Recovery of SMOS Salinity Variability in RFI-Contaminated Regions

Fabrice Bonjean, Jacqueline Boutin, Jean-Luc Vergely, Philippe Richaume, Roberto Sabia

> TGRS-2024-00401 <

Recovery of SMOS Salinity Variability in RFI-Contaminated Regions

Fabrice Bonjean, Jacqueline Boutin, Jean-Luc Vergely, Philippe Richaume and Roberto Sabia

***Abstract*— The Soil Moisture and Ocean Salinity (SMOS) satellite mission, operational since 2010, relies on an L-Band microwave interferometric radiometer to generate brightness temperature images along the swath, with global coverage every 3 days. These images are then used to derive sea surface salinity (SSS) with an effective resolution of less than 50 km. However, signal acquisition in some ocean regions is intermittently and significantly disrupted by radio-frequency interferences (RFI) from various terrestrial military or civilian sources worldwide. We develop a new methodology based on principal component and regression analyses to extract the RFI signatures in time and space, thereby enabling the construction of a corrected SSS estimate along the swath. This method successfully filters out many disruptive features characterized by long and wide branches occurring around the RFI sources, hence recovering SSS variability as demonstrated in comparison to in situ reference data. This correction methodology is an alternative to separate filtering procedures that were applied on brightness temperature at Level 1. Independent information indicating the probability of RFI occurrence on land areas or nearby is used to verify the timing of oceanic RFI contamination inferred by the correction process. The methodology performs particularly well in areas where the probability is close to 1 for a significant and contiguous portion of the entire period. Already applied with significant improvement in three selected regions, this correction method is a starting point for expanding and systematizing the methodology to treat as many RFI-polluted regions as possible and to recover SMOS SSS variability.**

***Index Terms*— Sea surface salinity, radio-frequency interference, L-band microwave radiometry, principal component analysis, regression.**

Fabrice Bonjean and Jacqueline Boutin are with the Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques–Institut Pierre Simon Laplace (LOCEAN/IPSL), CNRS, IRD, MNHN, Sorbonne Université, F-75005 Paris, France (e-mail: fabrice.bonjean@locean.ipsl.fr, jb@locean.ipsl.fr).

Jean-Luc Vergely is with ACRI-ST, F-78280 Guyancourt, France (e-mail: jean-luc.vergely@acri-st.fr).

Philippe Richaume is with the Centre d'Etudes Spatiales de la BIOsphère (CESBIO) (CNES/CNRS/UPS/IRD), 31401 Toulouse Cedex 9, France (e-mail: philippe.richaume@cesbio.cnes.fr).

Roberto Sabia is with ESA, ESRIN, 00044 Frascati, Italy (e-mail: roberto.sabia@esa.int).

I. INTRODUCTION

OCEAN salinity plays a crucial role in global climate dynamics, influencing ocean circulation [1], serving as a fundamental marker for changes in the Earth's water cycle [2], and affecting the life cycles of marine organisms [3]. The Soil Moisture and Ocean Salinity (SMOS) mission, launched by the European Space Agency (ESA), has been instrumental in providing global observations of sea surface salinity (SSS). This satellite mission uses an L-Band (1.41 GHz) interferometric radiometer to infer brightness temperatures (BT), which are then used to retrieve SSS [4].

However, the accuracy of these measurements can be significantly affected by Radio Frequency Interferences (RFI, [5]). Although the L-Band is a protected band, RFI contamination often arises from illegal transmissions or malfunctioning equipment, including human-made sources such as radars, TV and radio stations, and satellite communication systems. These signals can contaminate the SMOS retrieved SSS (e.g. Figure 2b and Figure 2e of [6], Figure 3 of [7], and Figure 1 of this article). SMOS uses an interferometric array fitted with 69 equally spaced receivers [8]. SMOS BTs are reconstructed from the cross-correlations of signals collected by each pair of receivers, known as the visibility function [9]. However, strong discontinuities in the observed BT scene, such as those induced by RFIs, lead to oscillations in the reconstructed BT around the RFI source, attributable to the Gibbs phenomenon (chapters 3 and 5 of [10]). These oscillations result in both positive and negative biases in the retrieved SSS. This underscores the unique challenge of addressing RFI in SMOS data, setting it apart from missions with real aperture radiometers and emphasizing the need for specialized RFI mitigation strategies to ensure accurate SSS retrieval.

In the past, significant efforts have been made to identify and filter RFI in SMOS data. The ensemble of methods for RFI detection in the SMOS mission operates across various signal processing domains [5]. At the Level 1 (L1) processing stage, RFI detection methods use a list of active RFI sources and flag the pixels around the sources according to the instrument impulse response characteristics. They also flag entire snapshots that contain significant RFI contamination. In the visibility domain (L1a), RFI treatment involves a detection and flagging algorithm that uses a user-defined threshold to identify and nullify RFI-contaminated visibilities, and applies a median filter to the correlated noise injection data to remove outliers caused by RFI [11], [12]. In the spatial domain (L1c), an RFI flagging mechanism is used to compute the RFI

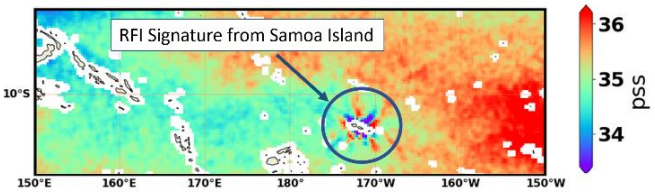

Figure 1: SMOS SSS from CCI SSS Level 3 Version 3 on August 15th, 2018, in the southwestern tropical Pacific.

contamination level for each pixel based on the difference between the measured and the estimated RFI response. The RFI contamination level flags are assigned to each pixel according to four levels, and are stored in the L1c products [13], [14].

At the Level 2 (L2) processing stage, both soil moisture and SSS RFI detection methods use screening and statistical behavior to detect the presence of RFI [15]. For SSS, they flag snapshots that contain pixels whose BT level exceeds the expected value by a certain threshold, or that have excessive spatial standard deviation (std). They also flag outliers based on the residuals of the model fit.

Global RFI probability maps are generated by applying detection methodology and can be used to provide information to the national authorities for locating and eliminating RFI sources (Figure 8 of [16]). Also, a database is maintained where the output of the RFI monitoring tools is stored (ERMIT, [17]). This includes information on RFI sources, SMOS passes, maps, statistics, and reports.

In this study, we focus on developing a new correction procedure for application in a post-processing stage, following the generation of L2 SSS data. While current RFI filtering effectively removes contaminated measurements, it also inadvertently eliminates SSS variability in the filtered areas. Thus, we face a choice between a) implementing stringent filtering that excludes SSS measurements across many areas, b) opting for moderate filtering, or c) creating a methodology to correct certain RFI contaminations. Presently, moderate filtering is employed in the SMOS L2 SSS processing. Despite the significant enhancements in L2 SMOS SSS through RFI detection and filtering noted in prior studies, RFI signatures persist in some SMOS SSS regions. Contamination by permanent RFI with constant intensities throughout the SMOS period can be managed by systematic corrections at the Level 3 (L3) or Level 4 (L4) stage (e.g. [18]), yet the challenge lies with intermittently occurring RFI signals. Our goal is to develop methods that can detect and correct variable RFI contamination in the processed L2 SMOS SSS data, acting as an effective post-processing step.

Section II of the paper presents the different datasets used for the creation and assessment of the corrections. The approach and methodologies are comprehensively explained in Section III. Section IV showcases the outcomes of implementing the correction methods. The paper concludes with Section V, which discusses the significance of the results and prospective future research, followed by Section VI, which provides the concluding remarks.

II. Data

A. SMOS SSS dataset

The SMOS mission, launched in November 2009, carries a passive microwave interferometric radiometer operating at L-Band. It operates in a sun-synchronous, polar circular, dawn-dusk orbit at a mean altitude of 764 km. It provides global observations of soil moisture and SSS. Due to the 18-day sub-cycle of SMOS, a specific ocean location is observed with the same SMOS measurement geometry approximately every 18 days.

We use the SMOS SSS dataset from the 7th reprocessing (RE07), conducted at the Centre Aval de Traitement des Données SMOS (CATDS). The prequalification product (L2P) is employed, which contains the retrieved SSS prior to any subsequent SSS bias correction that results in the publicly delivered 'L2Q' product. These products are arranged daily for both ascending and descending orbits separately. For a comprehensive description of the CATDS RE07 processing, please refer to [19].

B. ISAS SSS dataset

The In Situ Analysis System (ISAS) SSS dataset we utilize is generated using the ISAS v6 algorithm, an optimum interpolation (OI) tool designed for synthesizing global Argo data sets [20]. This tool creates monthly gridded salinity fields from in situ measurements, with spatio-temporal length scales that range from 300 km to four times the Rossby radius and from 3 weeks to 1 week. The extent of smoothing applied by the ISAS OI largely depends on the availability of in situ measurements, which, on average, is every 10 days and every 3° × 3° pixel [21].

C. TAO mooring data

Measurements from the long-term salinity records of the Global Tropical Moored Buoy Array (GTMBA) are used. Each buoy in this array continuously measures salinity near the sea surface (~1 m) in tropical regions, providing high temporal resolution data throughout the entire SMOS period. The GTMBA dataset includes contributions from the Tropical Atmosphere Ocean (TAO) project in the Pacific [22], the sole moored array referenced in this study. These buoys record salinity every 10 minutes, with the raw data subsequently post-processed into hourly increments, achieving an accuracy of 0.2 pss. For satellite data comparison, only the highest quality surface salinity measurements (at 1 m depth, QC flag = 1) were selected. This study employs the monthly mean dataset and focuses on time series analysis at a specific location within the southwestern Pacific (170°W, 8°S). Time series from other locations are also used, but associated results are provided only as supplementary material.

D. CESBIO probability dataset

The probability dataset, derived from SMOS data, provides global daily maps of land RFI detection occurrences (see Appendix A for a more detailed description). These maps are

separated into ascending and descending orbits to account for the SMOS's forward-tilted attitude effect. The RFI detection makes use of the L1c BT contaminated by RFI, as identified in Level 2 Soil Moisture (L2SM) UDP/DAP products [15]. The probabilities are computed as the ratio of the number of BT detected as contaminated by RFI to the total number of observed BT over a seven-day period. The RFI situation awareness reflected in these maps should be viewed as a best-case scenario, with the actual level of contamination likely to be worse. The maps also indicate the occurrence of detected and removed BT as the probability level rises.

We construct an aggregated dataset from these global daily maps, resampling them onto a grid with 1°×1° resolution every 7 days. At each time step, we select the maximum probability of occurrence within each grid box. These data are henceforth referred to as the RFI probability.

III. Methods

Our approach operates on the SMOS SSS dataset presented in II.A. It uses principal component analysis (PCA, [23]) and linear regression to identify RFI signatures in both time and space, thereby enabling the construction of a correction term. This correction term is subtracted from the original dataset to yield the corrected SMOS SSS dataset. Building on this general approach, we have developed two methods: the Regional Method (RM) and the Pointwise Method (PM). The processes for these methods are outlined in Figure 2 and thoroughly explained from III.A to III.I.

Due to SMOS image reconstruction, contamination from an RFI source extends over a large region. Consequently, SSS biases in contaminated pixels vary temporally in phase (or in opposite phase) with the RFI source. This property has been one of the bases for the development of RM; it is not used in PM. SSS biases depend on the latitude (ϕ) and longitude (λ) of the pixel: they generally decrease as the distance to the source increases. Furthermore, at a given location, SSS biases depend on the geometry of the SMOS measurements used for SSS retrieval. This geometry, referred to as x_{swath} in the following, is characterized by the pixel's location across the satellite swath and by the satellite orbit orientation (x_{orb}, which can be either ascending (A) or descending (D)). True SSS is independent of x_{swath}. Both RM and PM exploit this property to estimate SSS biases from systematic differences between SSS retrieved in one geometry and another.

RM targets region-wide RFI signatures from a RFI source Ω, by examining variance across the four dimensions: time (t), ϕ, λ, and x_{swath}. It uses the covariance between time series from different pixels within a limited 2D region, $\mathbb{C}$, surrounding Ω, and across various x_{swath}. This process is used to derive a characteristic time series, $U_1(t)$. The region $\mathbb{C}$ is enclosed within a larger region $\mathbb{R}$, which is targeted for comprehensive RFI treatment. The $U_1(t)$ series, which captures the temporal patterns of RFI within $\mathbb{C}$, is extrapolated to represent RFI's temporal patterns within the entire region $\mathbb{R}$. This extrapolation is utilized for the RFI mitigation throughout $\mathbb{R}$. By using $U_1(t)$ and a spatial pattern derived from regression analysis, we separate RFI signatures associated with the source Ω from the true SSS signal.

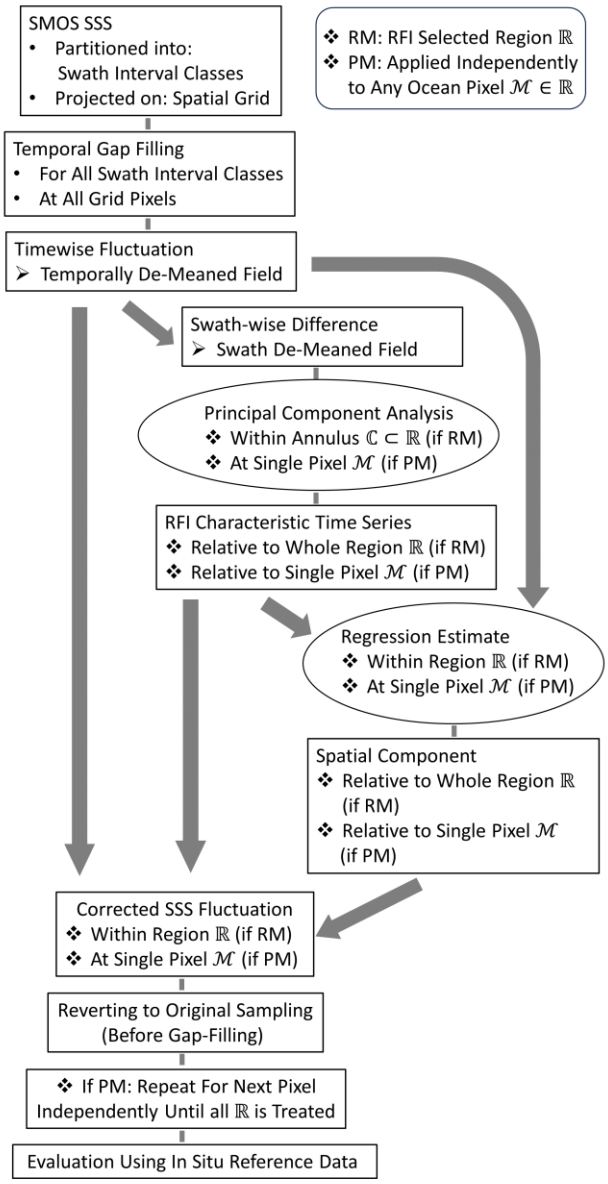

Figure 2: Schematic diagram of the RM and PM approaches.

Conversely, PM processes each pixel $\mathcal{M}$ independently of its neighbors, analyzing only variations in t and x_{swath}. This process isolates a characteristic time series specific to $\mathcal{M}$'s location, using it in a targeted regression to remove RFI signatures at $\mathcal{M}$. When PM successively applies this method across all pixels in the entire region $\mathbb{R}$, it offers an alternative to RM's 4-dimensional approach. PM emerged to address RM's limitations in handling multiple RFI sources and in defining its operational domain, as is discussed later.

In Section IV, we examine three regions as examples of both methods applied to areas contaminated by RFI (TABLE I). While it is not necessary to pinpoint the exact locations of RFI sources, we always focus our analysis on a region surrounding a recognized RFI source. This is important for RM's implementation. These sources can be inferred from SMOS SSS maps (as mentioned in Section I) or from RFI-

specific tools such as the RFI probability dataset (Section II.D) or ERMIT [17].

A. 4- dimensional (S) and 2-dimensional (s) SSS functions

The SSS data discussed in Section II are categorized based on (t, x_{swath}), where t is discretized $(t = 1,2,\dots,132)$ and signifies the specific month from January 2010 to December 2020, and $x_{\text{swath}} = -400, -375, \dots, +375, +400$ km for both ascending and descending orbits separately (a total of $N_{\text{swath}} = 66$ intervals). SSS for each class is then mapped onto an Equal-Area Scalable Earth Grid (EASE), resulting in a 4-dimensional $S(t, \phi, \lambda, x_{\text{swath}})$ field. This comprehensive field is used for RM. When the coordinates (ϕ, λ) correspond to any grid pixel $\mathcal{M}_i(\phi_i, \lambda_i)$ analyzed using PM, we denote this as the 2-variable function $s_i(t, x_{\text{swath}})$, which equals $S(t, \phi = \phi_i, \lambda = \lambda_i, x_{\text{swath}})$, where ϕ_i, λ_i specify a grid location by index i. For simplicity, we will not mention the index i when there is no doubt. Refer to TABLE II for all method notations.

TABLE I
SPECIFICATIONS OF THE RFI REGIONS INVESTIGATED.

Region Name	Samoa	Barbados	Guinea Gulf
RFI Source Coordinates (Ω)	13.9°S 171.7°W	13.23°N 59.55°W	5°N 3°W
Annular Area (ℂ) for RFI Time Series (RM)	100 km -500 km	150 km - 600 km	100 km - 500 km
Region Limits (ℝ) for RFI Correction (RM & PM)	30°S - 5°N 170°E - 155°W	5°S - 30°N 75°W - 45°W	10°S - 20°N 20°W - 10°E

B. Gap-filling procedure: S^f and s^f

The S field and s function as introduced above are not fully populated since they are composed of swath data that do not cover the entire space-time domain for any single x_{swath} and x_{orb}. To help identifying the spatial and temporal patterns of the RFI signature, we initially fill in the missing values for each x_{swath} and x_{orb} using a temporal Gaussian convolution with a scale of $T = 2$ months; we use a gaussian std given by $\sigma_t = T/(2\sqrt{2\log 2})$. Note that this Gaussian convolution is not used for smoothing, but solely for gap filling [24], [25]. This process results in the fully completed field $S^f(t, \phi, \lambda, x_{\text{swath}})$ and function $s^f(t, x_{\text{swath}})$. However, for both S^f and s^f, pixels within the semi-permanent mask of the SSS data that cover land and sea ice are still excluded. On average, the filled values constitute about 59% of the total number of values in a time series (132). We employ a one-dimensional filling process in time, rather than a multidimensional approach involving latitude/longitude and swath distance, to preserve the sharpness of the gradient along the spatial dimensions.

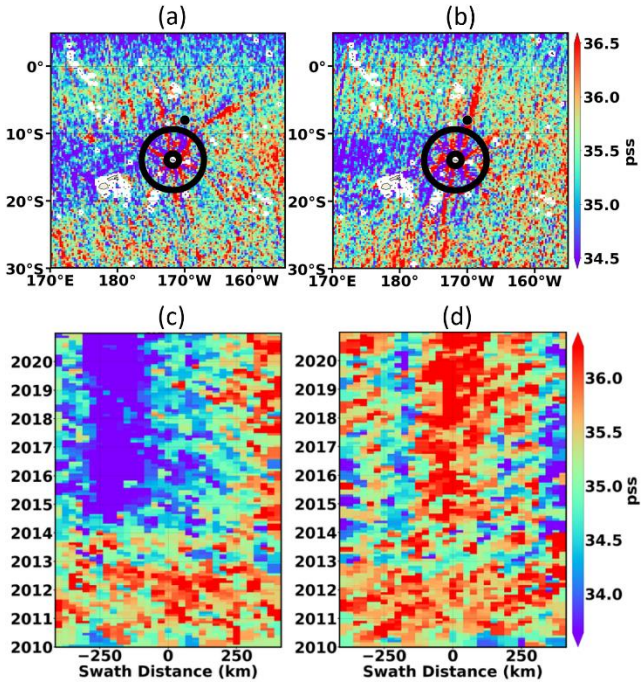

Figure 3: (a) Filled SMOS S^f field for September 2018 around the Samoa RFI source, displayed for $x_{\text{swath}} = -75$ km, $x_{\text{orb}} = \text{'A'}$. Two concentric circles with radii 100 km and 500 km are drawn, centered on the RFI source. The domain between the two circles, annulus ℂ, is used for the subsequent RM PCA (see text). A black dot is also shown at 8°S, 170°W that indicates the location used for subsequent illustrations. (b) Same as (a), but for $x_{\text{orb}} = \text{'D'}$. (c) Time series of the s^f field at the 8°S, 170°W location (black dot in (a) and (b)), for $x_{\text{orb}} = \text{'A'}$. (d) Same as (c), but for $x_{\text{orb}} = \text{'D'}$.

Figure 3 provides examples of filled fields near the Samoa islands, in September 2018, and across the entire period and swath at a specific location and orbit orientation. The significant RFI contamination is evident in panels a and b, where SSS branches are approximately 1 pss saltier or fresher than the surrounding area, diverging from the source. The RFI signature can also be observed across the swath in panels c and d. Prior to 2014, the SSS appears rather homogeneous across the swath. However, from 2014 onward, the variability in SSS is strongly dependent on swath classes.

C. Decomposition of SMOS SSS signal

The S^f field is assumed to comprise the following components:

$$S^f(t, \phi, \lambda, x_{\text{swath}}) = S_{\text{true}}(t, \phi, \lambda) + E_{\text{RFI}}(t, \phi, \lambda, x_{\text{swath}}) + E_{\text{other}}(t, \phi, \lambda, x_{\text{swath}}) + \varepsilon. \quad (1)$$

Here, S_{true} represents the actual "true" SSS at a given time, latitude, and longitude. It is independent of swath distance and orbit orientation. E_{RFI} is an uncertainty component that characterizes the RFI contamination and is dependent on all variables. E_{other} is an additional component that characterizes unspecified sources of uncertainty. Lastly, ε represents a

TABLE II
OVERVIEW AND NOTATIONS OF THE METHODOLOGICAL COMPONENTS

	RM	**PM**
Main Principles		
Time Series Characterization	In an annular region ($\mathbb{C}$) (see TABLE I)	Pixel per pixel ($\mathcal{M}$)
Applied Correction across the Swath	In a large region ($\mathbb{R}$) (see TABLE I)	Pixel per pixel ($\mathcal{M}$)
Notations		
$t, \phi, \lambda, x_{\text{swath}}$	time, latitude, longitude, swath distance for each orbit orientation	
SSS	$S(t, \phi, \lambda, x_{\text{swath}})$	$s_i(t, x_{\text{swath}})$ or $s(t, x_{\text{swath}})$ (pixel index i often omitted for simplicity)
	$S(t, \phi_i, \lambda_i, x_{\text{swath}}) = s_{(i)}(t, x_{\text{swath}})$	
Filled SSS	$S^f(t, \phi, \lambda, x_{\text{swath}})$	$s^f_{(i)}(t, x_{\text{swath}})$
	$S^f(t, \phi_i, \lambda_i, x_{\text{swath}}) = s^f_{(i)}(t, x_{\text{swath}})$	
SSS Fluctuation	$\dot{S}(t, \phi, \lambda, x_{\text{swath}})$	$\dot{s}_{(i)}(t, x_{\text{swath}})$
	$\dot{S}(t, \phi_i, \lambda_i, x_{\text{swath}}) = \dot{s}_{(i)}(t, x_{\text{swath}})$	
Filled SSS Fluctuation	$\dot{S}^f(t, \phi, \lambda, x_{\text{swath}})$	$\dot{s}^f_{(i)}(t, x_{\text{swath}})$
	$\dot{S}^f(t, \phi_i, \lambda_i, x_{\text{swath}}) = \dot{s}^f_{(i)}(t, x_{\text{swath}})$	
Swath-Averaged Field	$\langle X \rangle\big(t, (\phi, \lambda)\big) = \frac{1}{N_{\text{swath}}} \sum_{x_{\text{swath}}} X(t, (\phi, \lambda), x_{\text{swath}})$ $X = S^f, s^f, \dot{S}, \dot{s}, \dot{S}^f, \dot{s}^f$	
Difference Field	$\Delta X(t, (\phi, \lambda), x_{\text{swath}}) = X(t, (\phi, \lambda), x_{\text{swath}}) - \langle X \rangle\big(t, (\phi, \lambda)\big)$ $X = S^f, s^f, \dot{S}^f, \dot{s}^f$ (only for filled fields)	
RFI Characteristic Time Series	$U_1(t)$	$u_1(t)$
RFI Spatial Component	$W_1(\phi, \lambda, x_{\text{swath}})$	$w_1(x_{\text{swath}})$
RFI Probability	$P_{\text{RFI}}(t)$	
Correction		
Additive Correction Term	$-U_1(t)\, W_1(\phi, \lambda, x_{\text{swath}})$	$-u_1(t)\, w_1(x_{\text{swath}})$
Corrected Filled SSS Fluctuation	$\dot{S}^f_c(t, \phi, \lambda, x_{\text{swath}})$	$\dot{s}^f_c(t, x_{\text{swath}})$
Corrected SSS Fluctuation Original Sampling	$\dot{S}_c(t, \phi, \lambda, x_{\text{swath}})$	$\dot{s}_c(t, x_{\text{swath}})$
Correction Evaluation		
Mean Swath SSS Fluctuation, see (3)	$\langle \dot{S} \rangle$ and $\langle \dot{S}_c \rangle$ depend on (t, ϕ, λ)	$\langle \dot{s} \rangle$ and $\langle \dot{s}^f \rangle$ depend on t
In Situ Observation Benchmark	$\dot{S}^{\text{obs}}(t, \phi, \lambda)$	$\dot{s}^{\text{obs}}(t)$
Timewise Std of Difference with Observations	$\Upsilon = std_t(\langle \dot{S} \rangle - \dot{S}^{\text{obs}})$ $\Upsilon_c = std_t(\langle \dot{S}_c \rangle - \dot{S}^{\text{obs}})$ depend on (ϕ, λ)	$\gamma = std_t(\langle \dot{s} \rangle - \dot{s}^{\text{obs}})$ $\gamma_c = std_t(\langle \dot{s}_c \rangle - \dot{s}^{\text{obs}})$ at a pixel i
Improvement/ Degradation of Correction	$\Upsilon - \Upsilon_c \begin{cases} > 0 \text{ improves} \\ < 0 \text{ degrades} \end{cases}$	$\gamma - \gamma_c \begin{cases} > 0 \text{ improves} \\ < 0 \text{ degrades} \end{cases}$
Methods Comparison	$\Upsilon_c(\phi_i, \lambda_i) - (\gamma_c)_i \begin{cases} > 0 \text{ PM better} \\ < 0 \text{ RM better} \end{cases}$ at a pixel i	

random noise component, the dependency of which is not specified; ε notably includes random noise contribution from the gap-filling process. The same type of decomposition is applicable to s^f:

$$s^f(t, x_{\text{swath}}) = s_{\text{true}}(t) + e_{\text{RFI}}(t, x_{\text{swath}}) + e_{\text{other}}(t, x_{\text{swath}}) + \delta. \quad (2)$$

s_{true}, e_{RFI}, e_{other}, and δ function analogously to S_{true}, E_{RFI}, E_{other} and ε, respectively, but are specific to each pixel.

D. Difference field

Let us consider $X(t, (\phi, \lambda), x_{\text{swath}})$ as a function that depends on time, swath distance, and optionally, on latitude and longitude. We can represent the average of X over all swath intervals as $\langle X \rangle$, and the departure from this average as ΔX:

$$\langle X \rangle\big(t, (\phi, \lambda)\big) \equiv \frac{1}{N_{\text{swath}}} \sum_{x_{\text{swath}}} X(t, (\phi, \lambda), x_{\text{swath}}) \quad (3)$$

$$\Delta X(t, (\phi, \lambda), x_{\text{swath}}) = X(t, (\phi, \lambda), x_{\text{swath}}) - \langle X \rangle\big(t, (\phi, \lambda)\big). \quad (4)$$

By applying the Δ operator to each term in (1), and noting that ΔS_{true} equals zero (as $\langle S_{\text{true}} \rangle = S_{\text{true}}$), we effectively eliminate S_{true}'s contribution:

$$\Delta S^f = \Delta E_{\text{RFI}} + \Delta E_{\text{other}} + \varepsilon'. \quad (5)$$

In this context, ε' represents a noise term that includes noise from SMOS SSS as well as SSS variability between successive satellite passes at intervals of less than 18 days. ΔS^f is derived from the data, and our working assumption is that it is predominantly influenced by ΔE_{RFI} in RFI-contaminated regions, with ε' being minor in comparison.

Under this framework, ΔS^f should not incorporate geophysical signal variability from the true SSS. It is crucial to effectively distinguish between the variability caused by RFI contamination and the inherent geophysical variability. The same considerations can be applied to s^f for any fixed location:

$$\Delta s^f = \Delta e_{\text{RFI}} + \Delta e_{\text{other}} + \delta'. \quad (6)$$

E. SSS fluctuation

Our primary objective is to detect and correct variable RFI contamination in SMOS SSS data. From this point forward, we focus exclusively on the variations from the temporal mean, which we refer to as "fluctuations", for any variable under study. We employ the following notation for this purpose: $\dot{X}(t) = X(t) - 1/N_t \sum_{t=1}^{N_t} X(t)$, where $N_t = 132$ denotes the total number of months in the given period.

F. RFI characteristic time series

In the context of PCA, it is implicitly assumed that RFI contamination is proportional to a singular temporal function that is shared across all swath distances. This assumption is applied in the PM approach at individual pixels, and this behavior can be partially attributed to the L1 reconstruction. The same assumption is extended to all pixels within a geographical domain in the RM approach.

Thus, applying PCA to $\Delta \dot{S}^f$ within a specific region $\mathbb{R}$ for RM or to $\Delta \dot{s}^f$ at a fixed pixel $\mathcal{M}$ for PM addresses two key questions: 1) Does a dominant mode control the variability of cross-swath SSS differences, which should originate solely from measurement artifacts rather than geophysical variability? 2) Does the timing of this dominant mode coincide with the RFI source's activity in region $\mathbb{R}$ for RM, or at a pixel $\mathcal{M}$ for PM?

1) **RM Method**

The development of PCA, where the covariance matrix is computed using a time-wise inner product [25] is as follows:

$$\Delta \dot{S}^f(t, \phi, \lambda, x_{\text{swath}}) = U_1(t) V_1(\phi, \lambda, x_{\text{swath}}) + \cdots + U_M(t) V_M(\phi, \lambda, x_{\text{swath}}). \quad (7)$$

Here, the U_k represent the principal components (PC, or time series, unitless), and the V_k the empirical orthogonal functions (EOF, or spatial patterns, in pss). The U_k are centered in time ($\sum_{t=1}^{N_t} U_k(t) = 0$), normalized ($\sum_{t=1}^{N_t} U_k^2(t) = 1$), and orthogonal to each other ($\sum_{t=1}^{N_t} U_k(t) U_l(t) = 0$ if $k \neq l$). The V_k are orthogonal to each other ($\sum_{\phi,\lambda,x_{\text{swath}}} V_k V_l = 0$ if $k \neq l$), and their norm is equal to the singular value Γ_k, which is equal to the variance square root of the mode, $\Gamma_k = \left(\sum_{\phi,\lambda,x_{\text{swath}}} V_k^2\right)^{1/2}$. PCA is a variance-maximizing procedure, and the modes are ordered such that $\Gamma_1 > \cdots > \Gamma_M > 0$. The percentage of variance of a mode k is given by $100 \times \Gamma_k^2 / \sum_{l=1}^{M} \Gamma_l^2$.

The total number of modes M is determined by the rank of the covariance matrix and is here equal to or less than the number of months in the monthly time series (132).

A first dominant mode is such that its variance is much larger than the variance from the second and all the higher modes: $\Gamma_1^2 \gg \Gamma_2^2$. If we assume that there is a primary dominant mode, then the corresponding PC $U_1(t)$ could represent the timing of RFI contamination in the region, and the associated EOF $V_1(\phi, \lambda, x_{\text{swath}})$ the RFI signatures in the region and along the swaths. To confirm that the RFI timing is captured by $U_1(t)$, we compare it to the time series of the RFI probability introduced in Section II.D, extracted from the grid box that encompasses the location of the RFI source. This probability time series is subsequently referred to as $P_{\text{RFI}}(t)$.

While PCA can be utilized across any geographical area, in RM it is confined to an annular area $\mathbb{C}$ centered around the localized RFI-source, as illustrated in Figure 3a-b and specified in TABLE I. The chosen annulus has a maximum radius on the order of 500 km, which is sufficiently large to accurately capture the RFI signal originating from the source, yet small enough to minimize the likelihood of interference from other RFI sources. The annulus' inner radius, typically over 100 km, acts as a barrier against other potential contaminations near the RFI source, such as land-sea contamination. This geographical limitation aims to achieve the most accurate RFI timing characterization associated with the specific RFI source.

2) **PM Method**

Similar to the RM approach and (7), the PCA of $\Delta\dot{s}^f$ at a fixed pixel (ϕ, λ) results in the following equation:

$$\Delta\dot{s}^f(t, x_{\text{swath}}) = u_1(t)v_1(x_{\text{swath}}) + \cdots + u_m(t)v_m(x_{\text{swath}}). \quad (8)$$

In contrast to RM, the RFI timing characterization is solely based on time variations across the swath at that pixel. The total number of modes, m, is determined by the number of swath distances (66), which is less than the number of months in the monthly time series (132). If the first mode is dominant, $u_1(t)$ could represent the timing of some RFI contamination at this pixel, and $v_1(x_{\text{swath}})$ the RFI signatures along the swaths. To verify if $u_1(t)$ captures some RFI timing, we compare it with $P_{\text{RFI}}(t)$, extracted from the grid box closest to the pixel (ϕ, λ) under consideration.

G. RFI spatial component

The time series $U_1(t)$ represents our timing characterization of RFI contamination emanating from the target RFI source in the treated region, while $u_1(t)$ represents our timing characterization of RFI contamination at a single examined location, independent of all other locations. We employ a regression procedure to generate the correction spatial component as follows.

Regression coefficients relative to $U_1(t)$ and $u_1(t)$ are computed as the direct projection of the initial dataset that we aim to correct, $\dot{S}^f(t, \phi, \lambda, x_{\text{swath}})$, onto $U_1(t)$:

$$W_1(\phi, \lambda, x_{\text{swath}}) = \sum_{t=1}^{N_t} U_1(t)\dot{S}^f(t, \phi, \lambda, x_{\text{swath}}) \quad (9)$$

and $\dot{s}^f(t, x_{\text{swath}})$ onto $u_1(t)$:

$$w_1(x_{\text{swath}}) = \sum_{t=1}^{N_t} u_1(t)\dot{s}^f(t, x_{\text{swath}}). \quad (10)$$

Since $U_1(t)$ is centered and normalized, it naturally follows that $W_1(\phi, \lambda, x_{\text{swath}})$ is the statistical regression coefficient at each location and for each swath interval of $\dot{S}^f$ onto U_1. Identically, $w_1(x_{\text{swath}})$ is the statistical regression coefficient of $\dot{s}^f$ onto u_1 for each swath interval.

H. RFI correction

We thus have the two elements U_1 and W_1 for RM, u_1 and w_1 for PM, that enable us to construct a corrected dataset $\dot{S}_c^f$ for RM within the entire region R, and $\dot{s}_c^f$ for PM at a single pixel $\mathcal{M}$:

$$\dot{S}_c^f(t, \phi, \lambda, x_{\text{swath}}) = \dot{S}^f(t, \phi, \lambda, x_{\text{swath}}) - U_1(t)\, W_1(\phi, \lambda, x_{\text{swath}}) \quad (11)$$

and

$$\dot{s}_c^f(t, x_{\text{swath}}) = \dot{s}^f(t, x_{\text{swath}}) - u_1(t)\, w_1(x_{\text{swath}}). \quad (12)$$

By design, the projections of $\dot{S}_c^f$ onto U_1 and of $\dot{s}_c^f$ onto u_1 are zero: we thus eliminate any covariance between the SMOS SSS dataset and the RFI characteristic time series in the treated region or pixel.

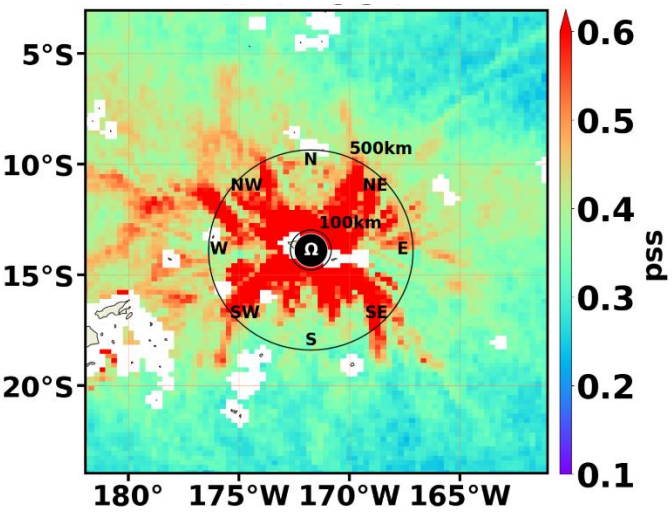

Figure 4: All-time std of $\langle\dot{S}\rangle$ at each location around the Samoa RFI source. The RFI source (Ω) and annular domain (100-500 km) are indicated as well as the rose angles N, NE, etc. used in the polar-Hovmöller representation.

I. Evaluation of the corrected datasets

1) **Preliminary Procedures**

In Section IV, we focus on characterizing the corrected dataset and evaluating its improvement over the initial dataset by benchmarking against in situ references.

To construct the final corrected dataset and facilitate comparisons with in situ datasets, we revert to SSS with the original time-space sampling prior to gap filling. Hence, we make comparisons using, for RM, the fields $\dot{S}$ and $\dot{S}_c$ derived from $\dot{S}^f$ and $\dot{S}_c^f$, and for PM, the functions $\dot{s}$ and $\dot{s}_c$ derived from $\dot{s}^f$ and $\dot{s}_c^f$, by applying the initial mask.

Before comparing with in situ datasets, we average the salinities across all swath classes, resulting in $\langle\dot{S}\rangle$ and $\langle\dot{S}_c\rangle$ for RM, and $\langle\dot{s}\rangle$ and $\langle\dot{s}_c\rangle$ for PM at all pixels, as defined in (3). These averaged salinities depend solely on time and geographical location.

2) **Visualization of the comparisons to ISAS**

The SSS differences with respect to ISAS are examined around the RFI source using a representation we will call "polar-Hovmöller." In this representation, the fields $\langle\dot{S}\rangle$ and $\langle\dot{S}_c\rangle$ are averaged over small angular sectors originating from the RFI source and encompassing the entire annular domain, as shown in Figure 4. This leads to time series for sub-angles between east (E), northeast (NE), north (N), etc. This allows for a comparison with the observational field dataset ISAS and the results from RM and PM, as demonstrated in the upcoming Section IV.

3) **Validation metrics**

The in situ observation dataset, denoted as $\dot{S}^{\text{obs}}(t, \phi, \lambda)$, or $\dot{s}^{\text{obs}}(t)$ at any pixel, serves as a benchmark for evaluating the methods. This dataset could be either ISAS SSS or mooring SSS, as discussed in Sections II.B and II.C respectively. The primary evaluation metric used is the timewise std of the difference between the SMOS SSS fluctuation and $\dot{S}^{\text{obs}}$. The

timewise std is calculated as follows (seeing that the variable $\dot{X}$ is already centered):

$$std_t(\dot{X}) = \left(\frac{1}{N_t}\sum_{t=1}^{N_t}\dot{X}(t)^2\right)^{1/2}. \quad (13)$$

Using this, we can define the evaluation metrics for the original and corrected datasets. For RM, the metrics are defined as:

$$\begin{cases} \Upsilon = std_t\left(\langle\dot{S}\rangle - \dot{S}^{\text{obs}}\right) \\ \Upsilon_c = std_t\left(\langle\dot{S}_c\rangle - \dot{S}^{\text{obs}}\right) \end{cases} \quad (14)$$

and for PM:

$$\begin{cases} \gamma = std_t\left(\langle\dot{s}\rangle - \dot{s}^{\text{obs}}\right) \\ \gamma_c = std_t\left(\langle\dot{s}_c\rangle - \dot{s}^{\text{obs}}\right) \end{cases} \quad (15)$$

iterated for all pixels. Henceforth, we will refer to these metrics as "std of difference".

This allows us to generate maps of these metrics for the region under study. The smaller the values of Υ_c or γ_c, the more effective the correction methods are compared to the observational benchmark. The differences $\Upsilon - \Upsilon_c$ and $\gamma - \gamma_c$ provide measures of the correction methods' effectiveness. Since the PM method is applied to all pixels within a region, we likewise obtain maps of the comparison metrics. Furthermore, by comparing $\Upsilon_c(\phi_i, \lambda_i) - (\gamma_c)_i$ for all pixels i, we produce a map illustrating the relative strengths and weaknesses of both methods across the region (refer to TABLE II for summary).

In addition to std of difference, we use the Pearson correlation at any pixel, defined by:

$$r = \frac{\frac{1}{N_t}\sum_{t=1}^{N_t}\langle\dot{s}\rangle(t) \times \dot{s}^{\text{obs}}(t)}{std_t(\langle\dot{s}\rangle) \times std_t(\dot{s}^{\text{obs}})}. \quad (16)$$

IV. Results

We conduct our study in three specific regions known for RFI contamination: the Samoa region surrounding the Samoa islands in the southwestern tropical Pacific; the Barbados region around Barbados island in the Caribbean Sea; and the Guinea Gulf region encompassing the Ghana coast and neighboring countries (TABLE I). In each region, we identified an RFI source initially through visible signatures in the SMOS SSS data, later confirmed by RFI probability data and ERMIT reports. The boundaries were chosen to centrally locate the RFI source, facilitating the testing and comparison of correction methods within these areas.

A. Samoa region

1) **Results from RM**

As outlined in Section III.F.1, we apply PCA to $\Delta\dot{S}^f$ in the annular region surrounding the RFI source in Samoa depicted in Figure 4. The first mode contributes to nearly 50% of the total variance and is largely superior to the second (5%) and all subsequent modes (Figure 5). The PCA time series associated with the first mode $U_1(t)$ distinctly indicates an RFI activation in 2014, remaining relatively stable before and after this period (Figure 6). This is consistent with the observational data in Figure 3c-d, where RFI signatures are evident after 2014. The RFI probability time series $P_{\text{RFI}}(t)$ further corroborates that the RFI source becomes fully active starting in early 2014, as it consistently hovers close to 1, while it fluctuates between 0 and 0.8 prior to 2014 (Figure 6). $U_1(t)$ and $P_{\text{RFI}}(t)$ are highly correlated with correlation of 0.90. Note that to ensure consistency across temporal scales, $P_{\text{RFI}}(t)$ was processed through a low-pass filter using a Gaussian convolution with a 1-month scale [24], [25]. EOF spatial patterns $V_1(\phi, \lambda, x_{\text{swath}})$, tied to the annular

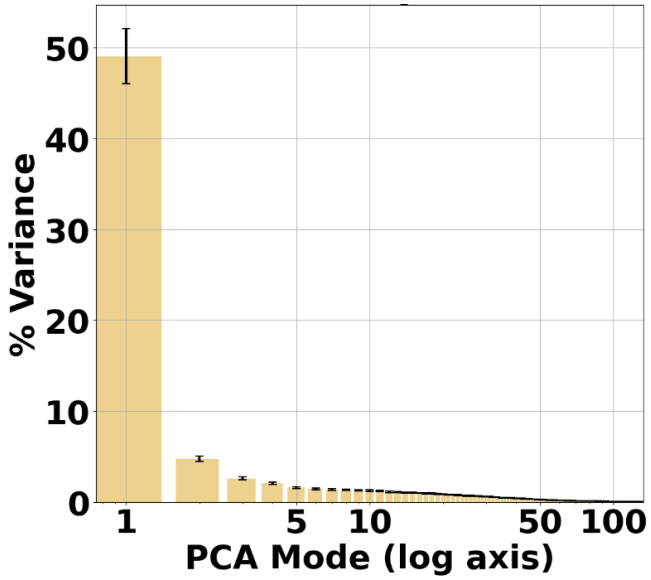

Figure 5: Distribution of the percentage variance among the PCA modes derived from the PCA of the filled differential dataset $\Delta\dot{S}^f$ within the Samoa annular region. The error bars, computed as per [26], provide a somewhat broad yet dependable indication of the separability of the modes.

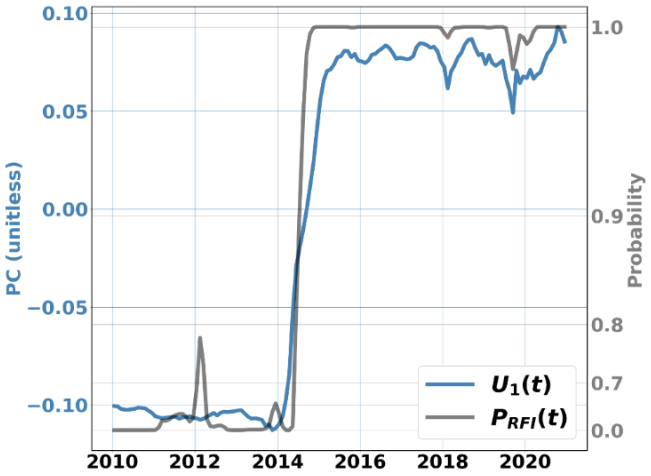

Figure 6: (Blue) Time series $U_1(t)$ (see text) associated to 1st mode of the PCA of $\Delta\dot{S}^f$ within the Samoa annular region as shown in Figure 3a-b. (Grey) RFI probability time series for a location near 13.9°S, 171.7°W which is identified as the RFI source location in this region. The probability axis undergoes a transformation by raising it to the power of 6 to highlight its similarity more effectively with the PC. Correlation of $U_1(t)$ with $P_{\text{RFI}}(t)$ is 0.90.

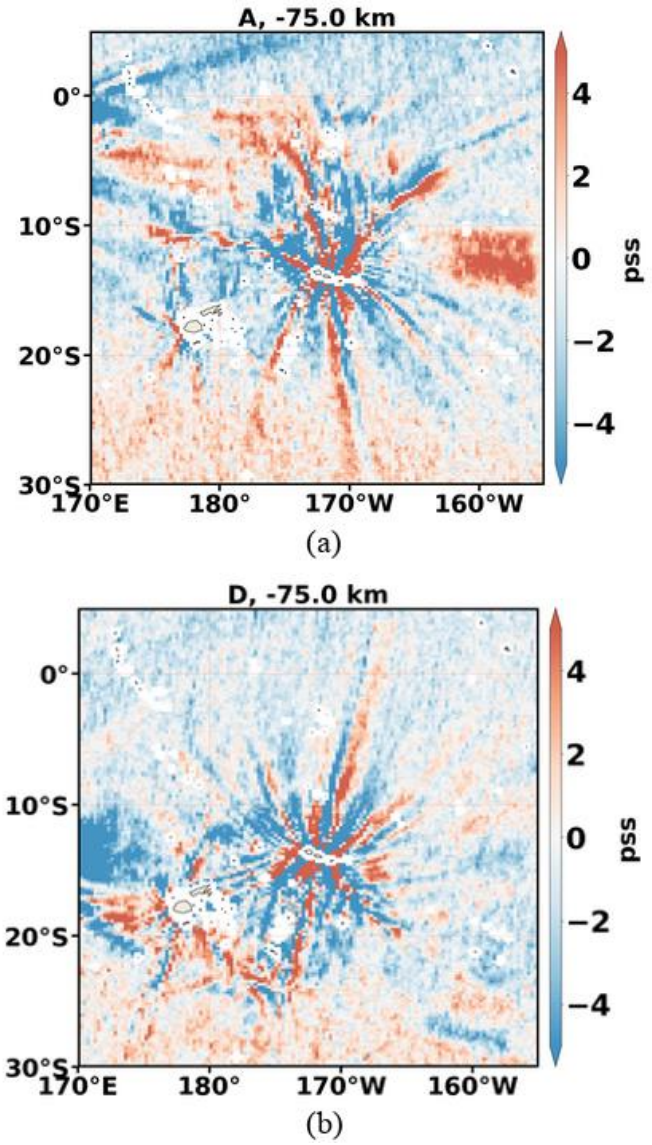

Figure 7: Projection $W_1(\phi, \lambda, x_{\text{swath}})$ of the 4-dimensional dataset $\dot{S}^f$ onto the RFI characteristic time series $U_1(t)$ for the Samoa region. Maps are shown for the case $x_{\text{swath}} = -75km$. (a) For $x_{\text{orb}} = 'A'$. (b). $x_{\text{orb}} =' D'$. The -75km interval is selected out of the 33 swath intervals for illustration.

domain and fundamental to PCA, are excluded from this presentation, as we do not use them in our methodology, and they do not add further insights beyond those offered by the regression-derived spatial components discussed later (figures are provided as supplementary material).

The spatial component $W_1(\phi, \lambda, x_{\text{swath}})$ characterizes RFI signatures that are prominently visible, with both positive and negative branches stemming from the source on the Samoa Islands (Figure 7). Some of these branches stretch over 1000 km, particularly those pointing towards the south, northeast and north-northeast. Other intense branches are noticeable in the northwest, originating from another RFI source outside the studied region. RFI signatures also radiate from the Fiji Islands, situated west-southwest of Samoa. These features from other sources are visible as their timing has some correlation with $U_1(t)$. Note that W_1 is similar in construction to V_1. However, V_1 is relative to the differential dataset $\Delta\dot{S}^f$ and to the annular domain, while W_1 is relative to the total dataset $\dot{S}^f$ within the large region which we aim to correct.

Before we explore the correction results, we briefly discuss the context of SSS variations in the Samoa region. A significant portion of this region falls within the South Pacific Convergence Zone (SPCZ), an area known for its strong interannual variability in SSS [6], [27]. In the area surrounding the RFI source, the ISAS SSS fluctuation ($\dot{S}^{\text{ISAS}}$) exhibits markedly positive values across the region prior to 2014, particularly during the latter halves of 2011 and 2012 (Figure 8a). Post-2013, $\dot{S}^{\text{ISAS}}$ reveals widespread and intense negative fluctuations.

From 2010 to 2013, the variability observed in the original SMOS SSS fluctuations, $\langle\dot{S}\rangle$ (Figure 8b), aligns with the variability identified by $\dot{S}^{\text{ISAS}}$ and is influenced by ENSO signals [6]. However, starting in 2014, the variability in $\langle\dot{S}\rangle$ no longer matches that in $\dot{S}^{\text{ISAS}}$ and is predominantly affected by directional variability around the RFI source. This period is characterized by pronounced RFI signatures with extremely high values, which persist and exhibit either positive or negative signs across various angular intervals.

Starting from 2014, the corrected SMOS SSS fluctuation $\langle\dot{S}_c\rangle$ no longer displays the overpowering RFI signatures (Figure 8c). The fluctuation range is approximately 1 pss, and the temporal variability now exceeds the variability across angles. This change is due to the elimination of the star-like RFI contamination pattern that previously affected the SMOS satellite measurements. Furthermore, the variations in $\langle\dot{S}_c\rangle$ are now largely in sync with $\dot{S}^{\text{ISAS}}$ for most of the observation period. In terms of magnitude, $\langle\dot{S}_c\rangle$ shows strong positive values over the region during specific periods, notably the latter halves of 2011 and 2012, similar to $\dot{S}^{\text{ISAS}}$. However, at the end of 2015 and beginning of 2016, and again late 2017 through late 2019, $\langle\dot{S}_c\rangle$ remains strongly positive, whereas $\dot{S}^{\text{ISAS}}$ shows more moderate values. Between these periods, we observe fresher water, though less fresh compared to $\dot{S}^{\text{ISAS}}$.

Single-location time series distinctly demonstrate the improvements made by the RM approach (Figure 9). For instance, severe RFI contamination, occasionally reaching up to 3 pss in proximity to the source (~250 km), is significantly reduced by the correction method (Figure 9a-c). Metrics comparing with ISAS generally indicate an improvement after correction, characterized by a reduction in the std of difference and an increase in correlation (TABLE IIIa-c).

However, not all RFI corrected locations see an increase in correlation, as shown by the data in Figure 9b and the corresponding values in TABLE IIIb (note that changes in the correlation coefficient are sometimes challenging to visually discern in time series plots, particularly when the most significant change occurs in the std of difference). A similar occurrence is observed in the Guinea Gulf region (location h), which will be discussed later. Considering the 95% confidence intervals associated with the correlation coefficients, changes in correlation, whether increases or decreases, are generally not significant. The exception is

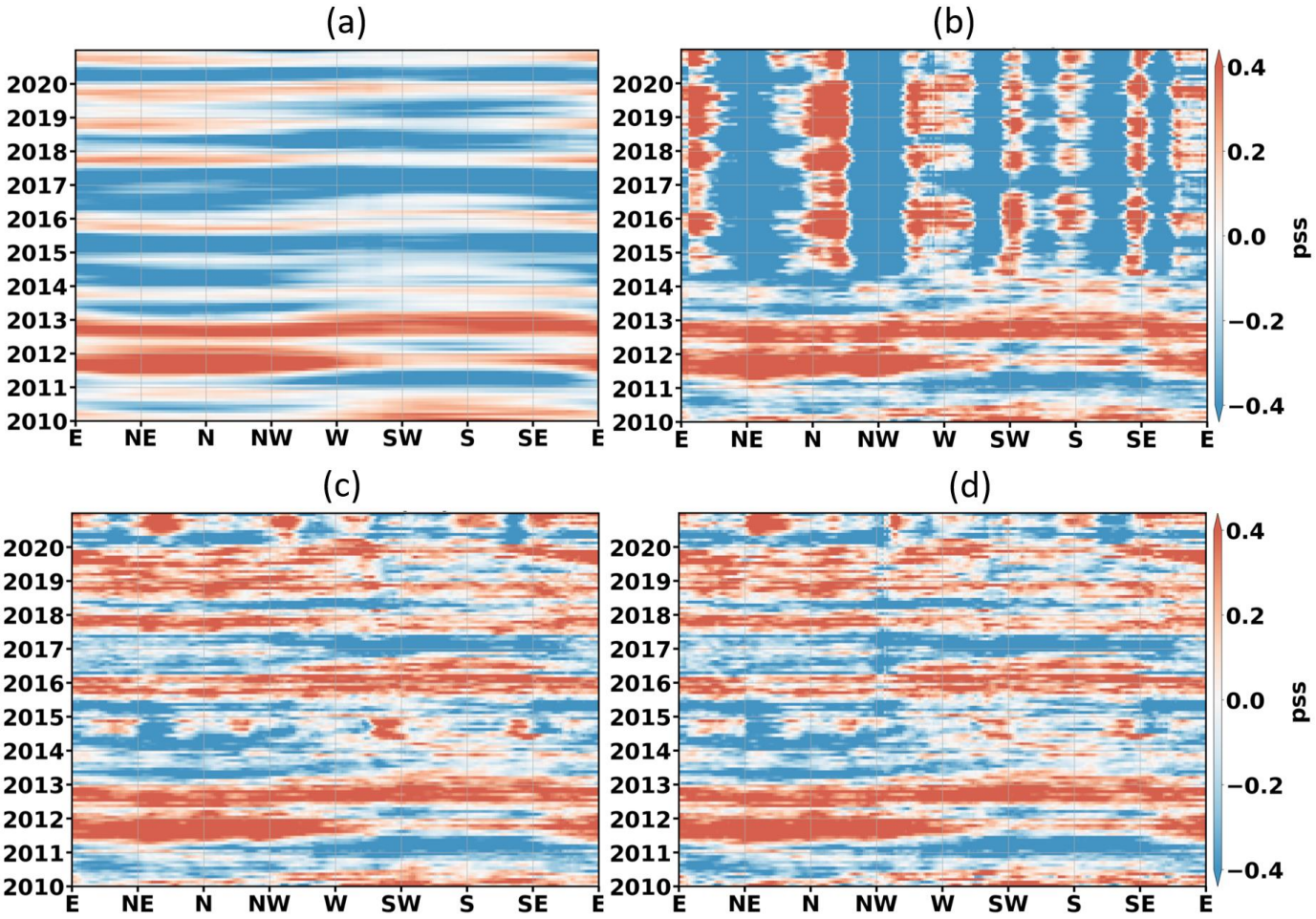

Figure 8: (a) ISAS SSS fluctuation ($\dot{S}^{ISAS}$) relative to the 2010-2013 period, presented as a polar-Hovmöller diagram, with the center at the RFI source as shown in Figure 4 (see text for details). (b) Same as (a), but for the swath-averaged initial SMOS SSS fluctuation ($\langle\dot{S}\rangle$). (c) Same as (a), but for the swath-averaged SMOS SSS corrected using RM ($\langle\dot{S}_c\rangle$). (d) Same as (a), but for the swath-averaged SMOS SSS corrected using the pointwise method at all pixels ($\langle\dot{s}_c\rangle$).

location a and, as seen later, e and f, where the correlation significantly increases (note also that the significance of a coefficient change is related to the overlap between confidence intervals before and after correction; the change is most significant when there is little or no overlap). Conversely, the std of difference significantly decreases in most locations impacted by RFI contamination.

As mentioned previously, strong deviations from ISAS, about 1 pss or more (in std of difference), are observed within a few hundred kilometers of the RFI source (Figure 10a). These deviations are considerably reduced due to the correction method (Figure 10b). Other RFI contamination features, such as the far-reaching branches, are also eliminated. The improvement achieved by the method is distinctly noticeable in Figure 10c, denoted by the blue domain. Intriguingly, a positive correction is also observed around the Fiji Islands, situated to the west-southwest of Samoa, where another RFI source is identified (ERMIT). The reason for this occurrence, whether due to Samoa's proximity or the chance that the RFI characteristic time series $U_1(t)$ captures some aspect of the timing associated with the Fiji RFI source, was not investigated in this section but will be revisited in the discussion of the PM results. Despite these effective corrections, some degradation, approximately -0.05 pss (red areas), is also detected at locations far from the RFI source.

Generally, few moorings are located near areas affected by RFI contamination. In the Samoa region, a TAO mooring at 8°S, 170°W lies within an RFI branch extending over 600 km from the Samoa source (indicated by a bullet in Figure 10). The RFI impact here is moderate compared to locations nearer the source. However, Figure 10c shows a patch of significant improvement in SMOS SSS using the RM method around this pixel, with ISAS as a benchmark. The $\langle\dot{S}\rangle$ time series closely matches both the mooring $\dot{s}^{\mathrm{MOOR}}(t)$ and $\dot{s}^{\mathrm{ISAS}}(t)$ up until early 2014 (Figure 9d; notice a gap in $\dot{s}^{\mathrm{MOOR}}$ between early 2012 and late 2014). From mid-2014 to 2020, $\langle\dot{S}\rangle$ shows a consistent negative bias of about 0.5 pss relative to both $\dot{s}^{\mathrm{ISAS}}$ and $\dot{s}^{\mathrm{MOOR}}$. The corrected series $\langle\dot{S}_c\rangle$ agrees more closely with in situ data during this period. Before the RFI activation, $\langle\dot{S}_c\rangle$ closely follows $\langle\dot{S}\rangle$, as the correction's RFI time series $U_1(t)$ is near zero, assuming its 2010-2013 mean is also removed (Figure 6).

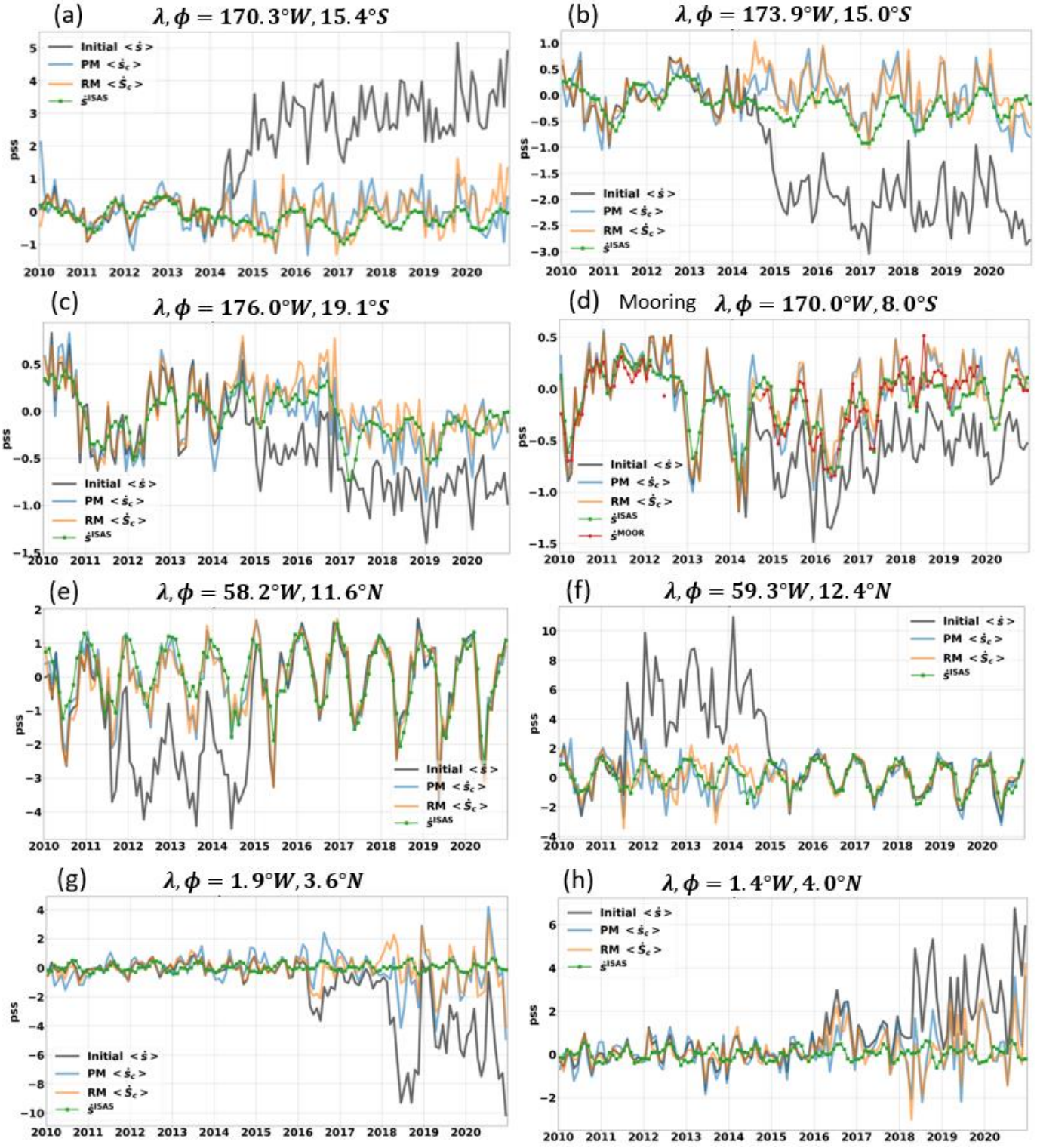

Figure 9: Time series of SMOS and observed SSS fluctuation, within the Samoa region at the locations: (a) 15.4°S, 170.3°W, (b) 15°S, 173.9°W, (c) 19.1°S, 176°W, (d) 8°S, 170°W; within the Barbados region at: (e) 11.6°N, 58.2°W, (f) 12.4°N, 59.3°W; within the Guinea Gulf region at: (g) 3.6°N, 1.9°W, (h) 4°N, 1.4°W. Location for (d) coincides with a TAO mooring. (Black) Initial SMOS $\langle \dot{S} \rangle$. (Blue) PM corrected SMOS $\langle \dot{s}_c \rangle$. (Orange) RM corrected SMOS $\langle \dot{S}_c \rangle$. (Green) ISAS $\dot{S}^{\mathrm{ISAS}}$. (Red) TAO mooring $\dot{S}^{\mathrm{MOOR}}$ (when available). Std of difference and correlation results are indicated in TABLE III.

TABLE III
STATISTICAL COMPARISON BETWEEN THE SMOS AND IN SITU (ISAS AND MOORING) SSS TIME SERIES AT SPECIFIC LOCATIONS, CORRESPONDING TO FIGURE 9. A 95% CONFIDENCE INTERVAL FOR THE METRIC, OBTAINED THROUGH A BOOTSTRAP METHOD [28], IS INDICATED IN PARENTHESES.

Region	Location	Time Series Pair		Std of Diff (pss)	Pearson Correlation
Samoa	(a) 170.3°W,15.4°S D(Source)=230km	ISAS	INIT SMOS	**1.69** (1.60,1.80)	**-0.23** (-0.37,-0.08)
			RM SMOS	**0.45** (0.40,0.52)	**0.53** (0.41,0.63)
			PM SMOS	**0.52** (0.46,0.61)	**0.40** (0.26,0.52)
	(b) 173.9°W,15°S D(Source)=263km		INIT SMOS	**0.94** (0.88,1.01)	**0.60** (0.48,0.70)
			RM SMOS	**0.38** (0.34,0.43)	**0.45** (0.32,0.57)
			PM SMOS	**0.38** (0.34,0.44)	**0.50** (0.37,0.60)
	(c) 176°W,19.1°S D(Source)=733km		INIT SMOS	**0.37** (0.33,0.40)	**0.64** (0.55,0.72)
			RM SMOS	**0.23** (0.20,0.26)	**0.73** (0.63,0.79)
			PM SMOS	**0.25** (0.22,0.28)	**0.72** (0.63,0.79)
	(d) 170°W,8°S D(Source)=683km		INIT SMOS	**0.30** (0.27,0.34)	**0.77** (0.70,0.82)
			RM SMOS	**0.21** (0.19,0.24)	**0.81** (0.76,0.86)
			PM SMOS	**0.21** (0.19,0.24)	**0.83** (0.78,0.87)
		MOORING	INIT SMOS	**0.33** (0.29,0.38)	**0.69** (0.56,0.78)
			RM SMOS	**0.19** (0.17,0.23)	**0.81** (0.71,0.87)
			PM SMOS	**0.21** (0.18,0.24)	**0.81** (0.72,0.88)
Barbados	(e) 58.2°W,11.6°N D(Source)=236km	ISAS	INIT SMOS	**1.41** (1.29,1.54)	**0.50** (0.39,0.60)
			RM SMOS	**0.66** (0.57,0.88)	**0.80** (0.70,0.85)
			PM SMOS	**0.61** (0.52,0.85)	**0.83** (0.72,0.87)
	(f) 59.3°W,12.4°N D(Source)=103km		INIT SMOS	**2.76** (2.40,3.15)	**0.31** (0.13,0.45)
			RM SMOS	**0.82** (0.70,0.97)	**0.69** (0.59,0.77)
			PM SMOS	**0.96** (0.83,1.17)	**0.65** (0.50,0.75)
Guinea Gulf	(g) 1.9°W,3.6°N D(Source)=192km		INIT SMOS	**2.63** (2.24,3.07)	**-0.11** (-0.29,0.08)
			RM SMOS	**1.04** (0.88,1.25)	**0.09** (-0.08,0.24)
			PM SMOS	**1.19** (0.99,1.47)	**0.19** (0.03,0.34)
	(h) 1.4°W,4°N D(Source)=205km		INIT SMOS	**1.56** (1.32,1.84)	**0.12** (-0.06,0.30)
			RM SMOS	**0.93** (0.77,1.19)	**0.04** (-0.13,0.23)
			PM SMOS	**1.02** (0.89,1.21)	**-0.04** (-0.19,0.14)

Metrics quantify RM's improvement: std of difference decreases from 0.33 pss to 0.19 pss, and correlation increases from 0.69 to 0.81 (TABLE IIId). Additional comparisons at TAO locations (supplementary material), not significantly affected by RFI, demonstrate the correction methods' performance in RFI-free areas. These analyses confirm from a local viewpoint that RM may reduce agreement with both ISAS and mooring data, especially far from the RFI source.

2) **Results from PM, and comparison with RM**

As earlier stated, the key distinction from RM is that the PM approach processes each location (ϕ, λ) individually. It is applied independently to all grid pixels within a region. We first demonstrate this method at the pixel that coincides with the previously introduced mooring location at 8°S, 170°W. Here, in accordance with Section III.F.2, the first mode is dominant, contributing to 27% of the total variance, while the second mode accounts for 5%. The time series $u_1(t)$ associated with this mode is shown in Figure 11, along with the RFI probability time series of the nearest RFI source $P_{\mathrm{RFI}}(t)$. The two series exhibit a correlation of 0.89, signifying a robust correlation. However, the visual alignment between the two series is not as pronounced as the RM results depicted in Figure 6. This discrepancy arises because, in RM, $U_1(t)$ integrates the RFI temporal signature from multiple pixels surrounding the source, while $u_1(t)$ is derived from a singular pixel.

At the location 8°S, 170°W, the cross-swath correction component $w_1(x_{\mathrm{swath}})$ for the descending orbit peaks at $x_{\mathrm{swath}} = -200\,km$ and increases, becoming positive for $x_{\mathrm{swath}} > 300\,km$ (Figure 12). For the ascending orbit, $w_1(x_{\mathrm{swath}})$ is positive between -100 and +200 km and turns negative outside this range. These patterns correspond with the variations in $\dot{s}^f$ post-2014 when the Samoa RFI perturbation is active, as shown in Figure 3c,d for $x_{\mathrm{orb}} =' A', 'D'$.

At 8°S, 170°W, PM operates a major rectification from 2014 onwards, similar to RM (Figure 9d), a pattern also observed at the other locations shown (Figure 9a-c). At 8°S, 170°W, the time series corrected by both methods are nearly identical

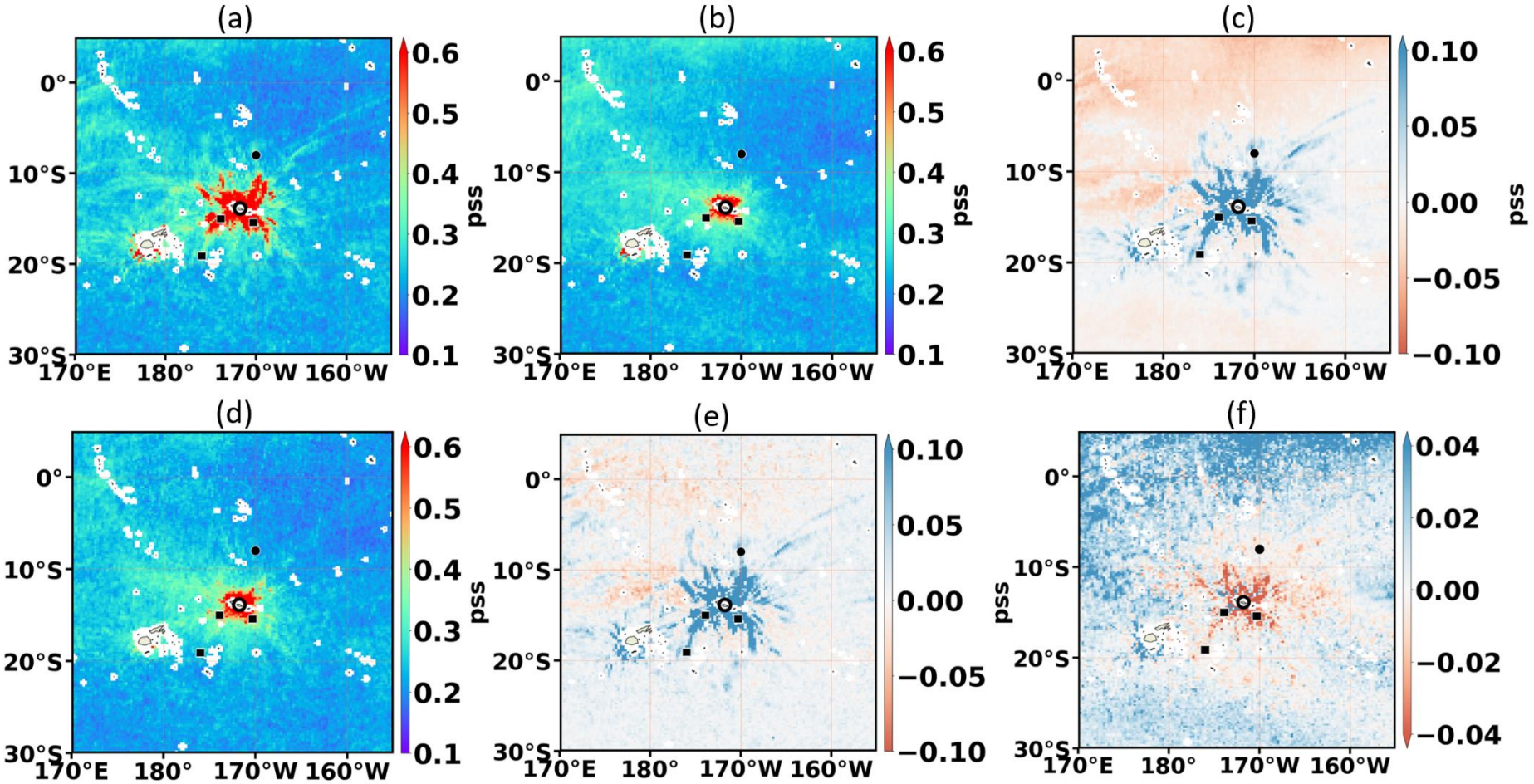

Figure 10: (a) Metric Υ (see TABLE II), equal to the timewise std of difference at each location between the swath-averaged initial SMOS SSS fluctuation $\langle \dot{S} \rangle$ and $\dot{S}^{\mathrm{ISAS}}$. (b) Υ_c, same as (a) but for the corrected SSS using RM $\langle \dot{S}_c \rangle$. (c) $\Upsilon - \Upsilon_c$, the algebraic difference between maps (a) and (b). (d) Metric γ_c at all pixels, which is the same as (b), but for the corrected SSS using PM $\langle \dot{s}_c \rangle$. (e) $\gamma - \gamma_c$ at all pixels, same as (c), but difference is between maps (a) and (d). (f) $\Upsilon_c - \gamma_c$ at all pixels, the algebraic difference between maps (e) and (c). RFI source (black circle), TAO mooring locations (black bullet) and other locations for time series evaluation (black squares) are indicated (see TABLE I and TABLE III).

except for a few periods of larger differences around mid-2011, mid-2012, and mid-2014

The comparison metrics, whether with ISAS or mooring data, are similar (TABLE IIIa-d). RM exhibits a higher correlation with ISAS, while PM shows a lower std of difference with mooring data. However, these PM/RM differences are moderated by the widths of the confidence intervals. At the non-mooring locations (a-c), PM and RM often show differences throughout the period, with RM's std of difference being either equal to or less than that of PM's, and RM achieving a higher correlation in two out of three cases. Specifically, at the 173.9°W, 15°S location (in b), both corrections occasionally degrade the correlation metric, possibly due to the introduction of noise to the SMOS SSS (but see the discussion on the significance degree of correlation changes in 1) above). These four pixels are specifically highlighted, but a broader comparison of PM and RM across the entire region is discussed below.

Upon examining the RFI source vicinity, PM significantly mitigates the RFI contamination, similar to the RM method (Figure 8d). Both PM and RM effectively eradicate major RFI disruptions in the SMOS SSS, thereby restoring variability consistent with that of ISAS (Figure 8a).

The effectiveness of PM across the entire region, as compared to ISAS, mirrors that of RM (Figure 10d,e). However, Figure 10f reveals some differences between RM and PM outcomes in the region. Considering ISAS as the accurate SSS reference, we observe that: 1) RM is more effective near the Samoa islands and the source, within a 200-300 km range, as evidenced by (slightly) better std scores for RM compared to PM; 2) In contrast, PM avoids the widespread (minor) degradation observed with RM far from the source, around 1000 km away; 3) PM also shows greater efficiency near the Fiji Islands, west-southwest of Samoa. These observations suggest PM's unique advantage of operating independently at each pixel, allowing for a distinct RFI characteristic time series $u_1(t)$ at each location, unlike RM, which uses a single series $U_1(t)$ for the entire region.

B. Barbados region

Due to the scarcity of mooring data in the Barbados and Guinea Gulf regions, comparisons are exclusively made with the ISAS observational field. Additionally, to maintain brevity in the article, evaluation illustrations are limited to general performance maps of the methods, as exemplified in Figure 10c,e,f for the Samoa region.

In the Barbados region, a significant RFI source was detected at approximately 13.23°N, 59.55°W, active from mid-2011 to the end of 2014. The RM PCA extracts a characteristic time series that correlates almost perfectly with the RFI probability, even though the variance of the first mode is only moderately dominant (Figure 13a). Note that in this case, the RFI probability is high (≥ 0.9) yet does not reach full saturation at a value of 1, contrasting with the situation in the Samoa region as depicted in Figure 6. As a result, the first PC effectively captures the variations in probability.

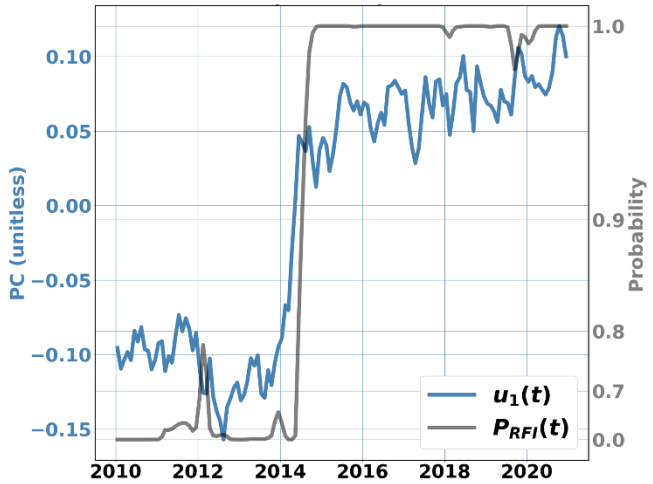

Figure 11: (Blue) Time series $u_1(t)$ (see text) associated to 1st mode of the PCA of $\Delta\dot{s}^f$ at the 8°S, 170°W location. (Grey) RFI probability time series for a location near 13.9°S, 171.7°W which is identified as the RFI source location in this region. The probability axis undergoes a transformation by raising it to the power of 6 to highlight its similarity more effectively with the PC.

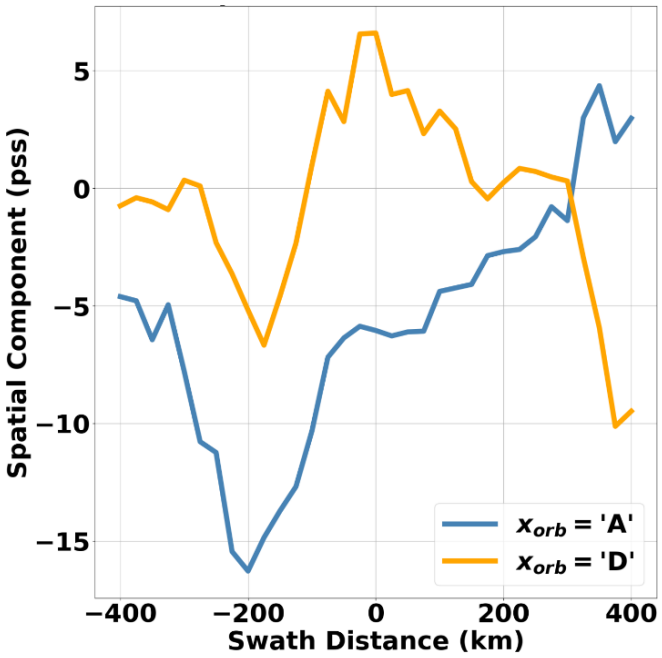

Figure 12: Projection $w_1(x_{\text{swath}})$ of the 2-dimensional salinity dataset $\dot{s}^f$ onto the RFI characteristic time series $u_1(t)$ at the location 8°S, 170°W. (Blue) Ascending orbit. (Orange) Descending orbit.

In pixels near the RFI source, as indicated in Figure 9e,f, and whose locations are shown in Figure 14a-c, both RM and PM significantly improve the reliability of SMOS SSS estimates. Notably, the substantial disruptions observed in the initial SMOS SSS series from 2011 to 2014 are eliminated following the application of these corrections, with the effect of aligning the corrected estimates closely with the ISAS series. This improvement is evidenced by a reduction in the

std of difference and an increase in correlation with ISAS, as detailed in TABLE IIIe,f.

The RM correction effectively mitigates the RFI contamination originating from the Barbados RFI source and extending along diverging RFI branches (Figure 14a). It seems as though some of these branches extend as far as the regions off French Guiana and the mouth of the Amazon River, a topic that is addressed from a more general viewpoint in the following paragraphs.

PM appears to improve the reliability of the SMOS SSS field over a wider area compared to RM (Figure 14b,c). However, these findings warrant careful interpretation: 1) Near the Barbados RFI source, RM and PM show similar effects, with RM performing better at specific pixels, a pattern also noted near the Samoa source; 2) PM effectively corrects for RFI around the islands of the Dominican Republic and Puerto Rico, northwest of Barbados, areas clearly influenced by other RFI sources. This corrective feature is not present in RM. 3) PM brings a significant portion of the SMOS SSS field, extending east of Barbados from about 20°N to the mouth of the Amazon River, into closer agreement with ISAS data. This pattern, reaching to the eastern boundary of our study, does not seem related to RFI contamination and is not present in RM results, except in the regions off French Guiana and the mouth of the Amazon River.

We conducted a more in-depth investigation, notably incorporating a Soil Moisture Active Passive (SMAP) L3 dataset into our analysis (refer to Appendix B). Our findings suggest that PM does not improve the accuracy of SMOS SSS measurements in this highly dynamic region, which is significantly influenced by the Amazon River discharge. The reasoning behind this is that PM inadvertently aligns SMOS SSS more closely with ISAS. However, considering that ISAS may not be a dependable benchmark in this region, it implies that PM might actually lead to a systematic deterioration in this area.

C. Guinea Gulf region

In the Guinea Gulf region, a strong RFI source was detected near 3°W, 5°N, which was intermittently activated at least four times for brief periods (Figure 13b). Another RFI source, located approximately 8° to the east, was constant but had a lower intensity level (not shown), and its impact was not as significant as the former. Utilizing RM, PCA generates a characteristic time series with a notable correlation to the RFI probability, although this correlation is not as strong as observed in the Samoa and Barbados regions (Figure 13b). Moreover, the variance percentage attributed to the first mode is marginally dominant.

However, as evidenced in selected pixels near the principal RFI source (Figure 14d-f), significant mitigation of RFI contamination is achieved following the application of either RM or PM, particularly from 2018 onwards (as shown in Figure 9g,h). In terms of the std of differences, RM outperforms PM slightly (TABLE IIIg,h). Yet, while PM significantly improves correlation in one of the two instances (i.e. g), the level of corrected SSS correlation is either diminished or remains marginal. Correlation is less reliable than the standard deviation of differences when dealing with noise, outliers, or uncertain reference data (see discussion). Therefore, reducing the standard deviation of differences is the preferred result of correction processes, even though any improvement in correlation is also beneficial.

From a broader perspective, RFI contamination associated with branches diverging from the RFI source is diminished, resulting in a more reliable SMOS SSS field when compared to ISAS (Figure 14d,e). Like observed in Samoa, in the Barbados region, PM provides more extensive corrections than RM, especially at locations distant from the RFI source at 5°N, 3°W (Figure 14f). Consistent with the patterns observed in both Samoa and Barbados, RM more effectively mitigates RFI contamination in areas close to the targeted RFI source. Conversely, PM effectively reduces RFI signatures around a

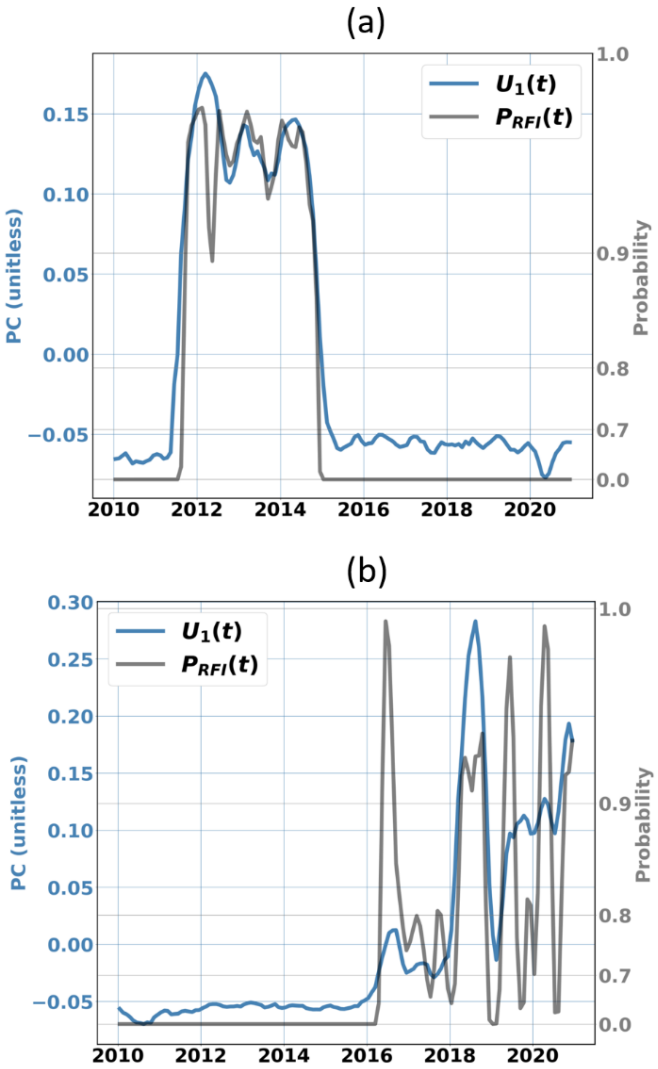

Figure 13: Same as Figure 6, but for: (a) The Barbados region and the 150km-600km annular region centered on 13.23°N,59.55°W; 1st mode variance percentage is 22%, 2nd mode, 6%: correlation of $U_1(t)$ with $P_{RFI}(t)$ is 0.99. (b) The Guinea Gulf region and the 100km-500km annular region centered on 5°N,3°W; 1st mode variance percentage is 17%, 2nd mode, 8%; correlation of $U_1(t)$ with $P_{RFI}(t)$ is 0.79.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2024.3408049

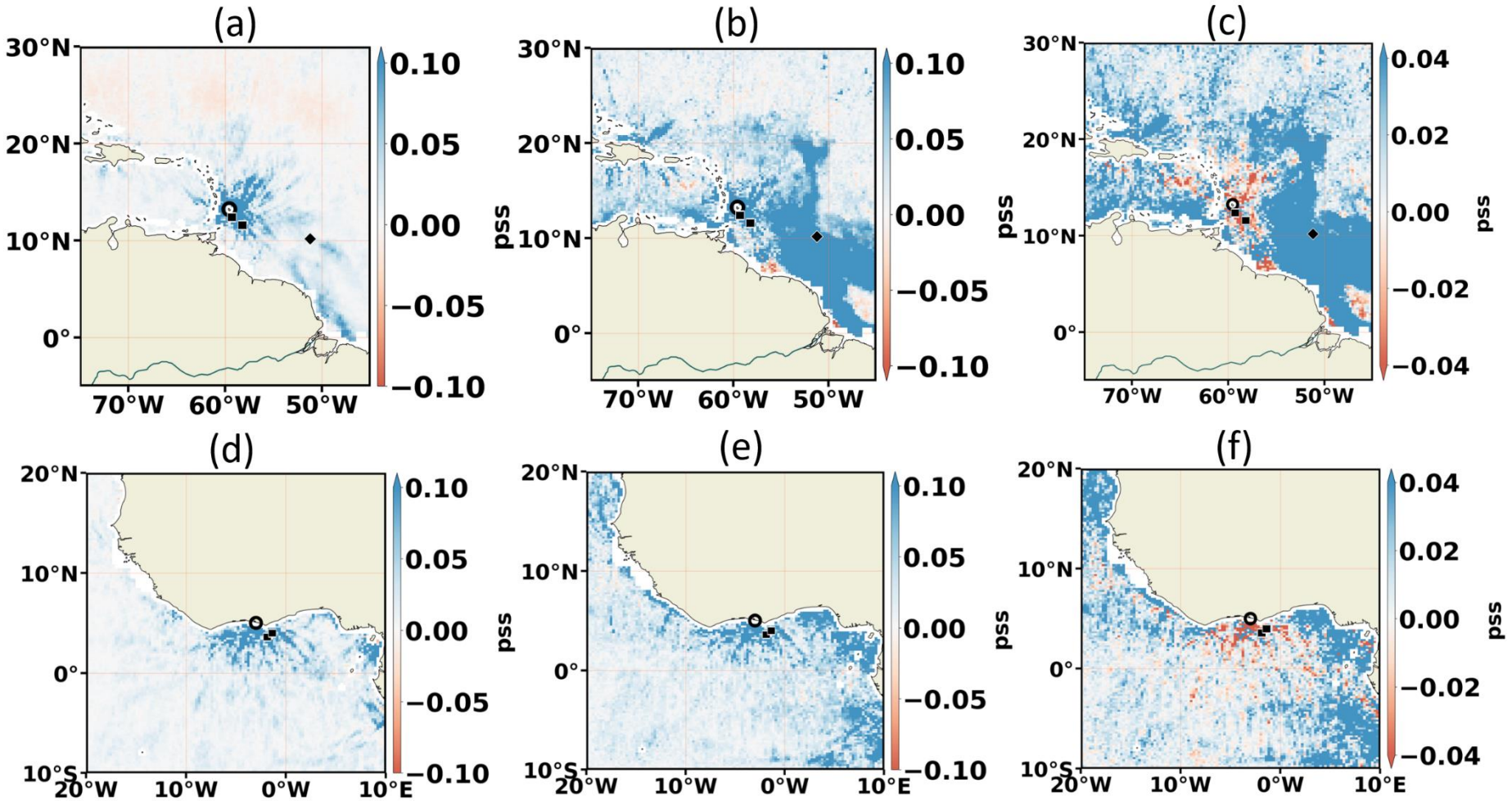

Figure 14: Same as Figure 10c,e,f, but for: (a), (b), (c) the Barbados region, and (d), (e), (f) the Guinea Gulf region. Hence, (a) and (d) show the metric $\Upsilon - \Upsilon_c$ (see TABLE II), (b) and (e) $\gamma - \gamma_c$, and (c) and (f) $\Upsilon_c - \gamma_c$. RFI source (black circle) and other locations for time series (black squares) are indicated (see TABLE I and TABLE III).

secondary RFI source located approximately 8° to the east.

Additionally, broad areas showing SSS modifications through PM are visible to the south, off the coasts of Cameroon, Equatorial Guinea, and Gabon, and to the north, from the coasts of Liberia to Mauritania. These features, which we do not immediately associate with improvement, may arise from causes such as those observed in the expansive correction areas in the Barbados region (see discussion about the reliability of ISAS in areas of strong SSS variability).

V. Discussion

We address several discussion points concerning the comparative advantages and disadvantages of the two methods outlined in the preceding sections.

On the one hand, RM can correct RFI contamination across a wide area, effectively restoring SSS variability that aligns with ISAS reference data. However, its application requires prior knowledge of the RFI source location, and assumes there is only one RFI source in the treated region. On the other hand, PM independently corrects RFI contamination at each location without any prior information about the RFI sources or affected regions. Nevertheless, PM may not be as robust as RM since it depends solely on cross-swath SSS variations and is more sensitive to uncertainties introduced from the gap-filling procedure and inherent noise in SMOS SSS data. RM demonstrates greater robustness near the targeted RFI source, as its approach to determining the RFI time series includes data from numerous locations around the source, alongside cross-swath variations at each pixel.

To reconcile these aspects, a potential solution could involve an as-yet-unimplemented hybrid method that combines the strengths of RM and PM. Like PM, this method would process each grid pixel (ϕ, λ) independently, but would also consider a surrounding area similar to RM to calculate a characteristic RFI time series and a spatio-temporal correction term, which would ultimately be applied solely to the (ϕ, λ) pixel. A circular area with a 400-km radius, encompassing the width of an SMOS swath, would serve as a suitable dimension for this "sliding region."

Both methods operate under the assumption that the RFI signature constitutes the primary mode of variability in the cross-swath SSS differences, as detailed in (5) and (6), and that this signature can be distinguished by a characteristic time series correlating with RFI probability. However, this assumption may not hold in certain regions or during periods where RFI contamination is either weak or sporadic, or where geophysical variability is particularly strong or complex within SMOS sub-cycle duration (18 days). For instance, Figure 15 illustrates variations across three regions of similar dimensions, demonstrating that the correlation between the PM-generated time series and RFI probability does not consistently exhibit the same level of coherence, making interpretation challenging.

Specifically, in the Samoa region (Figure 15a), despite the presence of multiple RFI sources (Samoa, Fiji, Gilbert Islands), areas of high correlation (> 0.8) are relatively well-defined. In contrast, within the Barbados and Guinea Gulf regions, where RFI contamination occurs more sporadically or with shorter durations, such pattern is not evident (Figure

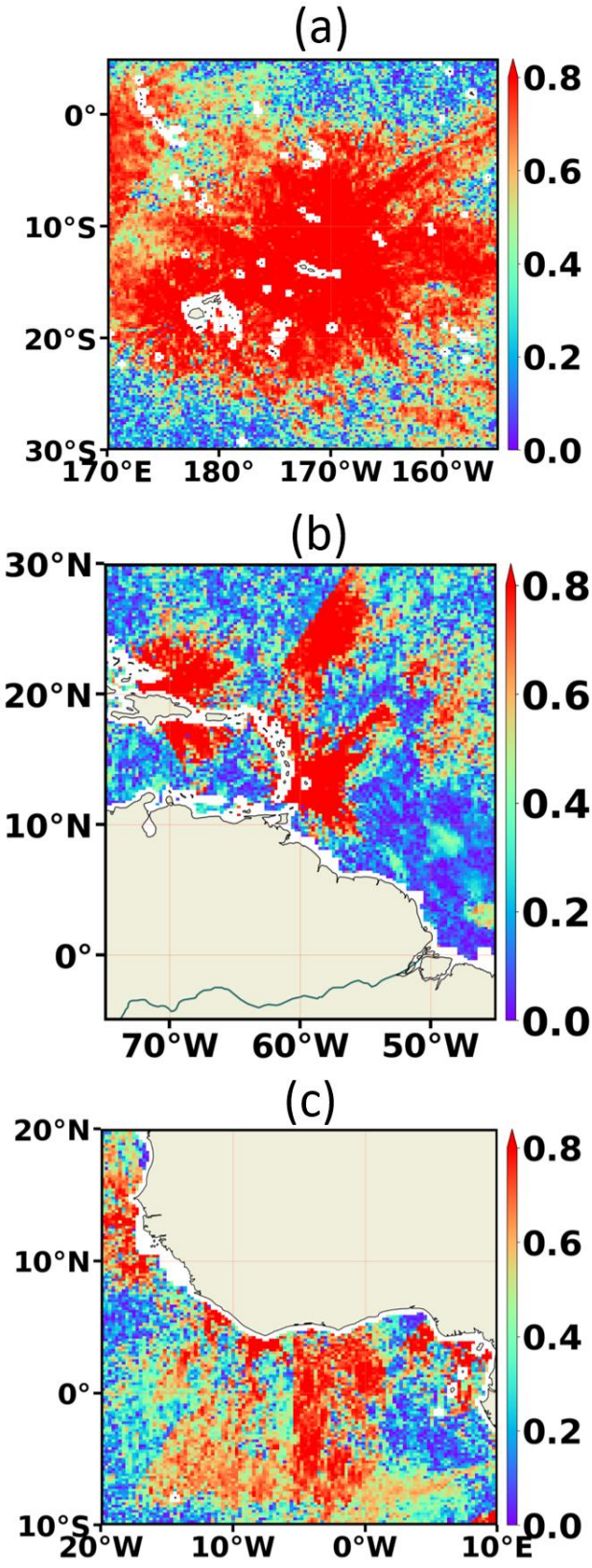

Figure 15: Absolute correlation value between $u_1(t)$, the PCA time series obtained from the first mode in the PM method and calculated at each pixel, and $P_{RFI}(t)$, the RFI probability time series from the closest identified RFI source.

15b,c). Notably, the region to the east of the Barbados RFI source, depicted in Figure 14b, exhibits negligible correlation in Figure 15b. Conversely, RFI sources along the African coast, west-northwest of the RFI source at 5°N,3°W, are responsible for the positive correction observed with PM (as shown in Figure 14e and Figure 15c). In this context, the implementation of a hybrid method, as proposed earlier, might improve coherence.

One significant challenge involves developing an intermediary step to determine whether to apply the correction. The primary obstacle is the absence of a completely reliable observational reference dataset for method training and evaluation. Nevertheless, treating ISAS as a potentially fully reliable benchmark, we experimented with a procedure using PM. Instead of relying solely on the first PCA mode for correction, we considered using the first n modes for multiple regression, where n could be zero if no correction is determined to be necessary, and $n \geq 1$ if it leads to improved results. The selection of n at each grid pixel was based on a comparison with ISAS data. This approach produced a map indicating integer values for n (not presented here). As expected, the "optimized" n was zero at many pixels where RFI contamination appeared absent. However, attempts to interpret n values across regions were not successful, and using only $n = 1$ emerged as the optimal compromise since using $n \geq 2$ might suppress some short-term SSS variability.

The use of the ISAS dataset as a reference for validation, which relies on sparse in situ observations, warrants further discussion. Being derived from Argo float measurements, the ISAS dataset may offer limited spatial and temporal coverage, especially in highly dynamic oceanic areas like coastal zones and river mouths (e.g. discussed in [29]). Moreover, it undergoes interpolation and smoothing processes that might introduce biases or inaccuracies into the SSS estimates. For example, ISAS might not accurately capture the sharp gradients or mesoscale features identifiable by SMOS. This consideration is particularly relevant for some correction features produced by the PM approach, such as those observed in the eastern Barbados region. Hence, in this specific area of significant variability, we also employed SMAP SSS as a benchmark for testing. Indeed, as demonstrated by [30], in highly variable regions, such as river plumes, SMAP and SMOS exhibit similar SSS variability with greater consistency compared to Argo interpolated products.

We evaluated our methods across three regions located within the tropical ocean's warm waters. In fact, the L-band radiometric signal-to-noise ratio is more favorable in warm waters, resulting in lower uncertainty in SMOS SSS measurements compared to those in colder regions. Future work should aim to assess the effectiveness of these methods in colder waters, which are also known to be affected by numerous areas of RFI contamination.

VI. Conclusion

Leveraging the SSS observations across different swath distances and instrument geometries, we demonstrated the feasibility of correcting specific RFI contaminations. We employed principal component and regression analyses to adjust the SMOS L2 SSS data for RFI contamination across various regions. Our methods focused on identifying and eliminating RFI signatures in both time and space, utilizing a characteristic time series derived from the PCA of the differential SSS field measured under different instrumental configurations. We explored two approaches for deriving and applying this characteristic time series: a Regional Method RM and a Pointwise Method PM.

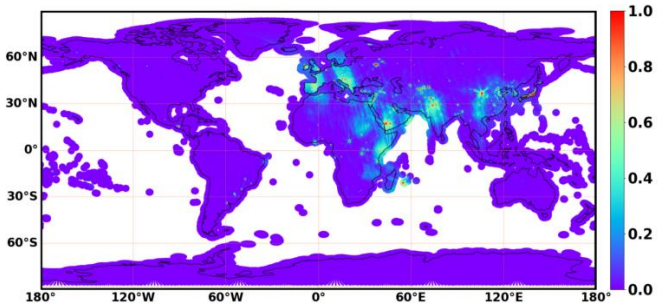

Figure 16: RFI probability map for ascending orbits on 2018, May 11th .

These methods were tested in three selected regions known for active RFI sources impacting SMOS SSS data quality: Samoa, Barbados, and the Guinea Gulf. We evaluated the methods by comparing the corrected SMOS SSS data against the in situ gridded datasets ISAS and TAO mooring time series.

The methods proved effective in mitigating the impact of RFI and improving the accuracy and variability of SMOS SSS, particularly in regions with high and consistent RFI probability. However, they also uncovered limitations, including dependence on precise RFI source location and timing, sensitivity to the specific region treated, and potential interference from other sources of uncertainty.

While RM was better at preserving short-term SSS variability, PM performed better in cases of contamination by several RFI sources. Future work should concentrate on finding a tradeoff between both methods.

The RM method has been implemented for generating version 4 of CCI SSS dataset, and both RM and PM have served as a starting point for expanding and systematizing RFI correction for SMOS SSS data.

Appendix

A. CESBIO probability dataset

The probability dataset consists of global daily maps, divided into ascending and descending orbits. This is to account for the SMOS satellite's forward-tilted attitude, which receives more incoming RFI signal from afar in the front than in the back. These maps display the probabilities of detecting land RFI occurrences, integrated over a seven-day sliding window (±3 days around the map's date), providing insights into RFI occurrences rather than their magnitudes. The computation method is based on an algorithm derived from detection processes, identifying L1c BT data contaminated by RFI that are excluded in the L2SM

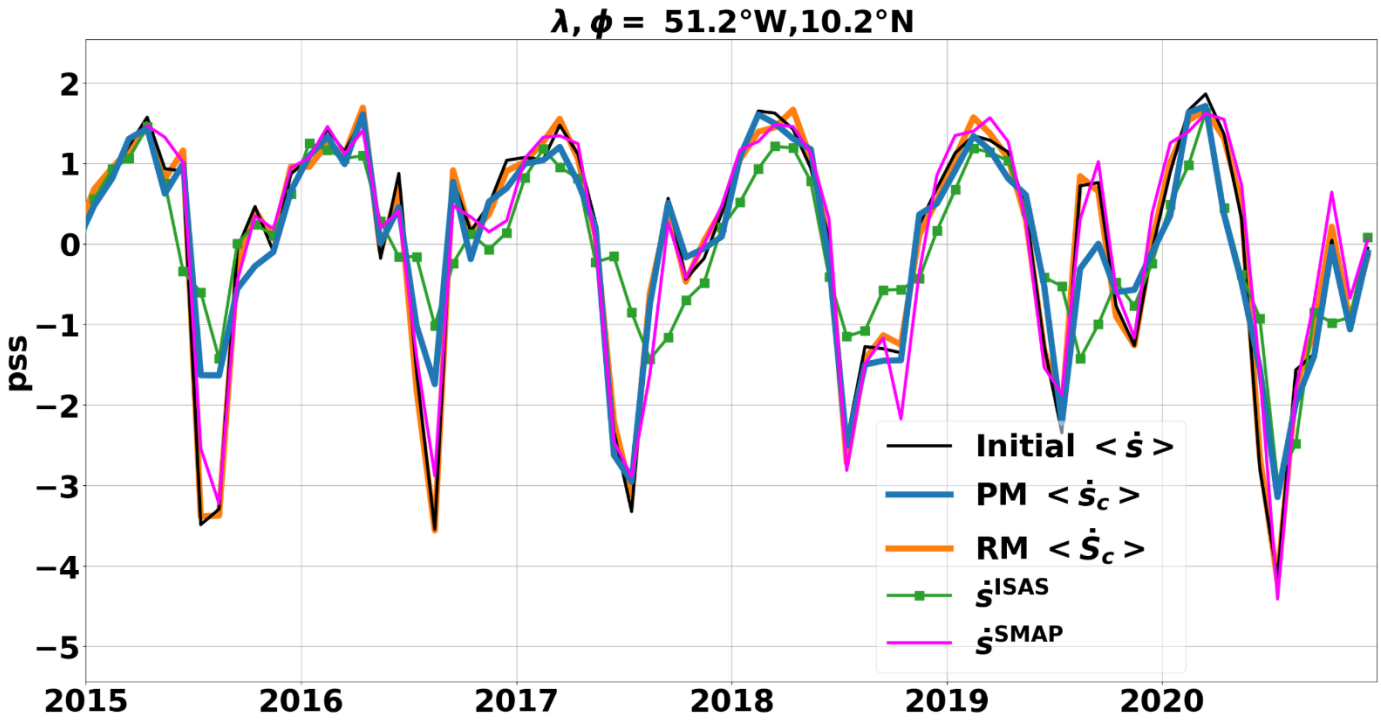

Figure 17: Time series of SMOS, SMAP, and observational SSS fluctuation, within the Barbados region at the location 10.2°N, 51.2°W. Time axis limited to 2015-2020 to align with the SMAP period and enhance visibility of discrepancies during fresh or salty peaks. (Black) Initial SMOS $\langle \dot{S} \rangle$. (Blue) PM corrected SMOS $\langle \dot{s}_c \rangle$. (Orange) RM corrected SMOS $\langle \dot{S}_c \rangle$. (Green) ISAS $\dot{S}^{ISAS}$. (Red) TAO mooring $\dot{S}^{MOOR}$. (Magenta) SMAP $\dot{S}^{SMAP}$.

UDP/DAP products [15]. Probabilities are calculated as the ratio of the number of BT detected as contaminated by RFI to the total number of BT observed over the seven-day period for each land grid-point on the SMOS map.

BT data are flagged as contaminated by RFI if either:

1) Their magnitudes fall outside the natural physical range, being either too low or too high, thus indicating the presenceof strong RFI sources; this applies to both co-polarized and cross-polarized measurements.

2) Their magnitudes lie within the physical range but exhibit an anomalous directional signature in the incidence angle of the Stokes-1 parameter, suggesting a lower level of contamination (a few to tens of K) from distant RFI sources, with signals propagating through the synthetic antenna sidelobes.

These detections rely on thresholds set to account optimistically for natural radiometric noise variances, with a preference for acceptance over rejection. Therefore, when BT data are detected and excluded, it is highly probable they are indeed contaminated. However, BT data that do not trigger these detection thresholds may still be contaminated. Consequently, the RFI situational awareness presented in these maps should be considered a best-case scenario, with the actual level of contamination potentially being more severe.

Localized sources are primarily detected through the criteria in item 1) above and appear as dots on the maps, while large, spatially diffuse contaminated areas are identified through the criteria in item 2) above. A probability of 0 indicates that no BT was detected as contaminated (Figure 16). This does not necessarily imply the absence of contamination but rather that any existing contaminations were below the detection threshold. A probability of 0.1 means that 10% of the BT observations were detected as contaminated. As the probability level increases to 0.15, 0.25, 0.5, and 0.95, the frequency of detected and removed BT correspondingly rises. Ultimately, a probability of 1.0 signifies that no BT observations passed the detection process, and therefore, no retrieval was attempted. This scenario is typically observed in the immediate vicinity of permanent RFI sources emitting continuous strong signals.

B. Comparison with SMAP SSS at a non-RFI location near Barbados

Launched by NASA in January 2015, the SMAP mission is an Earth-observing satellite mission. Like SMOS, it measures BT over the ocean to derive SSS. The SMAP L3 processing by Remote Sensing Systems (RSS) generates a smoothed SSS product with an approximate spatial resolution of 70 km. This smoothing is achieved by averaging the higher-resolution product, leading to reduced noise. Consequently, this 70 km resolution product is considered the standard for scientific applications and is appropriate for studies of both open ocean and coastal areas. The monthly-averaged version of this product is used in our analysis, with the time series fluctuations over the period from April 2015 to December 2020 denoted by $\dot{s}^{\mathrm{SMAP}}(t)$.

In this appendix, we explore SSS variations in the Barbados region, focusing on the area east of the RFI source where PM adjustments bring SMOS SSS measurements closer to ISAS (Section IV.B. and Figure 14b). Through analysis of time

series at a specific location (Figure 17, corresponding to the diamond in Figure 14b), we observe that SMOS and SMAP, providing entirely independent measurements, exhibit excellent agreement in this RFI-free zone, which is marked by strong seasonal SSS variability. This variability is notably influenced by the proximity to the Amazon mouth, featuring peak-to-peak amplitudes of around 5 pss or more. In contrast, ISAS data show a significantly attenuated seasonal amplitude, particularly regarding the cyclical freshening during the Northern Hemisphere summer. Interestingly, in late summer 2019, both satellites recorded a salty peak, again with very similar values, an unusual event that is completely absent in the ISAS data. Such discrepancies suggest that ISAS may not serve as a reliable reference in this specific region. Furthermore, it indicates that it may be preferable not to adjust SMOS SSS using PM (or RM) corrections in this context. The inadvertent smoothing of pronounced seasonal freshening episodes in the corrected SMOS SSS by PM, thereby aligning it more closely with ISAS, brings to light the necessity of an automated decision-making process concerning whether to apply these correction schemes. This issue is further discussed in Section V.

ACKNOWLEDGEMENT

This work was supported by the European Space Agency (ESA) CCI contract 4000123663/18/I-NB and by the French CNES/TOSCA SMOS-Ocean project.

REFERENCES

[1] G. Siedler, J. Gould, and J. A. Church, Eds., Ocean Circulation and Climate: Observing and Modelling the Global Ocean, (p. 715), vol. 103, 1st ed. Academic Press, 2001.

[2] P. J. Durack, S. E. Wijffels, and R. J. Matear, "Ocean salinities reveal strong global water cycle intensification during 1950 to 2000," Science, vol. 336, no. 6080, pp. 455–458, 2012. https://doi.org/10.1126/science.1212222.

[3] T. Röthig, S. M. Trevathan-Tackett, C. R. Voolstra, C. Ross, S. Chaffron, P. J. Durack, L. M. Warmuth, and M. Sweet, "Human-induced salinity changes impact marine organisms and ecosystems," Global Change Biology, vol. 29, pp. 4731–4749, 2023, doi: 10.1111/gcb.16859.

[4] N. Reul, S. A. Grodsky, M. Arias, J. Boutin, R. Catany, B. Chapron, et al., "Sea surface salinity estimates from spaceborne L-band radiometers: An overview of the first decade of observation (2010–2019)," Remote Sensing of Environment, vol. 242, 111769, 2020.

[5] R. Oliva, E. Daganzo, P. Richaume, Y. Kerr, F. Cabot, Y. Soldo, E. Anterrieu, N. Reul, A. Gutierrez, J. Barbosa, G. Lopes, "Status of Radio Frequency Interference (RFI) in the 1400–1427MHz passive band based on six years of SMOS mission," Remote Sensing of Environment, vol. 180, pp. 64-75, 2016, doi: 10.1016/j.rse.2016.01.013.

[6] A. Hasson, T. Delcroix, J. Boutin, R. Dussin, and J. Ballabrera-Poy, "Analyzing the 2010–2011 La Niña signature in the tropical Pacific sea surface salinity using in situ data, SMOS observations, and a numerical simulation," Journal of Geophysical Research: Oceans, vol. 119, no. 6, pp. 3855-3867, 2014, doi: 10.1002/2013JC009388.

[7] J. Boutin, N. Martin, N. Kolodziejczyk, G. Reverdin, "Interannual anomalies of SMOS sea surface salinity," Remote Sensing of Environment, vol. 180, pp. 128-136, 2016, doi: 10.1016/j.rse.2016.02.053.

[8] Y. H. Kerr et al., "The SMOS Mission: New Tool for Monitoring Key Elements of the Global Water Cycle," in Proc. IEEE, vol. 98, no. 5, pp. 666-687, May 2010, doi: 10.1109/JPROC.2010.2043032

[9] I. Corbella, N. Duffo, M. Vall-llossera, A. Camps, and F. Torres, "The visibility function in interferometric aperture synthesis radiometry," in IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, pp. 1677-1682, Aug. 2004, doi: 10.1109/TGRS.2004.830641.

[10] B. Picard, "Télédétection de la surface terrestre par un radiomètre imageur à synthèse d'ouverture: principe de mesure, traitement des données interférométriques et méthodes de reconstruction régularisées," Ph.D. dissertation, Université de Versailles - St Quentin, Ecole Doctorale des Sciences de l'Environnement d'Ile de France, 2004.

[11] Y. Soldo, F. Cabot, A. Khazaal, M. Miernecki, E. Slominska, R. Fieuzal, and Y. H. Kerr, "Localization of RFI Sources for the SMOS Mission: A Means for Assessing SMOS Pointing Performances," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 2, pp. 617-627, Feb. 2015, doi: 10.1109/JSTARS.2014.2336988.

[12] J. Barbosa, A. Gutierrez, "SMOS L1 Processor L0 to L1a Data Processing Model," Code: SO-DS-DME-L1OP-0007, Issue: 2.24, DEIMOS Engenharia, 2018, https://earth.esa.int/eogateway/documents/20142/37627/SMOS-L1a-Data-Processing-Models.pdf

[13] Y. Soldo, A. Khazaal, F. Cabot, and Y. H. Kerr, "An RFI Index to Quantify the Contamination of SMOS Data by Radio-Frequency Interference," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 4, pp. 1577-1589, April 2016, doi: 10.1109/JSTARS.2015.2425542.

[14] A. Gutierrez, R. Castro, P. Vieira, G. Lopes, and J. Barbosa, "SMOS L1 Processor L1c Data Processing Model," Code: SO-DS-DME-L1OP-0009, Issue: 2.17, DEIMOS Engenharia, 2017, https://earth.esa.int/eogateway/documents/20142/37627/SMOS-L1c-Data-Processing-Models.pdf

[15] "Algorithm Theoretical Basis Document (ATBD) for the SMOS Level 2 Soil Moisture Processor Development Continuation Project," ESA No.: SO-TN-ARG-L2PP-0037, Issue: 4.0, ARGANS No.: ASC_SMPPD_037, Date: 9th September 2019. [Online]. Available: https://earth.esa.int/eogateway/documents/20142/37627/SMOS-L2-SM-ATBD.pdf

[16] R. Oliva et al., "SMOS Radio Frequency Interference Scenario: Status and Actions Taken to Improve the RFI Environment in the 1400–1427-MHz Passive Band," in IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 5, pp. 1427-1439, May 2012, doi: 10.1109/TGRS.2012.2182775.

[17] E. Uranga, Á. Llorente, J. González, A. de la Fuente, R. Oliva, Y. Soldo, and F. Jorge, "SMOS ESA RFI Monitoring and Information Tool: Lessons Learned," Remote Sensing, vol. 14, no. 21, p. 5387, Oct. 2022, doi: 10.3390/rs14215387.

[18] J. Boutin, N. Reul, J. Koehler, A. Martin, R. Catany, S. Guimbard, F. Rouffi, J. L. Vergely, M. Arias, M. Chakroun, G. Corato, V. Estella-Perez, A. Hasson, S. Josey, D. Khvorostyanov, N. Kolodziejczyk, J. Mignot, L. Olivier, G. Reverdin, D. Stammer, A. Supply, C. Thouvenin-Masson, A. Turiel, J. Vialard, P. Cipollini, C. Donlon, R. Sabia, and S. Mecklenburg, "Satellite-Based Sea Surface Salinity Designed for Ocean and Climate Studies," J. Geophys. Res. Oceans, vol. 126, no. 11, Art. no. e2021JC017676, Nov. 2021, doi: 10.1029/2021JC017676.

[19] J.-L. Vergely, J. Boutin, N. Kolodziejczyk, "SMOS OS Level 3 and Level 4. Algorithm Theoretical Basis Document," 2022. [Online]. Available: https://doi.org/10.13155/92266

[20] F. Gaillard, T. Reynaud, V. Thierry, N. Kolodziejczyk, and K. von Schuckmann, "In Situ–Based Reanalysis of the Global Ocean Temperature and Salinity with ISAS: Variability of the Heat Content and Steric Height," J. Climate, vol. 29, pp. 1305–1323, 2016, doi: 10.1175/JCLI-D-15-0028.1.

[21] D. Roemmich, M. H. Alford, H. Claustre, K. Johnson, B. King, J. Moum, et al., "On the future of argo: A global, full-depth, multi-disciplinary array," Frontiers in Marine Science, vol. 6, 2019, doi: 10.3389/fmars.2019.00439.

[22] M. J. McPhaden, "The Tropical Atmosphere Ocean Array Is Completed," Bull. Am. Meteorol. Soc., vol. 76, pp. 739–741, 1995.

[23] R. W. Preisendorfer and C. D. Mobley, Principal Component Analysis in Meteorology and Oceanography. Amsterdam: Elsevier, 1988.

[24] S. W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing, 2nd ed. California Technical Publishing, 1999. ISBN: 0-9660176-7-6, 0-9660176-4-1, 0-9660176-6-8.

[25] "Convolution and Filtering in python, "The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package"," The Astrophysical Journal, vol. 935, no. 2, 2022, doi: 10.3847/1538-4357/ac7c74.

[26] G. R. North, T. L. Bell, R. F. Cahalan, and F. J. Moeng, "Sampling Errors in the Estimation of Empirical Orthogonal Functions," Mon. Wea.

Rev., vol. 110, pp. 699–706, 1982, doi: 10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2.

[27] J. R. Brown, M. Lengaigne, B. R. Lintner, et al., "South Pacific Convergence Zone dynamics, variability and impacts in a changing climate," Nat Rev Earth Environ, vol. 1, pp. 530–543, 2020, doi: 10.1038/s43017-020-0078-2.

[28] Efron, B., "Nonparametric estimates of standard error: The jackknife the bootstrap and other methods", Biometrica, vol. 68, pp. 589-599, Dec. 1981.

[29] G. Reverdin, L. Olivier, C. Cabanes, J. Boutin, C. Thouvenin-Masson, J.-L. Vergely, N. Kolodziejczyk, V. Thierry, D. Khvorostyanov, and J. Jouanno, "Missing Argo float profiles in highly stratified waters of the Amazon river plume," J. Atmos. Oceanic Technol., 2024, doi: 10.1175/JTECH-D-23-0072.1.

[30] S. Fournier and T. Lee, "Seasonal and Interannual Variability of Sea Surface Salinity Near Major River Mouths of the World Ocean Inferred from Gridded Satellite and In Situ Salinity Products," Remote Sens., vol. 13, 728, 2021, doi: 10.3390/rs13040728.