Earth Syst. Sci. Data Discuss., referee comment RC3
https://doi.org/10.5194/essd-2021-311-RC3, 2021

Comment on essd-2021-311

sandy thomalla (Referee)

Referee comment on "Application of a new net primary production methodology: a daily to annual-scale data set for the North Sea, derived from autonomous underwater gliders and satellite Earth observation" by Benjamin R. Loveday et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-311-RC3, 2021

This paper presents a multi year (19 month) dataset of glider and complimentary ocean colour net primary production (NPP) data with a variety of complementary ancillary products for a study in the North Atlantic. The data set in and of itself is new (although similar approaches have been implemented elsewhere to produce comparative data sets in support of physical-biogeochemical research). I found the supporting article to be appropriate and well articulated. Although I was unable to access the data itself, from the accompanying text it appears that the data set is of high quality with appropriate metadata and consideration of constraints and limitations. It is however not clear to me whether all steps in the data processing are retained in the data set to allow for a user to edit one or more of the authors steps in their post processing if need be e.g. to implement a different quenching correction method or an adjusted Kd PAR calculation or whatever the case may be. I believe that this data set meets the ESSD criteria of producing a data set that can be further reused for the benefit of Earth system science research, in particular by providing additional insight into that characteristics of regional, seasonal and sub-seasonal variability in the biological response to physical drivers at high resolution. That being said, to do so effectively would, I think, require the accompanying Temperature and Salinity profile data from the respective glider deployments, which from Table 3, do not seem to be included in data the set. This needs to be rectified.

Below I have provided a number of general comments and in some cases queries that require clarification or could be better addressed in the text (to prevent others from being uncertain).

The first and most prominent point being that this publication represents a data set and with the accompanying text, showcases the processing steps. But to me, the publication should be representing both the processor itself that was used to generate the data set as much as (and perhaps even more so) than the data set itdelf. The ability of this processor to be run in delayed mode as well as both near real time and to be able to

switch between different input data files (based on their availability) and to choose between different quenching correction approaches is something that could definitely benefit the broader scientific community. I would highly recommend that this processor be published alongside the data set if at all possible. If so, then I think that the title needs to better reflect the publication of both the data set and the processor, even if links to the processor are published elsewhere but referenced within the text.

More detailed comments:

Line 5: Here we introduce a powerful new technique

To be fair, the publication is actually presenting the data set (not the technique as such….. although as mentioned above I think that it should be both). In addition, the technique of merging in situ glider profiles and satellite Earth Observation measurements to produce NPP is not unique. You are not the first (and unlikely to be the last) to find oneself in a situation where a PAR sensor is absent from a glider (or float) deployment and thus has to resort to satellite PAR to run an NPP algorithm. I appreciate that your exact processor as a whole might be unique but the approach is not especially. This does not however detract from the value of the data set. I simply think that the uniqueness of the approach should not be overemphasised.

Line 8: (Loveday and Smyth, 2020)

I tend to prefer not including references in an Abstract if possible, but perhaps in this instance it may be necessary. I would nonetheless appreciate a few words that elaborate on what that particular study was about? That being said, I tried to look up the reference and it appears to be a link to a published data set which I was unable to access (i.e. not a research publication), which made me wonder how is that published data set with BODC different to the one being proposed to be published here with ESSD?

Line 9 and elsewhere in situ

Maybe I am wrong but I thought in situ had to be in italics

Line 10; “bloom phenology, mesoscale phenomena and mixed layer dynamics”

Has someone from the programme already used this data set to investigate any of these processes? If so, please elaborate so that a reader knows what the original data set was used for (i.e.primary research questions and findings).

Line 20 " This is in contrast to the international Argo float programme"

I am fairly certain that you are wanting to highlight the operational reach of Argo floats relative to ships, but since the previous sentence starts with " …the adoption of autonomous platforms has improved the operational reach of traditional research vessels" and this sentence starts with "This is in contrast to the international Argo float programme" it feels as though you are contrasting autonomous platforms with argo floats which I don't think was your intention. Perhaps this can be slightly reworded.

Line 33

Please include a brief description of additional limitations of satellites to correctly interpret ocean colour in coastal regions due to heavy sediment load, shallow bathymetry, super high chlorophyll that is masked out of standard algorithms etc.

Line 55-56 " For the first time we uncover the considerable temporal and spatial variability in NPP,"

This is not the first time that anyone has uncovered considerable temporal and spatial variability in NPP from a glider deployment. Maybe downplay the novelty of the approach a bit… it does not detract from the novelty of the specific dataset being generated and its potential research impact.

Line 56-57: driven by seasonal succession, fronts (Miller, 2009) and topographical features,

You do not provide any references to back up the statement that the temporal and spatial variability in NPP observed in your specific data set can be attributed to fronts and topographical features. If this has been done by others then references are required. Alternatively, rather highlight that this is how the data set can be used by others, which will further showcase it's value to the community (but as mentioned before, please ensure that temperature and salinity are included so that the role of physical processes such as fronts, stratification, mixed layer dynamics can be investigated in tendem with the NPP).

Line 64 "back-scattering"

figure 1 legend refers to backscattering (no hyphen) and throughout the text it interchanges with backscatter (best to be consistent).

Line 65 and 66

Measurements plural.

Line 67 figure 1

Maybe not necessary, but perhaps fluorescence sensor is more accurate than chla sensor in figure legend?

Line 69 on to

Entirely possible I am wrong here but should it not be onto (i.e. one word)

Line 76 NPP processor

I like the concept of the "NPP processor" in essence being the sum of all the code that defines figure 2. Perhaps the first time that you use the description "NPP processor' you could explain that that is what you are referring to?

Line 103 " Here, when required, SEO-based PAR data is used in lieu of in situ measurements.'

This sentence caused concern when I first read it as I have dealt with instances in the Southern Ocean where a simple swap out of satellite PAR for in situ PAR would not be possible as the differences in derived NPP were too different. Although I have since read further and realised that you have dealt with it in more detail later, I nonetheless think that this should be mentioned here, so that others like me won't preempt unnecessary concern.

Line 134 " Mixed layer depth (MLD) is then calculated from the density gradient according to Holte and Talley (2009)"

Please elaborate on criteria (i.e. how is it different from de Boyer Montégut)

Line 135 “If the MLD calculation fails, the MLD from the previous profile is used.’

Is there a limit on this … i.e. for how many profiles in a row can the MLD calculation fail and it will simply use the last measurable one? Also why/ when does the MLD calculation fail?

Line 138-141

I would suggest rather deleting these 4 sentences from here and instead retaining all PAR related processing to section 3.4.1 and all chla related processing to the relevant section on chla processing.

Line 138 “ The in situ [Chl-a] data is similarly treated.”

Rather specify that it is converted from fluorescens in units of volts to units of chla by multiplying by the scale factor (calibration coefficient) specific to the sensor and subtracting the manufacturer provided dark count.

Line 139 “In the latter case, [Chl-a] data is discarded where the calibrated value exceeds 1e5 mg m−3 .

Am I understanding this correctly …. I.e. it is discarded when chla is > 100 000 mg m-3? Surely that can’t be right… why would it ever be that high?

Line 140 "In the case of DM processing, post mission calibration factors can also optionally be applied, though none were applied in this Case."

In my experience this can pose a really big problem and should be discussed and addressed in a bit more detail e.g. consider maybe matching glider chla to sat chla to show that the manufacturer provided calibration coefficients per glider fluorescence sensor generated a chla concentration that in general matched satellite within some error range (being cognisant of expected deviation from spatial averaging of satellite versus glider etc.). I have experienced instances where I would simply not have been able to merge two gliders in a time series if I had had to rely purely on manufacturer coefficients (luckily I had in situ chla to calibrate individual glider sensors). If that doesn;t work then maybe show the similarities in chla between gliders for all swap over instances. I.e. that there are no step changes evident in chla profiles that could be the result of an unsuitable manufacture calibration coefficient of the sensor.

Line 149 sub-surface

I think that near-surface or just below the surface would be a better term to use than sub-surface (which implies any depth below the surface and not close to the surface).

Line 150-151 " using the method described in Hemsley et al. (2015)"

Not really a method…. More so an equation. I would suggest to rather say equation and provide the equation (from Hensley equation 2) and insert it as your equation 1. Then provide detail on what value you used for irradiance reflectance R.

Line 151 " Fresnel reflectance is"

I see Hensley et al (2015) used a set value of 0.48. Please can you elaborate as to why you instead calculated it (as opposed to using standard value as in Hensley et al., 2015)

and how your values of Fresnel reflectance derived from wind speed, relative humidity and mean sea level pressure compared to 0.48? And what the actual equation is to calculate Fresnel reflectance?

Line 154 "in the previous step"

The previous step does not actually explain how to get E0-. Rather it explains that the method of Hensley et al., 2015 was used to derive E0+ from E0-. I think better to be specefic.

Line 154-155 " SEO KdPAR , calculated from SEO Kd490 Saulquin et al. (2013), is then used to project broadband PAR into the subsurface across the glider depth record.'

This step is not clear to me.

Saulquin et al. (2013) states that existing models calibrated in open ocean waters tend to underestimate the attenuation of light in coastal waters. They investigate two relationships between KdPAR and Kd490 for clear and turbid waters using MERIS reflectances and the spectral diffuse attenuation coefficient Kd(λ) developed by Lee (2005).

Please can you clarify which equation you used to determine KdPAR from SEO Kd490?

Line 156 "Where SEO Kd490 is not available"

Why would SEO Kd490 not be available if you have SEO PAR? Are they not produced by the same satellite product?

Is it possible to compare / provide some indication of the comparison of KdPAR calculated from KD490 (when you have it) and KDPAR calculated from chla? Do they provide similar answers in KdPAR? Are they interchangeable? If different, how much does it affect final NPP calculations? Which is better? Why?

It is entirely possible that I have completely misunderstood this step but if that is the case then it is possible that others will too, so please explain with a bit more clarity. Thanks.

Line 157 "PAR(0)"

Do you instead mean PAR E0+ as described in the section immediately above? Same comment applies to equation (1).

Line 159-169

Should equation 2 that defines Zeu not come before equation 1 that utilises Zeu?

Line 161-162: ii) substantial gaps in the [Chl-a] data,

Not sure I understand this completely. Assuming this is not relevant to PAR from glider but only PAR from SEO? Even so, this is I think only relevant to instances where SEO Kd490 is not available as only then is chla used in the calculation of Zeu, which is then used to generate KdPAR? More clarity required please.

More detail is also required on what threshold constituted "substantial gaps"?

Line 163-164; The latter criteria prevents the glider from deriving NPP estimates from [Chl-a] readings that may have been gathered at depths where particle re-suspension in likely to make them unreliable.

Change in likely to is likely

Why is this exactly? I realise that you would get enhanced scattering from sediment load but why would that drive incorrect fluorescence readings? Does shallow topography tend to overestimate fluorescence? Or are you anticipating viable chlorophyll to be present in the sediments that were resuspended e.g. from dissolved organic matter?

165-166: Euphotic depth (Zeu : defined as the 1% of the surface light depth) is subsequently calculated for all good profiles, and is selectively used in the correction of the [Chl-a] profile.

Delete “the” in “defined as the 1%” i.e. to read defined as 1%

Also, Zeu is not 1% of the surface light depth, but rather the depth at which PAR is 1% of surface PAR.

It is also a bit confusing that Zeu is defined as both the depth at which PAR is 1% of surface light (i.e. 1% of PAR E0- or is it PAR E0+?) and as 34 x chl^0.39 in equation 2. Do you use one approach for glider PAR and the other for SEO PAR?

Please clarify Zeu being used in the correction of the chla profile to being used in some quenching correction approaches (discussed further in section 3.4.3). Indeed, the Xing et al., method that you ended up using for your data set being published here did not require Zeu for the correction of the chla profile.

168: Ed.

I believe that Ed is the incorrect term in this instance as you are comparing surface PAR (defined here as E0+ and not Ed) as per y axis title in figure 3 a) and b).

169: ~50 W m-2

How did you calculate this? Is this simply the average midday surface PAR of Glider minus MODIS? Also, please clarify what a difference of ~50 W m-2 amounts to in terms of %. For example, a typical underestimate of ~50 W m-2 from MODIS derived surface PAR relative to a glider surface PAR of 250 W m-2 is a 20% reduction, which to me feels substantial rather than insignificant.

If the MODIS surface PAR typically underestimates glider surface PAR is it not best to try to correct the MODIS derived product e.g. by applying a % increase to surface PAR generated from a regression between the two products?

I was wondering whether perhaps a good comparison to accentuate the similarities/ differences between the two methods would be to compare daily integrated PAR as this is in essence what is used in the NPP model?

Strange that for 454 Cabot the typical reduction in midday MODIS surface PAR relative to glider surface PAR was not evident in August. Any idea why this would be? As opposed to a more consistent midday offset for the rest of the timeseries?

I am particularly interested in these results because, as mentioned earlier, we have tried this approach in the past but had little success primarily because the offset between glider

and satellite PAR was too large and inconsistent. I am trying to remember if perhaps it was because I only used midday PAR for daily NPP calculations? It would be great to be able to better understand why the differences were inconsequential in your case and how this approach may be better implemented in other ocean regions and platforms (e.g. BGC-Argo).

171-172: cycle, with a mean ¯E+o that falls within 7% of the in situ broadband value.

What exactly is meant by broadband value?

172: that

Delete that

174: 3.4.2. Using the optical backscatter profile

back-scatter, backscatter, backscattering, please be consistent. My personal preference is for 'backscatter'

What about dark correcting? Was this done? By subtracting both the manufacturer provided dark count and presumably an in situ derived dark count?

180: with measurements of [Chl-a] < 0.0 mg m−3 discarded.

Could be that your dark count was too high?

181-182: To account for bio-fouling, the record is also discarded where there is a consistent "step change" of > 5 mg m−3 in [Chl-a] at depths below both the MLD and Zeu , as compared with the initial deployment value.

Interesting way to account for biofouling. I have never come across this. My first question would be why would you expect biofouling to generate a step change? My understanding is that biofouling would be gradual and reflected in a drift in the in situ derived dark count, which can subsequently be used to correct for biofouling (to a point).

I see no mention of the dark correction being applied to the glider fluorescence profiles, this should be specified in the section which notes the application of calibration coefficients. Over and above the manufacturer derived dark correction an in situ derived dark should be applied (to account for ageing of sensor drift or biofouling or change in dark current when integrating sensors onto gliders). This can be generated from each glider mission in delayed mode (if no drift is evident during the deployment e.g. from biofouling) or per profile (which can be implemented in near real time). This is typically done by generating a per profile minimum (see â□□â□□ Wojtasiewicz et al. 2018).

Note however that the fluorescence minimum should only be estimated for depths below the Zeu where quenching is not occuring and that this approach can be tricky in shallow low oxygen waters where fluorescence can increase rather than decrease at depth due to DOM.

183: Where the interpolation fails due to lack of data

What defines a failure? Trying to think why there would be a scenario where there was

not enough fluorescence data within a single profile to facilitate a depth interpolation?

196: – The Swart et al. (2015) method

It is impossible for me not to ask this, but why didn;t you rather test the Thomalla et al., 2018 method? The Thomalla et al., 2018 method was an improvement based on the Swart et al., 2015 method, which we would never have developed had the Swart et al. (2015) method (or any of the others for that matter) worked for that particular glider time series which displayed a significant subsurface chla maxima. In addition, the 2018 paper shows that none of the other methods (Xing, Biermann, Swart or Hemsley) worked well for that glider time series highlighting it as an apporporiate option to have compared in your processor. The key approach that the Thomalla method takes, which the others do not, is to try to figure out the actual depth range of quenching where the correction needs to be applied.

That being said, given how well the Xing method is shown to perform for your glider time series together with your description of the failure of the Swart and Hemsley methods that rely on backscatter, it would not surprise me if the Thomalla et al 2018 did not perform any better. Nonetheless, I would prefer to see it rather than or in addition to the Swart method (since it is an upgrade on a similar concept) and even if you choose not to include the Thomalla method I would very much like to see an additional panel included in Figure 4 to show the comparative performance of the Swart method.

209-215: The Xing et al. (2012) method clearly outperforms the other methods tested

I think it would be good to include a description of why and when this method has been shown to fail in other circumstances e.g. when chla max is below MLD but above Ed. This information is especially useful if the processor is published and other users have the option to implement the different quenching options.

213-215: In this case, quenching corrections using euphotic depth as a maximum depth limit (e.g. (Biermann et al., 2015)) over-correct as they tend to encapsulate the DCM in

the quenching correction process, extrapolating erroneously high [Chl-a] to the surface.

Perhaps worth including a sentence to explain that this approach was indeed designed to account for sub surface chla maxima but in open ocean regions where the MLD is typically deeper than the Ed.

218: for each using the solar irradiance

For each glider profile using the solar irradiance model

219: the model runs using the [O3],

What does the [O3] mean?

231-232: The processor retains the ability to implement this method, as detailed extensively in Hemsley et al. (2015) and represented in Figure 2 by the "Case 1" decision box.

Similar comment as previously regarding the fact that statements such as this are referring to the processor itself but not the data set being published here.

238: calculated from the corrected [Chl-a] and depth profiles and spectral downwelling PAR

Needs to be revised e.g: calculated from the corrected [Chl-a] and spectral downwelling PAR profiles

244: uncorrected

As in raw uncalibrated quenched profile?

Why is NPP calculated for the uncorrected chla profile?

Line 239 says that Net primary production is calculated from the corrected [Chl-a] profiles only.

248: subsequently

I think should rather be subsequent

248: Figure 7

I have to say I am amazed at the relatively miniscule difference in daily integrated NPP given the relatively big difference in midday PAR between the two methods. We tried a similar approach in the Southern Ocean with a different NPP model but our results were not comparable and we had to adjust Satellite PAR to get similar rates of NPP.

The periodic drops in daily integrated NPP to zero for both methods when no data is available (e.g. on 4 occasions during 454 Cabot) is i think not real. Perhaps you can rectify this in the presentation of the time series either by interpolating between profiles or rather just have gaps in the time series for where there is no data (rather than periods of a presumably "fake" zero NPP).

252-253: Two peaks are captured in April/May 2018 and April 2019, corresponding with the timing of the regional spring bloom, and with the latter event significantly more intense.

I am simply not seeing this?

I don;t see a double peak in April/ May 2018… maybe May/June 2018 if I try hard? But more likely what I would consider a highly variable system and not a two peak (double bloom) phenology. Similarly I do not see 2 peaks in April 2019… maybe a second peak in May but certainly not one that is significantly more intense.

253-255: Independent [Chl-a] estimates, derived from v4.2 of the ESA Ocean Colour Climate Change Initative (OC-CCI) data set (Sathyendranath et al., 2019), indicate that the 2018 spring bloom was relatively intense, suggesting that 494 (Stella) failed to fully capture this event.

Initiative spelt incorrectly

You need to show this data somewhere i.,e. Compare sat chla and glider chla for the entire time series and then discuss further, highlighting spatial / temporal averaging etc.

257-259: However, the divergence between the signals recorded by the concurrently deployed gliders 455 (Orca), 497 (Humpback) and 454 (Cabot) strongly suggests the presence of significant spatial heterogeneity in the region north of the Dogger Bank.

Where exactly is the Dogger Bank? can you maybe identify it on Figure 1? It is hard for me to be able to isolate evidence of significant spatial heterogeneity in the region north of the Dogger Bank when all the gliders seem to sample a similar region but evidence of spatial heterogeneity appears to be only in May-July. Perhaps this is more suggestive of seasonal heterogeneity rather than regional? Notably 478 Eltanin and 477 Dolomite covered very different regional domains but nonetheless had very similar NPP possibly providing evidence of the absence of patchiness and a more homogenous system at this point in time (phase of seasonal cycle)?

275: (1) bloom growth or decay

Decay refers to only one loss term e.g. does not include grazing, maybe needs a term that encompasses all e.g. bloom growth or loss (e.g. respiration, decay, grazing)?

277: exceeds 5 mg m-2

More like 4 mg m-2 I would say?

278: night-time cases uncorrected cases is

Delete one of the cases

296-297: The annual cycle and mean annual NPP rate agrees well with contemporaneous values interpolated from the monthly mean OC-CCI based NPP climatology for 1998-2018 (Kulk et al., 2020).

What NPP model was used for satellite derived NPP?

Were those for the same region, can you provide details on spatial averaging? What were the actual numbers compared to yours? Why not rather compare your results with satellite derived NPP for the same time period (and spatial region) as the glider deployments?

298: the April NPP peak in the glider data is somewhat lower than its OC-CCI counterpart,

It is not clear to me what figure in Kulk et al., 2021 you are comparing your results to?

301-303: The glider-based annual mean NPP value 98 gC m−2 a−1 is comparable with the 119 gC m−2 a−1 measured through extensive surveys carried out over ICES Region 7 (north of Dogger Bank) by Joint and Pomroy (1993), as well as with observation based estimates of 125 gCm−2 a−1 for the northern North Sea (van Beusekom and Diel-Christiansen, 1994).

Can you provide more information e.g. method of measuring NPP (14C?). How did they get to annual? measure monthly and integrate?

309: Similar modification of the glider based NPP measurements to a GPP estimates

How were these done?

315: sky conditions

As in cloudiness? Isn;t this mostly accounted for in your NPP calculations?

329: profile

Change to profiles

360: weeks re-emerge

Change to weeks to re-emerge

363: an 19-month

I think should be a 19-month

370-371: present in the production in region

Change to.. present in the rates of NPP for the region perhaps?

371: and highlight the advantages of using autonomous systems to persistently monitor the shelf-seas, especially in tandem with remote sensing based approaches.

I think that this could be elaborated further to highlight the value of the data set that has been generated here.