Interactive comment on ‘VODCA2GPP – A new global, long-term (1988-2020) GPP dataset from microwave remote sensing’ by Benjamin Wild et al.

Anonymous Referee
Received and published: 22 July 2021

Reviewer Comments
Reply to comments

This study produced a global GPP product (VODCA2GPP) using satellite-based vegetation optical depth (VOD) datasets from 1988-2020. Basically, the authors used the method proposed by Teubner et al. (2019), and evaluated the product accuracy based on the measurements of multiple eddy covariance towers globally. In addition, the authors compared the spatial and temporal differences of VODCA2GPP with MODIS and FLUXCOM GPP products. There are several main concerns in my mind to make me suggest rejecting the manuscript because of the important and inherent flaws of VOD data.

Response (1): Dear Referee,

Many thanks for reviewing our paper and outlining various concerns. We believe that most of the mentioned issues can be clarified which we will try to do in this reply.

To begin with, we would like to emphasize that the method that was used for producing the VODCA2GPP dataset is not solely based on Teubner et al. (2019) as suggested by your review but, even more importantly, on the refined methods from Teubner et al. (2021). The main difference between these two modelling approaches is that Teubner et al. (2021) additionally include air temperature in their model formulation which improves model performance substantially, especially in terms of temporal variability which is explained in detail in the manuscript.

First, VOD indicates the vegetation biomass, a large carbon pool, but the GPP indicates one of the carbon fluxes, much smaller compared to biomass. Teubner et al. (2019) and this study used the changes of VOD to simulate GPP, which requires the high performance of VOD to indicate biomass changes, a huge challenge. GPP is sensitive to leaf area index or leaf biomass, not wood and root biomass. However, leaf biomass is quite small fraction compared to other two components. It is obviously using VOD to estimate GPP has much large uncertainty.

Response (2): VOD is generally a strong indicator for vegetation water content which also makes it closely related to vegetation biomass. However, the sensitivity of VOD to different biomass compartments depends also on the frequency of the microwave band that is use to retrieve VOD. While longer wavelengths (e.g. L-Band) are mainly sensitive to large plant structures (i.e. tree trunks and large branches), shorter wavelengths (C/X/Ku-Band) provide information on smaller vegetation structures such as leaves and twigs (Frappart et al., 2020; Woodhouse, 2005). Hence high-frequency VOD is also strongly related to leaf area index and NDVI (Li et al., 2021), which are widely used predictors (among others) for vegetation productivity. Thus, in theory, high-frequency VOD and changes in VOD (ΔVOD) are good indicators for GPP. Teubner et al. (2018) demonstrated in their data-driven study that GPP is indeed highly correlated with high-frequency VOD and ΔVOD from X- and C-band. Based on these findings they developed the VOD2GPP model (Teubner et al. 2019, 2021) which incorporates VOD, ΔVOD, air temperature and the temporal median VOD per grid cell. The latter helps to deduct a static vegetation component which makes the model even more sensitive to small-scale changes in biomass (more on the mdn(VOD)-term in Response 5).

Especially based on the relationship between high-frequency VOD and canopy biomass, it is not “obvious” that VOD-based GPP estimates have much higher uncertainties but it is a new and
promising method. Especially, the use of VOD might overcome limitations of vegetation variables from optical sensors (e.g. cloud cover) in observing GPP and provide the scientific community with GPP estimates that allow to gain new insights on the global carbon cycle as VODCA2GPP is largely independent from existing products.

We agree with the reviewer that the used VOD datasets require a high performance. Therefore, we used the novel VOD Climate Archive (VODCA) dataset which has been used in several studies and is highly appreciated in the scientific community (e.g. Li et al., 2021; Frappart et al., 2020). For the production of VODCA2GPP, we specifically used VODCA from the high-frequency C-, X- and Ku-bands in order to account primarily for changes in the canopy biomass. Additionally, we are using very short time intervals (mostly only a few days) for the computation of ΔVOD. With such small time steps, we are further reducing the influence of woody components as they are not expected to show as strong changes as leaf biomass.

From the validation results, we found low model performance of VODCA2GPP shown as fig. 3a. VODCA2GPP only reproduce spatial variations of GPP about 25% (pearson r = 0.515), and I can’t trust the information provided by VODCA2GPP as such low performance.

Response (3): We agree that the performance of VODCA2GPP in comparison to the mean annual GPP at site level is lower compared to the established optical products MODIS and FLUXCOM. However, this performance of VODCA2GPP strongly depends on land cover (Fig A1 (new)). VODCA2GPP shows a good agreement with mean annual GPP at sites in wetlands, woody savannas, mixed forests, and (remarkably) evergreen broadleaved forests. The performance is rather low at sites with deciduous broadleaved forests, grasslands and open shrublands (Fig A1 (new)). Furthermore, differences in performance regarding absolute deviations (expressed in RMSE) are small. Therefore, we do not agree that VODCA2GPP cannot be generally trusted, but on the contrary we believe that based on our results VODCA2GPP delivers very reliable and largely independent information on global productivity. Additionally, we again want to highlight the overall good correspondence with the temporal dynamic of GPP from MODIS and FLUXCOM (Fig. 4) which shows that VODCA2GPP is trustworthy and of similar quality as the optical products.

In addition, limited by VOD products, the spatial resolution of VODCA2GPP is only 0.25°, which is not far enough for scientific applications, as MODIS Landsat GPP has reached to 30 meter spatial resolution.

Response (4): The applicability of the VODCA2GPP product depends on the type of application. Analogously like in remote sensing of soil moisture, we all know that soil moisture shows a very high spatial variability depending on topography and soil conditions but still satellite-derived soil moisture datasets at 0.25° spatial resolution receive a wide scientific application and are especially suitable for long-term climate-oriented studies (e.g. Dorigo et al. 2017). Despite the availability of some GPP products at higher spatial resolution, to the best of our knowledge there does not exist a global GPP dataset with a spatial resolution of 30 meters. The GPP product with the finest resolution that is available globally is MODIS GPP with 500m. FLUXCOM reaches 10km and the TRENDY models are available at 1°. Hence the spatial resolution of the VODCA2GPP product, might be well suited for comparisons with simulations from global vegetation and carbon cycle models.

Second, I am confused to model algorithms proposed by this study. The algorithms of this study look different a little bit with those of Teubner et al. (2019). Equ. 3.2 represents the model algorithms, the authors added the third term the temporal median of VOD (mdn(VOD)) derived from the complete time series which serves as a proxy for the landcover. Unfortunately, the authors did not explain how mdn(VOD) can indicate landcover. And I also concern why the authors need this term for simulating GPP.
Response (5): The model algorithms in this study are different from those in Teubner et al. (2019) because we make use of the refined model in Teubner et al. (2021). However, the mdn(VOD) term was already incorporated and described in the model algorithms from Teubner et al. (2019) representing a static vegetation biomass component for each pixel which we translated as ‘landcover proxy’. We agree that this term is a slight simplification that can lead to misunderstandings. We will replace the term ‘landcover’ with ‘vegetation density’.

The reason for incorporating mdn(VOD) in the model is to subtract larger structural components making the model more closely related to smaller structures like leaves, which increases model performance (Teubner et al., 2019). We will include a more detailed explanation on the mdn(VOD) term in a revised version of the manuscript.

Besides, the authors separated GPP into Ra and NPP components, so the model can provide the simulations of Ra (as least maintenance respiration) and NPP respectively, which may help us to judge if the structure of model is reliable. However, the manuscript also missed this kind important information.

Response (6): The separation of GPP into Ra and NPP in Eq. 3.1 is done to introduce the theoretical basis of the VOD2GPP model and is intended to support the theoretical understanding of our modelling approach. The VOD2GPP model itself is implemented using a data-driven machine learning model (GAM) which we trained with the input variables against in-situ GPP observations. Thus, the current model formulation is indeed not constructed to separate GPP into NPP and Ra. But it has the potential to provide the basis for such a modelling approach of explicitly representing Ra and NPP components in the future.

Another important model flaw, if the changes of biomass is equal to GPP. Obviously, the answer is NO! When the biomass lost partly at the 0.25 degree pixel, the GPP should be positive assuming the rest vegetation in this pixel still live, but lost biomass may larger than actual GPP contributed by the rest vegetation.

Response (7): We would like to point out that changes in biomass are not considered equal to GPP in our model. GPP is represented as the sum of both terms, Ra and NPP, with biomass contributing to the maintenance part of Ra and changes in biomass relating to the growth part of Ra as well as to NPP. For further considerations regarding biomass changes, please also see our Response (2) above.

Then, VODCA2GPP product may produce negative GPP.

Response (8): Our modelling approach produces almost entirely positive estimates of GPP. In some very rare cases (2.5% of all data points) we obtained GPP that is slightly negative. However, this is related to the extrapolation capacities of the trained GAM model and is common to most machine learning-based GPP estimates (also FLUXCOM includes a few negative values because of invalid extrapolations). We set all negative GPP values to zero in the final dataset.

Third, as the dataset description paper, I majorly pay much more attention to validation of dataset. The authors actually conducted very weak model validations, only validate the performance for reproducing spatial variations of GPP (low performance as the above comment). I would like to see if the product can reproduce the seasonal and interannual variations of GPP against the measurements of eddy covariance towers.

Response (9): Regarding our model validations we understand the concern that large parts of the validations are not based on in-situ GPP from eddy-covariance towers. This is due to the very limited availability of long-term and continuous in-situ observations of GPP (cf. Table B1) which drastically complicates a comprehensive in-situ validation of VODCA2GPP, and all other existing global GPP
datasets. Nevertheless, we conducted an analysis of long-term GPP variability where FLUXNET station data availability allowed it (Fig. A4) and analyzed the overall performance using FLUXNET data (Fig 1d, Fig 3, Fig A1). For the other validations we relied on state-of-the-art products of which the quality is assumed to be good. We will, however, add a 10-fold cross validation analysis (VODCA2GPP vs. FLUXNET) in a revised version of the manuscript.

In addition, the authors seem to hide the magnitude of global VODCA2GPP GPP. From fig. 6, VODCA2GPP estimates for global mean GPP intensity is almost twice than that of MODIS. If I am correct, the global GPP derived from MODIS is 108 Pg C yr⁻¹, thus the global magnitude derived from VODCA2GPP may reach to 200 Pg C or larger? This is a new and the maximum estimates of global GPP that I ever knew. The authors need provide enough robust evidences to prove this number is correct, else I can’t trust the VODCA2GPP, which will lead us to a wrong way.

Response (10): We do not hide anything: In all our figures it is clearly visible that VODCA2GPP has a positive bias compared to other global GPP datasets. We transparently discussed this in section 5.4 and we also show that the bias is much smaller when comparing VODCA2GPP to eddy-covariance measurements (Fig 1d, Fig 3 (new), Fig A1 (new), Fig A4). In fact, for most landcover types VODCA2GPP exhibits only very mild overestimation or even underestimation (Fig A1 (new)). It is true that VODCA2GPP derived estimates of yearly GPP reach 200 ± 2.2 Pg C yr⁻¹ which is indeed much higher than yearly GPP derived from MODIS. However, the comparison with MODIS/FLUXCOM GPP alone is not very meaningful as MODIS GPP, as well as FLUXCOM GPP, are known to substantially underestimate global GPP (Turner et al. 2006; Wang et al., 2017). Other studies suggested values that are much closer to those of VODCA2GPP. Welp et al. (2011) for example came to the conclusion that current estimates of GPP are too low and they estimated the actual yearly GPP to be in the range between 150-175 Pg C yr⁻¹. Koffi et al. (2012) came to similar conclusions (146 ± 19 Pg C yr⁻¹). We agree that VODCA2GPP has a tendency for overestimating GPP in very arid regions ((woody) savannas and open shrublands; Fig A1 (new)) but given the high uncertainties for global annual GPP estimates and the missing consensus regarding this value among the existing datasets we do not see this as an inherent flaw of the VODCA2GPP dataset.

Figure 3 (new): Mean annual in-situ GPP (FLUXNET) plotted against mean annual GPP from VODCA2GPP, FLUXCOM and MODIS for the respective grid cells. Mean annual GPP was computed from all available overlapping years and thus each station is represented by one dot. Red lines indicate the best linear fits determined by ordinary linear regression and the black lines represent the 1:1 lines.
Figure A1 (new): Scatterplots of mean annual GPP for the period 2002-2016 per vegetation type. Vegetation types indicate the predominant IGBP-vegetation type at the respective FLUXNET station. Abbreviations: CRO: Croplands; ENF: Evergreen Needleleaf Forests; DBF: Deciduous Broadleaf Forests; WET: Permanent Wetlands; WSA: Woody Savannas; MF: Mixed Forests; GRA: Grasslands; OSH: Open Shrublands; SAV: Savannas; EBF: Evergreen Broadleaf Forests.

We added the amount of global annual GPP (200 ± 2.2 Pg C yr⁻¹) in the revised manuscript.

Forth, I felt disappointed and the writing is very poor. There are some low-level mistakes in writing and format. All sections of the manuscript look very rough. Especially, the introduction and discussion are meaningless mostly, and the authors kept some sentences to repeat rather than discussing important scientific questions. Several paragraphs are very short consisting of only 2-3 sentences, which failed to provide useful information.

Feeling so bad when I was reading.
Response (11): We are sorry to read that the reviewer did not like our style of writing. We will ensure that the format and the writing is improved in a revised version of the manuscript.

Some minor comments.
line 49-54: this paragraph is not necessary.

Response (12): We added this paragraph to introduce the most important source of GPP data. We find it necessary to shortly introduce FLUXNET as it is used both for training and validation. We will, however, consider making this introduction more concise.

It seems the subtitle of “4. Results” missed.

Response (13): Thank you for making us aware of the missing subtitle. We added the subtitle in the revised manuscript.

There are no GPP values over the desert area in fig. 1c, but there are uncertainties of GPP in fig. 2. These two maps should keep constant for non-vegetated area.

Response (14): The reason why there are no values over some areas in the maps from Fig 1 is that for these comparisons we only used data that are available in all datasets in order to allow comparability between the products. For the uncertainty map we used all data that is available in VODCA2GPP which is why there are also values over the desert. In order to avoid confusion we added an additional figure which is consistent with the uncertainty map:

![Figure 2](image.png)

Figure 2 (new): a) Mean annual GPP as derived from VODCA2GPP for the period 1988-2019. b) Standard deviation of mean yearly annual GPP (1988-2019) as obtained by the uncertainty analysis.

Fig. 3 shows the poor performance of VODCA2GPP compared to fluxcom and modis products. So why do we need VODCA2GPP product?

Response (15): As explained above, Fig 3 indeed shows lower performance of VODCA2GPP compared to FLUXCOM and MODIS. However, Fig 3 should be analyzed in combination with Fig. A1 which shows that there are certain landcovers that substantially drive the lower performance while others perform well. Keeping these differences in mind, VODCA2GPP can be seen as a reliable data source. Furthermore, VODCA2GPP utilizes microwave remote sensing for predicting GPP and is thus largely independent from the mentioned optical remote sensing based datasets. The advantage of this independence is, for example, visible in the long-term trend analysis (Fig. 7) where some trends are
captured by VODCA2GPP which are also visible in TRENDY but not in MODIS or FLUXCOM. Also, the comparison of global anomalies (Fig 5, Table 2) suggests that VODCA2GPP is able to give additional insights in GPP that might be hidden in MODIS or FLUXCOM. Apart from these advantages, VODCA2GPP is also not prone to cloud cover which enables the continuous monitoring of frequently clouded regions like the tropics.

4.2 section. The authors only compared temporal variability of VODCA2GPP with modis and fluxcom datasets. How about VODCA2GPP to reproduce the interannual variability of GPP compared to fluxnet observations. Actually, the comparison with fluxnet observations is more important for this data description paper to let the readers and potential users to know if VODCA2GPP can reproduce the long-term or interannual variability of GPP.

Response (16): We will add a cross-validation analysis for VODCA2GPP vs. FLUXNET.

4.3 section. Why did the authors add the trendy simulations here, not compared previously?

Response (17): We mainly aimed to compare VODCA2GPP with FLUXNET site observations and other data-driven global estimates (FLUXCOM, MODIS) and not with models. However, it is known that e.g. the FLUXCOM product does not consider the CO2 fertilization effect and trends are not realistic (Jung et al. 2020). On the other hands, vegetation models do account for the CO2 fertilization effect and might provide more realistic estimates of GPP trends. Therefore, we decided to include the TRENDY simulations only in analyses of GPP trends. We will make this clearer in a revised manuscript.

4.4 section. Totally confused. Sometimes, the authors made comparison from 2002 to 2016, and otherwise 2003-2015? Can not understand the purpose of the authors.

Response (18): We agree that this is confusing and we apologize for making this not clear enough in the manuscript. The reason why the observation periods differ slightly is the fact that for the used MODIS and FLUXCOM data the observations only start in mid-January 2002 and end in mid-December 2016. Since for the long-term trend analysis we used the yearly median to compute Theil-Slope, we could not assure that these missing observations do not impact the results of the trends. Therefore, we decided to exclude the first and last year in these computations. For the other comparison (e.g. Fig 1), single missing observation do not have any noticeable influence on the metrics which is why we decided to use the fully available periods between 2002 and 2016. We will clarify this in a revised version of the manuscript.

Line 396-399: can’t accept this explanation for poor model performance. The sparse observations in arid and semi-arid regions will reduce the performance of modis and fluxcom, but VODCA2GPP showed the obvious bias.

Response (19): This explanation is of course not the only reason for uncertainties in VODCA2GPP but, as for all data-driven GPP products, the limited availability of in-situ GPP observations is known to decrease model performance, also for MODIS and FLUXCOM. Other, VODCA2GPP specific reasons for uncertainties in our product are discussed in section 5.1 and 5.4.

5.2 section. What is the objective of this section? I did not get any meaning points. The manuscript seems to be rough at the initial stage, and the authors need much more work to improve the writing.
In this section, we want to make the reader aware that validations of global GPP datasets, also of VODCA2GPP, are difficult due to the uneven distribution and generally low availability of eddy-covariance measurement sites (see also Jung et al. (2020)). Furthermore, the objective of this section is to point out that since MODIS and FLUXCOM also heavily rely on FLUXNET data, they cannot be viewed as fully independent data sources although they are using fundamentally differing modelling approaches. Therefore, this section is primarily meant to transparently inform the reader about a potential but unavoidable bias in the validations with FLUXCOM and MODIS due to the partly shared data sources. We will consider merging this paragraph section with section 5.4.

5.5 section. I can’t agree to this point. There are many factors may enhance the global gpp, how the author can claim the increasing trend imply the CO2 fertilization? There are many conclusions that the authors reached without robust evidences. By the way, this section is basically repeat of results.

Response (21): As we already wrote “there are several potential drivers for long-term increases in GPP”, but various studies concluded that CO2 fertilization effects are probably the most important drivers for increasing GPP (e.g. Haverd et. al, 2020; Walker et al., 2020; Campbell et al., 2017; Schimel et al., 2015). We are clearly stating that our results cannot be seen as evidence for this specific effect. We will rewrite this paragraph to avoid potential mis-understandings.

5.6 section. Again, a meaningless paragraph and section. It looks like a conclusion or summary

Response (22): We do not agree that this section is meaningless but the reviewer is right that parts of this section would indeed fit better in section 7 (Conclusion). We revised the manuscript accordingly.

References


