Triple Point Collocation Code and Example Use Case for L3 Satellite Salinity Data
Pi-MEP NASA-ESA Collaboration and Salinity Validation Data System: 3.1.2 Integration of Triple Collocation Analysis Tools

1. Background

Triple point collocation is a method to determine errors for three datasets of the same variable without making *a priori* assumptions about which dataset is the truth. Triple point collocation is frequently used in the analysis of satellite observation errors (e.g., Gruber et al., 2016). The triple point collocation code provided here is based on the original code of Hsun-Ying Kao as developed for NASA Aquarius Salinity Validation Analysis (Kao et al., 2018)). It has been adapted to use the covariance method outlined in Gruber et al. (2016). The triple point collocation code is flexible for use with any combination of three, two-dimensional datasets. An example use case with monthly, L3 SMAP, SMOS, and Argo data is provided to illustrate implementation of the method.

2. Method

The triple point collocation analysis method implemented in the provided code is based on the covariance notation described in Stoffelen (1998) and Gruber et al. (2016). Using covariance notation, versus other notations, allows for more flexible implementation since *a priori* assumption about which dataset is the "reference" and rescaling of other datasets to that reference are not required. The unscaled error variance (σ_ε^2) for each dataset is:

$$\sigma_{\varepsilon_X}^2 = \sigma_X^2 - \frac{\sigma_{XY}\sigma_{XZ}}{\sigma_{YZ}}$$

$$\sigma_{\varepsilon_Y}^2 = \sigma_Y^2 - \frac{\sigma_{YX}\sigma_{YZ}}{\sigma_{XZ}}$$

$$\sigma_{\varepsilon_Z}^2 = \sigma_Z^2 - \frac{\sigma_{ZX}\sigma_{ZY}}{\sigma_{XY}}$$

where, σ_i^2 is the dataset variance and σ_{ij} are the dataset covariances. RMSD is $\sqrt{\sigma_\varepsilon^2}$ for positive error variances. A detailed discussion of the underlying assumptions (signal error linearity, signal and error stationarity, error orthogonality, zero error cross-correlation, and dataset representativeness) can be found in Gruber et al. (2016).

3. Covariance Triple Point Collocation Code

The covariance triple point collocation code (`covariance_triple_point_collocation.m`) provided is flexible to be used with any combination of 2-dimensional data with uniform dimensions.

```
function [errorvar_d1,errorvar_d2,errorvar_d3]=...

covariance_triple_point_collocation(data1,data2,data3,largeflag)
```

Inputs:

Dataset 1: dimensions (1 by n, m by 1, m by n)
Dataset 2: dimensions (1 by n, m by 1, m by n)
Dataset 3: dimensions (1 by n, m by 1, m by n)
Large Flag: 1 = exclude large dataset differences (>5PSU) from analysis, 0=include all dataset differences regardless of size

Outputs:

Dataset 1 Unscaled RMSD: dimensions (1x1)
Dataset 2 Unscaled RMSD: dimensions (1x1)
Dataset 3 Unscaled RMSD Variance: dimensions (1x1)

Error checking for the correct number of input variables in completed. Then, the location of NaNs within each dataset is made uniform. If the flag is enabled, differences between each combination of datasets is computed, and if where differences are > 5, dataset values are replaced with NaNs. Finally, dataset variances, covariances, and unscaled error variances are calculated.

4. Example Use Case: *Monthly, L3 Satellite Salinity and in situ observations*

A use case of the triple point collocation code is to evaluate the error associated with a monthly L3 satellite salinity product. Rather than assuming one dataset is a perfect representation of sea surface salinity and any dataset differences due to error, the triple point collocation covariance method evaluates error for each of the 3 input datasets simultaneously. In the example use case monthly, L3 data products are used to determine 1) spatially averaged, monthly RMSD and 2) temporally averaged, spatial RMSD for 3 datasets. Details of the L3 datasets and use case code implementation, including additional MATLAB packages required, are provided below.

L3 Datasets:

<u>Dataset 1:</u>	SMOS Satellite Salinity, CATDS-CPDC L3 OS/SSS product L3Q Monthly (CATDS-PDC_L3OS_3Q_mixed)
Time:	January 2010 to May 2021 (RE07) June 2021 to January 2022 (OPER)
Resolution:	50km (Latitude) x 50km (Longitude)
Coverage:	81.9831°N to 83.5171°S 179.8703°W to 179.6109°E
Data Access:	https://www.catds.fr/Products/Available-products-from-CPDC www.catds.fr/Resources/Documentation

ftp://ext-catds-cpdc:catds2010@ftp.ifremer.fr/
Accessed on 25 February 2022

Dataset 2: SMAP Satellite Salinity, RSS L3 V4.0 Monthly (SMAP_RSS_L3_SSS_SMI_MONTHLY_V4)
Time: April 2015 to January 2022, No data for July 2019
Resolution: 0.25° (Latitude) x 0.25° (Longitude)
Coverage: 89.875°N to 89.875°S
0.125°E to 359.875°E
Data Access: https://podaac.jpl.nasa.gov/dataset/SMAP_RSS_L3_SSS_SMI_MONTHLY_V4
https://podaac-tools.jpl.nasa.gov/drive/files/allData/smap/L3/RSS/V4/monthly/SCI
Accessed on 25 February 2022

Dataset 3: Roemmich-Gilson Argo Climatology
Time: January 2004 to January 2022
Resolution: 1° (Latitude) x 1° (Longitude)
Coverage: 79.50°N to 64.5°S
20.50°E to 379.50°E
Pressure: 58 levels, 2.5db - 1975.0db
Data Access: http://sio-argo.ucsd.edu/RG_Climatology.html
Accessed on 25 February 2022

Use Case implementation:

Implementation of the use case is controlled by MATLAB script `triple_point_usecase_monthlyL3_v1.m`. This script is only valid for the dataset versions described above and provided in the `triple_point_usecase_monthlyL3_DATA.zip` file. The following MATLAB external packages are required and provided in `triple_point_usecase_monthlyL3_MATLAB_PACKAGES.zip`.

- Climate data toolbox function cdtarea.m is used to weight spatial error variances
  https://www.mathworks.com/matlabcentral/fileexchange/70338-climate-data-toolbox-for- matlab
- M_Map is used to plot spatial error variances
  https://www.eoas.ubc.ca/~rich/map.html
- cmocean is used for plotting colorbars
  https://www.mathworks.com/matlabcentral/fileexchange/57773-cmocean-perceptually-uniform-colormaps

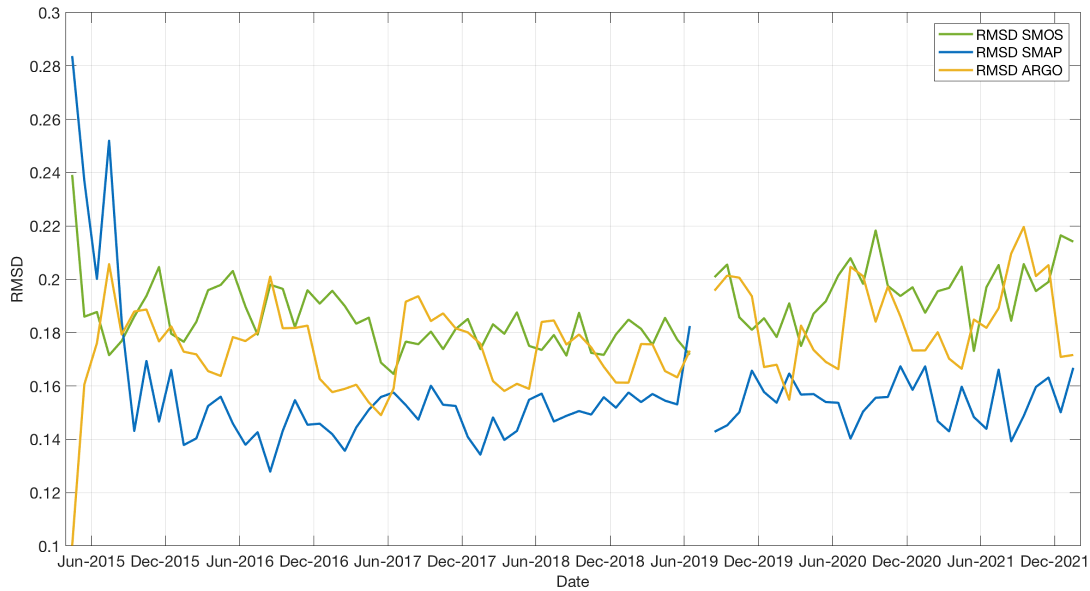
All user variables that may need to be edited for local environment and/or chosen run options are defined in the first section of script. Data location file paths should be edited to reflect the location of the data files provided once unzipped on ESA's machines. Additionally, start and end dates for the analysis can be edited. Valid dates are from 1 April 2015 to 31 January 2022. Finally, the option to include or exclude grid points with differences larger than 5 PSU between datasets from the triple point collocation analysis is given. These user defined variables were chosen to allow mapping to location of datasets in Pi-MEP and potential end-user-controlled GUI selections ESA may wish to implement.

Prior to calling the `covariance_triple_point_collocation.m` function, date, location, and salinity data for the 3 example datasets must be loaded and prepared. Data is reduced to the user defined time period. Common satellite salinity data quality flags were applied to SMAP data. Following guidance provided with the V4.0 product, data with a land fraction greater than 0.04 by antenna gain or 0.001 by 3dB contour and data with ice fraction greater than 0.003 were removed. Additionally, regions with surface temperatures less than 278.15K were removed. No additional quality flags are applied to SMOS data in the use case, however as a requirement of the triple point method, any grid point with a NaN for one dataset is excluded from the triple point analysis. Once all datasets are loaded, they are spatially regridded using linear 2-D interpolation to the coarsest dataset. For the use case, SMOS and SMAP are regridded to the RG Argo climatology spatial grid (1°x1°). Then, two different RMSD calculations are completed using `covariance_triple_point_collocation.m`. The decision to make `covariance_triple_point_collocation.m` two-dimensional was made to ensure maximum flexibility of the code provided. Finally, plots of both RMSDs are provided.

Spatially averaged, monthly RMSD:
For each month within the specified time range, a MxN (Lon, Lat) matrix of salinity values is passed to the triple point collocation function and a spatial mean monthly RMSE of each dataset returned. For spatial mean RMSD, a time series of errors for each dataset are plotted.

Inputs:
SMOS SSS: dimensions (M, N)
SMAP SSS: dimensions (M, N)
RG-Argo SSS: dimensions (M, N)

Outputs:
SMOS RMSD: dimensions (1x1)
SMAP RMSD: dimensions (1x1)
RG-Argo RMSD: dimensions (1x1)



Temporally averaged, spatial RMSD:

For each Lon, Lat pair, a MxN1 (Time, 1) matrix of salinity values is passed to the triple point collocation function and a temporal mean RMSE of each dataset returned. For temporally averaged measurements, spatial patterns of the temporal mean RMSE are plotted.

Inputs:

SMOS SSS: dimensions (M, N)

SMAP SSS: dimensions (M, N)

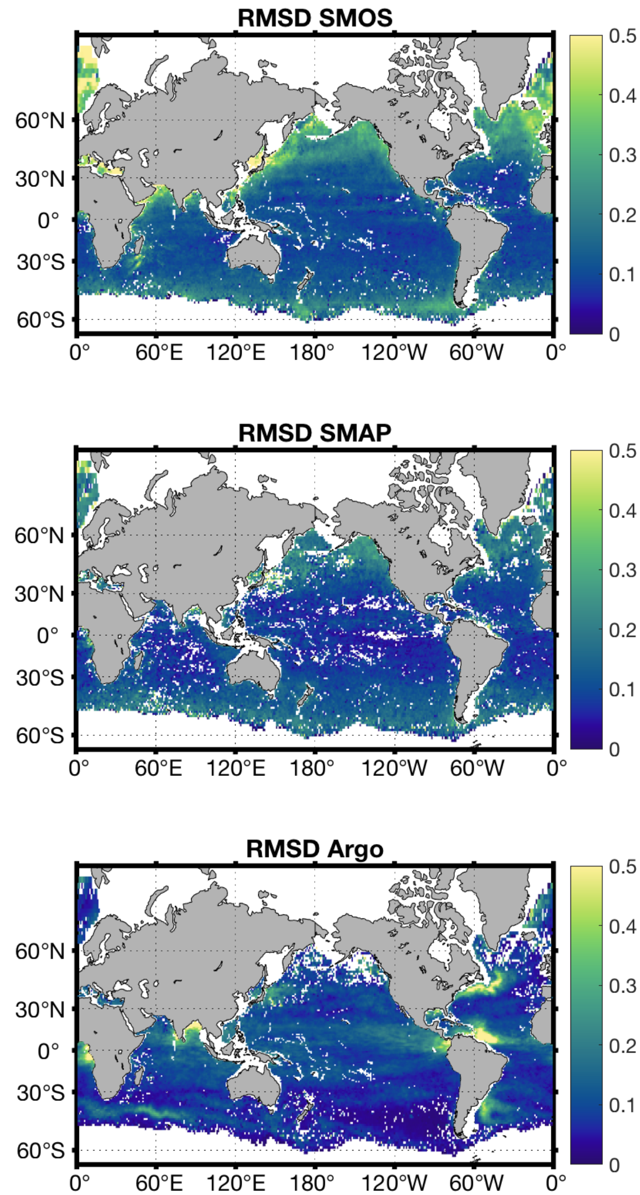
RG-Argo SSS: dimensions (M, N)

<u>Outputs:</u>

SMOS RMSD: dimensions (1x1)

SMAP RMSD: dimensions (1x1)

RG-Argo RMSD: dimensions (1x1)



Citations and Data Acknowledgments:

Boutin, J., Vergely, J. L., Marchand, S., d'Amico, F., Hasson, A., Kolodziejczyk, N., ... & Vialard, J. (2018). New SMOS Sea Surface Salinity with reduced systematic errors and improved variability. *Remote Sensing of Environment*, *214*, 115-134.

Gruber, A., Su, C. H., Zwieback, S., Crow, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation*, *45*, 200-211.

Kao, H. Y., Lagerloef, G. S., Lee, T., Melnichenko, O., Meissner, T., & Hacker, P. (2018). Assessment of aquarius sea surface salinity. *Remote Sensing*, *10*(9), 1341.

Meissner, T. and F. J. Wentz, 2019: Remote Sensing Sytems SMAP Ocean Surface Salinities, Version 4.0 validated release. Remote Sensing Systems, Santa Rosa, CA, USA.

Roemmich, D. and J. Gilson, 2009: The 2004-2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. Progress in Oceanography, 82, 81-100.

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of geophysical research: oceans*, *103*(C4), 7755-7766.