# Comparing Satellite Salinity Retrievals with *In Situ* Measurements: A Recommendation for Aquarius and SMAP

Julian J Schanze[1] (jschanze@esr.org), David M Le Vine[2] (david.m.levine@nasa.gov), Emmanuel P Dinnat[2] (emmanuel.dinnat@nasa.gov), Hsun-Ying Kao[1] (hkao@esr.org)

[1] Earth & Space Research, Seattle, WA

[2] Goddard Space Flight Center, Greenbelt, MD

## Abstract

The validation of sea surface salinity (SSS) retrievals from satellite measurements through comparisons with *in situ* data is important for evaluating the retrieval, improving the underlying geophysical model functions, and producing reliable estimates of measurement uncertainty. The evaluation of *in situ* and satellite data is complicated by the different spatial and temporal characteristics of the two measurements: satellite observations from Aquarius and SMAP are continuous and global but with spatial resolution on the order of 50 – 100 km, while *in situ* data (e.g. from Argo floats) are point measurements in both space and time. Although satellite data are global, it may take several days to cover the entire globe. Hence, the probability of an exact match in time and space between an *in situ* measurement and a satellite overpass is small. Here, we define a meaningful definition of a matchup between satellite and *in situ* measurements, primarily through the example of Aquarius/SAC-D, SMAP, and Argo float data. **The analysis suggests that useful matchups between satellite and *in situ* salinity are obtained by averaging all satellite observations within a space-time window with a radius of 50 km and ± 3.5 days centered on the *in situ* report.** This document is partitioned into two parts: In Section I, the background of satellite salinity remote sensing is used to develop a rationale for the chosen match-up criteria, while Section II explores examples using Aquarius/SAC-D and SMAP in sufficient detail to generate a match-up algorithm.

## 1. Background

### 1.1. Remote Sensing of Sea Surface Salinity from Space

Ocean salinity is a crucial variable in understanding ocean circulation and the global hydrological cycle, as 87% of evaporation and 78% of precipitation occur over the ocean (Schanze et al., 2010). Recently, sensors such as SMOS (Kerr et al, 2010), deployed by the European Space Agency (ESA), as well as Aquarius (Le Vine et al, 2007) and SMAP (Entekhabi et al, 2010), deployed by the National Aeronautics and Space Administration (NASA), have demonstrated the feasibility of measuring sea surface salinity (SSS) from space (e.g. Dinnat et al, 2019; Le Vine et al, 2018). These sensors operate in the spectral window around 1.413 GHz

(L-band), which is protected for passive use only. This window is near the narrow peak in the response of microwave radiometers to ocean at tropical and sub-tropical temperatures (Le Vine et al, 2010). The response is strong enough to make remote sensing possible, but because of the low frequency and limitations on antenna size for sensors in space, the instruments have spatial resolution on the order of 40-100 km. The coverage of these instruments is global with revisit times from 3 – 7 days. Validation of the satellite observations is an important part of documenting the reliability of the salinity product and testing the algorithm to find paths for improvement. However, comparing *in situ* and satellite data to obtain meaningful statistics is complicated by the very different spatial and temporal characteristics of the two measurements. Research is reported in the following sections that addresses the issue of defining a match-up between satellite and *in situ* measurements that can be useful for evaluating the satellite salinity product.

### 1.2. *In Situ* Measurements

Several sources of *in situ* measurements are available for the validation of salinity retrievals from satellite measurements. These include Argo floats (Fig. 1), thermosalinograph observations from research vessels and ships of opportunity, sea surface salinity (Surface Velocity Program, SVP) drifters, conductivity-temperature-depth (CTD) stations from research vessels, oceanographic moorings, as well as marine mammals instrumented with specialized measuring equipment.

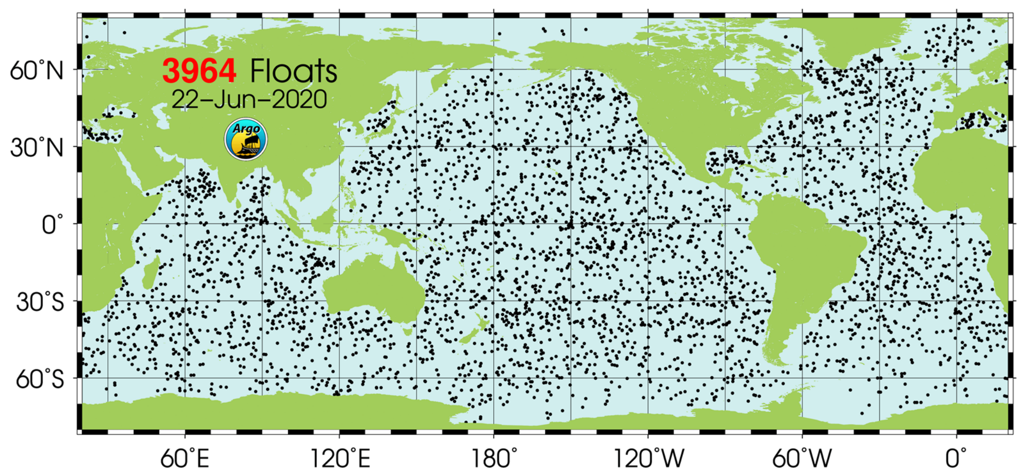



*Figure 1: Current distribution of Argo floats. The nominal coverage is 2.5° in latitude and longitude, from the UCSD Argo site* [http://www.argo.ucsd.edu/](http://www.argo.ucsd.edu/)

This sampling is neither homogeneous in space nor time, which is a potential source of bias in the validation of salinity satellite measurements. Most platforms do not sample salinity at the surface but instead sample at depths of approximately 1-10 m (Fig. 2), while L-Band retrievals integrate the uppermost 1-2 cm of the ocean (Swift, 1980). This difference in depth of the measurement is the vertical component of the 'representation' error. The horizontal component of the representation error is sub-footprint variability, which is the variability of salinity due to fronts, patches, filaments, and other gradients in an area that is equivalent in size to a satellite instrument footprint.

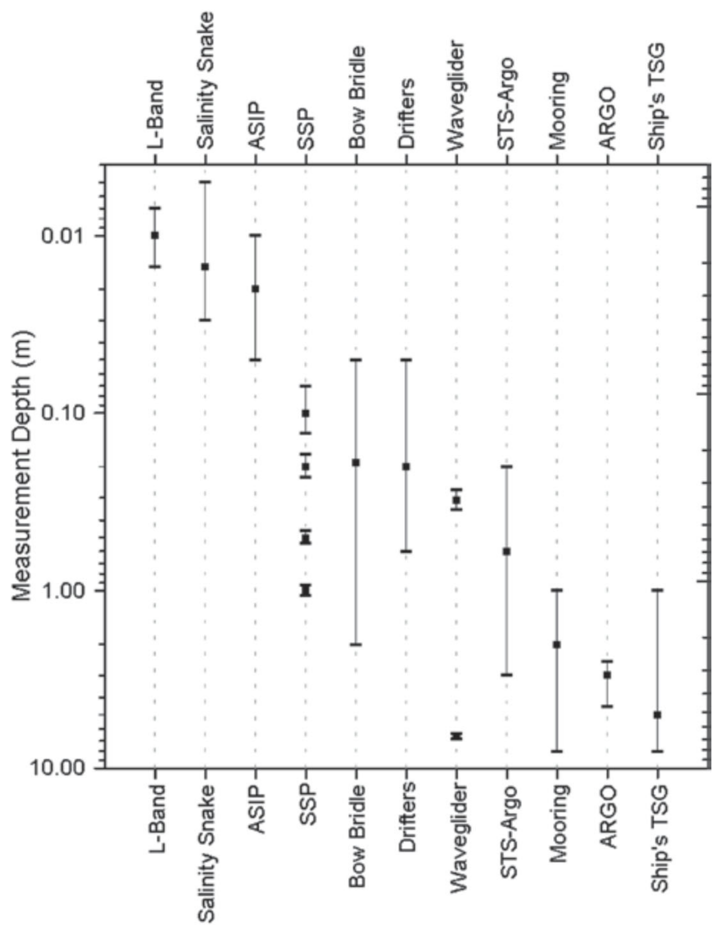



*Figure 2: Typical sampling depth of several in situ salinity observing platforms. The most commonly used platforms have observing depths between 1-10 m. Listed instruments from left to right: L-band satellite retrievals, salinity snake instrument, Air Sea Interaction Profiler (ASIP), Surface Salinity Profiler (SSP), Bow Bridle instrument, Drifters such as those used in the SVP, Waveglider, STS Argo floats with two salinity sensors, conventional Argo floats, and thermosalinograph (TSG) systems used on ships. From: Boutin et al. (2016).*

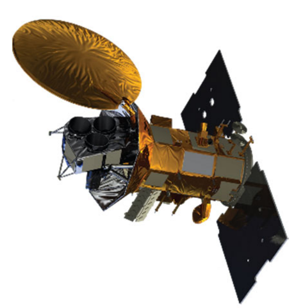
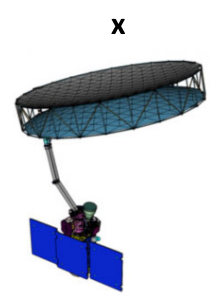
While it is possible to measure salinity directly using laboratory equipment, the platforms used in validation all use a combination of a conductivity cell and thermistor to measure conductivity and temperature, respectively. Salinity is then derived using an empirical formula, such as the Thermodynamic Equation Of Seawater - 2010 (TEOS-10). The problems associated with this measurement are two-fold: Biofouling of the conductivity cell or hysteresis in the pressure measurement can lead to long-term drifts, while differences in temperature between the seawater inside the cell and the thermistor cause erroneous salinity measurements. The temperature differences are a superposition between thermistor response time, thermal mass of both the thermistor and the conductivity cell, and the flow rate, which causes ‘spikes’ in the measurement. While the measurement error of *in situ* salinity measurements tends to be much smaller (generally less than ±0.01 g/kg) than that associated with satellite retrievals, it does exist and needs to be considered in a comparison. A unique sensor identification should thus be tracked throughout a matchup procedure to identify potential persistent biases.

When using *in situ* measurements for satellite validation, some of the points that need to be considered are:

- Inhomogeneities in spatial and temporal sampling must be addressed lest certain areas receive more weight in producing global statistics such as bias and root-mean-square difference (RMSD).
- Satellites and *in situ* platforms can be very accurate but produce different salinity measurements. Representation error is always present. The vertical component is caused by near-surface salinity stratification, while the horizontal representation error is caused by sub-footprint variability.
- In situ measurements, while generally more accurate than satellite retrievals, may still have systematic biases due to such things as biofouling or sensor hysteresis, or random errors due to mismatches in temperature and conductivity measurements or bubble intrusion.

### 1.3. Satellite Measurements

*Table 1: Overview of Passive L-Band Radiometers, with published instrument and orbital parameters*


| | SMOS (2009-now) | Aquarius/SAC-D (2011-2015) | SMAP (2015-now) |
|---|---|---|---|
| **Type** | Interferometer | 3-Feed Push-Broom | Conical Scan, 14.6 rpm, 40° |
| **Footprint** | 35-50 km | 85/100/122 km | 40 km |
| **Altitude** | 765 km | 657 km | 685 km |

| | | | |
|---|---|---|---|
| **Inclination** | 98.44° | 98.01° | 98.12° |
| **Swath** | 1000 km | 390 km | 1000 km |
| **Repeat** | 149 (18 sub) days | 7 days | 8 days |
| **Revisit** | 2-3 days | 7 days | 2-3 days |
| **Integ. time** | 1200 ms | 1440 ms | 16 ms |
| **NEDT** | 4 K | 0.12 K | 1.1 K |

Microwave radiometers on satellites operating in the protected spectral window at 1.413 GHz retrieve sea surface salinity (SSS) by measuring the thermal emission (brightness temperature) of the ocean. Changes in salinity change the conductivity of seawater which changes the emission (e.g. Le Vine et al, 2010),  but other factors also modulate the emission and contribute to the signal received by a sensor in space. These include sea surface roughness, sea surface temperature (SST), attenuation in the atmosphere, Faraday rotation in the ionosphere, and signal from the galaxy and Sun and Moon as well as contamination by ice and land appearing in the antenna sidelobes (Yueh, 2002; Font et al, 2004). To date, three satellites have been launched that measure sea surface salinity: ESA's Soil Moisture and Ocean Salinity (SMOS) mission (2009-present), NASA's Aquarius (2011-2015) instrument on the Aquarius/SAC-D satellite, and NASA's Soil Moisture Active Passive (SMAP) satellite (2011-present). Of these, Aquarius was the only instrument designed primarily to measure SSS, with soil moisture being the primary measurement goal of both SMOS and SMAP. Among the features included in the design of Aquarius to improve accuracy for remote sensing of salinity were careful thermal control, flying on the day-night terminator while looking away from the Sun, frequent sampling to help detect and remove RFI, and an exact repeat orbit to improve averaging to reduce noise. Spatial averaging was part of the plan to reduce noise. Aquarius reported retrieved SSS (e.g. in the L2 data files) each 1.44 sec. This integration time is still vastly oversampled compared to the time needed to move one along-track beam width (about 80 km or 11 s)

With an integration time of 16.8 ms, oversampling is also used in the SMAP retrieval compared to the size of the antenna footprint. Additional averaging is being done in the salinity product (Meissner et al, 2019) at Level 1B and beyond. In particular, in the case of the SMAP Remote Sensing Systems (RSS) V 4.0 product L1B data, the samples are first binned in integer angles (0-359 degrees) over each scan and then interpolated using the Backus-Gilbert optimal interpolation onto a 0.25° x 0.25° grid and separated into fore- and aft-looks (Meissner et al, 2019 Appendix A). Each 0.25° x 0.25° grid point is populated with samples of retrieved salinity. In the standard release, the '70-km' product, the 8 grid points surrounding any given point are combined with that point to produce an average salinity (i.e. the 9 values, comprised of the point and its nearest neighbors, are averaged). While this product has an approximate resolution of 70 km, it is presented on a 0.25° x 0.25° grid. Similarly, averaging is used to varying degrees in all SMOS products, as there is significant noise associated with the retrieval from the interferometric measurement.

Consequently, it may be tempting to think of Level 1B or 2 data as being independent measurements, while some form of averaging is used in virtually all retrieval algorithms to reduce noise. This raises the question of what information is conveyed by a matchup at these levels. In general, more averaging results is better matchups, despite discarding potentially useful resolution in areas of strong gradients. While still being a useful indication of measurement accuracy, matchups of single L1B/L2 satellite retrievals should not be compared to each other for these reasons. For a meaningful comparison between datasets, the same amount of temporal and spatial smoothing should be applied to the satellite data in the form of averaging or binning over a given range of space and time, which we aim to define here.

## 2. Definition of a Matchup

The fundamental problem in defining a matchup is that the *in situ* measurements are essentially point measurements, whereas satellite measurements are integrated over the antenna footprint. *In situ* data such as reported by Argo (which will be used here as the nominal example) are point measurements in time and (horizontal) space, whereas the satellite data are typically over-sampled with many values relatively close together in time. To accommodate the size difference, a matchup will be defined to occur whenever the antenna boresight is within a circle with radius, R, of the location of the *in situ* (i.e. Argo) report. Another way of looking at this is that a matchup occurs whenever an *in situ* report is within the footprint of the sensor.

Time is also another parameter that is different for the *in situ* and satellite measurements. The *in situ* data occur at distinct, usually well separated points in time (e.g. Argo reports) and even thought the satellite data cover the globe, the probability that a given *in situ* report will coincide in time with a satellite over pass is very small. To obtain a sample space large enough to be meaningful it is necessary to include in the definition of a matchup those cases where the satellite is within a time ±T of the time of the *in situ* report.

That is, a match up will consist of one *in situ* report and all the satellite observations within a space-time box defined by R and T. Computationally, this done by first sorting the *in situ* and satellite data by date. A given sub-range of both are loaded into memory, and for each *in situ* datum provided in that range, satellite data must be found that match the temporal and spatial criteria. Locating distances in time is straightforward, and while the concept of distance in space is relatively simple also, it is more computationally expensive because they are on a spheroid or geoid. After these two operations, a logical operation is performed to find all values of satellite data for which both temporal AND spatial constraints are fulfilled.

The value of the match up will depend on the size of the space-time box. Studies conducted during the planning for Aquarius (Lagerloef *et al.*, 2008) addressed the issue of space and time scales and although variability occurs on all scales it was concluded that significant ocean science could be accomplished with spatial scales on the order of 100 – 200 kilometers and time windows of weeks. The spatial scales that are appropriate for sensors in space are the antenna footprint (50 – 100 km) and the revisit time (7-8 days). The revisit time is the time until the satellite returns to the same position, which is suggestive of the time between independent samples. These values (e.g. R = 50 km and T = ±3.5 days) become the initial guess for the space-time box.

There is another time to be considered, which is the sampling rate of the satellite data. As mentioned, both Aquarius and SMAP report oversampled data. Because it is oversampled, it is noisy (or, at least, noisier than it could be). Hence, some averaging is needed. The possibilities of averaging are listed below. The goal of the matchup is to assess the accuracy of the satellite data using the *in situ* observation as a reference. Hence, the parameter of interest is the difference between *in situ* and satellite salinity and the possibilities are listed below as "salinity differences". The possibilities include:

- Average all points that match the space-time criteria (all salinity difference **ASD**)
- Use the closest point in space (single sample difference space, **SSDS**)
- Use the closest point in time (single sample difference time, **SSDT**)
- Use a weighted function to find a point that is close in both space and time (weighted single sample difference **WSSD**)

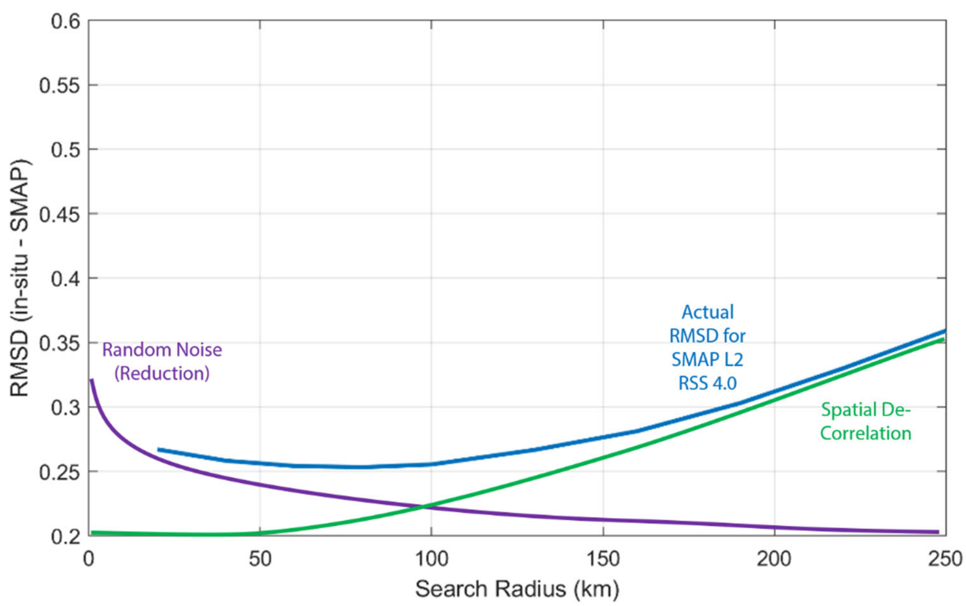



*Figure 3: Illustration of RMSD Noise Reduction (purple, for illustration only) and spatial de-correlation (green, estimated and for illustration only) along with the global RMSD for SMAP RSS L2 V4.0 at a fixed ΔT=±3.5 days and varying ΔS (x-axis). The optimal values of the search radius for noise reduction before spatial de-correlation becomes dominant are between 50-80km. While the blue line is the computed RMSD, the purple and green lines are a conceptual drawing to illustrate the competing processes and are not computed from data.*

- Average N closest points in space, time, or weighted (NSDS, NSDT, NSDW, etc)

Using the closest point in space (SSDS) or time (SSDT) is included in the list even though, as mentioned above, it is a problem for satellite sensors like Aquarius and SMAP, where the data is reported at a rate highly over-sampled compared to the time to move one footprint and the time required to achieve the intended radiometric accuracy. The problem is different with SMAP because averaging is built into the retrieval algorithm (Meissner et al, 2019), which leads to intrinsic problems in comparing these values.

## 2.1. Single Sample Difference

The single sample differences in space (SSDS) or time (SSDT) which matches one satellite measurement with one *in situ* observation is an obvious first choice for evaluation of the retrieved salinity. However, there are some issues with this choice. For example, they will have high standard deviations (STD) for Aquarius and SMOS because Level 2 salinity data is retrieved and reported at a highly oversampled rate. In the case of Aquarius, the radiometric noise in the 1.44 sec samples is high compared to the value needed to achieve an RMS accuracy on the order of 0.2 g/kg. In fact, the sensor was designed expecting an average on the order of 5 sec to reduce radiometric noise (Le Vine et al, 2007). These single sample differences may be informative as an indication of the noise, but they are not indicative of the accuracy intended in the design of the sensor. Either SSDT or SSDS yields an RMS value of about 0.65 g/kg for

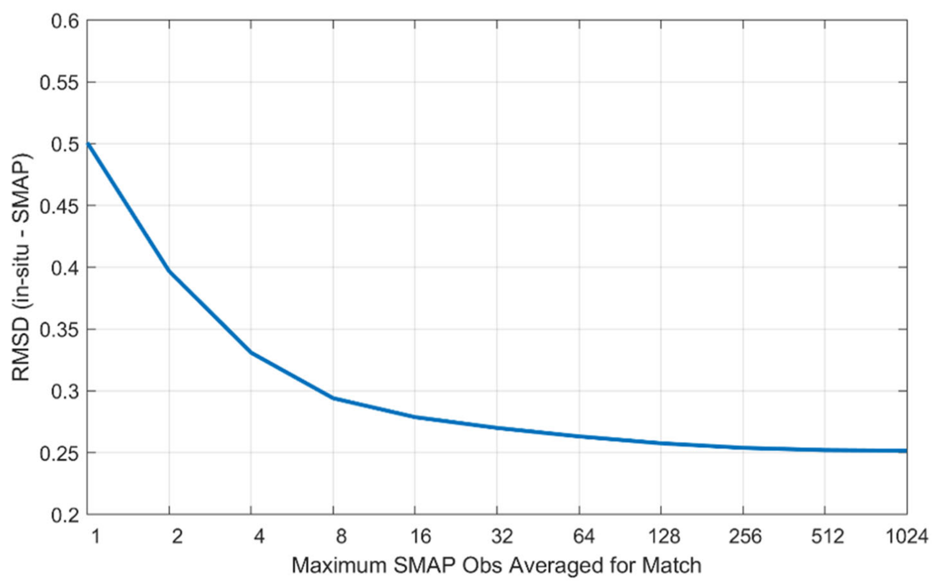



*Figure 4: Space-centric match-up with criteria ΔT<±3.5 days and ΔS<50km. The global RMSD for SMAP RSS L2 V4.0 is provided for the mean of the N closest points, where the case N=1 is equivalent to a single sample difference (SSDS).*

Aquarius. Some of the issues associated with using a single sample and the need for averaging are illustrated in the discussion below.

Here, we conduct an analysis using salinity from Remote Sensing System (RSS) SMAP SSS V4.0 (Meissner et al, 2019) at Level 2, using the official 70 km product and a space-time box with R = 50 km and T = ±3.5 days. The time window approximately matches the time required for Aquarius to cover the globe and is close to the repeat cycle of SMAP (8 days). The space window is consistent with an average Aquarius footprint and about twice as large the SMAP footprint, to insure that sufficient matches be found.

Fig. 4 reports the RMSD (vertical axis) as a function of the number of samples chosen (horizontal axis), starting with the closest in space and including the next closest and so on. The value “1” on the horizontal axis represents the single closest in space (i.e. the SSDS). The value “2” means that the closest and next closest sample are averaged in each space-time box and then the RMSD is computed for all such pairs at intervals of 2^N up to N=10. The RMSD for the single sample is high (0.5 g/kg) and decreases as more samples are included in each box. The decrease is rapid and slightly slower than it would be for independent samples. *Fig. 5* reports the same analysis but using points closest in time (i.e. SSDT). In this case, the RMSD for one sample is slightly higher than for SSDS and decays much more slowly. The latter is likely a reflection of the oversampling present in the satellite data (i.e. a sample closest in time will have nearest neighbors almost equally close in time for the same satellite pass but because of

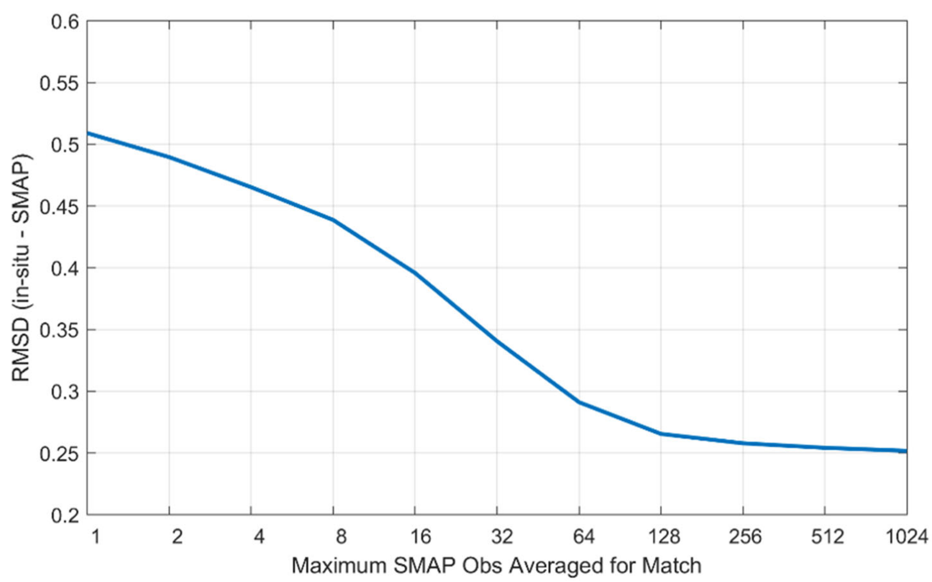



*Figure 5: Time-centric match-up with criteria ΔT<±3.5 days and ΔS<50km. The global RMSD for SMAP RSS L2 V4.0 is provided for the mean of the N closest points, where the case N=1 is equivalent to a single sample difference closest in time (SSDT).*

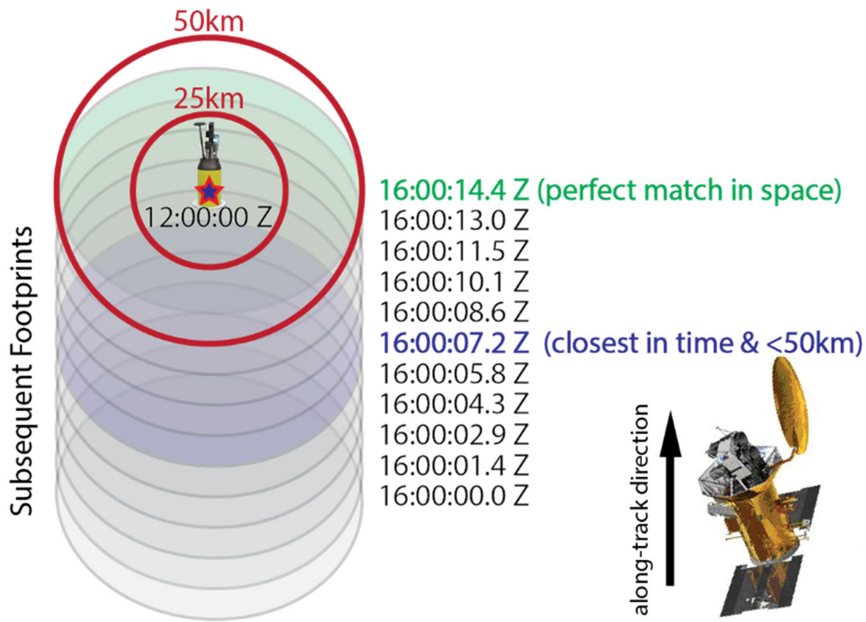



*Figure 6: Illustration of likely mismatches when using a time-centric approach to SSD matches. The ellipses indicate the 100x85km beam from Aquarius/SAC-D for 10 1.44s integrations as it approaches an* in situ *measurement, denoted by a star. The first 1.44s integration that matches the criteria ΔT<±3.5 days and ΔS<50km, while minimizing ΔT is indicated by a blue ellipse. This is the SSDT matchup. The green ellipse denotes a perfect match in space that occurs 7.2 s after this first match and would be chosen in a space-centric approach. This is the SSDS matchup.*

oversampling, they are not likely to be independent). The slightly higher RMSD for the single sample difference in time (Fig. 5) is likely a reflection of the fact that closest in time does not guarantee close in space (i.e. the satellite pass closest in time could be at the spatial edge of the space-time box). This problem is illustrated using a (fictitious) example in Fig. 6, in which an Argo measurement occurs at 12:00:00Z, followed by a satellite overpass approximately 4 hours later around 16:00:00Z. Using criteria of ΔT<±3.5 days and ΔS<50km, The SSDT matchup occurs at 16:00:07.2Z. While closest in time, it is maximally far away from the *in situ* observation. The SSDS matchup follows 7.2 s later and is a perfect match in space. While this example is somewhat contrived, a time-centric approach will always result in the selection of the farthest point in space unless the two measurement times coincide perfectly.

**2.2. Averaging**

The first issue to address in evaluating averaging is the size of the space-time box. The space-time box will determine the number of satellite observations available to be averaged. While the inclusion of more satellite estimates reduces random noise, de-correlation due to oceanic processes such as ocean currents, eddies, fronts, and general salinity gradients can lead to degradation of the estimate when large search windows are used in either time or space.

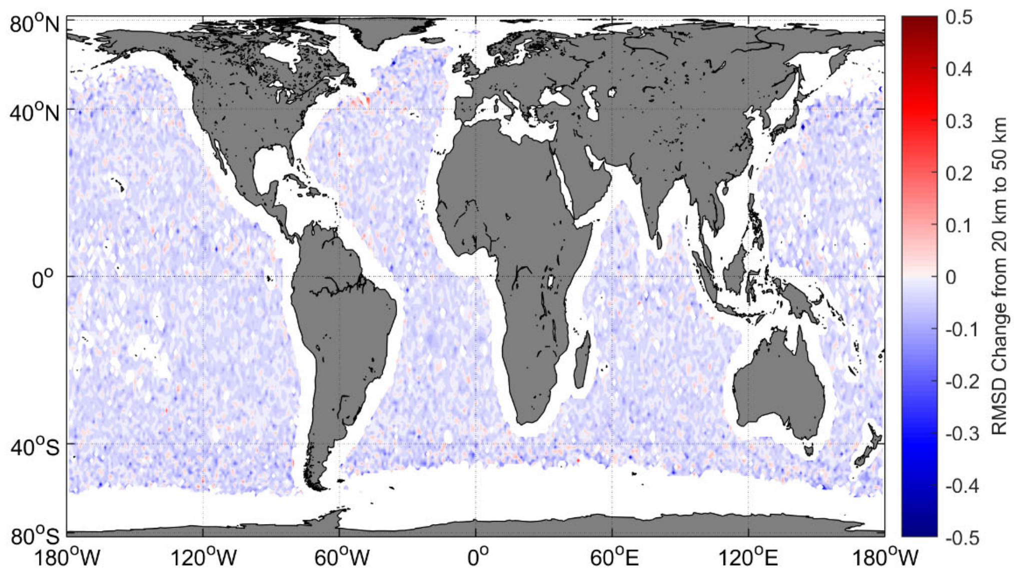



*Figure 7: Change in RMSD from a 20 km search radius to a 50 km search radius. With the exception of a small part of the Gulf Stream, the RMSD is reduced significantly.*

To address this issue, matchups were performed using SMAP RSS L2 V4.0 salinity data with a fixed time window of ±3.5 days, while the spatial search radius was varied from 20 km to 250 km. All values of satellite data in the space-time box were averaged to form a single salinity difference for each *in situ* data (Argo float). The results are shown in Fig. 3, which reports the RMSD on the vertical axis and the size of the box radius on the horizontal axis. Fig. 3 suggests that a range of 40-80 km is a reasonable choice that affords a minimum in RMSD and box size (i.e. to avoid spatial de-correlation in ocean salinity). To examine the effect of changing the search radius from 20 km to 50 km, a global map of change in RMSD is shown in Figure 7. From this, it is evident that the improvement in noise reduction is almost universal, with the exception of a small part of the Gulf Stream.

A similar parameter search was performed for the temporal search window, ranging from ±1 day to ±7 days in increments of ±0.5 days. The results are shown in Fig. 8. Although there is no clear minimum and the RMSD continues to decrease asymptotically, the decrease is much slower after about 3-4 days.

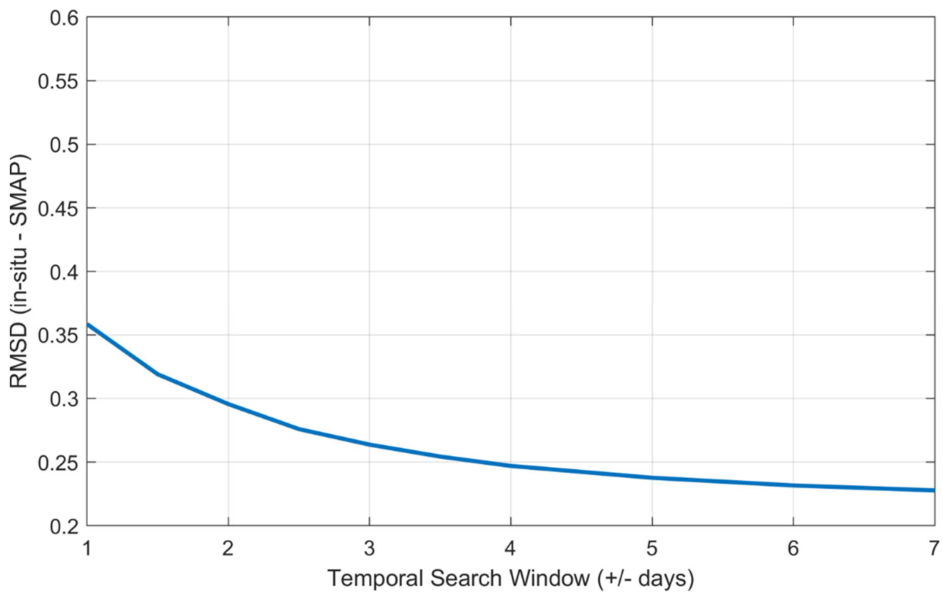



*Figure 8: Global RMSD noise reduction in SMAP RSS L2 V4.0 as a function of the temporal search window ΔT. The spatial search radius is kept fixed at ΔS=±50 km*. Values are computed at intervals of ±0.5 days.

## *3. Aquarius*

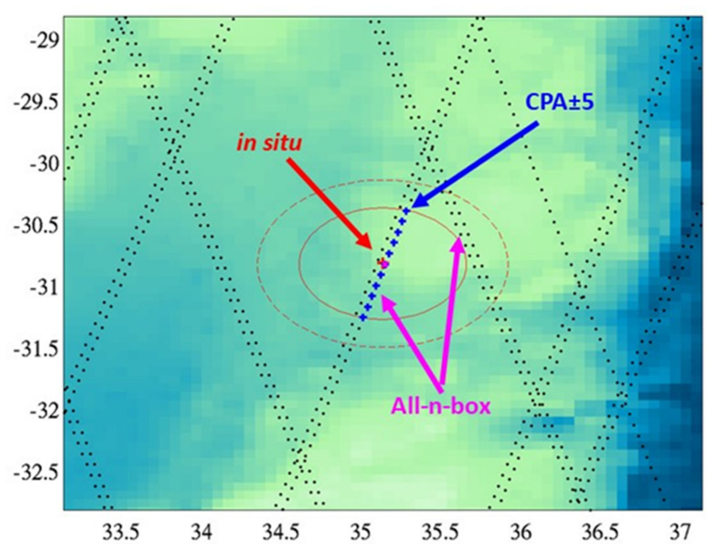



*Figure 9: Example of Aquarius tracks near an Argo float (red arrow) in the Indian Ocean in 2011. Blue dots represent 1.44s integration times of Aquarius, while the red dot indicates an Argo observation.*

To further illustrate the issues associated with choosing a space-time box to compute matchups, a real example of an Aquarius overpass is provided in Fig. 9. The dotted lines represent 1.44s integrations of Aquarius data for one Aquarius beam. The red dot (arrow) is the location of an Argo report and the circles around this point represent areas with radii of 50 km and 75 km. After 7 days (i.e. ± 3.5 days from a reference time) the Aquarius orbits repeat. The figure illustrates two features of the satellite matchup: First, the blue dots are from one pass (orbit) and are separated by about 1.44 sec. If the spatial box is big enough, it is possible that another orbit will be included adding another group of samples which could be as much as 7 days distant in time. The number of samples in the space-time box depends on the location of *in situ* data relative to the satellite overpass and could be a maximum as in the orbit shown through the Argo location or could be many fewer as in the case of an orbit nearly tangential to the spatial box.

### 3.1. Aquarius Sampling

This structure of the samples within the space-time box is illustrated in Fig. 10 which reports histograms of the number of samples for a combination of box radius (ΔS = 25 km and 75 km) and time window (ΔT = ±0.5, ±3.5, ±15 days). The data is for the Atlantic Ocean (north and south) and includes all matchups with Argo during the lifetime of the Aquarius mission. The data can be found under “data” at https://neptune.gsfc.nasa.gov/csb. The histogram for the space-time window (R = 25 km; T = ±0.5 days) is similar to the potential density function (PDF) that one obtains for the number of samples in a single pass through the box with uniformly

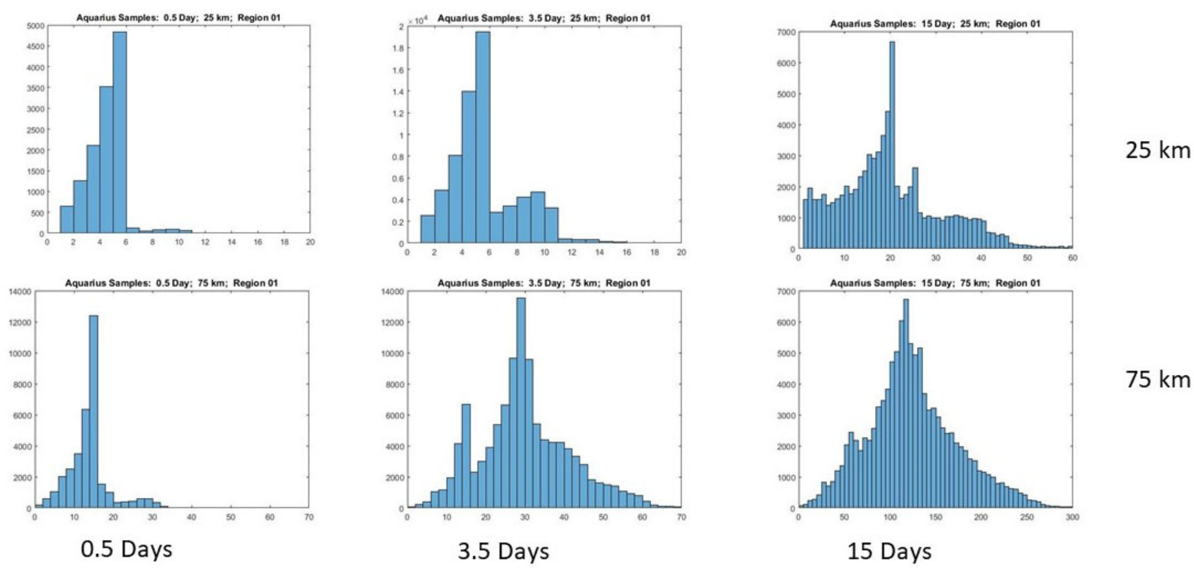



*Figure 10: Histogram showing the frequency of occurrence of the number of samples in the space-time box as function of the box size for ΔT = 25 and 75 km and ΔT = ±0.5, ±3.5 and ±15 days. The data are from the Atlantic Ocean and during the entire Aquarius mission (2011-2015).*

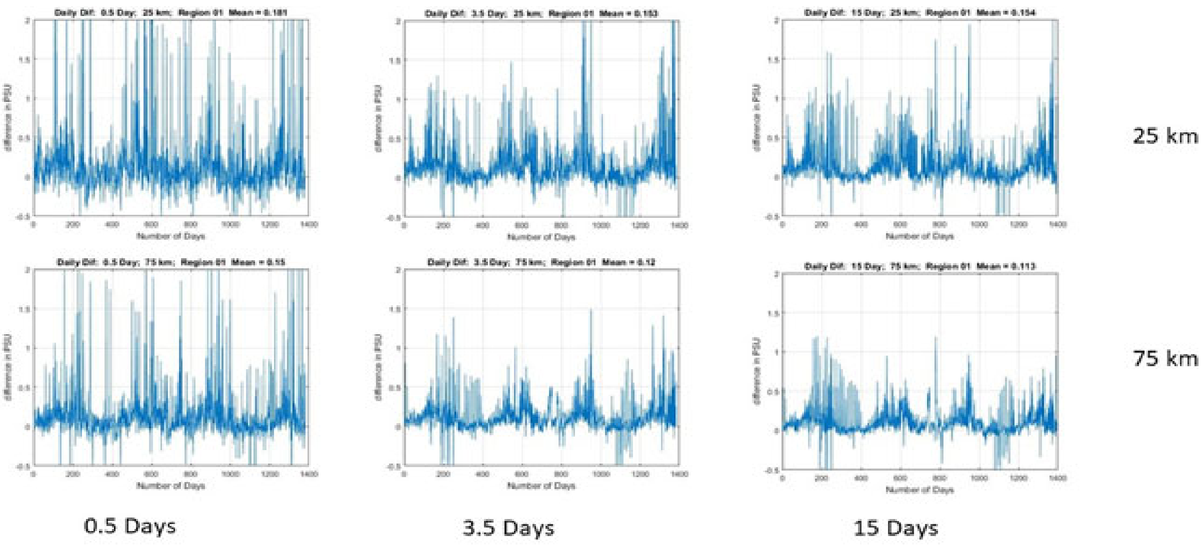



*Figure 11: The daily difference between satellite and in situ salinity averaged over the Atlantic Ocean plotted as a function of time for the Aquarius mission. The difference is computed by first averaging all the satellite observations in the space-time box associated with the* in situ *observation (Argo). No attempt has been made to eliminate outliers (large spikes).*

distributed distance of the path from the center. As the temporal search window increases, the probability of a second pass increases (with additional matchups evident at ΔT < 25km and ΔT = ± 3.5). Furthermore, as the radius increases, the probability of more than one pass increases (bottom row in Fig 10). At ΔT = ±3.5 days there are a large number of samples even for the 25 km box and somewhere between 25 km and 75 km it is no longer obvious how many passes are included in the histogram.

The statistics also seem to stabilize in this range of boxes. This is illustrated in Fig 11 which shows the daily average of the difference between the salinity reported *in situ* and the average of all those satellite measurements within the associated space-time box averaged over the entire data set. No attempt has been made to eliminate outliers. The noise decreases significantly in moving from ΔT = ±0.5 days to ±3.5 days and then not much more for increased time window. There also is some improvement in increasing the search area's radius from 25 km to 75 km.

### 3.2 Averaging Choices for Aquarius Data

The space-time window identified above will include multiple values of satellite salinity at each *in situ* observation (e.g. Argo report) as shown above in Fig. 2. It is reasonable to average the satellite retrievals matching the search criteria associated with a given *in situ* observation before computing the difference between the satellite and *in situ* salinity. Again, the question of how many samples to average is raised here. One approach is to take the sample of closest

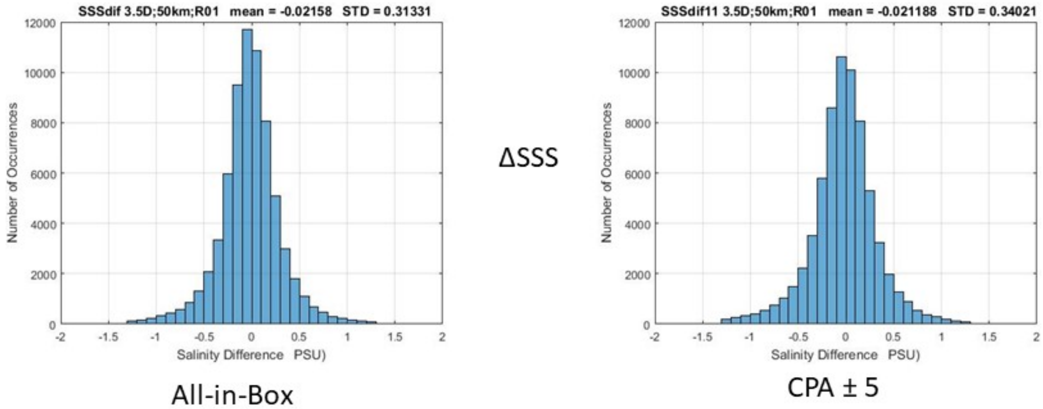



*Figure 12: Histogram showing the frequency of occurrence of salinity differences between the satellite retrieval from Aquarius and the value of Argo for the Atlantic Ocean computed over the entire Aquarius mission (2011-2015). Left: Averaging all samples in the space-time box; Right: Averaging only the sample of closest approach and 5 samples before and after*.

approach in space or time and average the 10 nearest neighbors. This corresponds to a travel time of about 100 km (10 x 1.44 sec x 7.5 km/sec) which is about the average cross track beam width. One could also argue for half this many (i.e. 2 samples per footprint which is equivalent to Nyquist sampling corresponding to a resolution of one footprint). The problem with this approach is the variability in the location of the nearest neighbors. For example, if the closest approach point lies at the edge of the box, it is possible that many of these points will not even fall within the space-time search criteria.

Another approach is to use all the samples matching the criteria, which results in only minor differences. This is illustrated for Aquarius in Fig. 12, which shows the histogram for the average salinity difference using the two approaches. The STD of the salinity difference is 0.31 using all the samples and is slightly higher at 0.34 using the 10 nearest neighbors. Using all samples in the box has the advantage of including slightly more samples, which results in a lower STD, and has the advantage that all the samples meet the space-time criterion. Figure 11 also shows that the space-time box (R = 50 km; T = ± 3.5 days) is near a noise minimum for Aquarius. Increasing the time window does not significantly decrease noise.

# 4. Conclusion

### 4.1 Space-Time box

Defining a matchup to be all values within a space-time box with the *in situ* observation at the center is necessary to accommodate the different space and temporal characteristics of the satellite and *in situ* data. It furthermore eliminates the difference between a time-centric and space-centric search. A natural choice is to choose a box with the spatial dimensions of the satellite footprint and a temporal window consistent with the revisit time of the spacecraft, but to facilitate comparisons between sensors, a 'lowest common denominator' should be picked. Selecting a circle of radius $\Delta S < 50$ km and a time window $\Delta T < \pm 3.5$ days assigns a search radius near the minimum in Fig. 3 and a time window near the point of diminishing returns as shown in Fig. 4.

### 4.2 Recommendations

In comparing SMOS, SMAP and Aquarius salinity, a common definition should be used. Using all samples in the space-time box appears to have some small advantages for Aquarius, and for lack of a better choice, it is recommended to use the same criteria for both instruments. This is not necessarily an issue when comparing different versions of the retrieval algorithm of one sensor with a new one for the same sensor. However, it can be an issue when comparing retrievals of Aquarius with those of SMAP because of the different levels of averaging inherent in the retrievals.

The recommended time window, $\Delta T < \pm 3.5$ days is close to the satellite revisit time and close to the time when the statistics seem to stabilize (e.g. Fig 5 and Fig 11). The spatial search radius, $\Delta S = 50$ km is consistent with the spatial resolution of both instruments and close to the value where the statistical variation seems to reach a minimum (e.g. Fig 3 and Fig 11). Averaging all samples in the space-time window has the advantage that all samples are consistent with the boundaries of the space-time box. Hence, the recommendation is made that the metric for the performance of SMAP and Aquarius salinity retrievals should be the difference between retrieved and *in situ* salinity using a common space-time box of $\Delta S < 50$ km and $\Delta T < \pm 3.5$ days and averaging all satellite samples within this box before computing the salinity differences. It is recommended that the evaluation be done at Level 2 for both Aquarius SMAP.

## References

Boutin, J., Chao, Y., Asher, W.E., Delcroix, T., Drucker, R., Drushka, K., Kolodziejczyk, N., Lee, T., Reul, N., Reverdin, G. and Schanze, J., 2016. Satellite and in situ salinity: Understanding near-surface stratification and subfootprint variability. Bulletin of the American Meteorological Society, 97(8), pp.1391-1407.

Dinnat, E.P., Le Vine, D.M., Boutin, J., Meissner, T. and Lagerloef, G., 2019. Remote sensing of sea surface salinity: Comparison of satellite and in situ observations and impact of retrieval parameters. Remote Sensing, 11(7), p.750.

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J. and Kimball, J., 2010. The soil moisture active passive (SMAP) mission. Proceedings of the IEEE, 98(5), pp.704-716.

Font, J., Camps, A., Borges, A., Martín-Neira, M., Boutin, J., Reul, N., Kerr, Y.H., Hahne, A. and Mecklenburg, S., 2009. SMOS: The challenging sea surface salinity measurement from space. Proceedings of the IEEE, 98(5), pp.649-665.

Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.J., Font, J., Reul, N., Gruhier, C. and Juglea, S.E., 2010. The SMOS mission: New tool for monitoring key elements ofthe global water cycle. Proceedings of the IEEE, 98(5), pp.666-687.

Lagerloef, G., Colomb, F.R., Le Vine, D., Wentz, F., Yueh, S., Ruf, C., Lilly, J., Gunn, J., Chao, Y.I., Decharon, A. and Feldman, G., 2008. The Aquarius/SAC-D mission: Designed to meet the salinity remote-sensing challenge. Oceanography, 21(1), pp.68-81.

Le Vine, D.M.; Lagerloef, G.S.E.; Colomb, F.R.; Yueh, S.H.; Pellerano, F.A. Aquarius: An instrument to monitor sea surface salinity from space. IEEE Trans. Geosci. Remote Sens. 2007, 45, 2040–2050.

Le Vine, D.M., Dinnat, E.P., Meissner, T., Yueh, S.H., Wentz, F.J., Torrusio, S.E. and Lagerloef, G., 2015. Status of Aquarius/SAC-D and Aquarius salinity retrievals. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(12), pp.5401-5415.

Le Vine, D.M., Lagerloef, G.S. and Torrusio, S.E., 2010. Aquarius and remote sensing of sea surface salinity from space. Proceedings of the IEEE, 98(5), pp.688-703.

Meissner, T., Wentz, F., Manaster, A. and Lindsley, R., 2019. NASA/RSS SMAP salinity: Version 4.0 validated release. Remote Sensing Systems Tech. Rep., 82219, p.55.

Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.J., Font, J., Reul, N., Gruhier, C. and Juglea, S.E., 2010. The SMOS mission: New tool for monitoring key elements ofthe global water cycle. Proceedings of the IEEE, 98(5), pp.666-687.

Schanze, J.J., Schmitt, R.W. and Yu, L.L., 2010. The global oceanic freshwater cycle: A state-of-the-art quantification. Journal of Marine Research, 68(3-4), pp.569-595.

Swift, C.T., 1980. Passive microwave remote sensing of the ocean—A review. Boundary-Layer Meteorology, 18(1), pp.25-54.

Yueh, S.H., West, R., Wilson, W.J., Li, F.K., Njoku, E.G. and Rahmat-Samii, Y., 2001. Error sources and feasibility for microwave remote sensing of ocean surface salinity. IEEE Transactions on Geoscience and Remote Sensing, 39(5), pp.1049-1060.

## Appendix A: Spatial Debiasing

If global values of RMSD, standard deviation, or offsets are to be computed, it is critical that the density of observations is considered. As is evident from Fig. 1, the sampling of Argo floats is highly inhomogeneous, as are ship-tracks providing thermosalinograph measurements. While various numerical methods may be used to address this problem, one of the easiest solutions is to use gridding techniques to produce global maps, similar to that in Fig. 4. The area covered by each grid point is then computed using geospatial routines and used as a weight for each grid point. When choosing a suitable coarse grid, ideally there will be no blank area aside from those that are routinely flagged due to their proximity to land or known radio frequency interference (RFI) sources.

Similarly, for a satellite-centric approach, polar areas are sampled much more frequently than equatorial areas due to their lower revisit intervals. As a result, spatial debiasing is also required in this case.

## Appendix B: Definition Details for Level 2 Aquarius and SMAP Salinity Matchup

### B.1 Summary:

For *in situ* data from Argo floats, it is recommended that the matchup consist of the average of all satellite observations falling in a space-time cylinder constrained by $\Delta S < 50$ km and $\Delta T <$ ±3.5 days centered on the location of the *in situ* salinity report. This might also be appropriate for other types of *in situ* observations (e.g. which are continuous in time and/or space such buoys or ship reports), but this has not been addressed in this report.

Recommended search window (for both SMAP and Aquarius)

R = 50 km
T = ±3.5 days
Search is centered on individual *in situ* reports (space and time)
All satellite observations with search window are averaged

Processing: For each Argo report, search satellite data 3.5 days before and after the time of the report to find all antenna boresight positions where Level 2 salinity retrievals are reported within a circle of 50 km of the position of the *in situ* location. Average all such satellite data.

**B.2 *In Situ* Data: Argo**

Data: US GODAE (http://www.usgodae.org/argo/argo.html)
Delayed mode for Aquarius
Quality flag
PSAL_ADJ_QC = 1 or 2: Adjusted salinity if available.
PSAL_QC = 1 or 2; If adjusted not valid or not available.
Do not use: If neither value above is available

**B.3 Aquarius Data Specifics:**

Level 2 data: file name Qyyyydddhhmmss.L2_SCI_V5.0
Parameter: SSS

Flags: The following flags are used to eliminate data when the flag is "set"
See AQ-014-PS-0006 for definition of flags.

Flag #3: Land contamination (moderate)
Flag #4: Sea ice contamination (moderate)
Flag #5: High wind speeds (moderate)
Flag #8: Radiometer RFI: |TA-TF|>0.3 Kelvin or bit overflow in
Flag #14: Full roughness correction not performed or no SWH
Flag #16: Pointing Anomaly Flag
Flag #17: rad_Tb_consistency > 0.40 or no SSS convergence
Flag #18: Cold water SST < 5°C (moderate)
Flag #19: TF-TA (moderate)
Flag #20: Non-nominal commanded sate
Flag #21: Reflected radiation from Moon or Galaxy exceeds threshold (moderate)
Flag #23: Unacceptable Asc/Dsc difference

Rain: rim_status_flag
1st bit: NOAA CMORPH is not available
2nd bit: CMORPH for IRR has been removed for low quality

**B.4 SMAP Data Specifics:**

Level 2 data: file name:
RSS_SMAP_L2C_rnnnnn_yyyymmddThhmiss_yyyydoy_FNL_V04.nc
Parameter: sss_smap_40km (preferred, less averaging)

sss_smap (nominal “70 km” product)

Flags: The following flags are used to eliminate data when the flag is “set”

Parameter: Level 2 Q/C flags in the SMAP V3.0 release
Filename: iqc_flag
For definitions see Remote Sensing Systems (RSS) Technical Report 101518, “NASA/RSS SMAP Salinity: Version 3.0 Validated Release”

If any of the following flags are set then the data is not used:

Flag #0: no valid radiometer observation in cell
Flag #1: problem with OI
Flag #2: Strong land contamination (gland>0.1)
Flag #3: Strong sea ice contamination (gice>0.1)
Flag #4: MLE in SSS retrieval algo has not converged
Flag #5: Sunglint
Flag #6: Moonglint
Flag #7: high reflected galaxy
Flag #8: moderate land contamination (gland>0.01)
Flag #9: moderate ice contamination (gice>0.001)
Flag #10: high residual of MLE in SSS retrieval algo
Flag #11: low SST (surtep – 273.15 < 5°C)
Flag #12: high wind speed (winspd > 15 m/s)
Flag #13: light land contamination (gland>0.001)
Flag #14: light sea ice contamination (gice>0.0005)
Flag #15: rain flag (IMERG rain-rate >0.1 mm/h)