SEA SURFACE SALINITY RETRIEVALS FROM AQUARIUS USING NEURAL NETWORKS

Yan Soldo[(1,2)], David M. Le Vine[(1)] and Emmanuel Dinnat[(1,3)]

[(1)] NASA Goddard Space Flight Center, Greenbelt, MD, USA
[(2)] USRA, Greenbelt, MD, USA
[(3)] Chapman University, Orange, CA, USA

ABSTRACT

Even though the Sea Surface Salinity (SSS) retrieved from Aquarius are generally very close to in-situ measurements, the level of similarity varies with the region and with the circumstances of the observations (wind speed, sea surface temperature, etc.). SSS is currently retrieved from the brightness temperatures measured by Aquarius and applying the current theoretical model for the propagation and emission of the natural thermal radiation. In this contribution we consider an alternative retrieval approach based on a Neural Network (NN) with the goal of improving the subsets of Aquarius SSS data that are in poorer agreement with in-situ measurements. The subset considered here are the SSS retrieved at latitudes higher than 30˚. The output of the NN approach are compared against in-situ measurements using four statistical metrics (correlation coefficient, bias, RMSD and 5% trimmed range). The output of the NN and the nominal Aquarius SSS are compared against SSS values from in-situ measurements and from ocean models. From these comparisons it appears that the output of the NN matches the in-situ measurements better than the nominal Aquarius SSS.

Index Terms— Sea surface salinity, Aquarius, neural networks

1. INTRODUCTION

The Aquarius instrument measured the Earth's natural thermal radiation for almost 4 years (2011-2015) in order to provide accurate measurements of the Sea Surface Salinity (SSS) [1] . Aquarius was equipped with three radiometers imaging in a push-broom fashion. In the Aquarius processing, the SSS is retrieved from the measured antenna temperatures using a retrieval algorithm based on the current understanding of the emission and propagation of thermal radiation from the ground up to the satellite altitude, with some empirical corrections [2] , [3] . The SSS retrieved with this approach are generally very accurate, in the sense that they match closely with in-situ measurements. However, the level of accuracy of the Aquarius SSS varies with the region.

In this contribution we investigate a different approach to retrieve the SSS from Aquarius TB (brightness temperature), which is based on a Neural Network (NN).

NNs have already been implemented for similar purposes by the SMOS (Soil Moisture Ocean Salinity) mission to retrieve soil moisture [4] and to retrieve SSS from simulated measurements [5] . They have also been applied to Aquarius data to calibrate satellite retrievals before assimilation by ocean models [6] but as far as we know not to retrieve SSS from the L-band brightness temperatures themselves.

2. MARGIN FOR IMPROVEMENTS IN THE AQUARIUS SSS

The first step of this investigation has been to identify the subsets of Aquarius data that we might be able to improve with the NN approach. We began with the difference between the SSS retrieved from Aquarius (SSS_Aquarius) and the SSS measured by the Argo floats (SSS_Argo). The colocations between measurements by Aquarius and Argo are provided in the latest release of Aquarius data (version 5) and do not include averaging. The differences SSS_Aquarius– SSS_Argo have been plotted as a function of several parameters (wind speed, sea surface temperature, etc.), in order to better understand which subsets of Aquarius data have clearer margins for improvements.

The clearest margin for improvement was noticed by plotting the SSS difference as a function of latitude (see Figure 1). The map in Figure 1 shows the colocations between Aquarius footprints and the surfacing of Argo floats for the entire duration of the mission. The color indicates the difference in salinity (SSS_Aquarius – SSS_Argo). The bottom plot was obtained by grouping the SSS differences in intervals of 5 degrees latitude and then by computing the mean and the standard deviation within each interval. The blue circles in Figure 1 represent the means and the error bars heights indicate ±1 standard deviation.

Some of the differences at high latitudes are likely to be due to low-level RFI (e.g., northeastern Atlantic and near Japan [7]). Attempts to detect these instances have been made (e.g., [8] , [9]), but they have not been implemented for the time being. Other differences (north Pacific, Southern Ocean) are probably unrelated to RFI and they correspond to regions with cold temperatures, low SSS and high winds, which are generally considered difficult conditions to retrieve SSS using theoretical approaches [10] . The bottom plot of

Figure 1 shows that bias is near zero in the range [-50°, 30°] and that the standard deviations increases south of -40° and north of 30°.
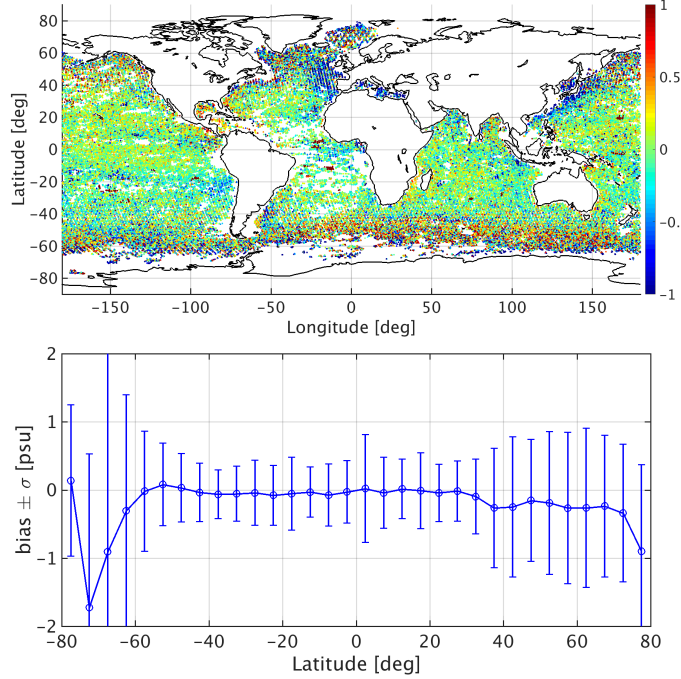

Figure 1: Difference between SSS_Aquarius and SSS_Argo. Map (top) and error bars (bias ± σ) as a function of latitude (bottom).

Based on these observations, we apply a NN approach to the Aquarius measurements at latitudes greater than 30° (in both hemispheres) to see whether we can obtain a set of SSS that would be more similar to our reference SSS, provided by the in-situ measurements of the Argo floats.

3. NEURAL NETWORKS

Three neural networks have been developed, one for each Aquarius radiometer. Each NN is a feed-forward NN with four layers (an input layer, two hidden layers and an output layer) and each layer is dense (every input is connected to every output). The input parameters of the NN are: TBs (H- and V-pol); sea surface temperature; significant wave height; wind speed; weighted fractions of land and sea-ice in the field of view; instantaneous rain rate; latitude and longitude. All these input parameters are read from the Aquarius data products (version 5). The TB used as input is the TB at the surface before roughness correction. Other choices of TB are also possible. For example, the NN could be trained to use the antenna temperatures or the TB at the "top of ionosphere" (i.e., before polarization rotation). However, using TB earlier in the processing chain would require that the number of input parameters increase to correct for propagation to the surface (e.g., polarization rotation angles, contribution from the direct and reflected signals from the galaxy, the sun and the moon, and others). The number of neurons in the hidden

layers is set to 15 in each layer, as above 15 no significant improvement was noticed. We also considered different options for the activation function (e.g., sigmoid, tanh) and found that ReLU (rectified linear units) provided slightly better results.

Training a NN consists in minimizing a cost function by adjusting the weights that correspond to the links between neurons. The NN used here was trained by minimizing the sum of the mean squared errors between the SSS from the NN (SSS_NN) and the SSS from Argo (SSS_Argo). The algorithm used for the minimization is the "Adam" algorithm [11] and the weights are initialized randomly. The three NNs (one for each beam) have the same architecture.

Each NN was trained using Aquarius version 5 data, applying the following criteria:
- valid match-up with an Argo buoy;
- |latitude| > 30° (except Mediterranean Sea);
- water fraction (geometric fraction of water in the field of view, weighted by the antenna gains) > 99%;
- only one Aquarius beam per NN.

The Mediterranean Sea is above 30° N; however, the typical issues affecting SSS retrievals at high latitudes (high winds, low SSS, low temperature) do not generally affect the Mediterranean Sea and therefore it was excluded.

The dataset of Aquarius-Argo colocations is divided into three subsets: the training, the validation and the test subsets. The first is used to train the NN. The second serves to assess the quality of the NN on data not used during training and it is used to determine the best architecture for the NN. The third is only used at the end to perform a final evaluation on a dataset that neither the NN nor the architects have seen before. Training, validation and test subsets correspond to 70%, 15% and 15% of the entire dataset, respectively, and the members are selected randomly. The entire dataset is made up of roughly 50,000 Aquarius-Argo colocations.

4. RESULTS

4.1 Comparison between SSS_NN, SSS_Aquarius and SSS_Argo

The results for the Aquarius inner beam are shown in Figure 2 (left), which is a density plot between SSS_NN and SSS_Argo and includes only the test subset.

Figure 2 (left) shows that SSS_NN corresponds well with SSS_Argo. However, the NN approach is justified only if the comparison between SSS_NN and SSS_Argo is better than the comparison between SSS_Aquarius and SSS_Argo. The comparison between SSS_Aquarius and SSS_Argo is shown in Figure 2 (right). The axes of Figure 2 (right) have been set equal to those of Figure 2 (left) so that they can be compared. As a result, there are a few data points with SSS_Aquarius below 30 psu and that are not visible in Figure 2 (right).

In order to assess these comparisons more quantitively, we have considered four statistical metrics: the correlation

coefficient (Pearson's *r*), the bias, the Root-Mean-Square Difference (RMSD), and the 5% trimmed range (range from the 5th to the 95th percentile). They are reported for all three beams in Tables 1 to 3.
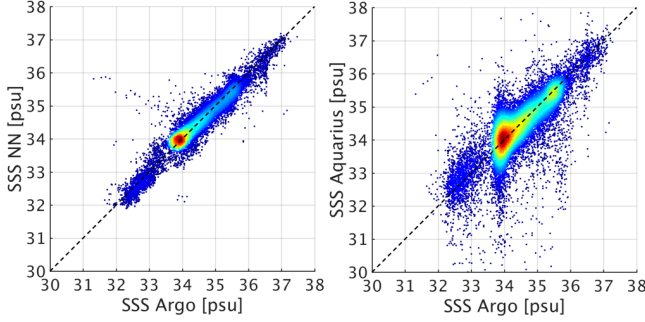

Figure 2: Comparison on the test set (inner beam) between: (left) SSS_NN and SSS_Argo; (right) SSS_Aquarius and SSS_Argo

	NN vs Argo	Aquarius vs Argo
r	0.949	0.697
bias	-0.034	-0.099
RMSD	0.290	0.869
5th percentile	-0.416	-1.103
95th percentile	0.296	0.812
5% TR	0.712	1.915

Table 1: SSS [psu] statistical comparison (inner beam)

	NN vs Argo	Aquarius vs Argo
r	0.960	0.676
bias	0.018	-0.093
RMSD	0.257	0.946
5th percentile	-0.308	-1.074
95th percentile	0.358	0.804
5% TR	0.666	1.878

Table 2: SSS [psu] statistical comparison (middle beam)

	NN vs Argo	Aquarius vs Argo
r	0.950	0.711
bias	-0.013	-0.096
RMSD	0.285	0.808
5th percentile	-0.395	-1.118
95th percentile	0.331	0.802
5% TR	0.726	1.920

Table 3: SSS [psu] statistical comparison (outer beam)

All metrics indicate that SSS_Argo is more similar to SSS_NN than it is to SSS_Aquarius.

4.2 Comparison of SSS_NN and SSS_Aquarius with other ocean datasets over one month of data

Once the NN for the three beams are defined, they can be used to routinely retrieve SSS from Aquarius measurements. The results (SSS_NN) are compared with SSS_Aquarius and with two other SSS datasets: the model HYCOM (HYbrid Coordinate Ocean Model) [12] and the climatology from the Scripps Institution of Oceanography [13] . Figure 3(a and b) show that, compared to the SSS from HYCOM and from Scripps, the SSS_Aquarius are noisier and with a negative bias near regions with strong RFI and near the transition with sea ice. On the other hand, the differences between SSS_NN and the SSS from HYCOM and Scripps, shown in Figure 3(c and d), have a lot less variability, meaning that SSS_NN is more consistently similar to both datasets. The regions where SSS_NN is most different from those datasets are near the outflow of the Uruguay river and off the western coast of Canada. These are dynamic regions of the ocean [14] , [15] and they correspond also to the largest differences between the two models, so that it is difficult to assess whether the SSS_NN is accurate there.

Also note that the Scripps dataset only provides SSS for |latitude| < 65˚, so in Figure 3(c, d and e) there is no data at the highest latitudes.

5. CONCLUSIONS

While the SSS retrieved from Aquarius are already very close to in-situ measurements (in this case, Argo), the level of accuracy varies depending on the region and on the state of the ocean. To improve the accuracy between Aquarius and Argo, a different approach, based on a NN, was developed. The results show that, at |latitude| > 30˚, the similarity between Aquarius and Argo SSS decreases, and using an empirical model such as NN can be a useful alternative.

6. REFERENCES

[1]   D. M. Le Vine, E. P. Dinnat, T. Meissner, F. J. Wentz, H.-Y. Kao, G. Lagerloef, and T. Lee, "Status of Aquarius and Salinity Continuity," *Remote Sensing*, vol. 10, no. 10, p. 1585, Oct. 2018.

[2]   Aquarius Project: Algorithm Theoretical Basis Document (ATBD), End of mission version, 2017. Available online at ftp://podaac-ftp.jpl.nasa.gov/allData/aquarius/docs/ v5/AQ-014-PS-0017_Aquarius_ATBD-EndOfMission.pdf. Accessed on Decemeber 2018.

[3]   D. M. Le Vine, E. P. Dinnat, S. Abraham, P. De Matthaeis, and F. J. Wentz, "The Aquarius simulator and cold-sky calibration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 9, pp. 3198–3210, 2011.
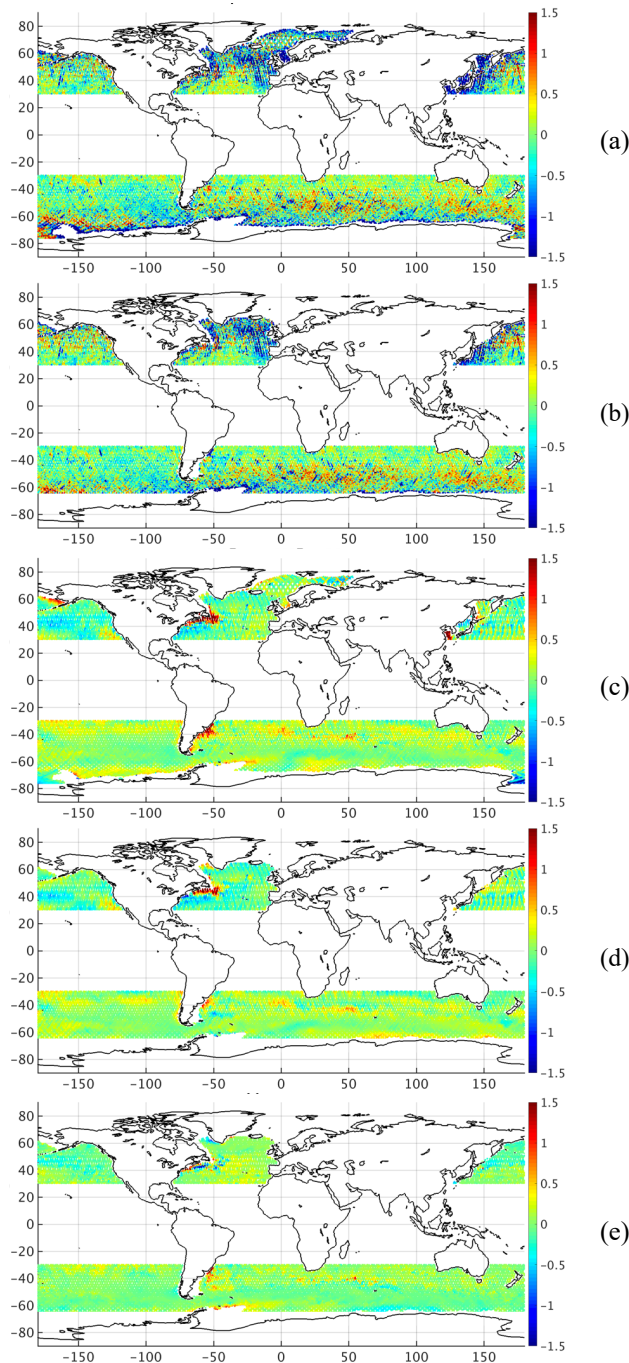
Figure 3: Difference maps for Jan 2016 (all three beams) between: (a) SSS_Aquarius and SSS_HYCOM; (b) SSS_Aquarius and SSS_Scripps; (c) SSS_NN and SSS_HYCOM; (d) SSS_NN and SSS_Scripps; (e) SSS_Scripps and SSS_HYCOM.

[4]   N.J. Rodríguez-Fernández, F. Aires, P. Richaume, Y.H. Kerr, C. Prigent, J. Kolassa, F. Cabot, C. Jiménez, A. Mahmoodi, and M. Drusch, "Soil moisture retrieval using neural networks: Application to SMOS," *IEEE Transactions on Geoscience and Remote Sensing*, 53(11), pp.5991-6007, 2015.

[5]   A. Ammar, S. Labroue, E. Obligis, C.E. Mejia, M. Crépon and S. Thiria, "Sea surface salinity retrieval for the SMOS mission using neural networks," *IEEE transactions on geoscience and remote sensing*, 46(3), pp.754-764, 2008.

[6]   G. Vernieres, R. Kovach, C. Keppenne, S. Akella, L. Brucker, and E. Dinnat, "The impact of the assimilation of Aquarius sea surface salinity data in the GEOS ocean data assimilation system," J. Geophys. Res. Ocean., vol. 119, no. 10, pp. 6974–6987, 2014.

[7]   T. Meissner, F. Wentz and D. M. Le Vine, "Aquarius Retrieval Salinity Algorithm," RSS Technical Report 120117, 2017.

[8]   D. D. Chen and C. Ruf, "A novel method to estimate the RFI environment," *IEEE Geoscience and Remote Sensing Symposium*, pp. 215-218, July 2014.

[9]   P. de Matthaeis and D. M. Le Vine, "Statistical Analysis of Aquarius Radiometer Radio Frequency Interference (RFI) and Implications on RFI Detection and Mitigation," NASA technical report, 2018.

[10]   J. Boutin, P. Waldteufel, N. Martin, G. Caudal and E. Dinnat, "Surface salinity retrieved from SMOS measurements over the global ocean: Imprecisions due to sea surface roughness and temperature uncertainties," *Journal of Atmospheric and Oceanic Technology*, 21(9), pp.1432-1447, 2004.

[11]   D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint:1412.6980*, December 2014.

[12]   E.P. Chassignet, H.E. Hurlburt, O.M. Smedstad, G.R. Halliwell, P.J. Hogan, A.J. Wallcraft, R. Baraille and R. Bleck. "The HYCOM (hybrid coordinate ocean model) data assimilative system," *Journal of Marine Systems*, 65(1-4), pp.60-83, 2007.

[13]   J. Holte, L.D. Talley, J. Gilson and D. Roemmich, "An Argo mixed layer climatology and database," *Geophysical Research Letters*, 44(11), pp.5618-5626, 2017.

[14]   R. P. Matano, V. Combes, A. R. Piola, R. Guerrero, E. D. Palma, P. T. Strub, C. James, H. Fenco, Y. Chao and M. Saraceno, "The salinity signature of the cross-shelf exchanges in the Southwestern Atlantic Ocean: Numerical simulations," Journal of Geophysical Research: Oceans, 119(11), pp.7949-7968, 2014.

[15]   M. Sena Martins, N. Serra, and D. Stammer, "Spatial and temporal scales of sea surface salinity variability in the Atlantic Ocean," *Journal of Geophysical Research: Oceans*, 120(6), pp.4306-4323, 2015.