0.094	0.482	0.180	0.128	0.803
	CCI4.4	-0.457	-0.083	0.008	0.100	0.481	0.176	0.125	0.807
	ISAS20	-0.278	-0.031	0.002	0.039	0.315	0.107	0.062	0.926

Note: The ISAS20 dataset, based solely on Argo data and included in all tables, occasionally outperforms all CCI+SSS versions in Atlantic (Table II), and more consistently in Arctic and Antarctic regions (Table III -Table IV).

2) Analysis of CCI+SSS V5.x performance in coastal and ice-edge regions

Regions near the coast (within 500 km in Atlantic and 250 km in Arctic and Antarctic) and close to the sea-ice edge (within 250 km) in Arctic and Antarctic (Figure 4, Table V - Table VII) generally show greater contrasts between V4.4 and V5.x. In Atlantic, coastal areas influenced by low-salinity waters, such as river mouths, display significantly larger differences in the upper SSS percentiles relative to in-situ data compared to offshore regions (Figure 4, Top). In Arctic, extreme percentiles, indicative of highly saline or fresh waters in the products, vary considerably across CCI versions (Figure 4, Bottom Left).

Variability among V5.x versions is more pronounced in these regions, though it is challenging to identify a consistently superior version in Atlantic and Antarctic. Notably, V4.4 often outperforms V5.x in several metrics, such as std of difference and correlation in Atlantic (Table V) and mean absolute difference in Antarctic (Table VII).

In Arctic, however, version 5.5 (and to a lesser extent, 5.4) demonstrates the best performance, surpassing V4.4 and other V5.x versions across most metrics (Table VI). Later results provide examples illustrating this superior performance.

Overall, version 5.5 offers significant improvement over V4.4 for regions near the coast or sea-ice edge in Arctic. In contrast, in Atlantic and Antarctic regions, V5.x versions do not show significant improvements over V4.4.

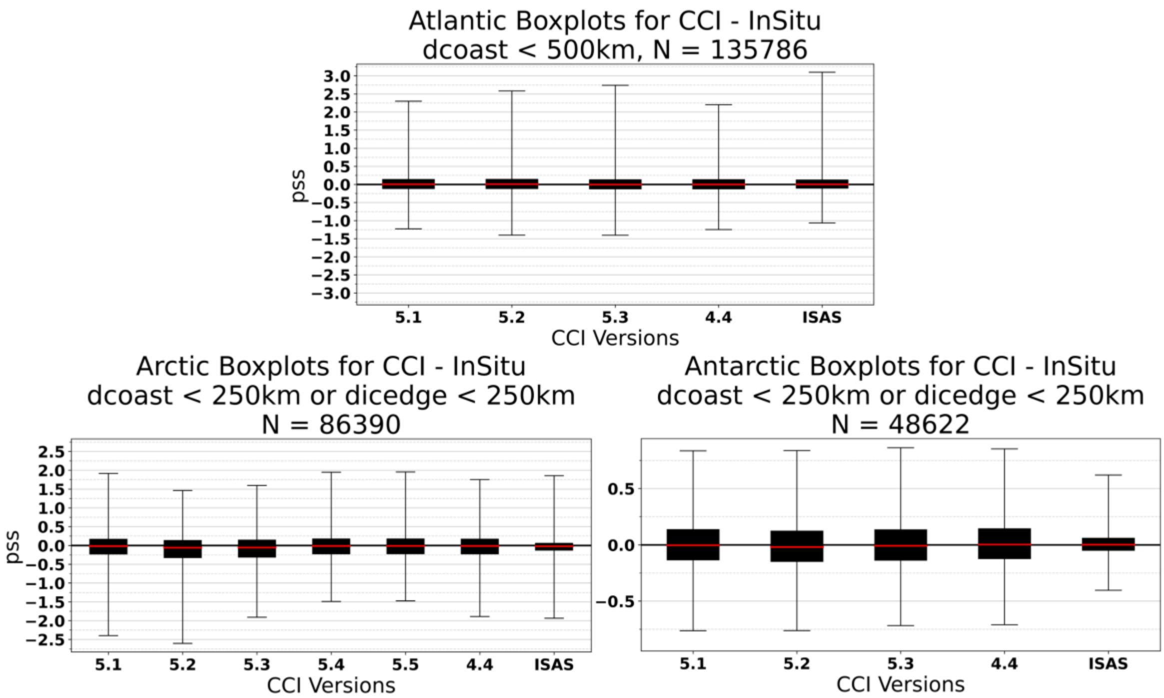

Figure 4: Same as Figure 3, but with distance criteria adjusted: for Atlantic, distances are less than 500 km from the coast; for Arctic and Antarctic, either the distance to the coast is under 250 km or the distance to the sea ice edge is under 250 km. Numerical values, including std of difference, mean absolute difference and correlation, are given in (Table V - Table VII).

Table V: Same as Table II, but with distance criteria adjusted: distances are less than 500 km from the coast. Percentile values coincide with the box plots in Figure 4, Top.

Region	Product	1st Percentile [pss]	Q1 (25%) [pss]	Median [pss]	Q3 (75%) [pss]	99th Percentile [pss]	StdDiff [pss]	MeanAbsDiff [pss]	r
Atlantic **40°S-40°N** **80°W-20°E** **Dcoast<500km** **N = 135,786** **Note: V5.1=V5.4=V5.5**	**CCI5.1**	-1.226	-0.117	0.005	0.136	2.283	0.696	0.269	0.858
	CCI5.2	-1.402	-0.116	0.007	0.140	2.584	0.743	0.286	0.835
	CCI5.3	-1.404	-0.126	-0.003	0.129	2.741	0.770	0.289	0.821
	CCI4.4	-1.247	-0.125	0.000	0.132	2.203	0.679	0.270	0.865
	ISAS20	-1.068	-0.105	0.002	0.122	3.102	0.806	0.286	0.810

Table VI: Same as Table III, but with distance criteria adjusted: distances are less than 250 km from the coast or less than 250 km from sea-ice edge. Percentile values coincide with the box plots in Figure 4, Bottom-Left.

Region	Product	1st Percentile [pss]	Q1 (25%) [pss]	Median [pss]	Q3 (75%) [pss]	99th Percentile [pss]	StdDiff [pss]	MeanAbsDiff [pss]	r
Arctic **45°N-90°N** **+/-180°** **Dcoast<250km** **or** **Dicedge<250km** **N = 86,390**	**CCI5.1**	-2.395	-0.229	-0.016	0.167	1.914	0.803	0.376	0.916
	CCI5.2	-2.605	-0.325	-0.059	0.131	1.464	0.755	0.399	0.929
	CCI5.3	-1.912	-0.311	-0.054	0.140	1.596	0.619	0.370	0.939
	CCI5.4	-1.490	-0.222	-0.013	0.170	1.945	0.598	0.339	0.942
	CCI5.5	-1.474	-0.220	-0.013	0.172	1.953	0.597	0.339	0.942
	CCI4.4	-1.892	-0.225	-0.015	0.169	1.752	0.651	0.351	0.939
	ISAS20	-1.937	-0.126	-0.022	0.056	1.854	0.746	0.285	0.923

Table VII: Same as Table IV, but with distance criteria adjusted: distances are less than 250 km from the coast or less than 250 km from sea-ice edge. Percentile values coincide with the box plots in Figure 4, Bottom-Right.

Region	Product	1st Percentile [pss]	Q1 (25%) [pss]	Median [pss]	Q3 (75%) [pss]	99th Percentile [pss]	StdDiff [pss]	MeanAbsDiff [pss]	r
Antarctic **80°S-45°S** **+/-180°** **Dcoast<250km** **or** **Dicedge<250km** **N = 48,622** **Note: V5.1=V5.4=V5.5**	**CCI5.1**	-0.763	-0.134	-0.003	0.134	0.836	0.358	0.200	0.575
	CCI5.2	-0.762	-0.147	-0.020	0.123	0.838	0.355	0.201	0.586
	CCI5.3	-0.718	-0.136	-0.009	0.134	0.862	0.354	0.201	0.581
	CCI4.4	-0.710	-0.124	0.003	0.142	0.852	0.354	0.197	0.574
	ISAS20	-0.404	-0.048	0.002	0.058	0.620	0.270	0.098	0.737

3) Regional and temporal evaluation of V5.x versions relative to V4.4

From 1) and 2), versions V5.1, V5.3, and V5.5 exhibit the most significant changes compared to V4.4 across most results, with V5.1 and V5.5 notably distinct in the Arctic. This summary now focuses on key findings comparing V5.1 and V5.3 to V4.4 in the Atlantic and Antarctic (where V5.1 is identical to V5.5) and V5.1 and V5.5 to V4.4 in the Arctic (where V5.3 contributes to V5.5).

Annual boxplots and metrics for V5.1 and V5.3 in the Atlantic open ocean confirm improvements for V5.1 over V4.4, particularly during the 2010–2011 period (Figure 5). Except in 2010, V5.3 shows a degradation in mean bias throughout the period. After 2011, V5.1 shows no consistent improvements over V4.4, while V5.3 exhibits slight degradation, particularly in mean absolute difference. In 2023, both V5.x versions display a pronounced negative systematic bias (~-0.05 pss for V5.1), the largest observed during the period.

In the Arctic region near coasts or the sea-ice edge (Figure 6), the median difference for V5.x is generally negative until 2013 (<-0.05 pss) and shifts to positive or near-zero from 2014 onward. Significant improvement appears in 2015 with V5.1 and V5.5, resolving an issue in V4.4 related to the SMAP data transition, consistent with major changes shown in Figure 2 (Bottom). V5.5 is confirmed as the best version in this region, with clear improvements over V4.4, while V5.1 shows degradation, particularly from 2010 to 2016. Beyond this, consistent improvements over V4.4 are absent except for V5.5 in the final (incomplete) year, 2022.

In Antarctic region near coasts or the sea-ice edge, V5.1 and V5.3 again show minimal differences, with occasional improvements over V4.4 (Figure 7). Notably, the median difference is closer to zero from 2010 to 2015. However, the bias is not improved afterward. Moreover, there is a slight degradation of both V5.x over the period, as seen from the other metrics

Overall, the V5.x products, like V4.4, exhibit notable interannual variability in bias relative to in-situ data.

In summary, while V5.x products show targeted improvements over V4.4 in specific regions and periods, particularly in Arctic and early years of the dataset, their performance across other regions and times remains variable.

4) Global analysis of bias normalized by satellite and mismatch uncertainty

This analysis provides validation insights for both product bias and uncertainty and is applied on the weekly (7-day) products. CCI (reduced-centered) biases from the CCI+SSS weekly products relative to in-situ data are similar for V5.1 and V5.5, but differ significantly from V4.4 (Figure 8). In 2010, V5.x shows substantial improvement, with a notable decrease in bias. From 2010 through 2013, V5.x alternately improves and degrades, while most of 2013 and 2014 shows no notable difference between V4.4 and V5.x.

A major improvement appears in 2015 across V5.x versions, significantly decreasing bias (as also noted in Figure 6). Post-2015, V5.x versions trend algebraically toward higher values: biases are sometimes larger than V4.4 when positive but closer to zero when V4.4 is negative. After 2015, bias in V5.x versions remains more stable with a narrower range. Although V5.x versions generally perform better, the minor differences between V5.1 and V5.5 curves make it difficult to conclude which version is superior.

5) Impact of climatological adjustments in Arctic regions

The differences between CCI V5.x are minimal, making it challenging to distinguish between them or to definitively favour one, a key objective of the PVASR. This summary here highlights the rationale for producing version 5.5, which incorporates a new constant climatological adjustment developed by Nicolas Kolodziejczyk to align with the latest ISAS climatological background (see Table I).

Systematic biases in CCI+SSS products are notably significant in the Arctic Ocean and surrounding areas. To address this, CCI V5.x versions after V5.1 include a new background adjustment based on the median of Kolodziejczyk's updated climatology. Figure 9 illustrates the effect of this adjustment, with improvements benchmarked against in-situ data.

We observe a reduction in mean bias within the Arctic Ocean in V5.5 (Figure 9, Top-Left) compared to V5.1 (Top Right), which previously exhibited a pronounced fresh bias relative to in-situ data (ΔSSS < -1 pss). The difference between V5.1 and V5.5 SSS partially reflects this updated climatological adjustment (Bottom).

However, the impact of this positive adjustment on the entire Arctic region is limited, as shown in previous results (Table III, Table VI). While Figure 9 presents encouraging results, they involve only a small subset of data points, limiting their overall influence on the broader statistical outcomes displayed in prior tables.

6) Conclusion: incremental improvements and future directions for V5.x versions

In conclusion, the V5.x versions of the CCI+SSS product show incremental improvements over V4.4, particularly from mid-latitudes to the pole in the Northern Hemisphere and during specific periods, while maintaining consistent performance across other regions. Among the V5.x products, V5.5 delivers the most consistent improvements over V4.4. However, despite targeted advancements such as enhanced climatological adjustments and sensor corrections, variability among V5.x versions and the limited scope of improvements underscore the need for further refinement to address systematic biases and achieve greater global accuracy.

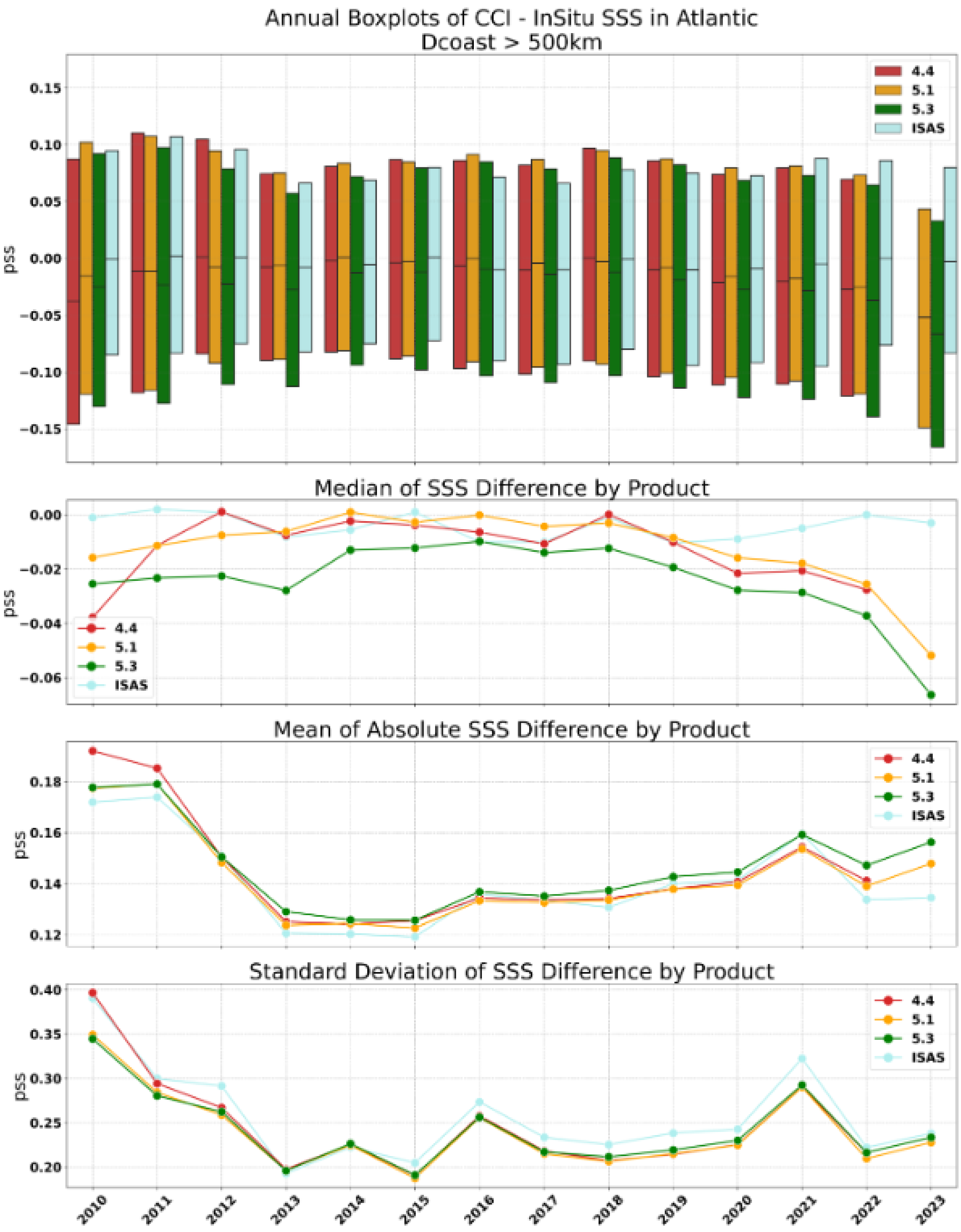

Figure 5: Annual boxplots and metrics of differences between collocated CCI and in-situ SSS for CCI versions 4.4 (Red), 5.1 (Orange), 5.3 (Green), and the reference ISAS20 dataset (Cyan) in Atlantic region, more than 500 km from the coast. Consistent collocated sampling is used across CCI versions each year for uniform percentile and metric calculations. Note: V5.x (5.1 and 5.3) fully covers 2023, while V4.4 data ends in 2022. Panels from Top to Bottom: Panel 1 shows annual boxplots with the 25th, 50th (median), and 75th percentiles; Panel 2 displays the median difference; Panel 3, the mean absolute difference (mad); and Panel 4, the std of difference (stddiff).

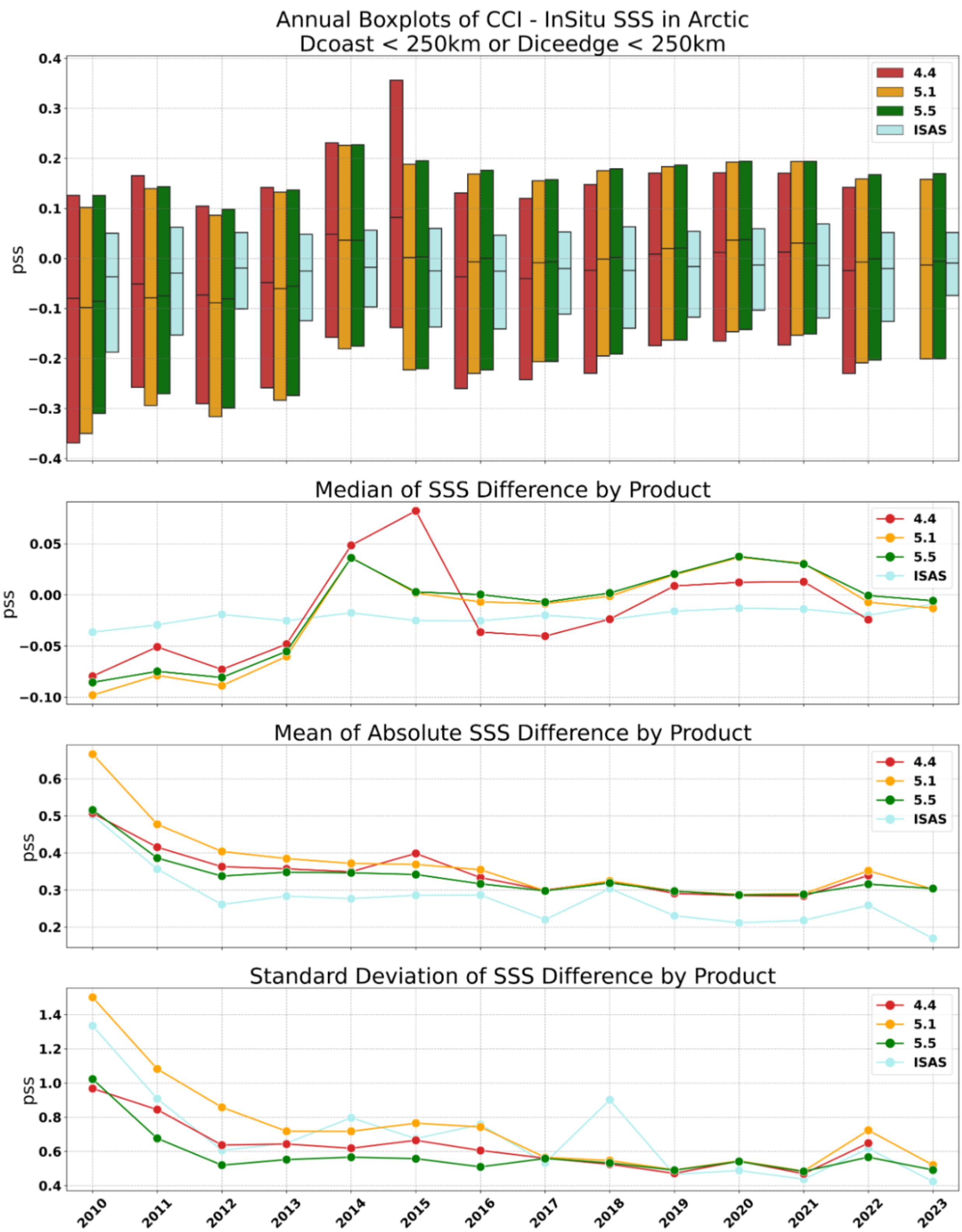

Figure 6: Same as Figure 5, but for Arctic region close to coast or sea-ice edge, and V5.5 is displayed instead of V5.3.

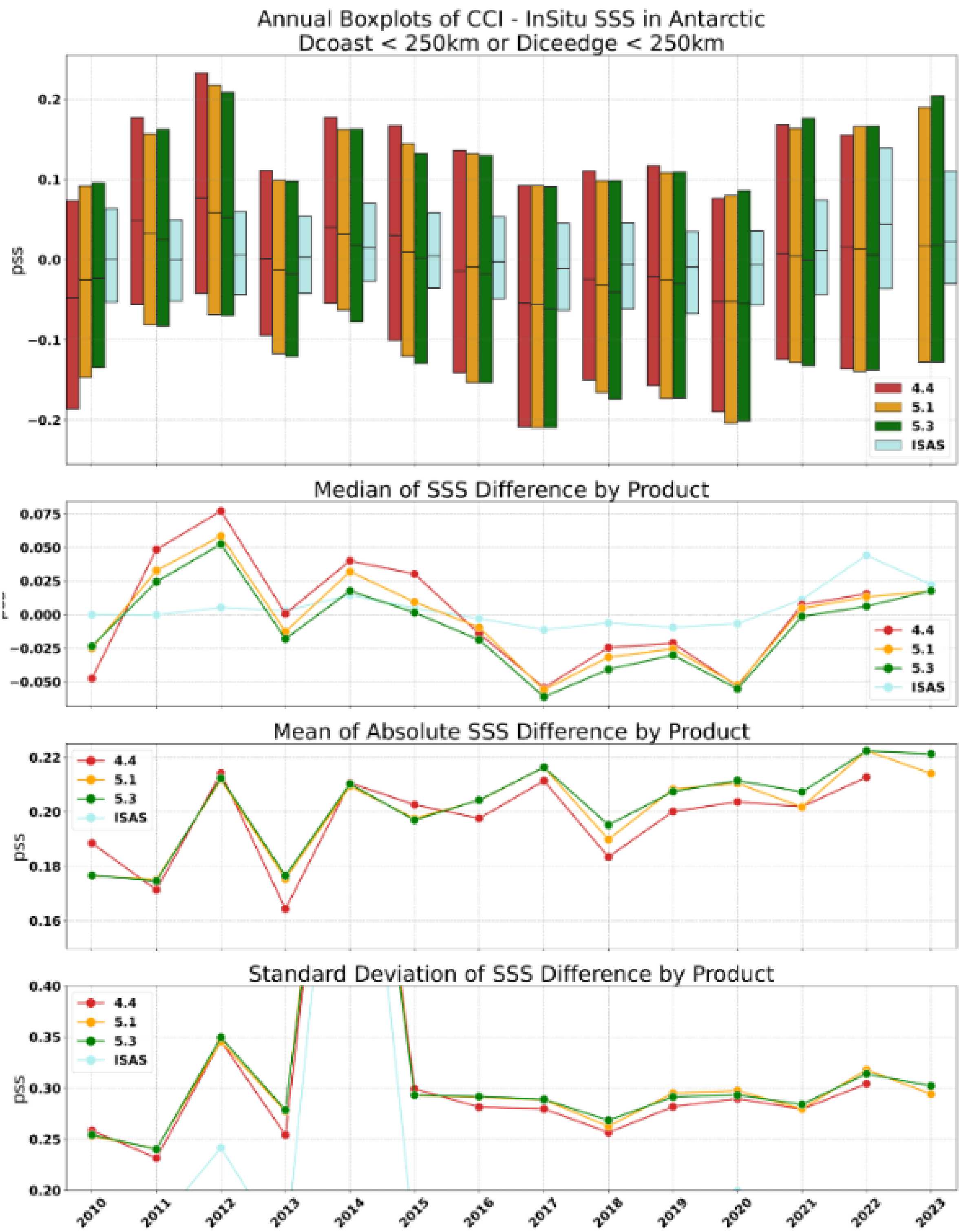

Figure 7: Same as Figure 5, but for Antarctic region near the coast or sea-ice edge. (Note: ISAS shows much lower mean absolute difference and standard deviation levels, making it partially or not visible within the axes. Additionally, results for 2014 are not relevant due to a low data count and the presence of in-situ data outliers.)

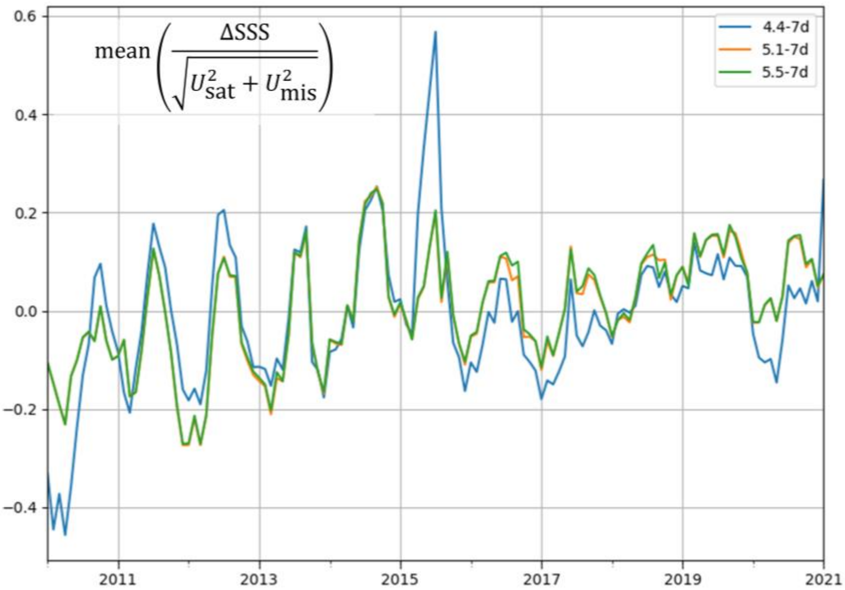

Figure 8: Normalized differences (ΔSSS) between in-situ measurements from Pi-MEP and weekly CCI data, accounting for satellite uncertainty (Usat) and sampling mismatch (Umis, scaled using SSS spectral analysis for variability below 1/12° resolution of GLORYS reanalysis), as a function of time. Measurement uncertainty is neglected. The global distribution is expected to be Gaussian with a mean of 0 and standard deviation of 1 (details in RD08). The curves show the mean normalized differences for: weekly CCI-SSS v4.4 (blue), v5.1 (orange), and v5.5 (green).

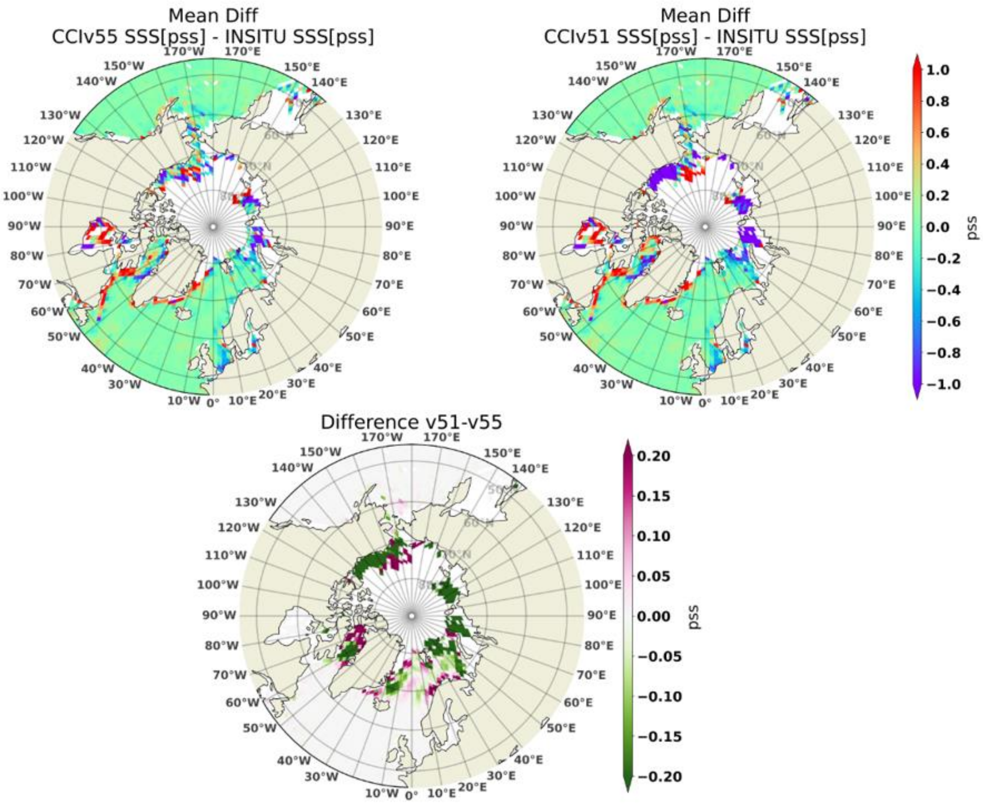

Figure 9: Maps of mean differences (systematic bias) between collocated CCI V5.x and in-situ SSS in Arctic region, displayed on a degraded Healpix grid (100 x 100 km²). Top Left: CCI V5.5. Top Right: CCI V5.1. Bottom: Difference between CCI V5.1 and V5.5.

2 Introduction

2.1 Scope

The report summarizes the results of the second round-robin algorithm comparisons for the CCI+SSS phase 2 project, evaluating the performance of CCI+SSS versions 5.1 through 5.5 global products. This round-robin exercise spans regions with particular emphasis on the tropical/mid-latitude Atlantic (referred to as the "Atlantic" region), Arctic, and Antarctic areas. The primary objective is to assess advancements since the previous phase, version 4.4. Additionally, the report tries to identify strengths and limitations within the products, offering insights to guide future improvements and development efforts. See Section 1 for a summary of the PVASR.

2.2 References

2.2.1 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-v5.0
AD06	End-to-End Essential Climate Variable Uncertainty Budget (E3UB)	SSS_cci-D2.3-E3UB-v5r0

2.2.2 Reference Documents

ID	Document	Reference
RD01	Alory G., T. Delcroix, P. Téchiné, D. Diverrès, 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. Deep Sea Res., 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). Biometry the principles and practice of statistics in biological research (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, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, doi: 10.1109/JSTARS.2013.2252602, 2013.	

ID	Document	Reference
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, Bulletin of American Meteorological Society, 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. Frontiers in Marine Science, 6(243), 23p. Publisher's official version : https://doi.org/10.3389/fmars.2019.00243.	
RD08	Thouvenin-Masson, C.; Boutin, J.; Vergely, J.-L.; Reverdin, G.; Martin, A.C.H.; Guimbard, S.; Reul, N.; Sabia, R.; Catany, R.; Hembise Fanton-d'Andon, O. Satellite and In Situ Sampling Mismatches: Consequences for the Estimation of Satellite Sea Surface Salinity Uncertainties. Remote Sens. 2022, 14, 1878. https://doi.org/10.3390/rs14081878	

2.3 Structure of the document

The PVASR is structured as follows:

- Section 1: Executive summary.
- Section 2: Introduction.
- Section 3: Definition of key terms used throughout the document.
- Section 4: Overview of tasks performed and comparison results.
- Section 5: Round Robin (RR) methodology:
    - 5.1: In-situ data details and collocation methods
    - 5.2: Metrics used in the RR tests.
- Section 6: Presentation of CCI+SSS satellite products evaluated against in-situ data.
- Section 7: Evaluation results, including:
    - 7.1: Verification across large regions encompassing all latitudes, refers to Section 1,
    - 7.2: Performance along high-latitude TSG tracks,
    - 7.3: Products evaluation summary
    - 7.4: Open issues and discussion.
- Section 8: Future perspectives for the CCI+SSS PVASR.

3 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 √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 10).

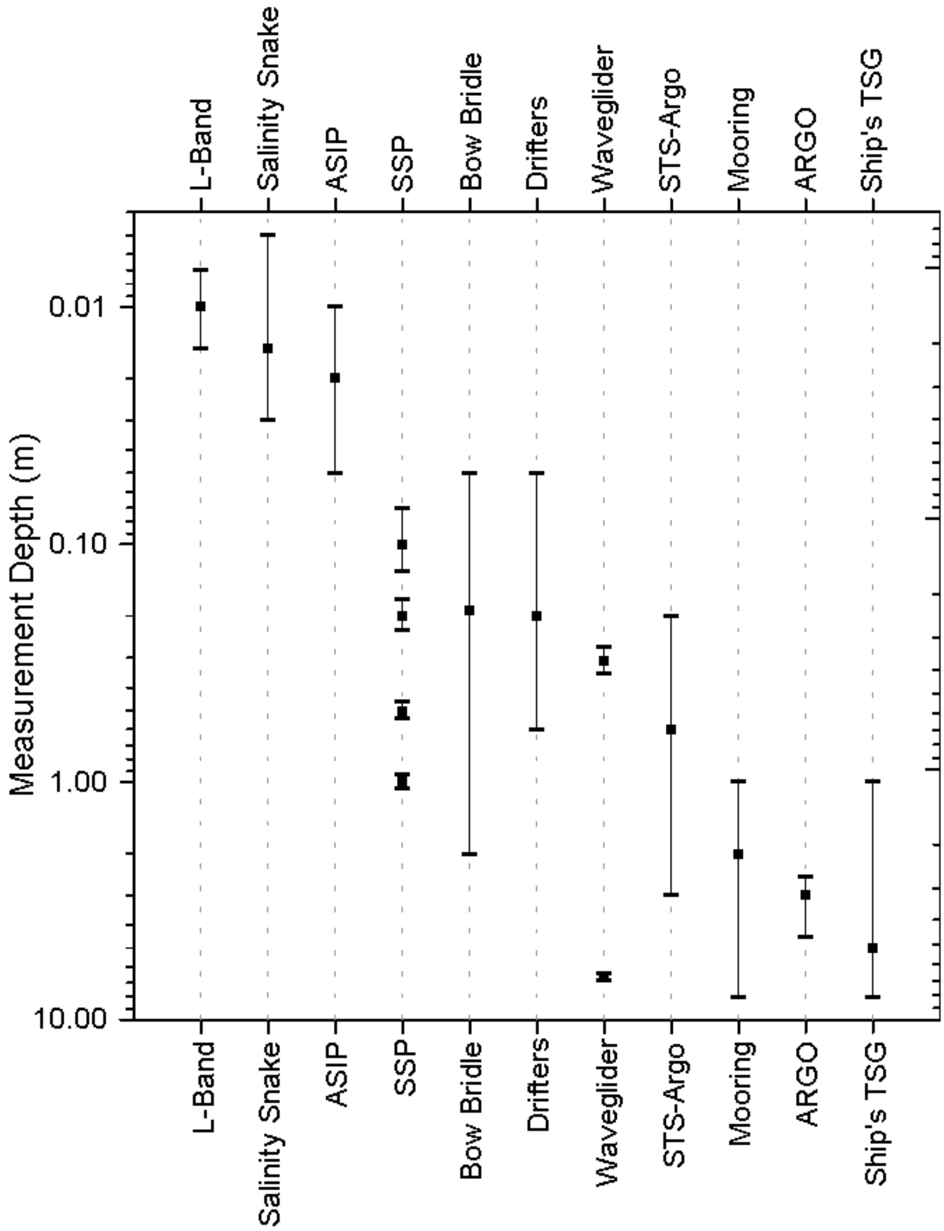

Figure 10: Scale portraying the typical depth at which near-surface salinity is measured by various sensors/platforms. The small squares show the average measurement depth and the capped lines show the range for that average. For profiling platforms (ASIP, Bow Bridle, STS-Argo, Argo) the range represents 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 ºC to 35 ºC (where the "mean" value shown in the figure is for 20 ºC and 35 pss). (Figure taken from [RD 06]).

4 Overview

Several metrics are used to evaluate the performance of the algorithms/products, each addressing uncertainties from distinct types of errors managed differently in satellite processing:

- **Bias**: Calculated as the difference between satellite and in-situ SSS, this metric helps identify systematic errors due to radiometer calibration, land-sea interference, sun contamination, RFI, etc.

- **Systematic Bias**: The mean or median difference between satellite and in-situ SSS reveals consistent bias between satellite products and in-situ SSS.

- **Percentiles:** The differences between satellite and in-situ SSS can be represented as a cumulative distribution, allowing extraction of quantiles or percentile values that capture the range of differences, from average values to more extreme cases. Box plots, which include the 1st, then 25th, 50th (**median**) and 75th (all within the interquartile range), and 99th percentiles, provide a concise visualization of these difference levels. This setup enables effective comparison of one product to others for ranking purposes, highlighting both typical and extreme difference values**.**

- **Standard Deviation**: Characterizes random errors resulting from measurement noise or other poorly characterized sources. The “robust” standard deviation, calculated using the median, enhances statistical reliability by reducing the impact of extreme values and outliers.

- **Absolute Bias**: Sometimes measured as the mean or median of absolute differences between satellite and in-situ SSS, highlighting absolute deviations.

- **Pearson Correlation (r)**: Measures the (square root) coefficient of determination between satellite and in-situ SSS, sensitive to strict filtering or smoothing of extreme values, such as low SSS in river plumes.

- **Distribution Properties**: The statistical distribution of the centered reduced variable provides insights into the appropriateness of CCI L4 SSS uncertainties alongside other performance metrics.

When evaluating metric significance, we sometimes use a bootstrap procedure to calculate a 95% confidence interval, estimating the sampling distribution but not accounting for observational and sampling uncertainties. Comparisons between confidence intervals across products allow for deeper evaluation and insights into satellite-derived SSS performance.

In this report, we used multiple datasets to comprehensively evaluate the satellite-derived SSS products available for producing CCI+SSS version 5. For this PVASR implementation, we focused on regions and global products (all CCI products are global in this version) and used various in-situ measurements. These include repetitive ship tracks across the North Atlantic and around Antarctica (TSGs) and a diverse ensemble of in-situ data from platforms collected in the Pi-MEP database, covering regions designated as Atlantic, Arctic, and Antarctic. Refer to Figure 1 for the specific geographical boundaries of Atlantic, Arctic, and Antarctic regions.

This report generally excludes very high latitudes of the Arctic Ocean, such as the Beaufort and Chukchi Seas, Nordic Seas, and Barents Seas, though some of our in-situ reference data are located there and used for comparisons. Specific validations in these areas will be performed separately by the validation team involved in the CCI+SSS option.

5 Round robin methodology

This section outlines the RR test methodology used to compare and validate various satellite products. The approach aims to ensure robust validation and avoid biases by evaluating CCI+SSS versions within three distinct regions: Atlantic, Arctic, and Antarctic. These regions capture unique salinity and environmental conditions while encompassing a global spectrum of surface salinity. To maintain objectivity, the global ISAS Argo-based dataset is excluded as a reference, as its climatology has been applied in CCI product calibration, particularly for mean bias corrections. Additionally, the updated climatology introduced by Nicolas Kolodziejczyk in versions 5.2 and later (Table I) further compromises ISAS's independence as a benchmark, potentially favouring certain CCI+SSS versions.

Regional validation: overview

All regions used for validation are independent of the SMOS OTT region.

1) Atlantic Region (40°S–40°N, 80°W–20°E):
   - All available in-situ data from various platforms, including Argo, are used for evaluations in the mid- and tropical latitudes of Atlantic Ocean.
   - As noted in Section 1, Table II and Table V, CCI versions V5.1, V5.4, and V5.5 are identical in this region.
2) Arctic Region (45°N–90°N):
   - In-situ data from multiple platforms, including Argo, are incorporated to evaluate performance in higher latitudes, accounting for the greater salinity variability typical of the Arctic.
   - In the Northern Atlantic, repetitive Thermosalinograph (TSG) track data are used, offering consistent monitoring of maximum and minimum SSS zones and capturing diverse SSS regimes.
   - Although all V5.x versions are distinct in the Arctic, the Northern Atlantic domain boxes used for TSGs lie south of 65°N, where V5.1, V5.4, and V5.5 are identical.
3) Antarctic Region (45°S–90°S):
   - Repetitive TSG tracks and all available in-situ data from multiple platforms, including Argo, are used to evaluate performance under the unique salinity conditions of high-latitude Antarctic regions.
   - As with the Atlantic, versions V5.1, V5.4, and V5.5 are identical in this region (Section 1, Table IV and Table VII).

Standardization and metrics

The RR methodology standardizes product validation by comparing multiple algorithms and related products against a common set of in-situ data, enabling direct intercomparison of CCI+SSS versions. Each algorithm is tested using consistent validation data and evaluated through standard metrics.

In this PVASR, the RR method employs box plot percentiles (1st, 25th, 50th, 75th, and 99th) to assess differences between satellite and in-situ data. Metrics such as bias, standard deviation, and correlation coefficients are also used to provide a comprehensive evaluation. These results inform algorithm selection recommendations (see Section 1 for major findings).

5.1 In situ data and collocation methodology

5.1.1 Atlantic, Arctic, Antarctic "all" in-situ data

1) Data

For the Atlantic, Arctic, and Antarctic regions, collocation of all available in-situ data with CCI products enables detailed comparisons between the products and reference data. Differences can be analyzed across all data or within specific spatial or temporal subsets, as already shown in Section 1. The data ensembles are sourced from the Pi-MEP facility (https://www.salinity-pimep.org/).

- Atlantic Region (40°S-40°N, 80°W-20°E) (illustrated in Figure 1, Top).
- Arctic Region (45°N-90°N) (Figure 1, Middle).
- Antarctic Region (45°S-80°N) (Figure 1, Bottom).

2) Collocation procedure

The CCI SSS products analysed in this PVASR are global, on a 0.25° x 0.25° grid (approx. 25 x 25 km² near the equator, much smaller at high latitudes). As the resolution of SMOS and SMAP, the primary sources for CCI products, is around 40-50 km, sampling products at half-resolution (25 km) is appropriate. To address grid bin size variation with latitude and maintain consistent spatial surfaces, we employ ~25x25 km² Healpix boxes (https://healpix.jpl.nasa.gov/) across all latitudes (virtually covering Earth with 786,432 pixels).

In-situ data from all platforms in each region are grouped into these Healpix boxes. Monthly bins are constructed by averaging in-situ data within each box for each month, yielding data for subsets of months per box. Within each box, the mean location of in-situ data per month is used for collocation with the nearest CCI product grid center.

This collocation links monthly CCI products with averaged in-situ data, enabling calculations of differences, correlations, and additional metrics. Further aggregations, such as spatially averaged "global" time series or resolution adjustments using larger Healpix boxes (e.g. Figure 9), are applied as described in subsequent sections.

5.1.2 Monthly binned ship tracks in North Atlantic

1) Data

These data are derived from measurements made by merchant ships using TSG along two transects: B-AX01 between southern Greenland and Denmark, and B-AX02 between Newfoundland and Iceland (Figure 11, red and black boxes), 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.

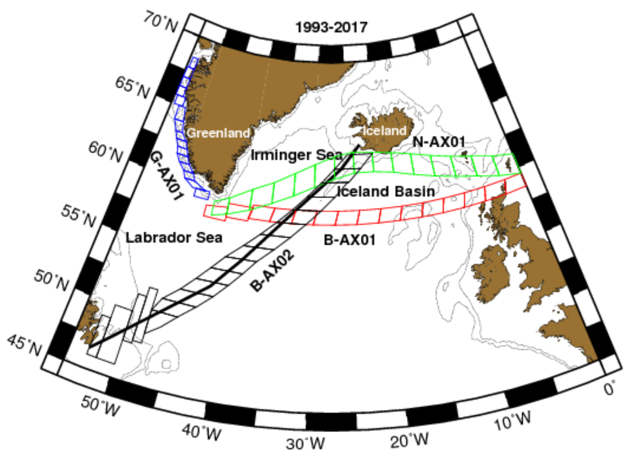

Figure 11: Map of boat transects and their division into regular boxes. An example of a ship track is shown as a solid black line along B-AX02 (Reverdin et al., 2018). In this work, only the B-AX01 and B-AX02 transects are used.

2) Collocation procedure

We averaged the satellite data over a month to compare satellite products with the monthly binned ship tracks dataset. Figure 11 describes the boxes (red and black) utilised during the averaging process.

5.1.3 Research vessel TSG in the Southern Ocean

1) Data

The Southern Ocean has several in-situ datasets, including recurrent measurements along specific paths over the years (Figure 12, Left). The temporal coverage of seasons and year-to-year variations are needed for the RR task, and the recurrence over the years provides this.

For this report, we have considered three ensembles of paths that are relatively well-sampled and only measure at depths above 11m (Figure 12, Right). The first path group is centred on a straight line near the 145°E meridian from Tasmania to Antarctica. The second path 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. The third path runs approximately from the tip of the Antarctic Peninsula to South America's Tierra Del Fuego.

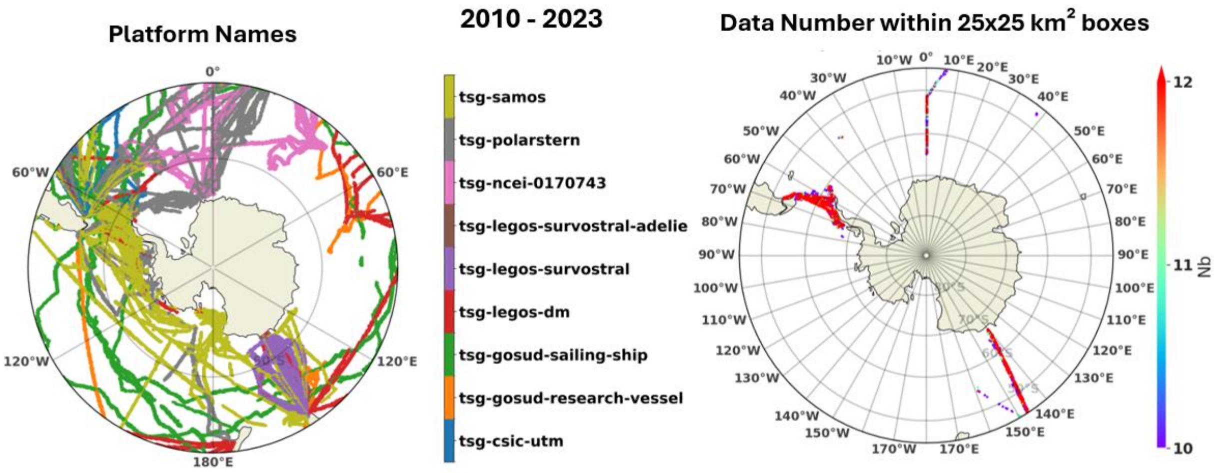

Figure 12: Left: TSG tracks spanning from 2010 to 2023, with each type of TSG cruise being represented by a different color. Right: Number of monthly-averaged TSG data within about 25x25 km² boxes along recurring ship tracks around Antarctica (minimum number of months is 10).

2) Collocation method

We follow the same procedure as Section 5.1.1, except only with the TSG data. Once collocations are completed within the ~25 x 25 km² pixels, metrics are computed within 150km radius discs.

5.2 Metrics

For selecting the algorithm, the metrics introduced in Section 4 are computed as described below (*horizontal bars indicate the mean over a set of measurements)*.

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

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

- Robust standard deviation (*std diff rob*):

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

- (Systematic) Bias:

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

- Mean absolute difference (mad):

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

- Correlation coefficient r or c the correlation coefficient.

$$r = \frac{\sum_{i=1}^{n}\left(SSS_{\text{sat}}^{i} - \overline{SSS_{\text{sat}}}\right)\left(SSS_{\text{in-situ}}^{i} - \overline{SSS_{\text{in-situ}}}\right)}{\sqrt{\sum_{i=1}^{n}\left(SSS_{\text{sat}}^{i} - \overline{SSS_{\text{sat}}}\right)^{2}} \cdot \sqrt{\sum_{i=1}^{n}\left(SSS_{\text{in-situ}}^{i} - \overline{SSS_{\text{in-situ}}}\right)^{2}}}$$

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

The SSS difference is divided by the time- and space-varying uncertainty magnitude calculated as:

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

for each measurement point, where $\text{sat}_{\text{uncertainty}}$ is the satellite uncertainty at that point (random uncertainty estimated from the L4 generation).

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

This approach, however, is a preliminary method for analyzing bias and satellite uncertainty. To refine this, as shown in [RD08], we also apply normalization using both satellite uncertainty and an estimate of sampling mismatch uncertainty. The sampling mismatch is calculated based on the GLORYS reanalysis, which provides a model-based SSS variability that accounts for spatial and temporal resolution differences between satellite and in-situ measurements. These combined analyses allow for a more comprehensive assessment of SSS variability and are presented in later results

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

As mentioned above, using the median calculation.

In these equations, $SSS_{\text{in\_situ}}$ corresponds to the salinity of the in-situ measurements, after the colocation processing described in Section 5.1., and $SSS_{satellite}$ corresponds to the salinity sensed by the satellite.

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

In this RR test, we evaluated five versions of the CCI+SSS products (see Section 1 and Table I), each a global product on a 0.25° x 0.25° grid. These versions were processed as part of the second iteration of CCI Phase 2, with the final version from Phase 2's first iteration (version 4.4) serving as the previous reference. Consequently, our report focuses on comparisons between CCI+SSS V4.4 (iteration 1) and the global V5.1 - 5.5 (iteration 2) products.

The CCI+SSS product versions are available as 7-day analyses, or monthly (30-day) averages sampled every 15 days. Unless otherwise specified, the evaluations focus on the monthly products.

7 Algorithm/Product evaluation

7.1 Verification across large regions encompassing all latitudes

This analysis, using in-situ data collocated with CCI+SSS V4.4 and V5.x products across the Atlantic, Arctic, and Antarctic regions, is presented and summarized in Section 1. The conclusion is that the V5.x versions of CCI+SSS show incremental improvements over V4.4, particularly in the Northern Hemisphere high latitudes, with V5.5 providing the most consistent gains, though further refinement is needed to address variability and systematic biases.

7.2 Performance along high-latitude TSG Tracks

1) Overall performance of V5.x along high-latitude TSG tracks

In the high-latitude North Atlantic along the TSG tracks, the new V5.1 (identical to V5.5) and V5.3 versions do not consistently reduce differences with in-situ data compared to V4.4, as shown in Figure 14 and Figure 15. The median, which is already near zero for many boxes in V4.4 (except near the track extremities), does not consistently improve in V5.1 and shows even less improvement in V5.3. The IQR for V5.x versions varies, being smaller (indicating improvement) or larger (indicating degradation) than in V4.4 across different boxes. However, V5.x versions display fewer outliers than V4.4, suggesting reduced volatility in these regions. Examples of time series are presented hereafter.

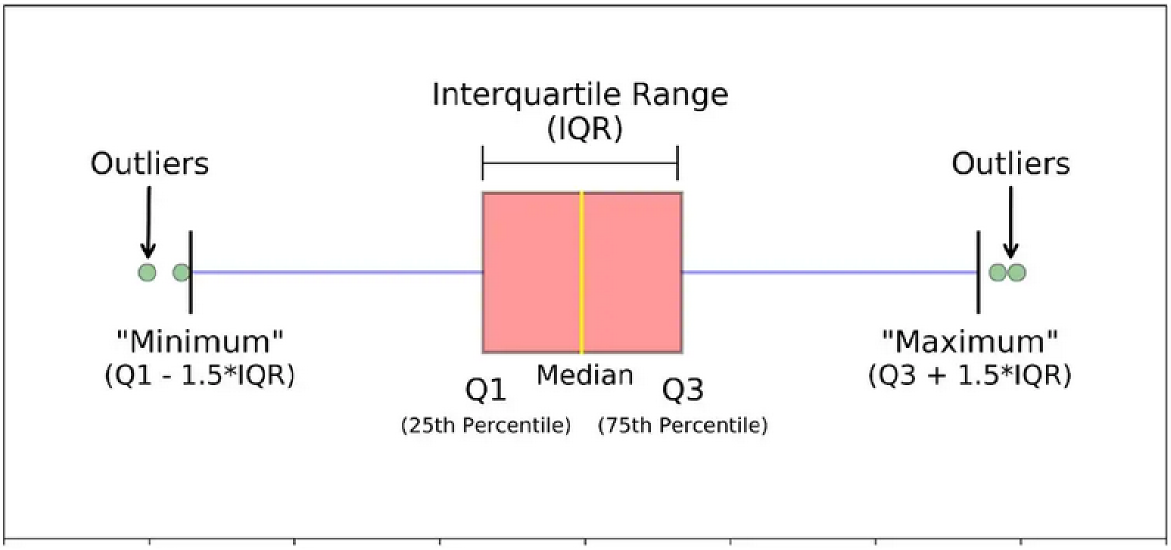

Figure 13: Interpretation keys for the box plots in the next figures

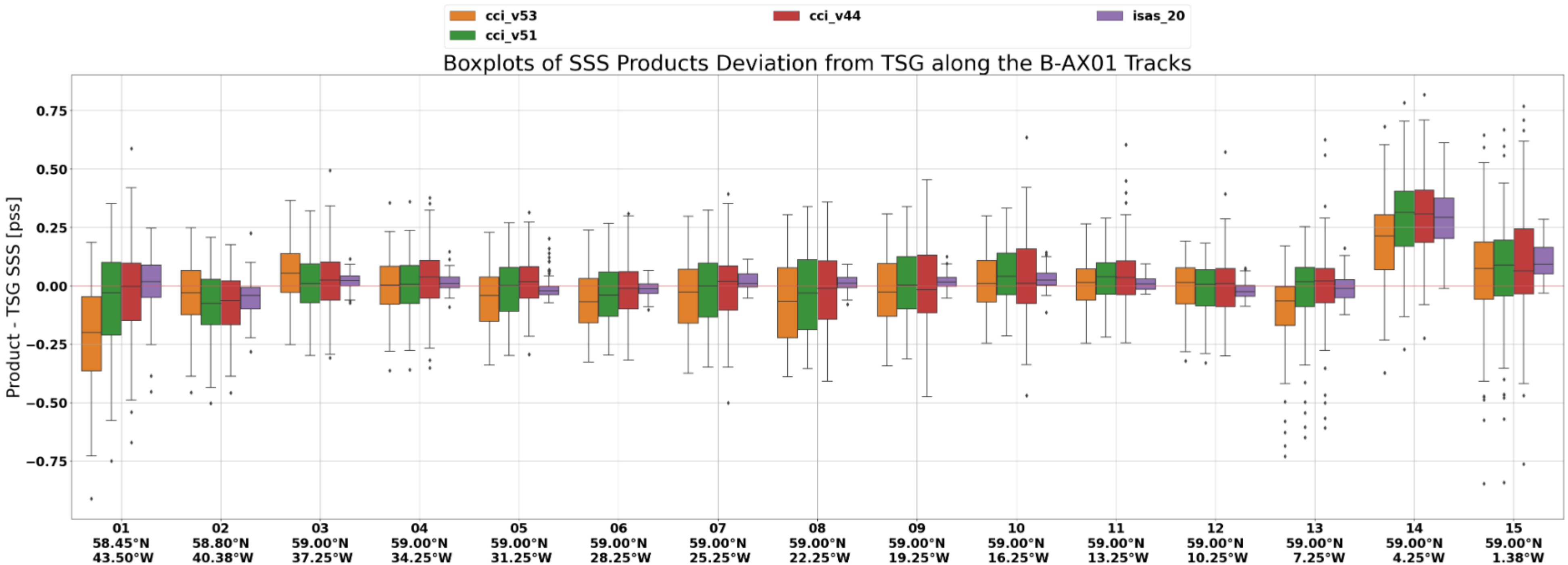

*Figure 14: Results of box plots for each bin along the North Atlantic B-AX01 TSG tracks, numbered from West to East (Figure 11). The plotted datasets for each bin result are arranged from left to right, namely: CCI V5.3, V5.1, V4.4, and ISAS20. The box for each product, as shown in Figure 13, represents the median and interquartile range (IQR) for the entire 2010-2022 period, and includes a standard outlier threshold (bars) set to 1.5*IQR.*

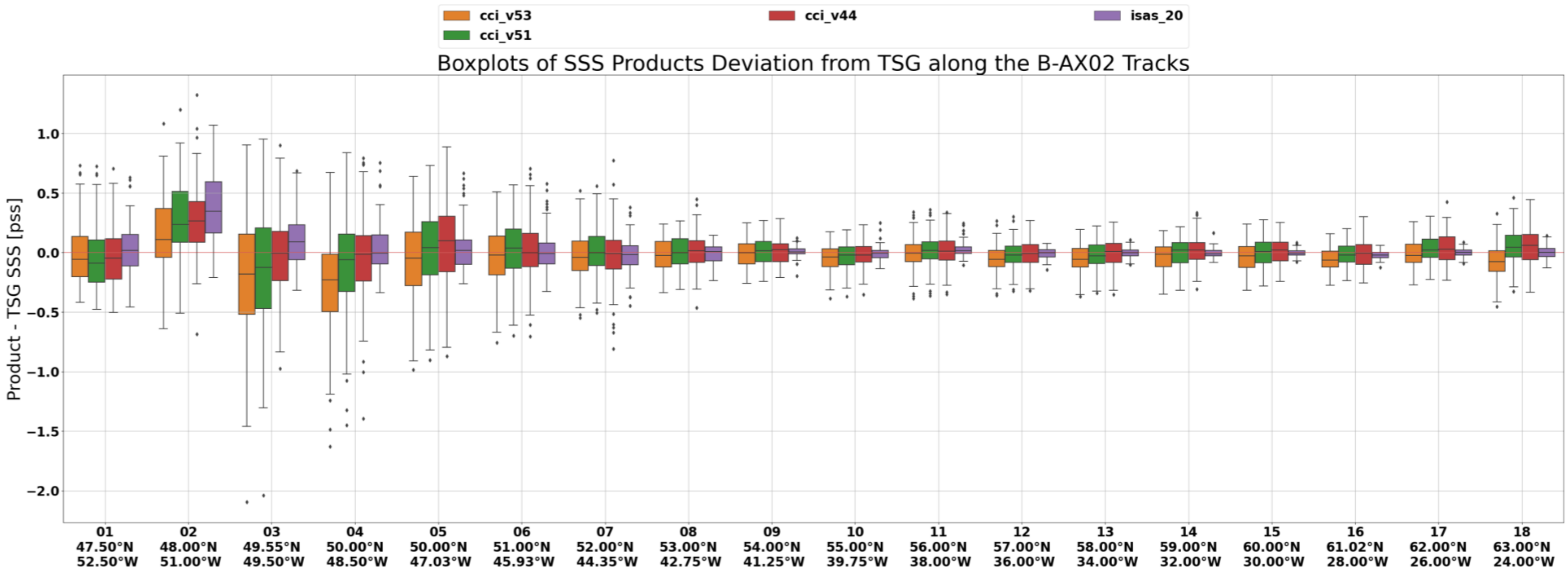

Figure 15: Same as Figure 14, but for B-AX02, and TSG tracks are numbered from South to North.

2) Observations from time series analysis

Although the previous boxplot statistics do not consistently show improvement of V5.x over V4.4, significant improvements are evident in individual time series, as shown in Figure 16 and Figure 17. Corresponding standard deviation of difference and correlation with the TSG data are provided in Figure 18Figure 21. Along B-AX01 in boxes 03, 04, and 09, a systematic reduction in differences with TSG SSS is observed in 2015, as reported in Section 1. This improvement is linked to updates in SMAP data processing (Figure 16). Additionally, the early period from 2010 to mid-2011 shows notable improvement in V5.x compared to V4.4, particularly along B-AX01 in boxes 04 and 09, and B-AX02 in boxes 05 and 14 (Figure 17), with much of V4.4's volatility corrected during this time.

Conversely, significant degradations are also observed in V5.x, such as mid-2011 in B-AX01 box 09, and B-AX02 box 14, as well as from 2010 to late 2011 in B-AX02 box 04.

The ISAS-20 series, based solely on Argo data, often agrees well with TSG series despite its independent data source. Since the TSG dataset has not been updated beyond December 2018, ISAS-20 serves as a reference for verification from 2019 onward. Notably, strong agreement between V5.x and ISAS-20 is seen in southern boxes of B-AX02 (e.g., boxes 04 and 05) where SSS seasonal variations are pronounced. However, V5.x does not consistently outperform V4.4 in these areas.

3) Statistical confidence

Quantitatively, the standard deviation of differences with TSG shows significant improvement of V5.x over V4.4 in B-AX01 boxes 10 and 11, where the confidence intervals do not overlap (Figure 18). These are the only instances of improvement supported by statistical confidence analysis. For other boxes, for B-AX02, and for correlation along both B-AX tracks, statistical significance is not achieved (Figure 18Figure 21). Conversely, the metrics for ISAS-20 consistently show superior performance, with confidence intervals that rarely overlap with those of the CCI products.

4) Seasonal and interannual variability analyses

The mean seasonal variation analysis of differences between CCI SSS versions and TSG data shows both improvements and degradations (Figure 24). V5.3 exhibits stronger degradations compared to V5.1, indicating less stability. Most improvements, though slight, occur around May–August, with some notable improvements (and degradations) in boxes near the track extremities. Overall, V5.3 appears less stable than V5.1.

The interannual variability analysis of differences between satellite-derived SSS and TSG data reveals some progress of V5.x over V4.4 (Figure 25). Both V5.x versions show clear improvements in interannual SSS during 2012–2013 and significant improvements in 2015 across most TSG track boxes. Some improvements are also observed at the start of the period (2010–2011), though interspersed with degradations. However, in B-AX01, recurring interannual degradations are evident after 2015.

In summary, the evaluation of CCI+SSS V5.x versions along high-latitude TSG tracks in the North Atlantic reveals mixed performance compared to V4.4. While V5.1 and V5.3 show reduced volatility and localized improvements, particularly in 2015 and early 2010–2011, degradations are also observed, with V5.3 exhibiting less stability overall. Statistically significant improvements in standard deviation of differences are limited to specific boxes, while ISAS-20 consistently outperforms all CCI products. Seasonal and interannual analyses further highlight both progress and shortcomings, emphasizing the need for continued refinement in V5.x versions to achieve more robust and consistent performance.

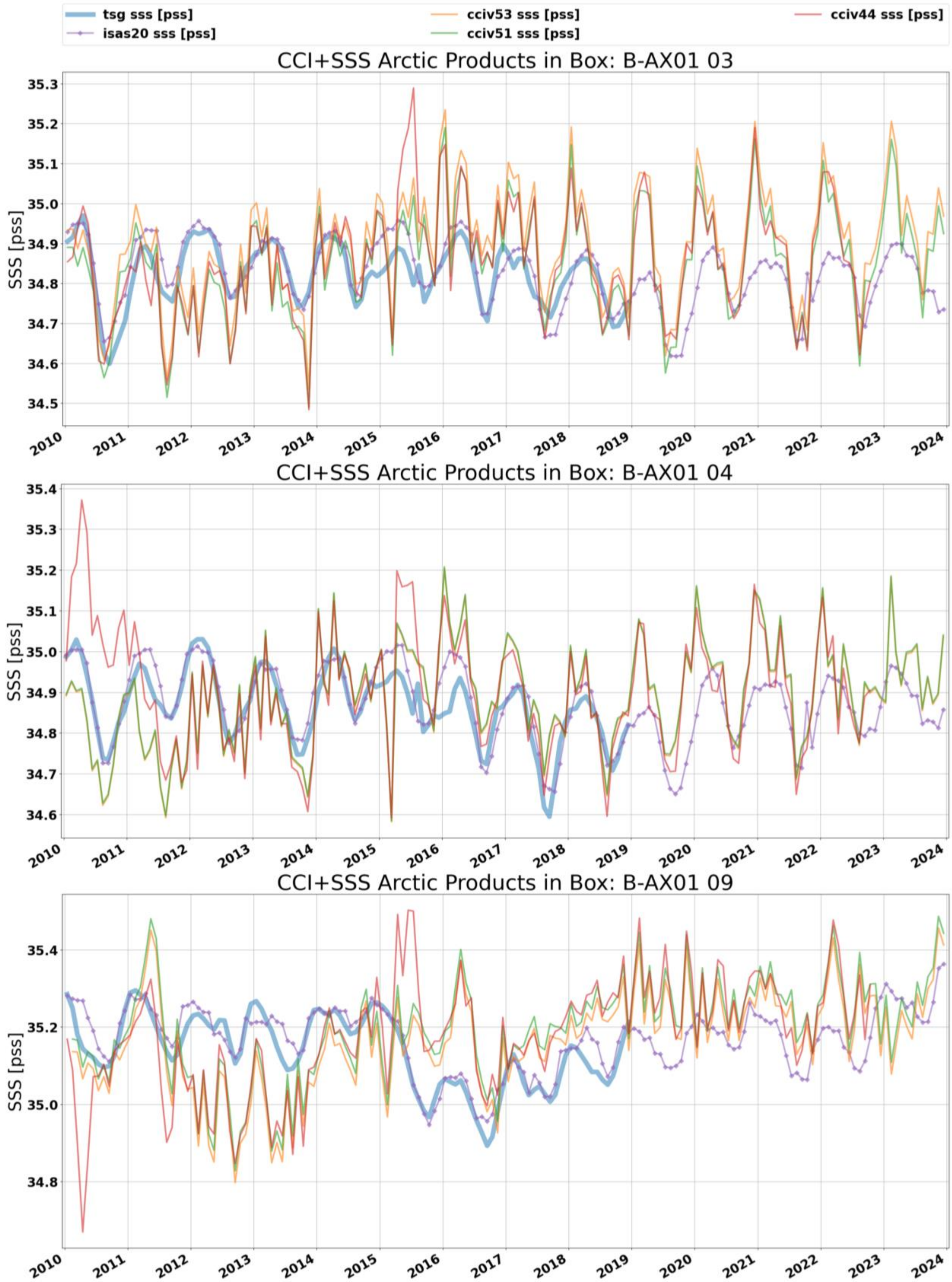

Figure 16: Time series of TSG data and products within different North Atlantic B-AX01 boxes.

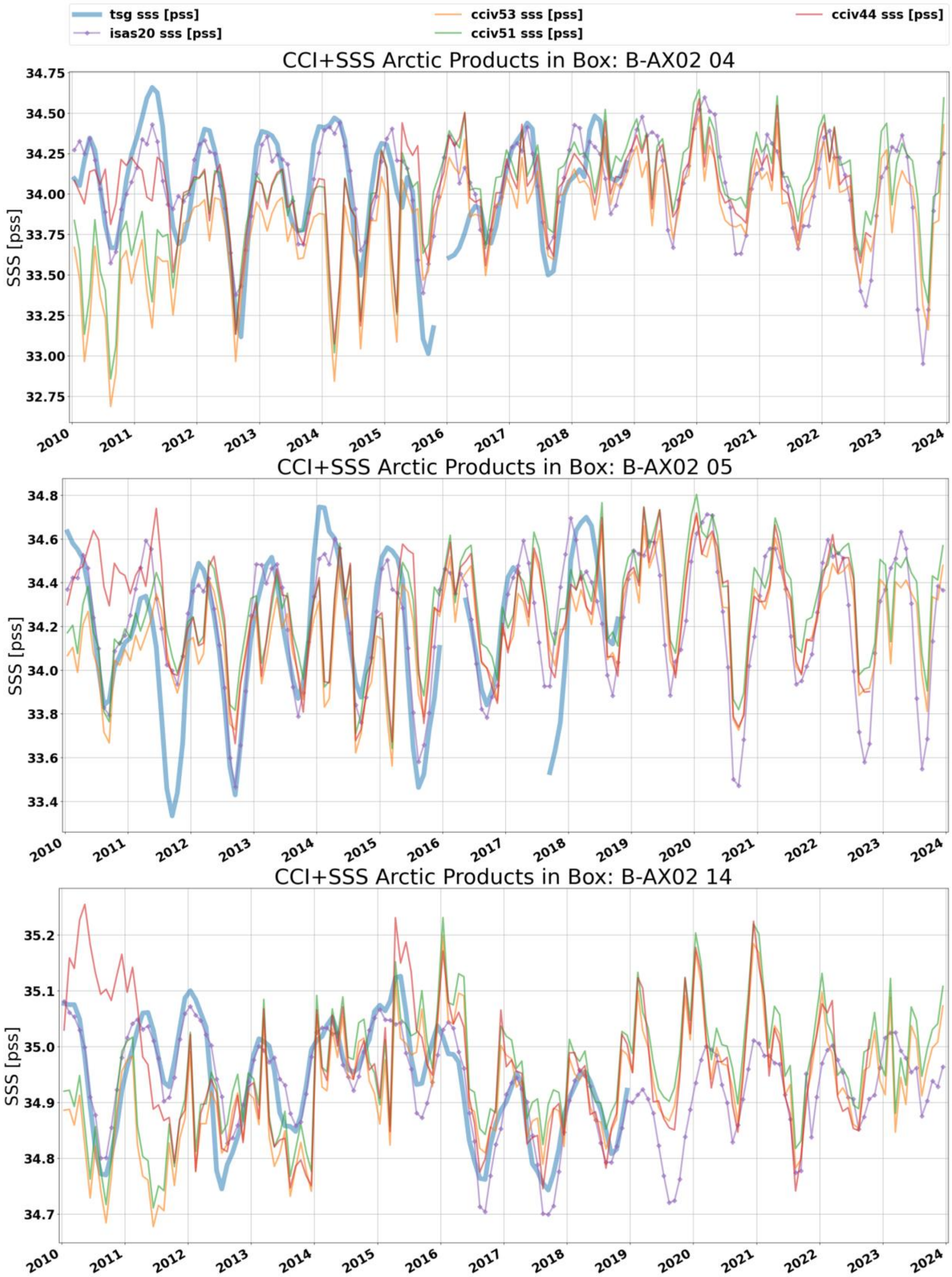

Figure 17: Time series of TSG data and products within different North Atlantic B-AX02 boxes

Figure 18: (Colour) Std of difference (stddiff) between each product and TSG SSS within each box for B-AX01. Number couples indicate the 95% confidence interval of stddiff using a bootstrap procedure.

Figure 19: (Colour) Std of difference (stddiff) between each product and TSG SSS within each box for B-AX02. Number couples indicate the 95% confidence interval of stddiff using a bootstrap procedure.

Figure 20: (Colour) 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.

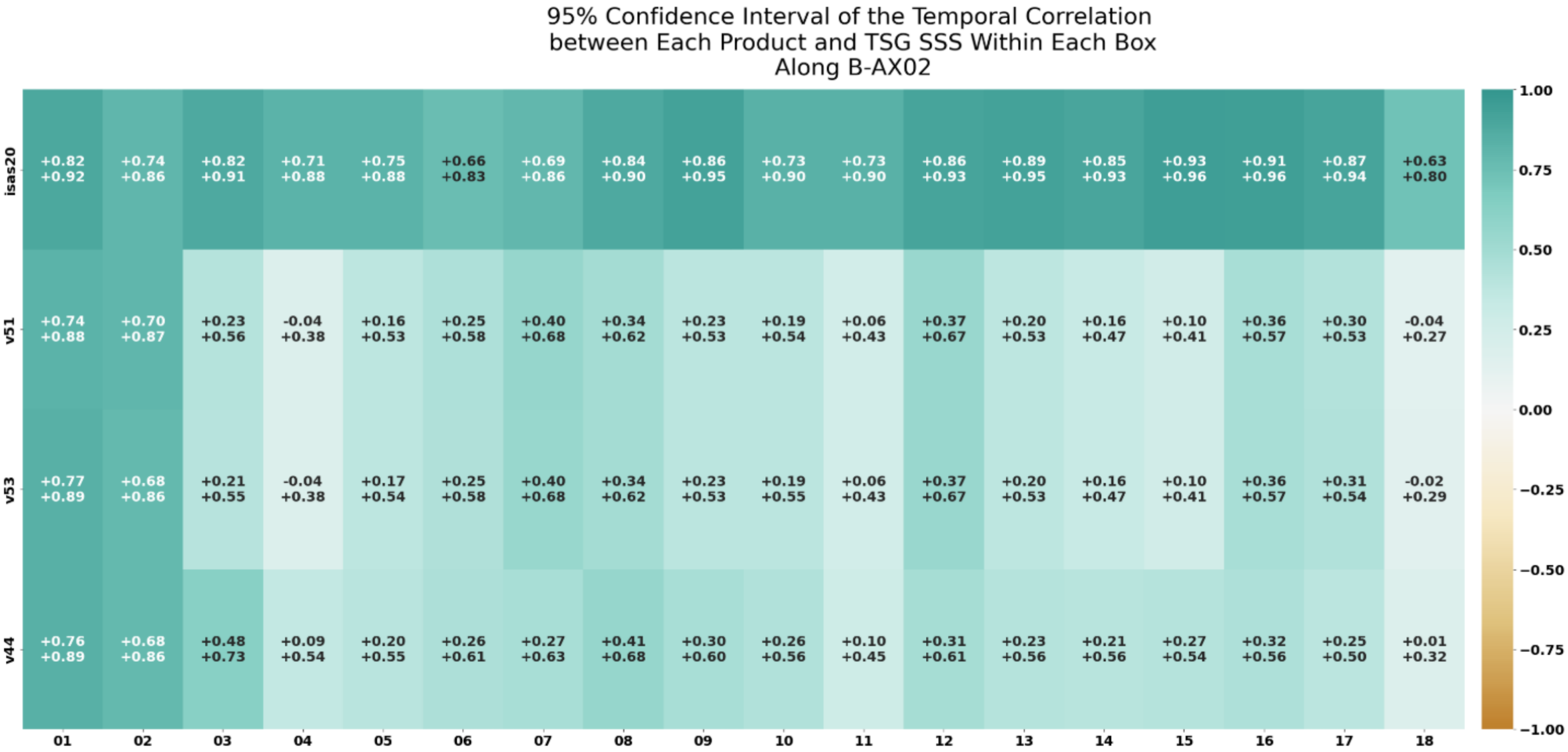

Figure 21: (Colour) Correlation between each product and TSG SSS within each box for B-AX02. Number couples indicate the 95% confidence interval of the correlation using a bootstrap procedure.

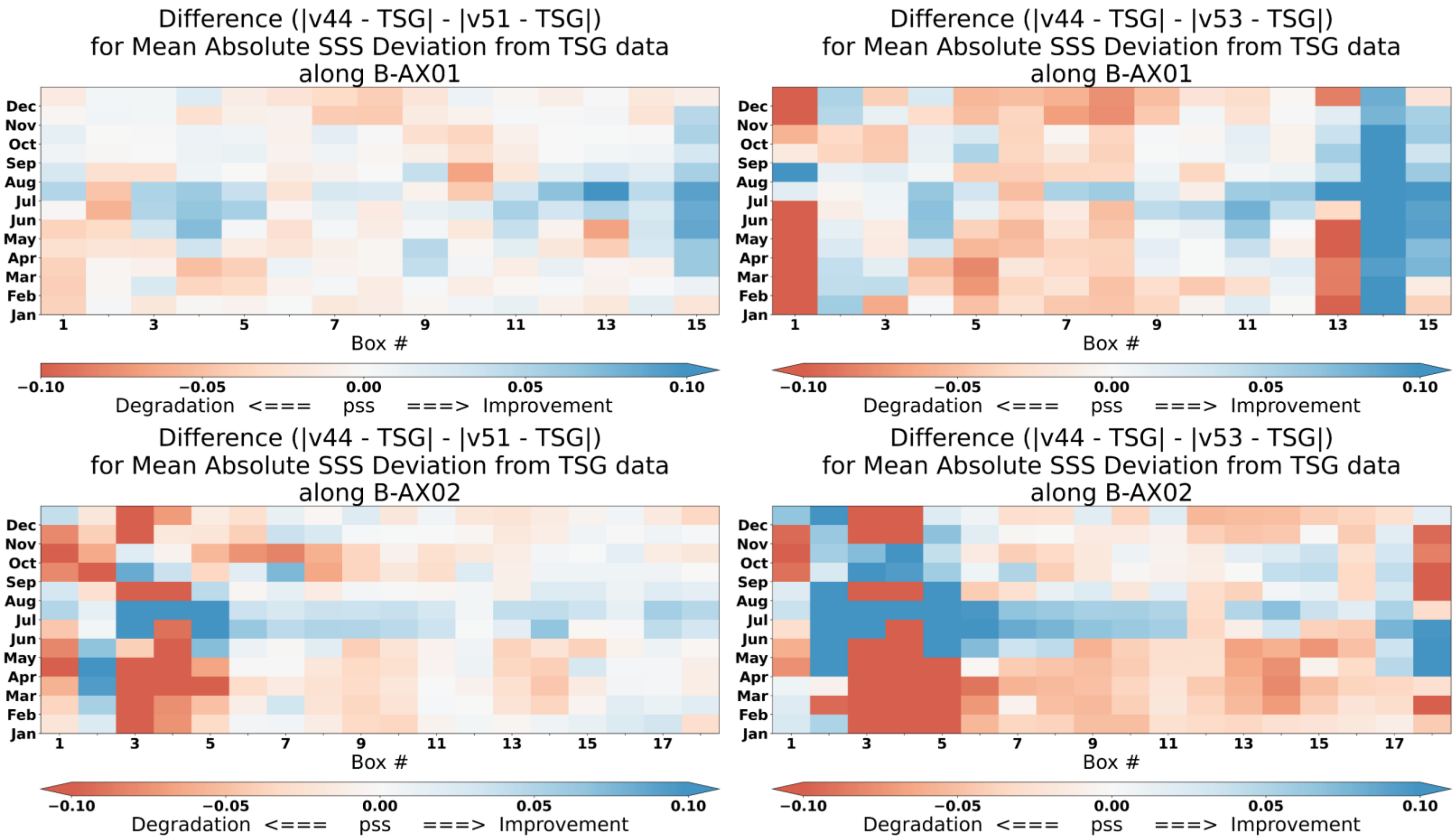

Figure 22: Hovmöller diagrams of the difference in mean absolute difference (MAD) between V4.4 and TSG, and V5.1 or V5.3 and TSG, for mean seasonal SSS in each box. Top panels show B-AX01, and bottom panels show B-AX02. Left panels evaluate V5.1 performance, while right panels evaluate V5.3. Positive values (blue) indicate improvement over V4.4, while negative values (red) indicate degradation.

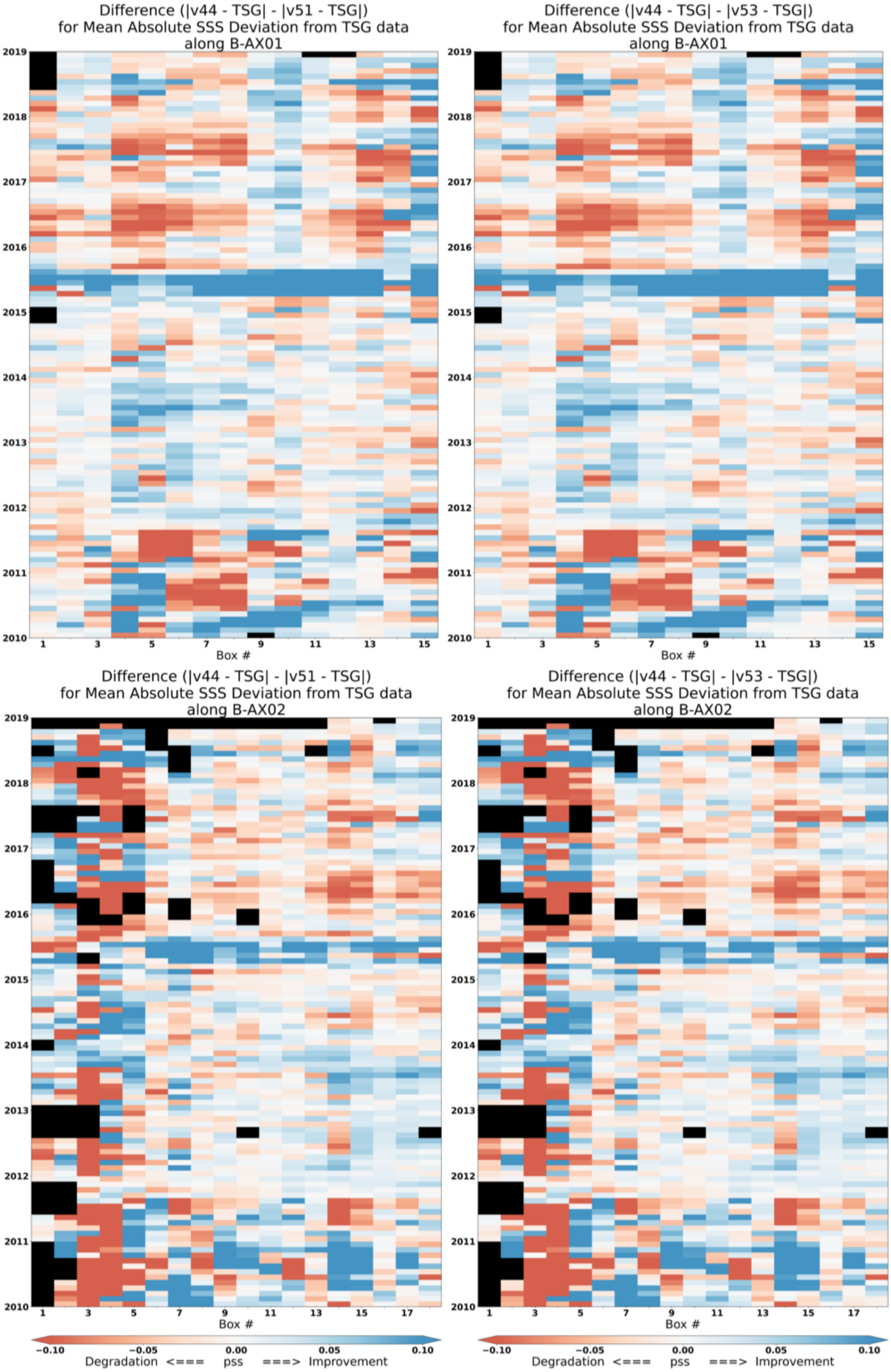

Figure 23: Hovmöller diagrams, as in Figure 22, but for interannual anomalies of SSS.

7.2.1 Southern Ocean comparisons

As noted in Section 1, no V5.x version significantly outperformed V4.4, particularly in the Antarctic region. This section provides a detailed comparison of V4.4 and V5.1 (equivalent to V5.5, the final standout version) using Antarctic TSG data as the benchmark. Across all TSG-collocated data (Section 5.1.3) in the Southern Ocean, the global comparison between V5.1 and V4.4 reveals no significant improvement. Joint distribution analysis shows some differences but no noticeable reduction in spread for V5.1 (Figure 24). By contrast, the ISAS-20 dataset demonstrates a much narrower spread, indicating superior accuracy.

1) Comparison metrics across Southern Ocean tracks

When examining metrics across the entire domain, V5.1 shows no consistent improvement over V4.4, with a slight degradation in the standard deviation of differences (Figure 25). While the collocated sampling ensures consistency across products, potential pre-collocation sampling differences may influence the intercomparison. The mean value analysis across the domain (Figure 26) highlights slight improvements for V5.1 during specific periods, including 2011 to mid-2012, late 2015 to mid-2016, and spring 2018. However, degradations are evident in 2010 and 2022.

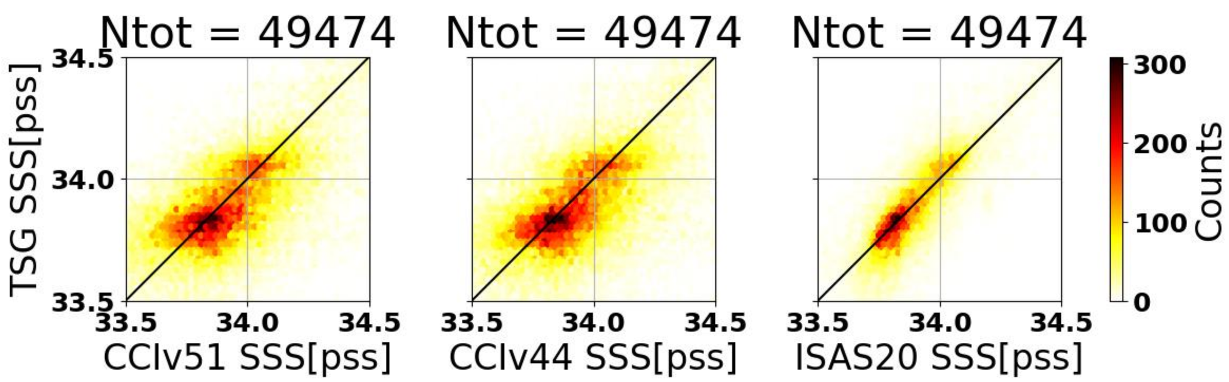

Figure 24: Joint distribution of each product with the TSG data within the entire domain around Antarctica.

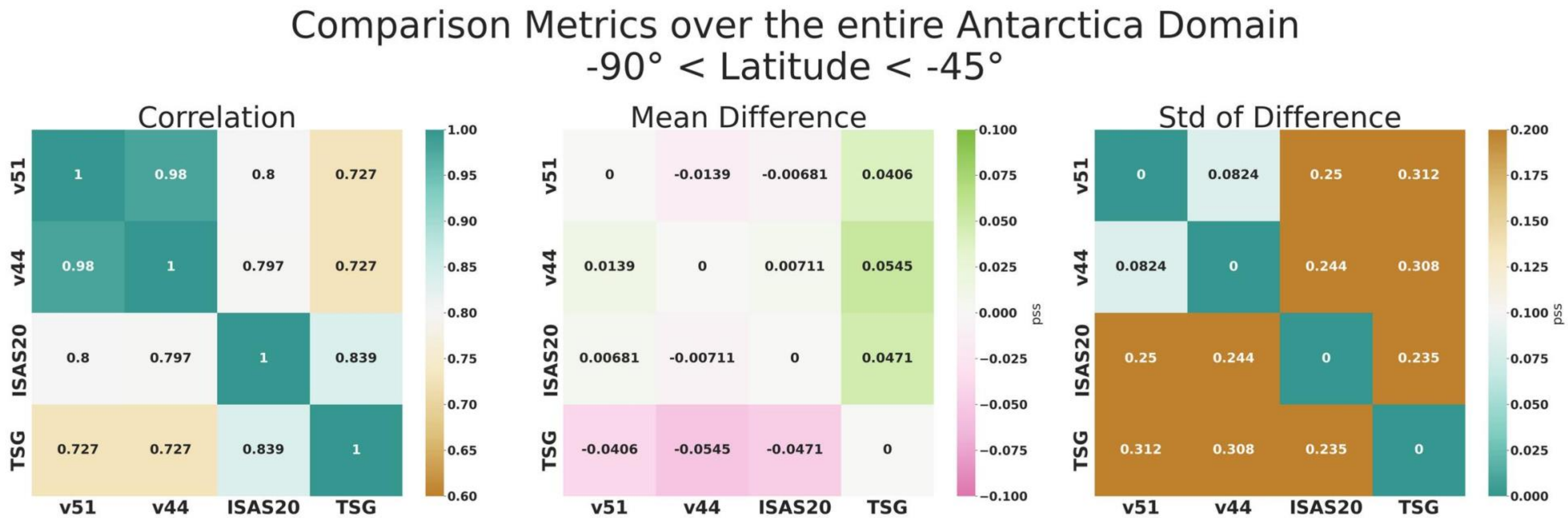

Figure 25: (Colour and Numbers) Correlation, mean absolute difference and robust std of difference between each product and the TSG data over the entire domain.

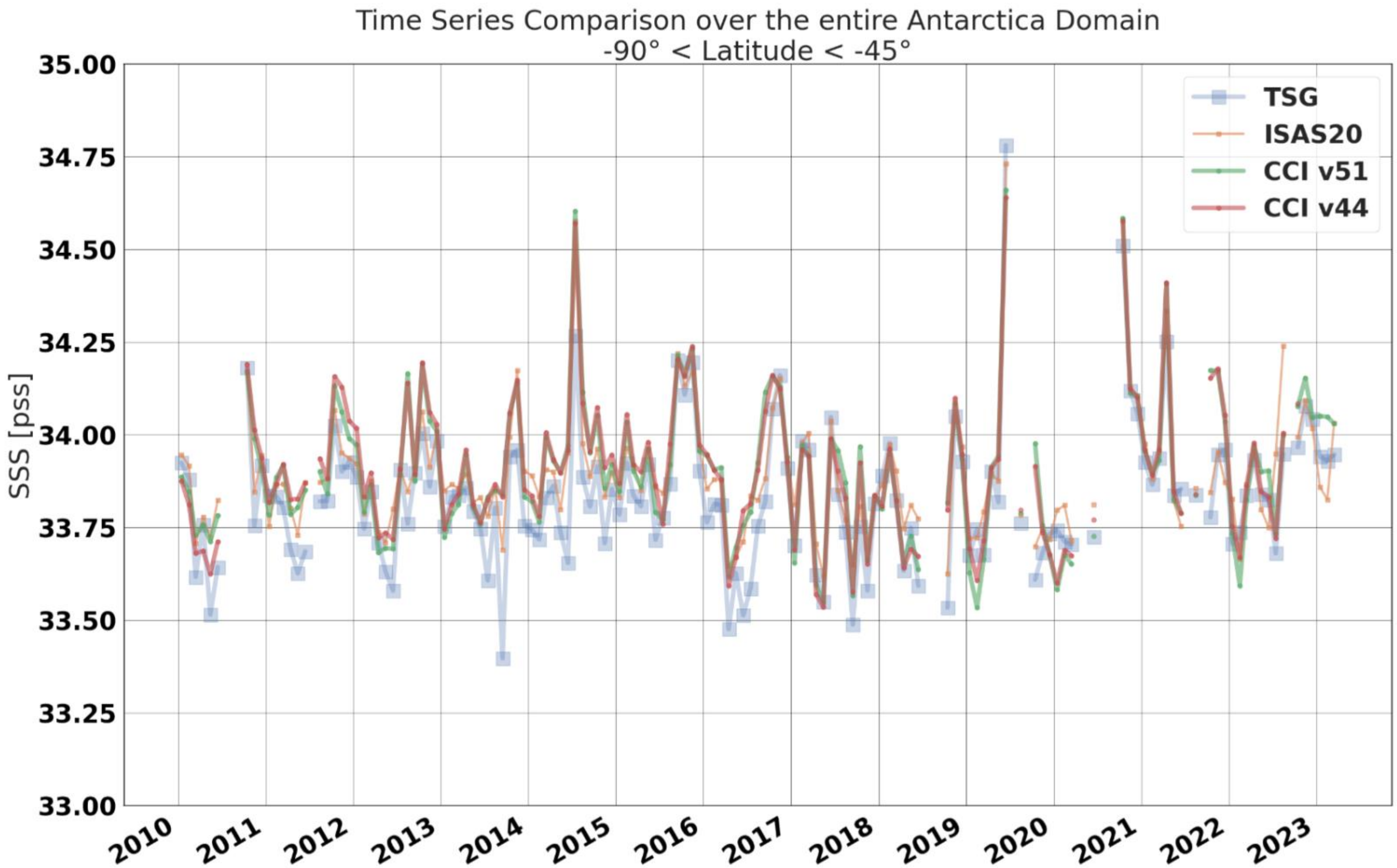

Figure 26: Time series of the averaged products and TSG data over the entire Antarctica domain where there are colocations.

2) Performance along specific TSG tracks

On the recurring TSG tracks between Tasmania and Antarctica, V5.1 shows slight improvement over V4.4, with reduced std of differences and improved correlation, particularly in the near-Antarctica box 7 (Figure 27). Along the South Africa-Antarctica tracks, the results are less favourable. Boxes 6 and 7 near Antarctica exhibit minor improvements in std of differences and correlation, but boxes farther from the Antarctic coast generally show degraded metrics compared to V4.4 (Figure 28).

For the newly added TSG tracks between the Antarctic Peninsula and South America, improvements are sparse (Figure 29). Notable positive results appear in box 2 near Tierra del Fuego, where both std of differences and correlation are better for V5.1. However, degradation is observed in boxes closer to the Antarctic coast, such as box 7.

In conclusion, while V5.1 shows localized improvements over V4.4 in specific areas and periods, particularly near Antarctica along the Tasmania track and during certain years, its overall performance in the Southern Ocean remains inconsistent. Degradations in several key metrics and limited advancements highlight the need for further refinement to achieve more robust results in this region.

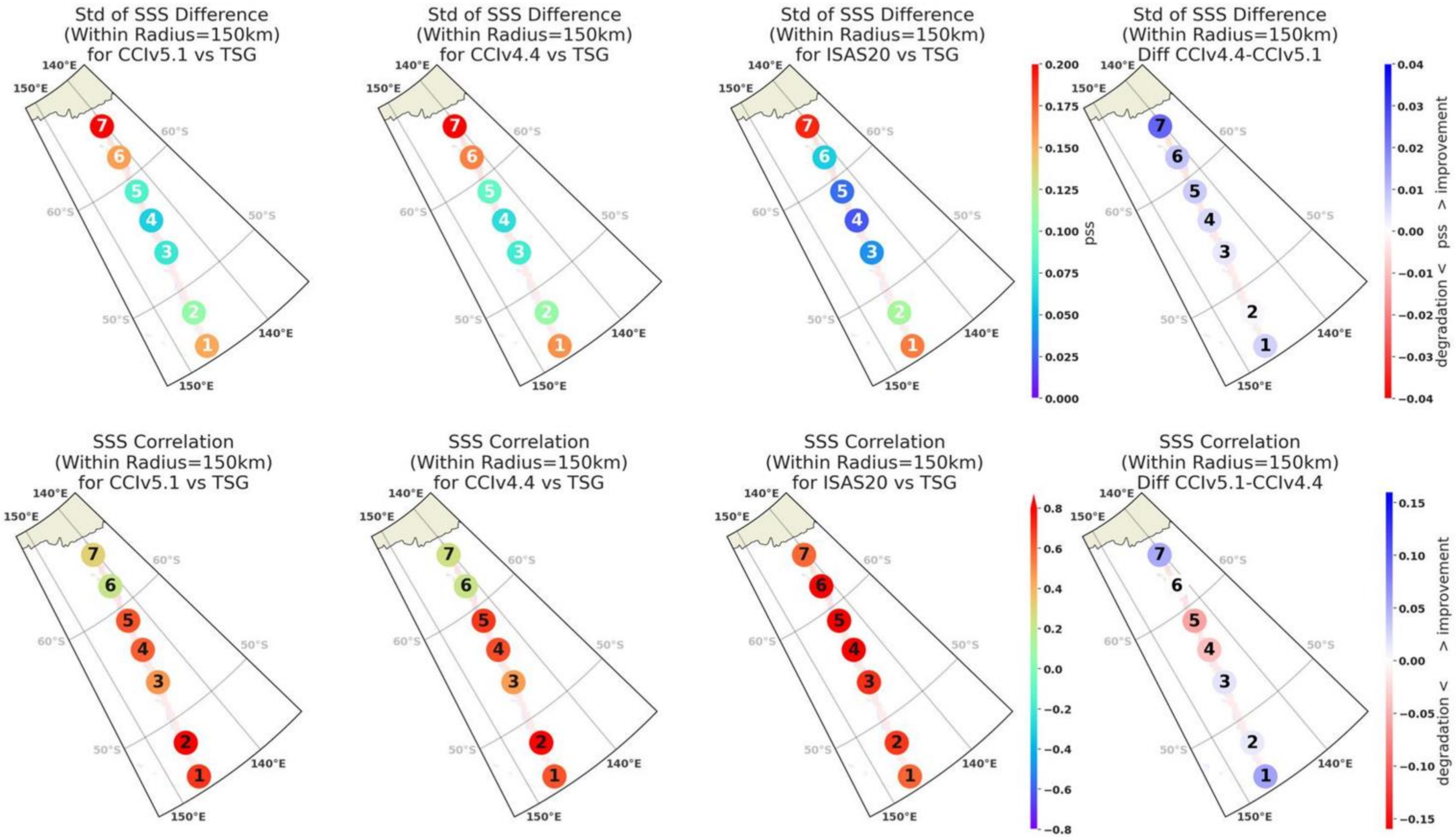

Figure 27: Collocation data number and comparison metrics between each product and the TSG data along ship tracks from Tasmania to Antarctica.

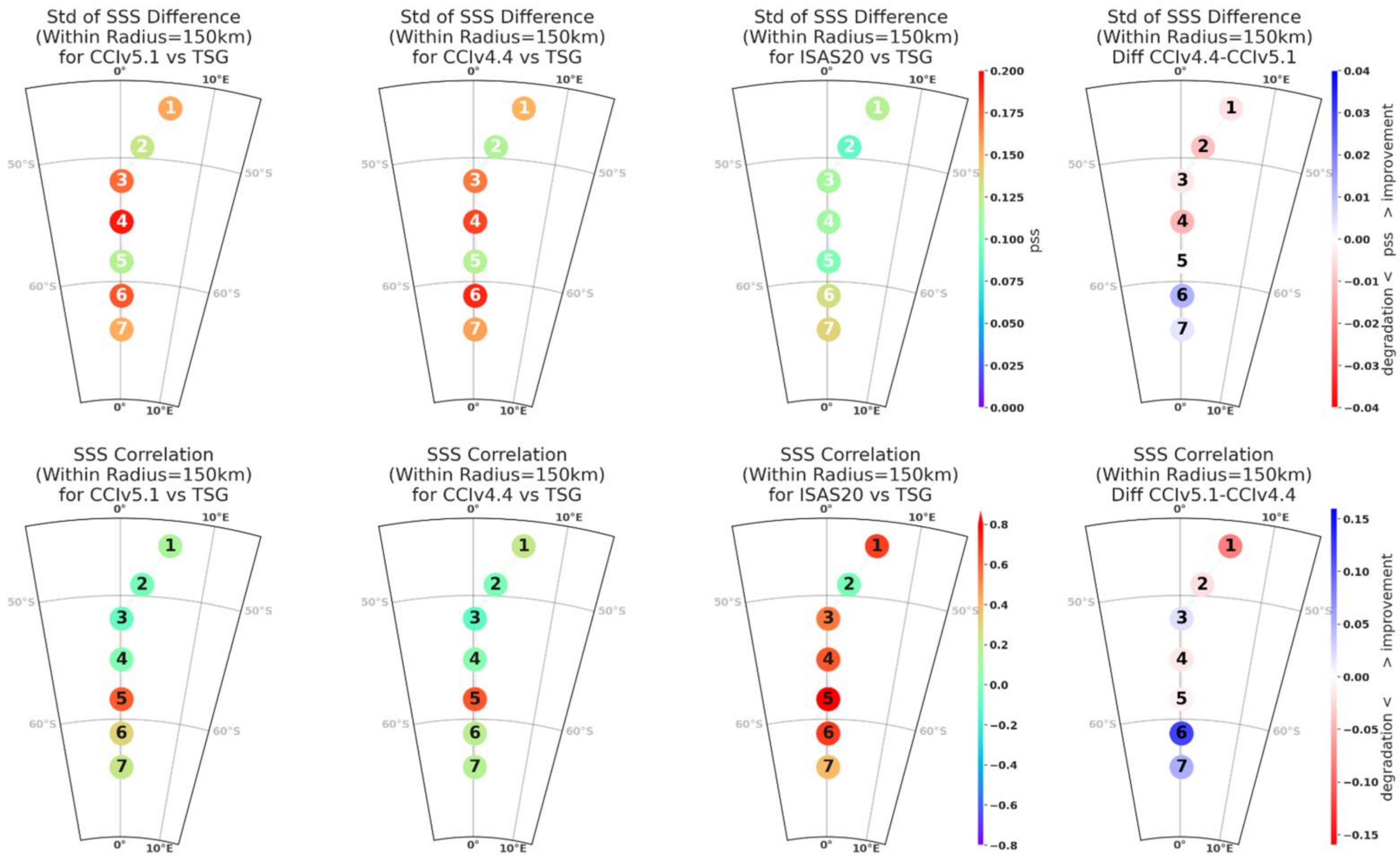

Figure 28: Collocation data number and comparison metrics between each product and the TSG data along ship tracks from South Africa to Antarctica.

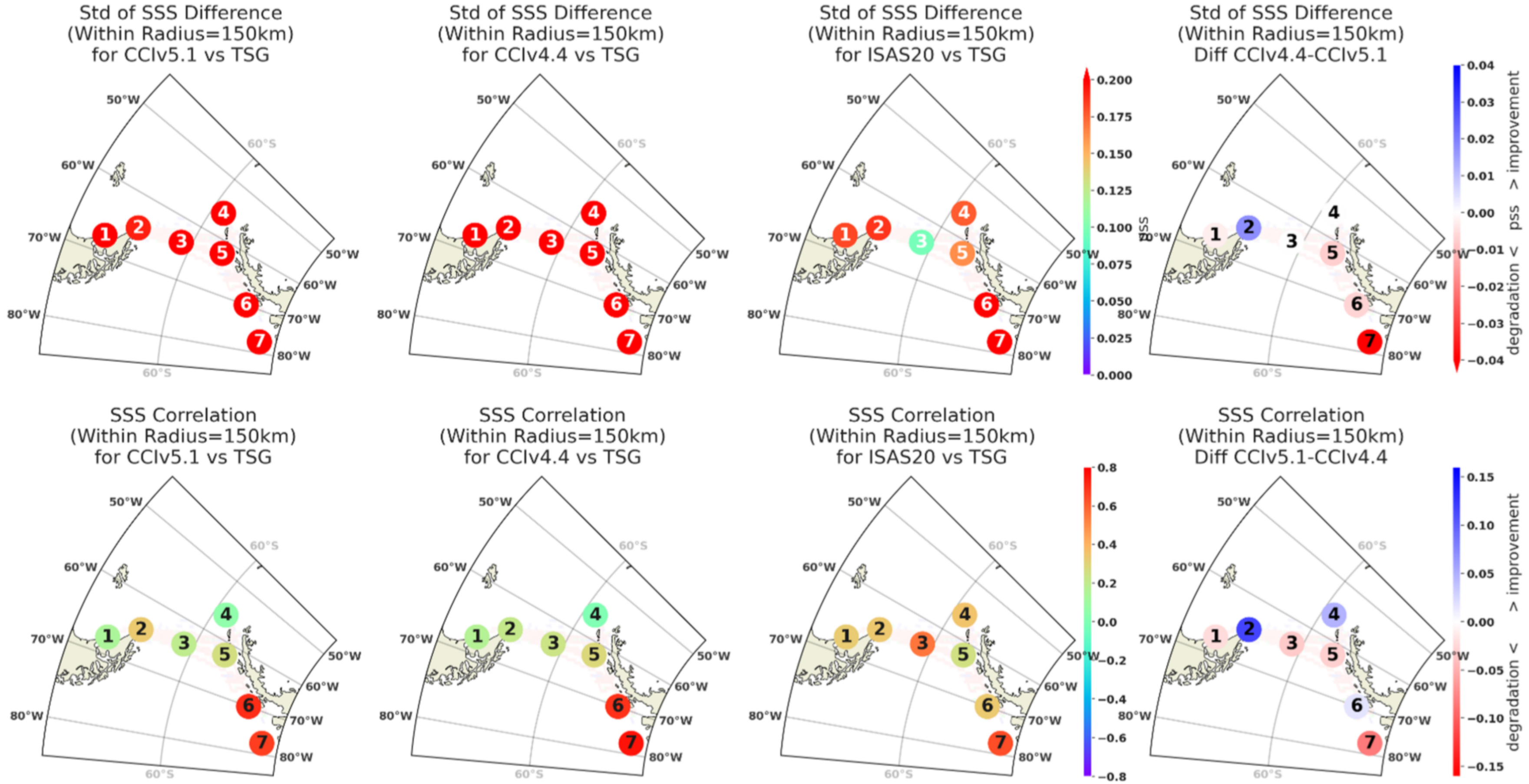

Figure 29: Collocation data number and comparison metrics between each product and the TSG data along ship tracks from Antarctic Peninsula to South America's Tierra Del Fuego.

7.3 Products evaluation summary

In conclusion, the V5.x versions of the CCI+SSS product exhibit incremental improvements over V4.4, particularly in mid- to high-latitude regions of the Northern Hemisphere and during specific periods. Notably, V5.5 demonstrates the most consistent enhancements among the V5.x versions, delivering improvements in areas such as reduced volatility and interannual variability during key periods like 2010–2011 and 2015. However, these gains are often localized, and degradations persist in certain metrics and regions, with V5.3 showing less stability overall.

In the Southern Ocean, V5.1 achieves occasional localized improvements, such as along the Tasmania-Antarctica track and during years like 2011–2012 and 2015–2016. Yet, these advancements are frequently overshadowed by inconsistencies and degradations, particularly in standard deviation and correlation metrics for tracks farther from the Antarctic coast. Meanwhile, ISAS-20 consistently outperforms all CCI+SSS versions in key metrics across regions, emphasizing the need for further refinement.

Overall, while the V5.x versions represent progress in addressing systematic biases and sensor corrections through targeted advancements like enhanced climatological adjustments, their variability and limited improvements underscore the necessity for continued development.

7.4 Open issues and discussion

A key uncertainty in product evaluations arises from differences in spatial coverage, particularly near coastlines and ice edges, after applying quality, land-sea, and ice-sea flags. For this PVASR, only collocated data at locations and times common across all products were used. However, this constraint may not fully ensure rigorous intercomparisons. A more robust approach would involve determining and applying a common mask to all CCI products before any evaluation processing, including collocation with in-situ data. To enhance future PVASRs, it is recommended to develop and apply such a mask dynamically during ongoing evaluations and to use a finalized common mask for the ultimate version and all previous versions comparison and PVASR documentation.

8 Conclusion and future work

The evaluation of CCI+SSS V5.x versions against in-situ data highlights incremental improvements over V4.4, particularly in the Northern Hemisphere's high latitudes and during specific periods. Version 5.5 emerges as the most consistent performer, offering localized improvements in bias, variability, and correlation metrics, especially in Arctic and select Southern Ocean tracks. However, performance across other regions and timeframes remains variable, with notable degradations in some metrics.

The results emphasize the need for further refinements to address systematic biases, enhance stability, and achieve greater global consistency. Future work should focus on:

- Developing and applying a common spatial mask to ensure rigorous intercomparisons across all versions during evaluations.
- Enhancing calibration techniques to improve performance in challenging regions, such as coastal areas and near ice edges.
- Expanding the use of diverse and updated in-situ datasets to improve validation coverage.
- Refining algorithms to reduce variability and ensure more uniform improvements across all global regions.

.

End of document