Zibordi and Berthon,"Coastal Atmosphere & Sea Time Series (CoASTS) and Bio-Optical mapping of Marine optical Properties (BiOMaP): the CoASTS-BiOMaP dataset"

Overall recommendation: Minor Revision

General comments

Public release of CoASTS and BiOMaP is exciting for the ocean color community as these comprehensive data sets have strong value for algorithm development and validation activities. This paper provides an overview of the data sets, methods used, an assessment of errors, and serves as a quick reference guide.

A question is how much of these data have been made publicly available previously in other compendiums such as the Vicente et al. (2019, 2022) and NASA SeaBASS. It is important to specify, moreover to ensure data is not duplicated in any future analyses.

Data sets extending 2 decades can be relevant to climate change studies. It would be useful to state this and use the term in key words.

Comments below are intended to compliment, not duplicate, the excellent comments by reviewer J Pitarch (JP), which I have read. I have also read the authors' replies. On these specific comments and replies, I only comment here where there may perhaps be some disagreement and I have a strong opinion.

The authors discuss some data being consistent with Case 1 or Case 2 waters. Since the authors mention the topic and it is relevant to intended applications, it would be useful to include some estimate of %Case 1 vs Case2 in Tables 1 and 2. While the practical application of Case 1 v Case 2 designations can be ambiguous, there are published quantitative metrics for this that would be very straightforward to implement. Even if approximate, providing these general water type estimates would be useful to many who will use these data.

Lines 189-208: the extrapolation and derivation of slopes for irradiance profiles is described as taking the log and fitting a line. The most accurate method is fitting the nonlinear exp relationship to the profile data. Derived slopes will be different between the two methods because assumed error distributions get skewed after taking the log, which is inaccurate. The authors appear (understandably) reluctant to revisit processing procedures for these very large data sets, but it would be a small effort to select a representative smattering of profiles from each campaign and apply both methods so an estimate of related biases in derived parameters such as Kd could be given.

It is also stated that spikes above 3 std due to wave focusing were rejected from radiometer profile data. However, there is nothing wrong with this radiometric data and it should be included in any fit; these spikes can make a significant difference. If a time series was collected at depth we would absolutely want to include the full time series in deriving average radiometric intensities. These spikes can be orders magnitude greater than average intensity at a particular depth (see Stramski's work on this). If light is being focused by a wave at any moment during a profile, surrounding data points will be affected by defocusing and thus be deficient in intensity relative to a time series average at that depth. Spikes due to focusing should be included. Again, maybe an analysis can be carried out on a subset of the data to gauge potential associated biases.

Similarly, some "tuned" automated outlier removal algorithm was apparently used for all the IOP data, removing measurements "exhibiting poor spectral and spatial (i.e., vertical) consistency" but neither the "filtering process," criteria for "consistency" or "extreme differences," or the approach to "tuning" are provided. These details are needed for a reader to understand how the data was processed. It is furthermore stated in line 293 that the filtering removed spikes from bubbles and large particles. If effects of bubbles are removed due incomplete air evacuation in water, this is absolutely appropriate and typically only occurs at the very beginning of data records, as the plumbing soon clears of air. However, if the filtering is also removing spikes during profiles of "large particles" and data "exhibiting pronounced differences with respect to those characterizing the mean of profile spectra," this can be highly problematic. There is no justification for removing spikes in IOPs from large particles. In some particle fields comprised mostly of large detrital aggregates or large colonial plankton, almost all the IOP signal can come from significant spikes associated with numerous large particles. These large particles are inevitably undersampled by the relatively small sample volumes of AC devices and bb sensors, so there is likely residual bias in our measurements relative to the GSD of a satellite unless long in-water time series were recorded, but removing spikes of good data from large particles would certainly exacerbate any bias. Similarly, significant work was done in the 1990's and 2000's on the optical properties of thin layers, which can be intense (order of magnitude higher than background) layers of particles less than a meter thick and have strong effects on ocean color (Petrenko et al. 1998; Zaneveld and Pegau 1998). These layers are common throughout the coastal and open ocean. Would your filtering approach remove these effects?

In section 4.2, it is stated that measurements were processed in accordance with guidance from the manufacturer (WET Labs 1996), but this guidance has always been insufficient and antiquated relative to the best methods agreed upon by the community. These best practices have been maintained in published IOCCG Protocols that have recently been updated. Methods here should cite relevant chapters from the Protocols and provide detail on any deviations with related impacts to data quality.

I strongly agree with JP that the a_nw(715) value should be published in these datasets. As JP states, many would argue the method for the scattering correction applied here is not the most accurate. Including a_nw(715) enables the community to apply other published scattering corrections and possibly other scattering corrections developed in the future.

As JP states, the 0.4 factor for the Hydroscat correction is problematic, but any value is guessing really. There is also no separation of a constant water background in Eq 4, which has always been inherently problematic. The only thing we can really do however is acknowledge what the realistic errors for this sensor are.

Was there replication for the TSM measurements? It looks like there was in some cases but was this standard practice? Please clarify.

I strongly suggest including histogram plots of *c*(490 or 532) and SPM as was done for Chl in Fig. 7. These are quick diagnostics for water types for your reader and contribute to the objective of this paper as an overview and guide for the data set.

I agree with JP that the inclusion of negative values for parameters such as bbp suggests a lack of rigorous QA/QC. I suggest if you choose to include, add a statement this is a conscious decision and that such negative values “remain within expected errors reported herein” (if you agree with that statement).

Regarding the plots, an IOP plot I find is a strong diagnostic of the quality of a data set while also being a strong proxy for particle composition is *bb/b*. This parameter incorporates *a* and *c* measurements from the AC device as well as *bb* from the Hydroscat and falls within a relatively narrow range of about 0.04 to 0.3. I would suggest the authors add this plot.

Moreover, more attention could/should be given to the robustness of the data, QA/QC, and error assessments here. In my opinion, addressing the quality of the data set in a rigorous manner is what elevates this paper to a peer-reviewed contribution as opposed to a simple introduction and guide to these data sets, which could just be posted as a readme online with the data sets. Reviewer JP suggests a closure analysis would be a straightforward means of assessing the inherent robustness of the data sets – I thought the same thing in reading the manuscript and strongly agree, this is a super idea. Such an analysis effectively boils down all disparate bias and random errors in the entire data set down to one error number. As such I disagree with the authors’ comment such an assessment is beyond the scope of the paper. Closure results can also be directly compared to a handful of other closure analyses with high quality data such as Pitarch et al. (2016) and Tonizzo et al. (2017) and would provide an immediate comprehensive gauge of quality. But not only did J Pitarch suggest such an analysis, I believe we are all indebted to JP for actually doing the assessment in his review! I was not able to access the figures from his review online, but he states the results appear good. At the very least, the authors should reference JP’s closure assessment in the online ESSD Discussion (I assume these stay online indefinitely?), provide the salient results, and make a statement as to how these results compare with previous closure assessments from the literature. Well done, Jaime, we all thank you, this is an important contribution! If the Editor is looking for Reviewer awards, you get my vote 😉.

Specific comments

Section 3 title: suggest “Measurements” should be “Measurements overview”

Section 3.f: I believe a_p, a_ph, and a_dt were measured. This sentence should be reworded to be precise.

Section 3.i: “Total suspended matter (TSM)” is not precise since a filter was used with some pore size cutoff, thus “total” particles were not assessed. The convention that is often used is “Suspended particulate matter (SPM).”