Error characterization of SSS products using triple collocation analysis

Nina Hoareau, M. Portabella, W. Lin, J. Ballabrera-Poy and A. Turiel
SSS: Essential Climate Variable

More than 9 years of remote sensing SSS data are available thanks to satellite passive L-band missions.

Satellite salinity measurements provide unprecedented spatio-temporal resolution/coverage as compared to any other observation system (Argo, CTD, TSG, moored buoys, etc.).

⇒ A comprehensive error characterization of the available SSS products is necessary.

In this study, we focus on the validation of SMOS-BEC and Aquarius v4 Level 3 products.
Limitations of direct comparison validation approach

In direct comparisons, *in situ* data are assumed to be true or perfect at satellite scales => only the relative error is estimated.

**Spatial and temporal representation** when comparing satellite gridded product versus in situ data:

- Vertical representation: first cm / ~5m
- Spatial representation: 0.25°/ point measurement
- Temporal representation: ~weekly / instantaneous (Argo)

=> The different spatial & temporal representation of the data will impact the direct comparisons and therefore needs to be accounted for during the validation process => **representativeness error**

**Absolute error** estimation requires at least 3 independent measurement systems.

=> Use of *Triple collocation Analysis*. 
Triple Collocation (Stoffelen, 1998)

- **Triple collocation** (TC) was conceived as a tool for **intercalibration** and **individual error assessment** of three different collocated WIND data sets (Stoffelen, 1998).

- Given 3 measurement systems with different spatial resolution (buoy, satellite, model), $s_i$, $i=1,2,3$, the measurement and its error are modelled by the following linear equation:

  $$s_i = a_i S + b_i + \delta_i$$

**TC model assumptions:**
- Errors are additive
- Error distributions are close to Gaussian
- The collocated data sources are independent
Triple Collocation (Stoffelen, 1998)

Representativeness error

Representativeness error ($r^2$) corresponds to the common variability resolved by Systems 1 and 2, but not resolved by system 3.
Triple Collocation (Stoffelen, 1998)

Representativeness error

- Suppose system 1 and 2 resolve smaller turbulent scales than system 3. The true variance common to these smaller scales is:

\[ r^2 = \langle \delta_1 \delta_2 \rangle \]

which is part of the measurement errors \( \delta_1 \) and \( \delta_2 \).

\( \Rightarrow r^2 \) is the **correlated part** of the errors of \( s_1 \) and \( s_2 \).

- Assuming that, since \( s_3 \) does not include these smaller scales, its measurement error \( \delta_3 \) is independent of \( \delta_1 \) and \( \delta_2 \), and:

\[ \langle \delta_1 \delta_3 \rangle = \langle \delta_2 \delta_3 \rangle = 0 \]

**Representativeness error** \((r^2)\) corresponds to the common true variance of Systems 1 and 2, not resolved by system 3.
TC algorithm

Initialization,

\( n = 0 \), \( i = [1,2,3] \)

\( S_i^n = \text{Collocated } i\text{-data set} \)

\( a_i^n = 1.0 \)

\( b_i^n = 0.0 \)

\[ a_i^n \leq 10^{-5} \]

\[ b_i^n \leq 10^{-5} \]

Convergence?

- No
- Yes

\( n = n + 1 \)

\( i = [2,3] \)

\( M_{ij}^n = \langle S_i^n S_j^n \rangle \)

\( a_2^n = \frac{M_{12}^n}{M_{13}^n} \)

\( b_2^n = \langle S_2^n \rangle - a_2^n \langle S_1^n \rangle \)

\( a_3^n = \frac{M_{23}^n}{M_{12}^n} \)

\( b_3^n = \langle S_3^n \rangle - a_3^n \langle S_1^n \rangle \)

\( S_i^{n+1} = a_i^n S_i^n + b_i^n \)

\( a_3 \delta_1 \text{ and } \delta_2 \text{ depends on } r^2 \)

Errors estimated after convergence and for calibrated data

\[ \varepsilon^2 = M_{12} - r^2 = M_{23} = M_{13} \]

\[ \langle \delta_i^2 \rangle = M_{ii} - \varepsilon^2 \]
**SSS data**

- **Space: The closest grid point to the in-situ location is used.**
- **Temporal: Collocation to the central day of Aquarius product.**

Total of 1456 collocations with the six products are obtained over the study period of 2013, in the Tropical band

=> Obtained sextuplets of TAO, SMOS, Aquarius, GLORYS2V3, WOA13, WOA09 collocated data.

**Period of study: 2013**
- All the SSS data sources are available at this period.
- 2013 is **not influenced by strong events** such as El Niño (2014-2015) or La Niña (2011-2012), which are known to be unresolved by the climatology, thus leading to strong biases in the latter.

<table>
<thead>
<tr>
<th>Product</th>
<th>Spatial Resolution</th>
<th>Temporal resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAO</td>
<td>Point</td>
<td>Daily average</td>
</tr>
<tr>
<td>GLORYS2V3</td>
<td>0.25º</td>
<td>Daily product</td>
</tr>
<tr>
<td>Aquarius v4</td>
<td>1º</td>
<td>7 days average</td>
</tr>
<tr>
<td>SMOS OA</td>
<td>0.25º</td>
<td>9 days average</td>
</tr>
<tr>
<td>WOA13</td>
<td>0.25º</td>
<td>Daily interpolation from monthly product</td>
</tr>
<tr>
<td>WOA09</td>
<td>1º</td>
<td>Monthly product</td>
</tr>
</tbody>
</table>
Representativeness error estimation: method

Until now, to estimate $r^2$ with sea surface wind data the methods have been based on:

- Integrating the difference between the scatterometer wind power density spectra (PDS) and those of the numerical model output (Vogelzang et al. 2011)
- Calculating the cumulative variance of scatterometer and model wind components (Vogelzang et al. 2015).

**Problem:** SSS PDS spectral slopes of the different products are sensitive to the presence of noise (see POSTER section this afternoon, based on Hoareau et al., TGRS, 2018).
Alternative approach based on TC intercalibration assumption (Lin et al., 2016):

**Assumption** that a successful TC provides three data sets well intercalibrated.

⇒ **TC calibration coefficient** $a_i, b_i$, are related to the value of $r^2$
⇒ Setting a wrong $r^2$ leads to a miscalibrated system 3 with respect to systems 1 and 2.

Therefore, an effective way of estimating $r^2$ is to repeat the TC analysis for different $r^2$ values until an optimal intercalibration of the different data sources is achieved.

=> Check the data scatterplots after each intercalibration
Representativeness error Estimation

Before TC

System 1

System 2

System 3

After TC

SSS data:
System 1 -> TAO
System 2 -> GLORYS
System 3 -> SMOS

If wrong $r^2$ ⇒ Not well calibrated

If correct $r^2$ => Well calibrated
Representativeness error Estimation: Results

*Slope values as a function of the representativeness error ($r^2$) for the triplets (Left) TAO-GLORYS-SMOS and (Right) TAO-Aquarius-SMOS.*

Correct $r^2$ values
Representativeness error Estimation: Results

**Acronyms:**
- T: TAO in-situ
- G: GLORYS model
- A: Aquarius satellite
- S: SMOS satellite
- 13: WOA 2013 climatology
- 09: WOA 2009 climatology

**Representativeness Error** ($r^2$) for the different triplets of SSS data.

<table>
<thead>
<tr>
<th></th>
<th>TGA</th>
<th>TGS</th>
<th>TA13</th>
<th>TS13</th>
<th>TA09</th>
<th>TS09</th>
<th>TAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r^2$</td>
<td>0.009</td>
<td>0.023</td>
<td>0.027</td>
<td>0.011</td>
<td>0.034</td>
<td>0.020</td>
<td>0.015</td>
</tr>
</tbody>
</table>

$r^2$ estimation with sextuplets => robust TC analysis results

$T_{TGS}^2 \approx T_{TGA}^2 + T_{TAS}^2$

$T_{TA13}^2 \approx T_{TAS}^2 + T_{TS13}^2$

⇒ $r^2$ values help to identify the systems having the finest and the coarsest effective spatiotemporal resolution:
TAO<GLORYS<Aquarius<SMOS<WOA13<WOA09
Random Error Estimation at Satellite resolved scales

The TAO error variation gives an indication of the \textit{uncertainty} of the proposed methodology about $0.01$

\begin{align*}
\text{At system 2 resolution: } \delta_{\text{TAO}} &= \sqrt{\delta_1^2 - r^2} \\
\text{At system 3 resolution: } \delta_{\text{TAO}} &= \delta_1
\end{align*}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
 & TAO & GLORYS2V3 & AV4 & SMOS & WOA13 & WOA09 \\
\hline
\textbf{Aquarius scale} & $0.18\pm0.01$ & 0.18 & $0.17\pm0.01$ & $0.24\pm0.01$ & 0.29 & 0.31 \\
\textbf{SMOS scale} & $0.22\pm0.01$ & 0.21 & $0.21\pm0.01$ & $0.20\pm0.01$ & 0.26 & 0.29 \\
\hline
\end{tabular}
\end{table}
Summary & Conclusions

The TC technique consists of using 3 independent, intercalibrated and collocated data sources to provide an estimate of their individual random error (SD).

1) The analysis has been carried out at the scales resolved by the two satellite products: SMOS Objective Analysis and Aquarius v4 Level 3.

2) The representative error has been accounted for during the TC validation of six different SSS products along the tropical band for the year 2013 => Sextuplets give robust TC analysis results.

3) The $r^2$ estimation method is based on the analysis of the intercalibration results.

4) It has been found that the representativeness error ($r^2$) contributes to 15% ~ 50% of the error estimates.

5) $r^2$ values help sorting the systems in terms of their effective spatiotemporal resolution:
   TAO < GLORYS2V3 < Aquarius v4 < SMOS OA < WOA13 < WOA09

6) The TC method developed here leads to an uncertainty of about 0.01 in the SSS error estimates.
The validation has been carried out at the satellite-resolved spatiotemporal scales.

It has been found that the TAO SD error at the Aquarius v4 and SMOS OA spatiotemporal scales is 0.18 and 0.22, respectively.

=> The error values include the contribution of the following representativeness errors:
- the horizontal scale difference between the point-wise observation and the 0.25°-1° grid sizes of the satellite products
- the vertical mismatch between TAO measurement at 1-1.5m depth and the satellite at 1 cm depth
- the different temporal resolution of TAO (1 day) and satellite products (7-9 days).

The partition of these error contributions remains a research topic in oceanography.