Reply to Reviewers and Action

The two outstanding Reviewers, Jaime Pitarch and Mike Twardowski, provided a number of valuable comments and suggestions to the manuscript. Their effort is fully appreciated.

An attempt has been made to take actions in reply to each comment when considered justified by the Journal policy. In fact, a main general suggestion by both Reviewers is to provide more analysis on the dataset. Actually, some more analysis have been included in the manuscript (*i.e.*, the distribution of *SPM* values and the plot of *bbp*/*bp* vs *Chla*). It is however recalled that ESSD papers (see https://www.earth-system-science-data.net/about/manuscript_types.html) *should highlight and emphasize the quality, usability, and accessibility of the dataset, database, or other data product and should describe extensive carefully prepared metadata and file structures at the data repository. … Although examples of data outcomes may prove necessary to demonstrate data quality, extensive interpretations of data – i.e. detailed analysis as an author might report in a research article – remain outside the scope of this data journal.*

Because of the above policy, the Authors feel that some of the analysis proposed by the Reviewers (with specific reference to the *Closure* one) are not essential and that the quality of the data is already proven by the elements included in the manuscript and also by the work documented in the former decades by the Authors themselves.

Also accounting for the suggestions provided by the Reviewers, the dataset has been revised: *i.* the absorption values determined with the AC9 at 715 nm and not corrected for the scattering offset, have been added; and ii. some data records mostly from the Mediterranean Sea not passing basic quality control criteria, but not flagged in the version formerly submitted to PANGAEA, have been removed. This implied a new submission of the dataset to PANGAEA. Because of this, the revised dataset has a new reference number https://doi.org/10.1594/PANGAEA.971945 and its finalization is ongoing.

The comments from the Reviewers are all itemized in the following sections. A reply and the related actions are then provided for each comment.

#1 Replies to the comments by Jaime Pitarch and actions

General comment by the Reviewer
I am very pleased to have been given the opportunity to review this manuscript as I am aware of the lifetime work of the authors in defining the highest standards and producing high quality reference data in the field of satellite ocean color. The monitoring programs CoASTS and BiOMaP have generated lots of publications, and a remaining question was where all the data was going to be after the finalization of such programs. So now it appears that a circle is closed.
I have read the paper and downloaded the dataset. Before publication, I have a number of comments of varying importance that, to my understanding, need attention.

Major comments by the Reviewer

Comment #1
Absorption from water samples is only provided at the Satlantic bands, which is regretful, as it was measured hyperspectrally. I do not know the reason to downgrade the data, and it definitely reduces its value for optical studies, also considering the growing interest in hyperspectral data (e.q., PACE). The authors are encouraged to submit the hyperspectral data.
Reply
The CoASTS-BiOMaP dataset submitted to PANGAEA was conceived to support bio-optical investigations with comprehensive multi-parametric and consistent near-surface quantities. Because of this, the laboratory absorption measurements were only provided at the center-wavelengths of the related multi-spectral field radiometric data.
It is definitively appreciated that *i.* the CoASTS and BiOMaP measurements can support a number bio-optical, methodological and instrumental applications beyond the strict and obvious bio-optical ones and that *ii.* hyperspectral measurements (when available) or also full profiles instead of the sole near-surface data, are relevant and desirable data. But this is something that goes beyond the objectives of the current work. An expansion of the dataset could be considered as a future task, but not for the current data submission.
Action
The objective of the work and of the related dataset is strengthened in the introduction.

Comment #2
Paragraph from line 175 to 182: on the above-water reference sensor, I see the correction for the imperfect non-cosine response. What about other uncertainty sources such as temperature and non-linearity, as it is recommended in above-water radiometry (e.g., Trios)? And are any of these corrections made to the in-water sensors?
Reply
In agreement with current know-how, multi-spectral radiometers rely on a much simpler design and technology with respect to the hyper-spectral ones. Because of this, they exhibit lower sources of significant uncertainties: the temperature dependence is negligible within the 410-700 spectral range (Zibordi et al. JTECH 2017); stray-lights are negligible assuming interference filters are of high quality and their out-of-band response is within specifications (Johnson et al. AO 2021). Because of this, some uncertainties due to the potential non-ideal performance of multi-spectral radiometers are commonly not included in uncertainty budgets. Exception is the non-cosine response of irradiance sensors, which heavily depends on the manufacturing process and optical materials used. This may lead to substantial differences across individual irradiance collectors.
Action
The above elements are now mentioned in the relevant section.

Comment #3

Line 193-196: the interval 0.3 m – 5 m looks arbitrary. Any comments on why this choice is appropriate? Does it relate to the unphysical Kd values that I report below?

Reply

The determination of subsurface radiometric values from profile data always requires the identification of a suitable near surface "extrapolation interval" exhibiting linear dependence of log-transformed radiometric data with depth. In the case of the CoASTS-BiOMaP data generally collected in coastal waters, the most appropriate extrapolation intervals were determined within the 0.3 and 5 m depth limits (these limits do not identify the extrapolation intervals themselves, but the values within which the extrapolation interval is generally located).

How the extrapolation intervals are determined is well stated in the manuscript. The process is definitively subjective (*i.e.*, the extrapolation interval is chosen by an analyst with the aid of a number of ancillary information), but for sure, it is not arbitrary.

The question on K_d is addressed in the reply provided to the comment #18.

Action

The text has been improved to avoid misinterpretation on the actual extrapolation intervals and the depth limits within which they are commonly located.

Comment #4

The transmission of upwelling radiance through the surface to form the water-leaving radiance is made with that 0.544 factor, which is not up to date with today's knowledge. I suspect that the reason to choose this value is because the difference in the final product would be minimal when using another one. However, I report evidence that this is not the case.

For a flat surface, the relationship between in-water and in-air upwelling radiance is: $Lw=\tau w,Lu(0-)$

Where $\tau w,=1-\rho nw2$

ρ is the Fresnel reflectance of the air-sea interface. Assuming unpolarized light, it has an analytical expression $\rho=12|\sin2(\theta a-\theta w)\sin2(\theta a+\theta w)+\tan2(\theta a-\theta w)\tan2(\theta a+\theta w)|$

θa and θw are the wave propagation angles in air and in water, respectively, and are related by $\sin(\theta a)=nw\sin(\theta w)$

For $\theta a=\theta w=0$, there is a singularity. One can apply the small angle approximations for the trigonometric functions, so in the limit, it is: $(\theta a=\theta w=0)=(nw-nanw+na)2$

It is commonly accepted now that it is inaccurate to assume a constant $\tau w,a$ for a given geometry due to the spectral dependence of nw (and secondary influence by temperature and salinity too). Such dependences are taken from the state-of-the-art values by Roettgers et al. (2016). Therefore, the theoretical curve for τw, can be seen in Fig. 1. In addition, I have made some Hydrolight simulations, in which the same nw values are used, but also, the transmission is affected by the surface roughness depending on the wind speed. What emerges from Fig. 1 is that increasing wind speeds reduces light transmission. In terms of the total error made by assuming $\tau w,a=0.544$, it may not seem much, but in reality, they are in the order of 1-2%, which accounts for about 20% of the total uncertainty reported for the final Rrs product. Therefore, to reduce total uncertainty I encourage the authors to consider updated look up tables for τw.

Reply

The spectral dependence of the water-air transmission factor for a flat sea surface in the interval of interest for the CoASTS-BiOMaP dataset is well within +/-1%. This is confirmed by the data provided by the Reviewer and also by the work of Voss and Flora (JTECH, 2017). Because of this, any carelessness on the spectral dependence of the water-air transmission factor does not appreciably affect the uncertainty budget of the derived radiometric quantities in the spectral range of the CoASTS-BiOMaP data.

The inclusion of the wind speed dependence in such a transmission factor adds the detrimental dependence on wind speed to L_w. In fact a properly determined subsurface L_u is marginally affected by sea state (and consequently by the wind speed). Thus introducing a wind speed dependence on the water-air transmission factor would add such a dependence to L_w. This is not a desirable dependence for data envisaged to support bio-optical modelling.

Action

The spectral dependence of the water-air transmittance is now well stated. The uncertainties affecting spectral L_w when not considering such a dependence, are also stated .

Comment #5

On the fit of simple analytical functions to variables like the Q factor and possibly others like *bb*, I did not understand if the actual data were replaced by the fits. If that is the case, I prefer then to have both data and uncertainty rather than their surrogate analytical forms.

Reply

The fit of Qn data is introduced to minimize the impact of any uncertainty affecting intra-band radiometric calibrations of multi-spectral radiometers or surface perturbations affecting the extrapolation process. In practice, any deviation indicated by the fit with respect to the actual Qn values is used to monitor the performance of the L_u and E_u sensors in the field (deviations of +/-1% are typical, variations exceeding +/-2% are a warning). The fitted Qn data are those saved and included in the shared dataset. Still, both L_u and E_u data are provided, any data user may re-compute the data at his preference if interested in actual Qn data or in assessing their uncertainties.

Action

More details have been provided on Qn fitting, but no additional action was taken. Quality controlled and "smoothed" Qn values are those expected to best serve the community.

Comment #6

On the bidirectional correction in lines 233-241, it is a bit disturbing to read in line 241 that in case 2 waters "this correction may be affected by large uncertainties". There are significant parts of the dataset in case 2 waters. How large are those uncertainties? Ongoing research has proven that is it better to apply Morel than not to apply any correction at all, and Morel has shown to provide surprisingly good results in case 2 waters, not because of the qualities of the model itself, but because all bidirectional correction models underestimate the correction to be made, but chlorophyll is overestimated in case 2 waters with the band ratio of Morel, which produces a higher correction, that ends up being beneficial. In any case, I believe that this part of the processing will need update to be in line with latest developments in bidirectional studies, knowing the interest of the authors in keeping the uncertainty budget as low as possible.

Reply

The sentence "this correction may be affected by large uncertainties" addressed to the Morel et al. (AO, 2002) corrections for bidirectional effects applied to non-Case 1 waters, is well supported by the work of Talone et al. (OE, 2018). The corrections applied to CoASTS–BiOMaP data rely on actual *Chla* values from HPLC analysis and not from any algorithm. It is emphasized that all the fundamental data required for producing alternative corrections for bidirectional effects are available: any user can thus implement its own ignoring that applied in the current dataset.

Action

The potential for producing high level radiometric data products with alternative corrections for bidirectional effects, is mentioned.

Comment #7

On the ac-9 measurements, I have several comments that follow.

How regular were the factory calibrations? It is said that instruments have to be calibrated before and after any campaign. Is this the case with the ac-9?

Reply

The two AC9s operated during the CoASTS and BiOMaP campaigns were factory calibrated on a yearly basis (obviously with a number of exceptions over almost three decades). Definitively, the instruments were sent to the manufacturer for maintenance and calibration each time there was evidence of sensitivity decay in a single band (implying the replacement of the related filter and detector). The pre- and post-campaign "calibrations" refer to measurements of the milli-Q water offset performed by the JRC team with the instruments in their deployment configuration. These measurements were intended to correct for any offset affecting the factory calibration coefficients over time (*i.e.*, between successive factory calibrations).

Action
Factory and field calibrations are now better explained.

Comment #8
The Zaneveld method does not correct the non-finite acceptance angle of the c detectors as it is stated (note that the "c" is missing in line 277), and in fact it is rarely corrected by anybody. To do that, one should have a guess of the VSF between 0 and 0.93 degrees, but in any case, the "real" $ct{-}w$ is higher than the measured than the factor that varies a lot, mostly between 1 and 2.
Reply
Thanks you for catching the inappropriateness of the statement on the correction for the non-finite acceptance angle.
Definitively, Boss et al., (2009) suggested corrections based on ct-w(AC9)/ct-w(LISST-F) = 0.56 (0.40-0.73). But it would have been speculative any correction not supported by dedicated VSF measurements.
Action
The text has been revised declaring that corrections are not applied for the non-finite acceptance angle of the c tube plus detector.

Comment #9
On the scattering correction method of the absorption data from the "a" tube, I also believe that the Zaneveld method questionable. Zaneveld overcorrects the absorption data, which leads to an underestimation. I see indirect evidence of it in Figure 5 from the manuscript, where the absorption comparison at 443 nm almost always shows negative biases with respect to the laboratory measurements (although the ac-9 provides better closure of Rrs than the water samples as I show below, so this is puzzling and needs to be addressed by the authors). I suggest using the method by Roettgers et al. (2013), that, if applied, is supposed to perform much better. This choice should be in line with authors approach of using only consolidated methods, approved by the two very good assessments by Stockley et al. (2017) and Kostakis et al. (2021).
Reply
Roettgers et al. (2013) showed that the Zaneveld et al. (1994) underestimates the absorption at wavelengths greater than 550 nm. In the blue and blue-green, and in particular at 443 nm, the agreement was shown quite good between the AC9 and the "true" absorption value from a PSICAM. Stockley et al. (2017) observed relative errors lower than 20% for the Zaneveld et al. (1994) correction in the spectral range 412-550 nm (lower than 10% for wavelengths 412-488nm). Thus, the negative biases observed at 443 nm documented in the manuscript and mentioned by the reviewer, cannot be explained only by the scattering correction method by Zaneveld et al. (1994).
Also, the hypothesis of negligible non-water absorption in the NIR was shown to be questionable for highly turbid waters (*e.g.*, Elbe River, Baltic Sea and North Sea), but acceptable for the oligotrophic Mediterranean Sea waters (Stockley et al., 2017).
It is agreed that the correction method proposed by Roettgers et al. (2013) and verified by Stockley et al. (2017) is definitively a progress with respect to Zaneveld et al (1994), in particular in the green and red spectral regions, but his universal applicability is not assured. An excerpt from Stockley et al. (2017) states: "*The performance of the empirical approach is encouraging as it relies only on the ac meter measurement and may be readily applied to historical data, although there are inevitably some inherent assumptions about particle composition that hinder universal applicability.*"
Also from Stockley et al. (2017): "*Methods experience the greatest difficulty providing accurate estimates in highly absorbing waters and at wavelengths greater than about 600 nm. In fact, residual errors of 20% or more were still observed with the best performing scattering correction methods*."
Considering the above findings, the AC9 data are provided with the correction originally proposed by Zaneveld et al. 1994, still appreciating it is far from being the most accurate. In the manuscript this is explicitly acknowledged through the comparison of absorption measurements from the AC9 with those from laboratory measurements performed on discrete water samples.

Action
Some of the above elements are now included in the manuscript to support the preference to process the AC9 data applying the correction scheme proposed by Zaneveld et al. (1994).
Additionally, accounting for a suggestion by the reviewers, the AC9 absorption values at 715 nm are provided without any correction applied for the scattering offset. This solution is intended to support the implementation of alternative corrections by any data user.

Comment #10
In fact, for research purposes, it is recommended that the authors share the absorption coefficient uncorrected for residual scattering, so it can be useful material to further investigate this matter.
Reply
This request was considered.
Action
The AC9 absorption values at 715 nm without any correction for the scattering offset are now provided. As already stated, this solution is intended to support the implementation of alternative corrections by any data user.

Comment #11
On the quantification of the uncertainties coming from the ac-9, certainly the value 0.005 $m{-}1$ is not a proper estimate. That is a rule of thumb estimate of the instrument precision in the user manual, which is accompanied by the 0.01 $m{-}1$ accuracy, also in the manual. There is no mention of uncertainty sources related to instrument absolute calibration, non-linearity, determination of the pure water measurement, correction of the temperature and salinity differences and correction of the residual scatter, some others related to the measurement protocol and the individual operator, and even some others that I may have missed. All these sources are likely to result in something bolder than the manufacturer user manual. The authors are expected and encouraged to investigate and comment on these aspects. Otherwise, how does one explain the differences that the authors find in their Figure 5?
Reply
Based on theoretical Monte Carlo computations, Leymarie et al. (2010) provided estimations of relative errors ranging from 10 to 40% for $c_{t\text{-}w}$ and generally lower than 25% for $a_{t\text{-}w}$ (5-10% when the absorption by in water optically active components is high), but up to 100% for waters showing high scattering.
Stockley et al. (2017) observed relative errors lower than 20% for the Zaneveld et al. 1994 correction for wavelengths in the range of 412-550nm (lower than 10% for wavelengths 412-488 nm) and more than 50% for wavelengths greater 600 nm. Twardowski et al. (2018) provided an estimate of the "operational" uncertainty (for example, considering 2 calibrated AC9 close one to the other) as low as 0.004 m^{-1} (not taking into account errors associated to the scattering corrections).
Action
Considering the above results the manuscript has been revised indicating that the uncertainties in AC9 absorption are larger than 0.005 m^{-1}, and can reach several ten percent in highly scattering waters with values more pronounced in the blue-green spectral regions.

Comment #12
I also have a few concerns about the Hydroscat backscattering data. First, in lines 331 and 332, what is exactly meant with the annual factory calibration "complemented" by pre-field calibration, in terms of determining the scale factor and the dark offset of the measurement?
Reply
Equivalent to the procedure put in place for the two AC9s used within the framework of the CoASTS-BiOMaP campaigns, also for the two HydroScat-6 instruments, there were regular factory calibrations tentatively performed on a yearly basis. The pre-field and post-field calibrations (leading to the determination of the spectral "Mu" response coefficients and gain ratios) performed in laboratory by the JRC team with a "calibration cube" and a spectralon reference plaque allowed to detect and correct sensitivity changes between successive factory calibrations.

Action
The difference between factory and pre-field calibrations has been clarified.

Comment #13
Equation (4) is the correction for absorption along the pathlength recommended by the manufacturer. However, after investigating on it, Doxaran et al. (2016) investigated on it and found that the “0.4” is a totally arbitrary number. They proposed a more accurate expression instead.
Reply
Doxaran et al. (2016) provided findings on the basis of measurements performed in: *i*. Río de la Plata turbid waters (Argentina, with total scattering coefficient greater than 20 m^{-1} at 550 nm and a tentative average value around 50 m^{-1}) and ii. Bay of Bourgneuf Waters (France, with total scattering coefficient greater than 10 m^{-1} at 550 nm with a tentative average around 40 m^{-1}).
Because of this, the empirical relationship provided by Doxaran et al. (2016) indicating K_{bb}-a_{nw}=4.34*b_b (see their Fig. 5b for the HydroScat-6) refers to values of b_b spanning between 0 and 2.5 m^{-1}.
The b_b BiOMaP values roughly range between 0.0005 and 0.1 m^{-1} (with values of a_{nw} lower than 1.0 m^{-1}). In this interval of b_b, the Fig. 5 by Doxaran et al. (2016) shows values following a relationship with a much higher slope than the empirical fit resulting from the whole range of b_b values. Thus, is that empirical fit really more appropriate for the low b_b values than the standard relationship used here? For sure, the problem is an open one.
Action
In the manuscript it is now stated that in the absence of any consolidated processing for HdroScat-6 data, the CoASTS-BiOMaP processing was made relying on the equations provided by the manufacturer.

Comment #14
Removal of pure water data is made after tabulated data by either salt water or fresh water by Morel, but the state of the art values are those given by Zhang et al. (2009). Their model is analytical and has an explicit dependency on salinity, so that one may use concurrent CTD data for obtain *bbw* accurately. Again here, the differences on the final products are likely to be small, but it is preferrable to replace old and biased values with updated ones at zero cost.
Reply
Zhang analytical values are for sure a general improvement with respect to the use of the values from Morel (1974) and fitted according to Twardowski et al. (2007). However, in the majority of cases it would have an almost negligible effect on the retrieval of CoASTS-BiOMaP b_{bp}. For instance, in the oligotrophic clear waters of the eastern Mediterranean Sea showing salinity values around 38.0-39.0, the difference in Beta(90degrees) between Morel (1974) and Zhang et al. (2009) is very low, *i.e.,* approximately 0.00004 m^{-1}.
Action
The manuscript now mentions the corrections for b_w relying on Zhang et al. (2009), alternative to the application of data from Morel (1974).

Comment #15
As for the ac-9 data, estimating an uncertainty of 0.0007 $m{-}1$ for bbp is wishful thinking. True uncertainties are much larger than that and are the result of a number of factors like those listed above. Can the authors look for a more realistic value based on their own research or in literature?
Reply
The value of 0.0007 m^{-1} was estimated by Whitmire et al (2007). The actual uncertainty is expected to be higher and dependent on many factors related to processing hypotheses (like the correction for the attenuation along the pathlength evoked above). A further uncertainty source is that related to the choice of the “chi” value for converting Beta140 into b_b: the standard value used here is 1.08 but Berthon et al. (2007) found that, for the Adriatic Sea, a more appropriate value (based on VSF measurements) is 1.15(+/- 0.04). Also in this case it can be said that more work would be needed.

Action

The uncertainty of 0.0007 m^{-1} is now stated to be a minimum value for b_b measurements, but likely to be much larger due to uncertainties intrinsic of the processing hypothesis.

Comment #16

On the absorption from water samples, the paragraph of lines 378-380 is confusing to me. Probably it needs rephrasing. Maybe the authors mean that the absorption of particulate material between 0.2 and 0.7 micron is negligible with respect to the fraction larger than 0.7 micron? If so, is there some evidence of that in data or literature?

Reply

The text simply states that the absorption budget misses some components that cannot be captured due to difference in pore-size of the filters used produce samples for dissolved and particulate matter absorption analysis. It is also added that likely the missing contribution is not big.

Action

The text has been slightly revised and a citation to Morel and Ahn (*J. Mar. Res. 1990*) *has been added.*

Comment #17

CDOM measurements - usage of a 10 cm cuvette inside of a spectrometer is known to be suboptimal in oligotrophic areas like the Mediterranean Sea, even the western basin and in winter. Water is simply too clear to provide a clean spectrum at visible wavelengths. I understand that there is nothing that the authors can do to overcome this issue in case they did not use better suited instruments (like Ultrapath), so at least, an acknowledgement is needed that measurements were performed in suboptimal conditions.

Reply

Full agreement with the Reviewer. This was known, accepted and justified by the standardization of measurements across CoASTS and BiOMaP programs regardless of the water type.

Action

It is acknowledged that the accuracy of CDOM measurements in oligotrophic waters is definitively challenged by the short path-lengths of the laboratory spectrophotometers used for absorbance measurements. It is now well stated in the introduction that some of the measurement methods primarily implemented for optically complex coastal waters, may not warrant a desirable high accuracy in oligotrophic clear waters.

Comment #18

Next type of comments is on the data present in the dataset. It is written (lines 507-511) that basic quality control criteria, like Kd to be higher than the clear water theoretical value (aw+bbw?), were required for a measurement to be included in the dataset, but I have plotted all Kd values and I see that many spectra are less than such value, and some even negative, see Fig. 2. I have repeated the analysis for KL and Ku and I have found the same issue (not shown). Same for some absorption data. Regarding bb, all values are positive, but when removing the water contribution following Zhang et al. (2009), many derived bbp values are negative. Although the number of bad spectra may be marginal, this reduces the confidence that this dataset aspires to; so this needs attention before making the public release.

Reply

Definitively, there were some stations exhibiting negative K_d values. These stations were not originally removed due to an inactive flag applied during the construction of the PANGAEA dataset. These data records have now been removed from the revised dataset being considered affected by a poor extrapolation process in the near surface water layer. These data records refer to measurement stations performed in clear waters during clear sky conditions, which challenge the determination of K_d in the near surface extrapolation layer due to impact of wave focussing.

The two quality indices provided for K_d and b_b spectra are obtained from the subtraction of a constant K_w value at 490 nm (0.0212 m^{-1} by Smith and Baker, AO 1981) and a constant b_{bw} value at 488 nm (0.001603 m^{-1} or alternatively 0.001233 m^{-1} for the Baltic and Black Sea waters) from the

corresponding $K_d(490)$ and $b_b(488)$ values. These indices do not have any impact on the data themselves, their negative value simply suggests some caution.
These indices were mostly introduced to support the use of data from clear waters by identifying questionable spectra challenged by the water type as well as by the applied measurement and data reduction methods. Any user can use, ignore or re-compute those indices and consequently drop whatever data record would be judged 'bad'. Still, the relatively small number of these spectra challenged by measurement or processing methods applied for a critical condition, cannot become the reason to question the dataset.

Action
The potential for determining underestimated values of K_d for critical measurement condition due to wave focussing, is now clearly stated.
The values of K_w and b_{bw} applied to determine the quality indices are also provided and some additional details are added.

Comment #19
On the phytoplankton absorption data and the chlorophyll concentration, I have plotted one against the other in Fig. 3 at 665 nm, with a highlight on the Eastern Mediterranean data. What I see is that there is the expected tight relationship, but I am concerned about a drop in sensitivity that I see in the lower end. The chlorophyll data has an evident trend towards saturation at about 0.03 – 0.04 $mg\ m{-}3$, which is too high to resolve the variability in the oligotrophic oceans. I have overplotted the public data by Valente et al. (2022) and, for the few dots in the lower part, I see that the general linear trend is continued. So, authors may try to explain, and if possible, solve this issue.
Reply
As already stated, the highly oligotrophic clear waters of the Eastern Med sea challenge the absorption and scattering methods applied. Clearly the same water type may also affect the accuracy of the derived *Chla* concentrations. This could certainly explain why a few (4-6 points) in the $a_{ph}(665)$ versus *Chla* plot provided by the Reviewer, suggest saturation for the lowest *Chla* values. This is what the CoASTS-BiOMaP dataset can provide for highly oligotrophic clear waters. Still, away from arguing with the Reviewer, his plot including an additional open access dataset, shows only 3 points out of thousands exhibiting *Chla* values below the questioned ones.
Action
A statement on quantification limits of Chla for the highly oligotrophic clear water conditions of the Med Sea is now specifically stated.

Comment #20
The dataset is optically complete, and therefore something that I am missing in the paper is an Rrs closure exercise. A high degree of closure helps to increase the confidence on the dataset. In the case that large differences appear, the individual sources have to be inspected. The authors have provided a closure exercise for absorption, which is appreciated, and where significant differences appeared. For Rrs, I have done the closure exercises myself for absorption both from the ac-9 and from the water samples. This is done in Figure 4, for the ac-9 and in Fig. 5, for the water samples. To calculate Rrs in both cases, Lee et al. (2011) model was used. Considering the radiometric data as reference, results seem to indicate that absorption from ac-9 delivers quite clean data and closure seems very good in general. On the other hand, there are clear differences when absorption from the water samples are used. The plot suggests that absorption from the water samples is much noisier at blue wavelengths and tends to underestimate the real value.
Reply
The Reviewer is acknowledged for his effort to produce closure exercises using the CoASTS-BiOMaP data and its considerations. A comparison between absorption coefficients from AC9 and the analysis of water samples is already provided. Any additional analysis, is considered beyond the objectives of the current work
Action
The authors consider a closure analysis beyond the objectives of the manuscript (see the introductory note to the overall reply).

Comment #21

Final comment is related to the data presentation in the article. It is nice to see the spectra and the ternary plots, and readers can have an idea of the water types that are represented. There are many ways to present the dataset, here just a few that might be of interest to the reader:

- Crossed relationships among IOPs
- Rrs vs. of chlorophyll, compared to the global relationship
- Kd vs. of chlorophyll, compared to the relationship by Morel
- Chlorophyll vs. the other two water constituents
- One Rrs band ratio vs. another one
- TSS vs. $Rrs(665)$

Reply

Thanks for all the suggestions, clearly feasible, desirable and hopefully interesting. However, the manuscript aims at presenting the dataset with some analysis, and not exploiting its content in any possible direction. Some analysis are presented but major extended analyses are not requested for a manuscript submitted to ESSD with the objective to introduce a dataset.

Action

Still, an additional scatter plots displaying bbp/bp vs Chla has been included as well as a new plot showing the distribution of SPM values.

Minor comments

I think it is a requirement that the link to the dataset is shown in the abstract too.
Line 21: “applied equal” → used equally.
Line 39: “benefited of” → benefited from.
Line 54: “moderately” →moderate
Line 91: “attempting”: very vague term. What does it mean in this context, precisely?
Line 96: probably a link to the IOCCG protocol will help here, for those interested.
Table 2: the two-letter country code chosen by the authors looks arbitrary. There is a standardized one named ISO 3166-1 alpha-2, which I advise to follow.
Line 124: talking about in situ vs. laboratory measurements is confusing. Laboratory measurements are made on part of the in situ data. I prefer to talk about field instrumentation vs. laboratory measurement of field samples.
Line 209: no need to say “so called”, as this name is well consolidated and known by everybody.
Line 343. “Wattman” → Whatman

Action

All relevant corrections have been made. Thanks.

#2 Replies to the comments by Mike Twardowski and actions

General comments by the Reviewer

Comment 1
Public release of CoASTS and BiOMaP is exciting for the ocean color community as these comprehensive datasets have strong value for algorithm development and validation activities. This paper provides an overview of the datasets, 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.
Reply
Some early L_{WN} and *Chla* (only) data were submitted to SeaBASS for SeaWIFS validation. However, those data are outside the temporal interval considered for the CoASTS-BiOMaP dataset accessible through PANGAEA. A few BiOMaP parameters (R_{RS}, *Chla*, a_{ph}, a_{dg}, b_{bp} and k_d) were also submitted to MEREMAID for MERIS validation. Some of those data, 33 stations out of the overall 695 performed in the Black Sea, were later included in the dataset assembled by Valente et al. in 2016 and successive versions.
Action
A note on the data included in Valente et al. (2016) and successive versions, has been added in the manuscript.

Comment 2
Datasets extending 2 decades can be relevant to climate change studies. It would be useful to state this and use the term in key words.
Reply
Full agreement.
Action
The keyword 'Climate change' has been included in the text.

Comment 3
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.
Reply
The Case-1 / Case-2 index, which is always determined according to Loisel and Morel (1998) during data processing, was not included among the CoASTS and BiOMaP quantities because its value is questionable in some marine regions such as the Baltic Sea. Still, any potential user will have the possibility to determine its own index considering the comprehensiveness of the CoASTS-BiOMaP dataset.
Action
Any direct mention to Case 1 waters is now mitigated across the manuscript, even though it cannot be avoided.

Comment 4
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 datasets, 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.
Reply
The classical extrapolation method based on the linear fit of log-transformed data was specifically chosen for the CoASTS-BiOMaP data publication to ensure consistency with any other similar dataset.
The two extrapolation methods mentioned by the Reviewer were already comprehensively investigated in D'Alimonte, D., Shybanov, E. B., Zibordi, G., & Kajiyama, T. (OE, 2013). Not being aware of any previous or successive equivalent investigation, that paper already provides the basis for satisfying curiosities on the method relying on actual exponential extrapolation of profile data (which was specifically applied to some BiOMaP radiometric profiles).
Action
None. A comparison of subsurface radiometric data for a number of stations obtained with the two extrapolation methods, is considered out of the scope for this work. In addition an analysis similar to that proposed by the Reviewer is already matter of a publication.

Comment 5
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.
Reply
The 3-sigma filter only affects the determination of the slope when the number of points per unit depth is low. This was prevented through the application of the multicast profiling method, which increases the number of points to several hundred in the few meter extrapolation interval and consequently increases the precision of the regression (see Zibordi et al. JAOT 2004). In conclusion, the 3-sigma filter allows to detect and remove very few extreme outliers without any appreciable impact on the extrapolation process.
Action
Some more details on the filtering scheme and the number of points per unit depth has been added.

Comment 6
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?

Reply

The importance for characterizing detrital aggregates or large colonial particles is appreciated. However, this was not one of the objectives considered for CoASTS-BiOMaP measurements. The AC9 measurements were always performed using the inlet filters provided by the manufacturer. This set up destroys any aggregate or colony.

The data filtering discussed in the manuscript acts on 'spikes': perturbations that generally affect one or various individual values in the profile and often just the 'a' or 'c' measurements. Spikes frequently occur in coastal waters and near the surface where, regardless of any effort to get read of the air in the measurement tubes, sometimes occasional bubbles or big particles present in the surface layer affect the measurements. Without removing these spikes, the average of the AC9 data collected near the surface and included in the CoASTS-BiOMaP dataset would not be representative of the typical water at the station and more than this would exhibit inconsistencies between 'a' and 'c' values.

Action

Some more details on the measurement and processing methodologies have been added in the revised manuscript.

Comment 7

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.

Reply

Full agreement.

Action

The limits of the applied processing are recognized. As an tempt to allow for alternative corrections for the scattering offsets, the AC9 absorption values at 715 nm not corrected for the scattering offset, have been include in the updated version of dataset.

Comment 8

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.

Reply

Full agreement.

Action

Also in this case the limits of the applied processing have been recognized.

Comment 9

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

Reply

Duplicates were always collected and analysed. The average of the two sample values was commonly taken as the final SPM value for each station. Occasionally, one of the two samples was excluded when the duplicates were showing differences tentatively exceeding 20%. Often a look at the filter allowed to identify the problem. If not, SPM values from temporally and spatially close stations were used to subjectively choose what sample to keep. Sometime, AC9 profile data were required to identify the affected sample.

Action

Some more details on SPM analysis and data have been added.

Comment 10

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 dataset.

Reply

The recommendation by the reviewer has been considered.

Action

Plots of SPM distributions have been added. Plots of c(490) distribution were also produced, but not included because considered not adding much to the manuscript (their inclusion would have further increased the already large number of figures).

Comment 11

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).

Reply

There is no lack of QA/QC, unless this only refers to the lack of extensive data analysis.

This is witnessed by the continuous efforts put in instrument calibration, verification of performance, inter-comparisons and data curation over almost 3 decades. Definitively negative b_{bp} indicate limits. But measurements are affected by uncertainties and in the case of the HydroScat-6 and AC9 data the impact of uncertainties is enhanced in highly oligotrophic waters. Actual values close to ‘zero’ of any quantity could be determined as negative due to measurement uncertainties. The objective of the flags is to put this forward for critical measurement conditions: those mostly performed in the Eastern Mediterranean Sea.

Action

A sentence has been added to state that negative values of the quality indices are expected to be explained by measurement uncertainties.

Comment 12

Regarding the plots, an IOP plot I find is a strong diagnostic of the quality of a dataset 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.

Reply

Also this recommendation has been considered.

Action

Scatter plots of bbp/bp vs Chla have been included in the manuscript and discussed with respect to data available in literature.

Comment 13
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 dataset in a rigorous manner is what elevates this paper to a peer-reviewed contribution as opposed to a simple introduction and guide to these datasets, which could just be posted as a readme online with the datasets.
Reply
QA indicates any action taken to ensure proper execution of measurements. QC is any effort addressed to quality check the quality of data products. This is what was done for each individual quantity included in the dataset as documented in this and previous works.
Action
Definitively, further extended data analysis may strengthen QC. But ESSD papers are specifically intended to introduce details on dataset shared with the community. They are not considered research articles (see also the statements provided in the introductory note to the overall reply).

Comment 14
Reviewer JP suggests a closure analysis would be a straightforward means of assessing the inherent robustness of the datasets – 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 dataset 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.
Reply
The Authors are obviously appreciating the effort by one of the Reviewer and are also pleased by the results. However, extended scientific analysis are outside the objective for ESSD papers (see the introductory note) and the differences between absorption measurements from AC9 and water samples are already amply presented in a dedicated figure. It should be recognized that addressing the reason for these differences is a work by itself.
Action
No action has been taken in reply to this recommendation.

Comment 14
Section 3 title: suggest "Measurements" should be "Measurements overview"
Reply
Agreed
Action
The title of the section has been changed.

Comment 15
Section 3.f: I believe a_p, a_ph, and a_dt were measured. This sentence should be reworded to be precise.
Reply
Agreed
Action
The sentence has been rewritten.

Comment 16

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).

Reply

Agreed.

Action

TSM has been renamed by SPM.