Supplement of Atmos. Meas. Tech., 17, 2739–2759, 2024
https://doi.org/10.5194/amt-17-2739-2024-supplement

Atmospheric Measurement Techniques
Open Access

Supplement of

Retrievals of aerosol optical depth over the western North Atlantic Ocean during ACTIVATE

Leong Wai Siu et al.

Correspondence to: Leong Wai Siu (leongwaisiu@arizona.edu)

Supplement

S1 Triple collocation analysis with synthetic datasets

In this study, two aerosol optical depth datasets RSP and HSRL-2 were obtained from the ACTIVATE field campaign. And then a third dataset MODIS from satellite was introduced for the triple collocation analysis. There are two purposes for the following analysis:

1. Test the sensitivity of the data quality of the third dataset on the result of the triple collocation analysis.

2. Test the sensitivity of the potential bias within the third dataset (e.g., due to the bias from combining two satellite datasets) on the result of the triple collocation analysis.

To generalize the results, it will be helpful to study synthetic datasets. This approach has been used in some triple collocation studies (e.g., Su et al. (2014)).

Assume that we know the ground truth, which is given as a sinusoidal function, say, one full cycle with mean = 0.1, minimum = 0, maximum = 0.2. For example, in python, such an array with 5000 data points can be created using the numpy library from the following code:

```
ds = np.sin(np.linspace(0, 2*np.pi, num=5000))*0.1 + 0.1
```

Then we can add some white noise to the ground truth to generate synthetic datasets. The amount of noise is controlled by a signal-to-noise ratio (SNR), which is simply the ratio of ground truth variance and noise variance. To give some context using real world data, we can estimate the SNR of the datasets in this study by $b_i^2 \mathrm{Var}[\Theta]/\mathrm{Var}[\varepsilon_i]$ (Gruber et al., 2016; McColl et al., 2014). The SNRs of RSP, HSRL-2, and MODIS (2344 triplets in total) are 1.74, 5.98, and 2.79, respectively.

We carry out experiments 1–3 with three synthetic datasets A, B, and C, each having 5000 data points (Table S1). The only difference is the SNR of dataset C (1, 5, and 25). Note that all datasets are re-generated for each experiment, so, for example, dataset A in experiment 1 is not the same as dataset A in experiment 2. The result shows that all TC metrics of datasets A and B are not very *sensitive* to the amount of noise in dataset C (i.e., the quality of dataset C).

Table S1: Summary of the synthetic experiments 1–3 with $n = 5000$ data points. Signal-to-noise ratio (SNR), error standard deviation σ_ε, and error correlation coefficient r with respect to the ground truth are provided for each synthetic dataset.

Expt	n	A			B			C		
		SNR	σ_{ε_A}	r_A	SNR	σ_{ε_B}	r_B	SNR	σ_{ε_C}	r_C
1	5000	5	0.0329	0.906	5	0.0311	0.915	1	0.0699	0.713
2	5000	5	0.0312	0.916	5	0.0314	0.915	5	0.0308	0.917
3	5000	5	0.0325	0.909	5	0.0315	0.913	25	0.0134	0.983

We now repeat the same set of experiments but with only 500 data points (Table S2). All TC metrics of datasets A and B now have larger variations. This shows that the robustness of the TC metrics is *more sensitive* to the number of data points than the data quality. Therefore, in this study we try to maximize the number of data points, even it may introduce some small biases.

One powerful benefit of this synthetic data approach is that since both the ground truth and errors are *known*, we do not need to compute all the TC metrics via the equations given in the

Table S2: Summary of the synthetic experiments 4–6 with $n = 500$ data points. Signal-to-noise ratio (SNR), error standard deviation σ_ε, and error correlation coefficient r with respect to the ground truth are provided for each synthetic dataset.

Expt	n	A			B			C		
		SNR	σ_{ε_A}	r_A	SNR	σ_{ε_B}	r_B	SNR	σ_{ε_C}	r_C
4	500	5	0.0344	0.891	5	0.0316	0.911	1	0.0735	0.671
5	500	5	0.0312	0.917	5	0.0319	0.909	5	0.0310	0.914
6	500	5	0.0311	0.918	5	0.0306	0.921	25	0.0175	0.972

manuscript using synthetic datasets. That is to say, we can compute all these TC metrics directly from the ground truth and errors.

To test the potential bias within the third dataset, we use the similar setup of experiments 1–3 but *randomly* add some bias, to 50% of data points (i.e., 2500) in the third dataset. In python, the indices of these data points can be generated using the numpy library with the following code:

```
ind = np.random.choice(5000, 2500, replace=False)
```

We now have experiments 7–9 with the required synthetic datasets A, B, and C with a bias of 0.05 on 2500 randomly chosen data points of C (Table S3). The TC metrics of datasets A and B are highly consistent in all cases. While the true bias between Terra and Aqua is not known, it should not be too big given that the expected error of the MODIS Dark Target retrieval algorithm is ~0.03–0.05 for the AOD range in the study region. Therefore, even if Terra and Aqua may be biased differently, the effect of this particular bias should be small on the triple collocation analysis.

Table S3: Summary of the synthetic experiments 7–9 with a bias of 0.05 on half of the dataset C and $n = 5000$ data points. Signal-to-noise ratio (SNR), error standard deviation σ_ε, and error correlation coefficient r with respect to the ground truth are provided for each synthetic dataset.

Expt	Bias	A			B			C		
		SNR	σ_{ε_A}	r_A	SNR	σ_{ε_B}	r_B	SNR	σ_{ε_C}	r_C
7	0.05	5	0.0323	0.908	5	0.0315	0.914	1	0.0751	0.692
8	0.05	5	0.0315	0.913	5	0.0318	0.913	5	0.0401	0.871
9	0.05	5	0.0311	0.915	5	0.0317	0.913	25	0.0288	0.925

Reference

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, International Journal of Applied Earth Observation and Geoinformation, 45, 200–211, doi: 10.1016/j.jag.2015.09.002, 2016.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophysical Research Letters, 41, 6229–6236, doi: 10.1002/2014gl061322, 2014.

Su, C.-H., Ryu, D., Crow, W. T., and Western, A. W.: Beyond triple collocation: Applications to soil moisture monitoring, Journal of Geophysical Research: Atmospheres, 119, 6419–6439, doi: 10.1002/2013jd021043, 2014.