

Earth Syst. Sci. Data Discuss., referee comment RC2  
<https://doi.org/10.5194/essd-2022-159-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on essd-2022-159**

Andre Belo-Couto (Referee)

---

Referee comment on "A compilation of global bio-optical in situ data for ocean colour satellite applications – version three" by André Valente et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-159-RC2>, 2022

---

In my opinion the paper is very well crafted and of high very quality, as it is the dataset.

The dataset is a unique and of high interest for the ocean colour community.

However,

I do have a concern,

This paper is regarding a 3rd version of the dataset, meaning that there are already 2 other papers published previously regarding the this dataset. Hence, the dataset is not new, but updated.

As such, this paper shows that the number of observations for the recent years as increased significantly when comparing with the previous version(V2).

This paper, however, is basically a copy from the V2 paper (Valente et al. 2019) with some additional paragraphs to let the reader know how much new data there is. Even the figures, 16 of them, are the same as in the previous paper (except figure 1). Which might be the intentional, as it is indeed an update on the description of a previous version of the same dataset.

I don't know the policy of the journal regarding the publication of papers that describe/discuss the updated of a dataset.

In my view, it would be more interesting, and beneficial to the reader (and public in general), to present and discuss how the new update changed the past version(s) of the dataset. Whereas there would be no need to maintain the same text/discussions/figures from previous papers. As this knowledge is already published, thus, it already has a doi to be referred to. Unless, of course these results/discussions changed due to the new update.

Further general comments below:

- . it would be useful to know up front, i.e. in the introduction, how much data from which project was added in this new version of the dataset.

- . the new dataset is improved by

- . v3 uses AERONET-OC v3

- . more observations

- . the description of a few datasets do not add anything from the previous dataset paper, for e.g., ARCSSPP does not add anything to a new version as it only provides data from 1954 to 2006

- . the same as above for:

- . GeP&CO it ran from 1999 to 2002

- . BARENTSSEA 1997 to 2013

. BIOCHEM 1997 to 2014

. ESTOC from 1994 to 2011

. Figure 1 doesn't seem to cover the beginning of the data range wavelengths, i.e., 313 nm, it seems to start between 360 and 340 nm

. ~15% increase of data from Valente et al 2019, from previous existing stations. This information should be in the introduction, so we know up front what the update is.

. Results section (copy from the v2 paper + some paragraphs stating that there is an increase of obs. in the current version)

4<sup>th</sup> paragraph - some general results on chla (as in V2 paper)

+ a sentence on increasing n obs between V2 and V3 (+5% fluor, +16% hplc)

. question: (not that it is significant, but) why are the limits for both types of chl methods different? I.e.,

$0.001 < \text{chla\_fluor} < 100$

$0.002 < \text{chla\_hplc} < 99.8$

6<sup>th</sup> paragraph - relationship between rrs ratio with chla

Id it change between different versions? If so, any mechanism that could explain it? (for e.g., predominance of El Nino/La Nina phases during the new coverage?)

. it would be interesting to see if some descriptive relationship statistics evolution between versions, for e.g. coefficient of correlation, between the nasa algorithms and the rrs/chla ratio found in the V1, V2 and V3

7<sup>th</sup> paragraph - general results of aph, adg, bbp

+ only new data for aph (+30%)

8th paragraph - Kd didn't change (as in V2 paper, no need to repeat, already discussed)

9th paragraph - as in V2 paper

F1 is new from V2 paper

F2 to F 16 are as in V2 paper (with some, F6, F12, F13, F16 showing statistic values slightly changed)