

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

Formatting as follows:

[Reviewer comments](#)

[Reply to comments](#)

[Changes to the manuscript](#)

### Anonymous Referee

Received and published: 22 July 2021

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.

*Please note: All references to materials (e.g. Figures, Tables, Sections) now refer to the numbering of the revised manuscript.*

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 used 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 ( $\Delta VOD$ ) 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  $\Delta VOD$  from X-

and C-band. Based on these findings they developed the VOD2GPP model (Teubner et al. 2019, 2021) which incorporates VOD,  $\Delta$ VOD, 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  $\text{mdn}(\text{VOD})$ -term (including changes to the manuscript) 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  $\Delta$ VOD. 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.

*Page 7, line 201: “The usage of short time intervals (in the order of several days) for the computation of  $\Delta(\text{VOD})$  is crucial since it reduces the influence of larger vegetation components (e.g., stems) and makes the model more sensitive to changes in 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 A3). 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 A3). 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. 3) which shows that VODCA2GPP is trustworthy and of similar quality as the optical products.

*We added a cross-validation and a more thorough comparison with in-situ GPP for different time-scales to give a better overview on the quality of VODCA2GPP (please see response 9)*

In addition, limited by VOD products, the spatial resolution of VODCA2GPP is only  $0.25^\circ$ , 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.

We added a paragraph in Sect. 5.2 where we discuss the limitations associated with the lower resolution and why VODCA2GPP can still be a valuable source of data:

*Page 19, line 427: “Another limitation of VOD products in general, and thus also of VODCA2GPP, is the limited spatial resolution (0.25°). The lower achieved spatial granularity from passive microwave remote sensing is presumably another reason for the slightly weaker performance of VODCA2GPP in comparison with optical remote sensing derived GPP since the VODCA grid cell might not always be well represented by the in-situ measurements. Furthermore, the lower spatial resolution of VODCA2GPP is disadvantageous for the analysis of local GPP as small-scale variations in GPP might be hidden in VODCA2GPP. However, due to its long-term availability and generally high reliability, VODCA2GPP can still serve as a valuable source of data for various other applications (Sect. 5.4), especially concerning long-term climate-related studies and climate model evaluation.”*

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 ( $\text{mdn}(\text{VOD})$ ) derived from the complete time series which serves as a proxy for the landcover. Unfortunately, the authors did not explain how  $\text{mdn}(\text{VOD})$  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  $\text{mdn}(\text{VOD})$  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.

The reason for incorporating  $\text{mdn}(\text{VOD})$  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  $\text{mdn}(\text{VOD})$  term in a revised version of the manuscript.

We replaced the term ‘landcover’ with ‘vegetation density’ and added a more detailed explanation of the  $\text{mdn}(\text{VOD})$  term:

*Page 7, line 182: (...) the temporal median of VOD ( $\text{mdn}(\text{VOD})$ ) derived from the complete time series which serves as a proxy of vegetation density. Specifically,  $\text{mdn}(\text{VOD})$  is incorporated to subtract larger structural vegetation components which makes the resulting model more closely related to biomass changes of smaller structural vegetation components such as leaves. It was shown that this increases model performance (Teubner et al., 2019).*

Besides, the authors separated GPP into  $R_a$  and  $NPP$  components, so the model can provide the simulations of  $R_a$  (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.

We added a more detailed explanation for the mentioned negative GPP:

*Page 8, line 239: "In rare cases (~2.5% of all data points) the machine learning model produced slightly negative results for GPP, due to the extrapolation capacities of the trained GAM. As negative GPP is not possible, such estimates were set to zero."*

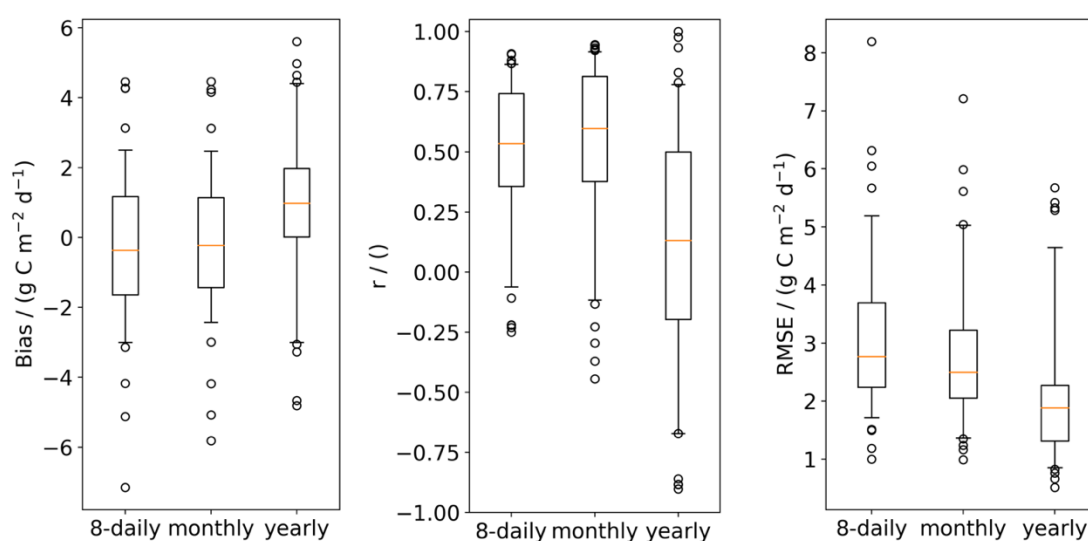
Third, as the dataset description paper, I majorly pay much more attention to validation of dataset. The authors actually conducted very weak model validations, only validate the performance for reproducing spatial variations of GPP (low performance as the above comment). I would like to see if the product 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. A6) and analyzed the overall performance using FLUXNET data (Fig 1d, Fig 4, Fig5, Fig A1, Fig A2, Fig A3). For the other validations we relied on state-of-the-art products of which the quality is assumed to be good.

In order to give a more thorough overview on the performance of VODCA2GPP on different time-scales we added several validations focusing on the comparison with in-situ GPP estimates from FLUXNET in the revised manuscript. Specifically, we added the following validations:

- A site-based cross-validation from which we computed the following error metrics: Root mean square error, bias and Pearson’s  $r$ . This analysis was conducted for different time-scales: 8-daily, monthly and yearly (Fig 5). The adapted methods section (3.5 Site-level evaluation and uncertainty assessment) reads as follows:

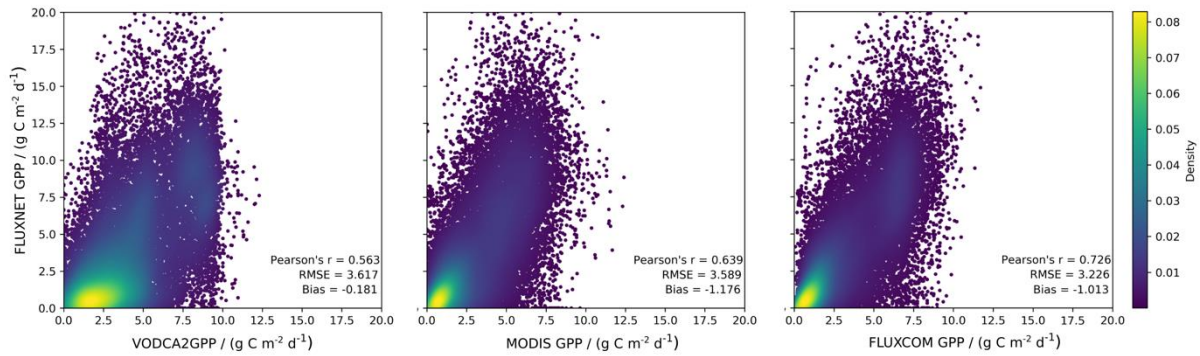
Page 8, Line 243: *“The robustness of the model was evaluated based on a site-based cross-validation analysis during which the influence of the selection of available in-situ stations on the GPP model was investigated. For the cross-validation 10 VODCA2GPP models were trained. Each of the 10 models was trained with 90% of the available FLUXNET stations, while the remaining 10% of the stations were retained for validation (Teubner et al., 2019). Every station was excluded exactly once which is why this approach is classified as pseudo-random. Model performance was assessed at all sites that were omitted in the respective model run by computing Root Mean Square Errors (RMSE), Bias, and Pearson’s  $r$  for different time scales (8-daily, monthly and yearly). [...]*



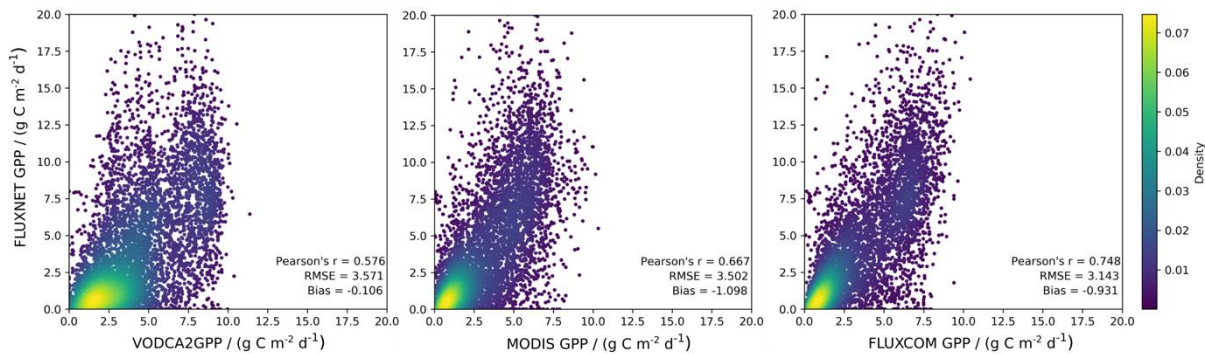
We added the following text in the results section:

Page 13, Line 321: *“The site-based spatial cross-validation also exhibits only a small (negative) bias of VODCA2GPP (Fig. 5) for monthly and 8-daily GPP values while the bias for annual variations is positive and slightly higher. High median Pearson’s  $r$  for 8-daily and monthly values indicates good model performance for interannual variations. Pearson’s  $r$  drops substantially at annual scale which is partly explained by the short period covered by most FLUXNET time series (mean observation time span: 7.27 years), and thus may show little interannual variability. The RMSE decreases with increasing observation length scales.”*

- Site-level comparisons with eddy covariance measurements for 8-daily (Fig 3) and monthly (Fig A2) sampling (in addition to the mean yearly values that were already included (Fig A1)). In order to allow comparability with the current state-of-the art we conducted these analysis not only for VODCA2GPP but also for MODIS and FLUXCOM.



**Figure 4: GPP from FLUXNET plotted against GPP from VODCA2GPP, MODIS and FLUXCOM for the period 2002-2016 with 8-daily sampling.**



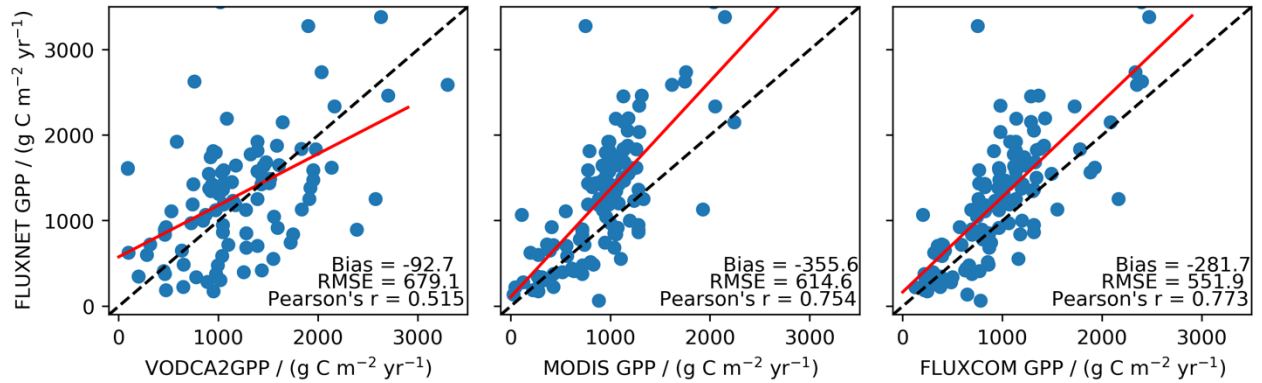
**Figure A2: GPP from FLUXNET plotted against GPP from VODCA2GPP, MODIS and FLUXCOM for the period 2002-2016 with monthly sampling.**

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<sup>-1</sup>, 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.

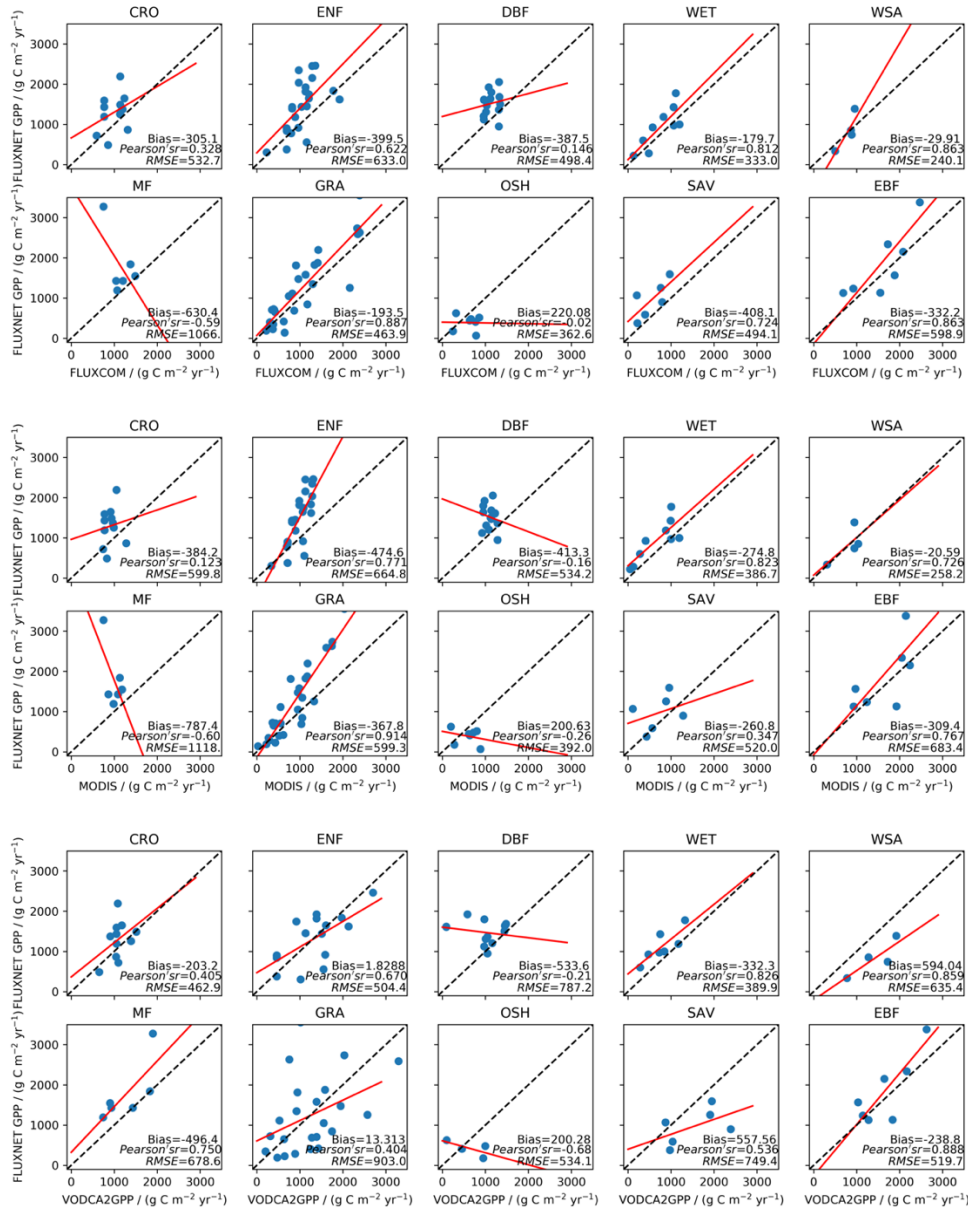
Response (10): We do not hide anything: In all our figures it is clearly visible that VODCA2GPP has a positive bias compared to other global GPP datasets. We transparently discussed this in section 5.4 and we also show that the bias is much smaller when comparing VODCA2GPP to eddy-covariance measurements (Fig 1d, Fig 4, Fig5, Fig A1, Fig A2, Fig A3). In fact, for most landcover types VODCA2GPP exhibits only very mild overestimation or even underestimation (Fig A3). It is true that VODCA2GPP derived estimates of yearly GPP reach  $200 \pm 2.2$  Pg C yr<sup>-1</sup> 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<sup>-1</sup>. Koffi et al. (2012) came to similar conclusions ( $146 \pm 19$  Pg C yr<sup>-1</sup>). We agree that VODCA2GPP has a tendency for overestimating GPP in very arid regions ((woody) savannas and open shrublands; Fig A3) 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.

- We added the amount of global annual GPP ( $200 \pm 2.2 \text{ Pg C yr}^{-1}$ ) in the revised manuscript (Page 11, Line 283).
- In Fig A1. and Fig. A3 (previously Fig. 3 and Fig A1), we adapted the units, replaced the 1:1 line and added the bias in the performance metrics:



**Figure A1: 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 A3: Scatterplots of mean annual GPP for FLUXCOM, MODIS and VODCA2GPP for the period 2002-2016 per vegetation type. Vegetation types indicate the pre-dominant 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**

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 that the reviewer did not like our style of writing.

We thoroughly revised and improved the manuscript. We rewrote most sections and merged/removed paragraphs when necessary. As our changes concerning the style of writing are



extensive we do not list them here explicitly but we would like to refer you to our submitted “Author’s Tracked Changes” version where every improvement is documented in detail.

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 made the introduction of FLUXNET more concise but we decided to keep this paragraph not only because we believe it is important for the readers but also to appreciate the FLUXNET initiative. The revised paragraph reads as follows:

Page 2, Line 48: *“Locally, GPP can be determined at in-situ flux towers, which measure the net exchange of carbon dioxide by means of eddy-covariances that are partitioned into component GPP and ecosystem respiration fluxes (Baldocchi, 2003). FLUXNET (Pastorello et al., 2020) is the global network of flux towers covering all major biomes, and provides the scientific community with harmonized and well-documented flux observations. FLUXNET stations, however, are sparsely and unevenly distributed, which complicates the derivation of GPP globally.”*

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

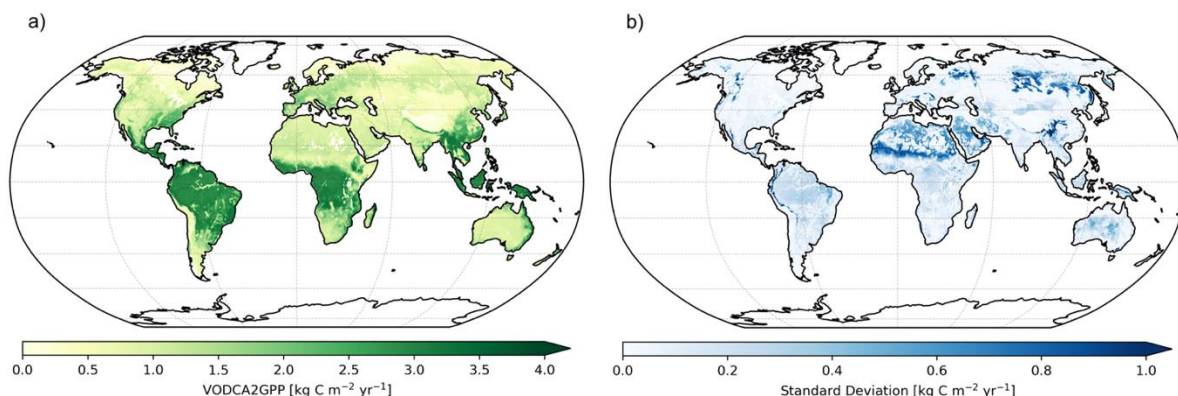
Response (13): Thank you for making us aware of the missing subtitle.

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 confusions we added map of mean annual VODCA2GPP (Fig 2a) which is consistent with the uncertainty map (Fig 2b):



**Figure 2: a) Mean annual GPP as derived from VODCA2GPP for the period 1988-2019. b) Standard deviation of mean annual GPP (1988-2019) as obtained by the uncertainty analysis based on site-level cross-validation.**

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 (now: Fig A1) 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 6, 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.

We added section 5.4 where we discuss potential applications of VODCA2GPP and highlight that VODCA2GPP, just like any other global GPP dataset, cannot be seen as “true reference” but can help to gain insights that might be hidden in other datasets:

*Page 20, Line 465: “5.4 Potential applications of VODCA2GPP*

*The validation results show that VODCA2GPP generally exhibits a high consistency with in-situ GPP observations and global state-of-the art GPP products indicating that VODCA2GPP can be used complementary to current global GPP products. For the analysis of global as well as regional GPP anomalies, VODCA2GPP can provide valuable insights that might be hidden in other observational products due to the fundamentally differing observation methods and the associated limitations related to saturation effects, cloud cover, and other atmospheric effects such as water vapour content or aerosols (Xiao et al., 2019).*

*Also, for the monitoring of global GPP trends, VODCA2GPP has the potential to serve as independent and reliable source of data. The long-term trend analysis suggests that the majority of biomes have increased their primary productivity since 1988. There are several potential drivers for long-term increases in GPP, the most important ones being global warming, land-use changes and elevated CO<sub>2</sub> concentration in the atmosphere (Piao et al., 2019). The observed long-term trends in GPP across the different products support the theory of elevated atmospheric CO<sub>2</sub> leading to an increased uptake of CO<sub>2</sub> (Haverd et al., 2020; Walker et al., 2020; Campbell et al., 2017; Schimel et al., 2015). The absence of trends in FLUXCOM does not contradict but rather supports this theory as FLUXCOM does not account for CO<sub>2</sub> fertilization effects (Jung et al., 2020). Due to the shortness of most in-situ GPP time series, it is, however, difficult to draw final conclusions on the existence, magnitude and reasons for long-term variations of GPP. Therefore, the global influence of atmospheric CO<sub>2</sub> on vegetation productivity remains uncertain, but VODCA2GPP allows to gain new perspectives on long-term GPP trends and might help to identify and quantify driving factors for increasing long-term primary productivity.*

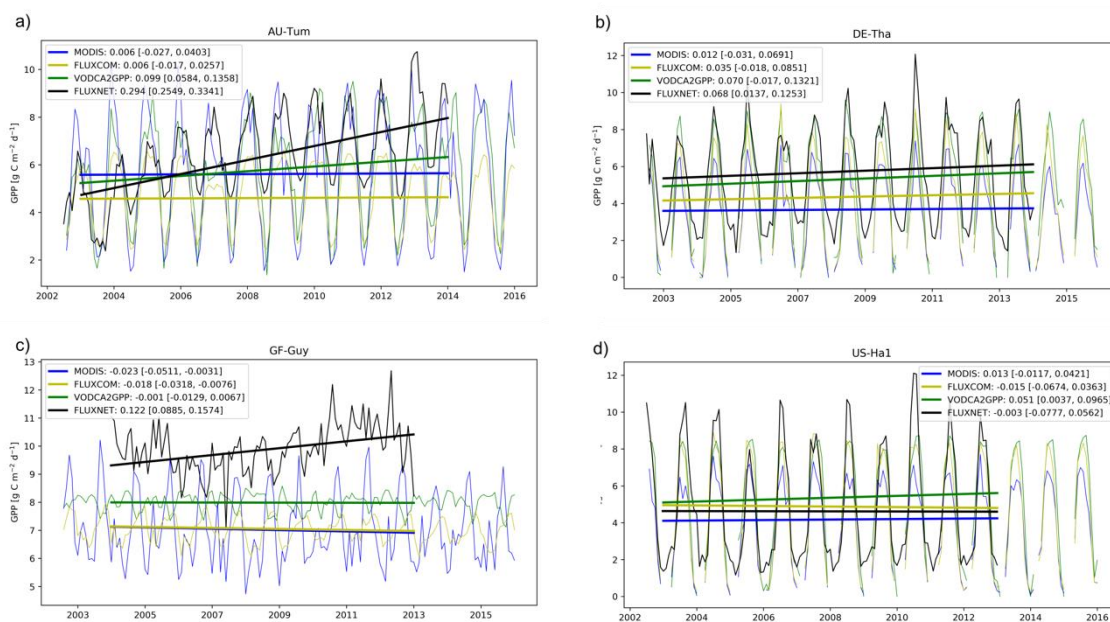
*Furthermore, VODCA2GPP can be used as a largely independent source of data for the intercomparison and validation of other existing or newly developed global GPP datasets and models. Currently, there exists a multitude of global GPP products and there exist large inconsistencies among*

products (Zhang and Aizhong, 2021). Similar to other global GPP dataset, VODCA2GPP cannot be seen as a true reference but using it as additional reference might help to acquire a more comprehensive picture on the performance of other datasets especially in context of long- and short-term variability in GPP.”

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 added a cross-validation analysis for VODCA2GPP vs. FLUXNET and included further comparisons with eddy covariance data on different time scales. Please find our improvements regarding the validations in response 9. Additionally to those, we updated Fig A6 and included the exact magnitude (incl. upper and lower confidence intervals) of the derived long-term trends:



**Figure A6: Exemplary collection of time series of in-situ FLUXNET GPP together with extracted time series from MODIS, FLUXCOM and VODCA2GPP. The stations were selected because of their high data availability for the respective landcover. The lines indicate the regression lines as obtained from the Theil-Sen slope estimation for yearly median GPP. The trends are computed for the common observation period with FLUXNET. The slope [g C m<sup>-2</sup> yr<sup>-1</sup>] is depicted in the legend together with the respective 90% lower/upper confidence intervals. The depicted stations are:**  
a) AU-Tum: Tumbarumba, Australia; Lat: -35.65 °N, Lon: 148.15 °E; Landcover: EBF  
b) DE-Tha: Tharandt, Germany; Lat: 50.96 °N, Lon: 13.57 °E; Landcover: ENF  
c) GF-Guy: Guyaflux, French Guiana; Lat: 5.28 °N, Lon: -52.93 °E; Landcover: EBF  
d) US-Ha1: Harvard Forest EMS Tower, United States; Lat: 42.54 °N, Lon: -72.17 °E; Landcover: DBF

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

Response (17):

We made this clearer in the revised manuscript:

*Page 3, Line 93: "In our analysis, we compare the VODCA2GPP dataset mainly with other data-driven products (FLUXNET, MODIS and FLUXCOM). However, FLUXCOM does not account for CO2 fertilization effects (Walker et al, 2020) which is why trends derived from FLUXCOM are not realistic (Jung et al., 2020). Therefore, we assess monthly anomalies and long-term trends in VODCA2GPP also against TRENDY models, which consider CO2 fertilization."*

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 clarified this in the revised manuscript by updating the description in Table 3:

*Page 16, Line 372: "Table 3: [...] The analysed periods are 2003-2015 which corresponds to the fully overlapping period for all datasets (for the period 2002-2016 there were some data gaps in the used MODIS and FLUXCOM data at the very beginning/very end. Since these data gaps could potentially impact the slope estimation, the slightly shorter period 2003-2015 was used) [...]"*

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 the sections 5.1 and 5.3.

We added various validations to give a more thorough overview on the performance of VODCA2GPP vs. FLUXNET (please see Response 9). We revised the discussion (Sect. 5.3) by including the insights gained from these analysis:

*Page 19, Line 435: "There is only a very small bias when comparing VODCA2GPP with eddy-covariance measurements from FLUXNET but relatively large discrepancies in absolute GPP exist between VODCA2GPP and other remote-sensing-based products, and to a lesser extent with process-based TRENDY models. [...]"*

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.

Response (20): 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 removed this section in the revised manuscript and instead rewrote this paragraph and included it in Sect. 5.3.

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 rewrote this section (now Sect 5.4) into a section which gives an overview on potential applications of VODCA2GPP. The derivation and comparison of trends is one potential application. Possible reasons for trends in GPP are still discussed but we once more highlighted that our results cannot be seen as robust evidence. Please see response 15 for the exact changes.

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 removed this section in the revised manuscript and instead included the information in other parts of the manuscript where it is better suited (mainly in Sect. 7 Conclusion):

Page 21, Line 498: *"[...] The analysis of monthly anomalies exhibited various extreme events in VODCA2GPP that are also found in one or more existing product indicating high plausibility of VODCA2GPP derived anomalies. Furthermore, trends derived from VODCA2GPP contain several plausible patterns that match those derived from the TRENDY-v7 simulations but are not visible in both observational products and vice versa. This suggests that the novel microwave-based approach in VODCA2GPP has the potential to reveal novel findings about temporal dynamics in GPP at large scales that are not yet captured by other GPP products."*

## References

Campbell, J. E., Berry, J. A., Seibt, U., Smith, S. J., Montzka, S. A., Launois, T., Belviso, S., Bopp, L., and Laine, M.: Large historical growth in global terrestrial gross primary production, *Nature*, 544, 84–87, , <https://doi.org/10.1038/nature22030>, 560 2017

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, *Remote Sensing of Environment*, 203, 185–215, <https://doi.org/10.1016/j.rse.2017.07.001>, 2017.

Frappart, F., Wigneron, J. P., Li, X., Liu, X., Al-Yaari, A., Fan, L., ... & Baghdadi, N. (2020). Global monitoring of the vegetation dynamics from the Vegetation Optical Depth (VOD): A review. *Remote Sensing*, 12(18), 2915.

Haverd, V., Smith, B., Canadell, J. G., Cuntz, M., Mikaloff-Fletcher, S., Farquhar, G., Woodgate, W., Briggs, P. R., and Trudinger, C. M. Higher than expected co<sub>2</sub> fertilization inferred from leaf to global observations. *Glob Change Biol*, 26(4):2390–2402, <https://doi.org/10.1111/gcb.14950>, 2020.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozi, D., Nabel, J. E. M. S., Nelson, J. A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, *Biogeosciences*, 17, 1343–1365, <https://doi.org/10.5194/bg-17-1343-620>, 2020, 2020.

Koffi, E. N., Rayner, P. J., Scholze, M., and Beer, C. (2012), Atmospheric constraints on gross primary productivity and net ecosystem productivity: Results from a carbon-cycle data assimilation system, *Global Biogeochem. Cycles*, 26, GB1024, doi:10.1029/2010GB003900.

Li, X., Wigneron, J. P., Frappart, F., Fan, L., Ciais, P., Fensholt, R., ... & Moisy, C. (2021). Global-scale assessment and inter-comparison of recently developed/reprocessed microwave satellite vegetation optical depth products. *Remote Sensing of Environment*, 253, 112208.

Schimel, D., Stephens, B. B., and Fisher, J. B. Effect of increasing co<sub>2</sub> on the terrestrial carbon cycle. *P Natl Acad Sci*, 112(2):436–441, <https://doi.org/10.1073/pnas.1407302112>, 2015.

Teubner, I. E., Forkel, M., Jung, M., Liu, Y. Y., Miralles, D. G., Parinussa, R., van der Schalie, R., Vreugdenhil, M., Schwalm, C. R., Tramontana, G., Camps-Valls, G., and Dorigo, W. A.: Assessing the relationship between microwave vegetation optical 725 depth and gross primary production, *Int. J. Appl. Earth Obs.*, 65, 79–91, <https://doi.org/10.1016/j.jag.2017.10.006>, 2018.

Teubner, I. E., Forkel, M., Camps-Valls, G., Jung, M., Miralles, D. G., Tramontana, G., van der Schalie, R., Vreugdenhil, M., Möisinger, L., and Dorigo, W. A.: A carbon sink-driven approach to estimate gross primary production from microwave satellite observations, *Remote Sens. Environ.*, 229, 100–113, <https://doi.org/10.1016/j.rse.2019.04.022>, 2019.730

Teubner, I. E., Forkel, M., Wild, B., Möisinger, L., and Dorigo, W.: Impact of temperature and water availability on microwave-derived gross primary production, *Biogeosciences*, 18, 3285–3308, <https://doi.org/10.5194/bg-18-3285-2021>, 2021.

Turner, D. P., Ritts, W. D., Cohen, W. B., Gower, S. T., Running, S. W., Zhao M., Costa, M. H., Kirschbaum, A. A., Ham, J. M., Saleska, S. R., and Ahl, D. E.: Evaluation of MODIS NPP and GPP products across multiple biomes, *Remote Sens. Environ.*, 102, 282–292, <https://doi.org/10.1016/j.rse.2006.02.017>, 2006.

Walker, A. P., De Kauwe, M. G., Bastos, A., Belmecheri, S., Georgiou, K., Keeling, R., McMahon, S. M., Medlyn, B. E., Moore, D. J. P., ... and Zuidema, P. A.: Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO<sub>2</sub>. *New Phytologist*, 1. <https://doi.org/10.1111/nph.16866>, 2020.

Wang, L., Zhu, H., Lin, A., Zou, L., Qin, W., and Du, Q.: Evaluation of the Latest MODIS GPP Products across Multiple 770Biomes Using Global Eddy Covariance Flux Data, *Remote Sensing*, 9, 5, <https://doi.org/10.3390/rs9050418>, 2017.

Welp, L. R., Keeling, R. F., Meijer, H. A., Bollenbacher, A. F., Piper, S. C., Yoshimura, K., Francey, R. J., Allison, C. E., and Wahlen, M.: Interannual variability in the oxygen isotopes of atmospheric CO<sub>2</sub> driven by El Niño. *Nature*, 477(7366):579–582, <https://doi.org/10.1038/nature10421>, 2011.

Woodhouse, I.H. (2005). *Introduction to Microwave Remote Sensing*. CRC Press.

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

Formatting as follows:

Reviewer comments

Reply to comments

Changes to the manuscript

### Anonymous Referee

Received and published: 11 August 2021

General comments:

This study introduced a new global GPP dataset using satellite-derived vegetation optical depth (VOD) dataset spanning 1988-2020 in spatial resolution of 0.25o. The method for relating GPP to VOD was proposed in Teubner et al (2019, 2021). This study also presented validation against FLUXNET observation and other global GPP datasets (MODIS, FLUXCOM, TRENDY models). The difference of spatial and temporal GPP dynamics from other GPP reference datasets were shown. However, given the lower VOD-GPP model performance and extremely high global amount of VODCA2GPP, I do have concern about the quality of this dataset and its advantage over other datasets. The writing also needs a lot of improvements since there are both phrasing mistakes and technical mistakes.

*Please note: All references to materials (e.g. Figures, Tables, Sections) now refer to the numbering of the revised manuscript.*

Response: Dear Referee,

Thank you very much for reviewing our paper and giving very constructive feedback. We understand the concerns regarding the quality of the VODCA2GPP dataset but we strongly believe that although VODCA2GPP shows a slightly weaker performance than current state-of-the-art GPP products when comparing mean annual GPP, VODCA2GPP can still be a very valuable data source for studying the global productivity.

It is true that VODCA2GPP GPP estimates are relatively high compared to estimates from current state-of-the-art products. On the other hand, MODIS and FLUXCOM are known to substantially underestimate GPP especially in very productive regions (Fig 4, Fig A1/A2/A3; Turner et al. 2006; Wang et al., 2017) which is why large differences between MODIS/FLUXCOM and VODCA2GPP do not necessarily implicate substantial overestimation of VODCA2GPP. In fact, Welp et al. (2011) suggested that current estimates of GPP are generally too low and they estimated the actual yearly GPP to be between 150-175 Pg C yr<sup>-1</sup>. Koffi et al. (2012) came to similar conclusions (146 ± 19 Pg C yr<sup>-1</sup>). The comparison of mean annual GPP between VODCA2GPP and FLUXNET GPP does not suggest an overall overestimation of VODCA2GPP but exhibits lowest bias for VODCA2GPP among all analyzed datasets (e.g. Fig 4). Nevertheless, we agree that VODCA2GPP has a tendency for overestimation for arid regions (e.g., (woody) savannas and open shrubland; Fig A3). However, due to the missing consensus among existing datasets (Anav et al., 2015) we do not see this as an inherent flaw of VODCA2GPP but rather as a systematic problem in the observation of GPP originating from generally low in-situ data availability.

We improved the style of writing by rephrasing large parts of the manuscript and we removed technical mistakes. Additionally to the changes listed below we would like to refer you also to the



“Author’s Tracked Changes” version of our manuscript where every improvement is documented in detail.

Specific comments:

1 The global total amount of VODCA2GPP datasets is not given.

The global total GPP as derived from VODCA2GPP amounts on average to  $200 \pm 2.2 \text{ Pg C yr}^{-1}$ .

We added this number in a revised version of the manuscript:

Page 11, Line 282: *“The mean global total GPP as derived from VODCA2GPP amounts to  $200 \pm 2.2 \text{ Pg C yr}^{-1}$ .”*

2 line 19, line31, it should be ‘at global scale’.

Thank you for this suggestion. We have revised this in the latest version:

Page 1, Line 18: *“Here, we introduce the new VODCA2GPP dataset, which utilizes microwave remote sensing estimates of Vegetation Optical Depth (VOD) to estimate GPP at global scale for the period 1988 - 2020.”*

Page 1, Line 31: *“A trend analysis for the period 1988-2019 did not exhibit a significant trend in VODCA2GPP at global scale but rather suggests regionally different long-term changes in GPP.”*

3 line 32, it should be ‘different’.

Thank you for this suggestion. We have revised this in the latest version:

Page 1, Line 31: *“A trend analysis for the period 1988-2019 did not exhibit a significant trend in VODCA2GPP at global scale but rather suggests regionally different long-term changes in GPP.”*

4 line 32-33, it should be ‘similar significant increase’. Please check the usage of two ‘for’.

Thank you for this suggestion. We have revised this in the latest version.

Page 1, Line 32: *“For the shorter overlapping observation period (2003-2015) of VODCA2GPP, MODIS GPP, and the TRENDY-v7 ensemble significant increases of global GPP were found.”*

5 line 43-44, please reformulate it.

Thank you for this suggestion. We have revised this in the latest version:

Page 2, Line 41: *“GPP therefore plays a key role in mitigating the negative effects of anthropogenic emissions.”*

6 line 46-47, please reformulate it.

Thank you for this suggestion. We have revised this in the latest version:

Page 2, Line 44: *“Quantifying GPP is essential to understand the effect of climate variability and changes in atmospheric CO<sub>2</sub> concentrations on the land carbon cycle (e.g., Baldocchi et al., 2016; Nemani et al., 2003).”*

7 line 51, please use clause.

Thank you for this suggestion. We have revised this in the latest version:

Page 2, Line 48: *“Locally, GPP can be determined at in-situ flux towers, which measure the net exchange of carbon dioxide by means of eddy-covariances that are partitioned into component GPP and ecosystem respiration fluxes (Baldocchi, 2003). FLUXNET (Pastorello et al., 2020) is the global network of flux towers covering all major biomes, and provides the scientific community with harmonized and well-documented flux observations.”*

8 line 53, what did you mean by ‘measurement bias’? Is it affected by the distribution of flux towers or just measurement techniques?

Here, we refer to the uneven distribution of flux towers which leads to the under/overrepresentation of certain landcovers/biomes in the training data.

We removed this from the introduction and instead left the note on the underrepresentation of certain biomes in the FLUXNET data in the discussion where it is more relevant:

Page 18, Line 408: *“[...] the observation bias which is introduced by unevenly distributed FLUXNET sites decreases the model’s robustness. GPP is measured in situ only at a few locations and these stations are mostly located in temperate regions (e.g., Europe and North America) while semi-arid and tropical forest regions are underrepresented in the training data.”*

9 line 236, please reformulate it.

Thank you for this suggestion. We have revised this in the latest version:

Page 8, Line 239: *In rare cases (~2.5% of all data points) the machine learning model produced slightly negative results for GPP, due to the extrapolation capacities of the trained GAM. As negative GPP is not possible, such estimates were set to zero.*

10 line 245, for site-level validation, you should also use cross validation method.

We included a site-based cross-validation in the revised manuscript. Please see (13) for the results of the cross-validation and all other improvements regarding the validation of VODCA2GPP.

11 line 264, the subtitle for section 4 is missing.

Thank you for making us aware of the missing subtitle.

We have added the subtitle in the latest version (Page 10, Line 268)

12 In figure 2, the uncertainty is very high in desert area (more than ten time of the GPP itself, why?), for example, northern Africa, which is blank in figure 1. Please use single-hue color series and check to ensure the consistency with figure 1.

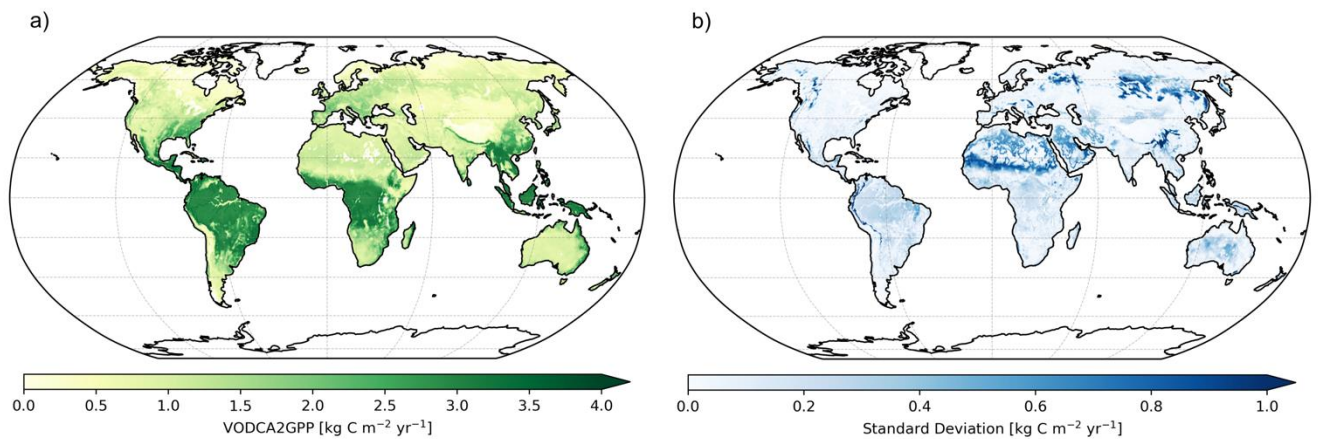
Unfortunately, the scaling of the uncertainty plot (Fig 2b) was not correct.

We adapted the scaling in the revised manuscript (Fig 2b) and apologize for this mistake.

Nevertheless, it is true that the uncertainties are very high in some very arid regions such as the Sahara. The exact reason for the low robustness of the model in these regions is unclear but we assume that these instabilities are partly explicable with missing training data from desert areas.

The reason for the missing consistency between Fig 1a and Fig 2b is that in Fig 1a we only used data that is available in all three products (MODIS, FLUXCOM, VODCA2GPP) in order to allow comparability.

We agree that this is confusing and added an additional figure (Fig 2a) which now includes all available VODCA2GPP data which is consistent with Figure 2b. Furthermore, we updated Figure 2b by using a single-hue color scale and merged it with Fig 2a.

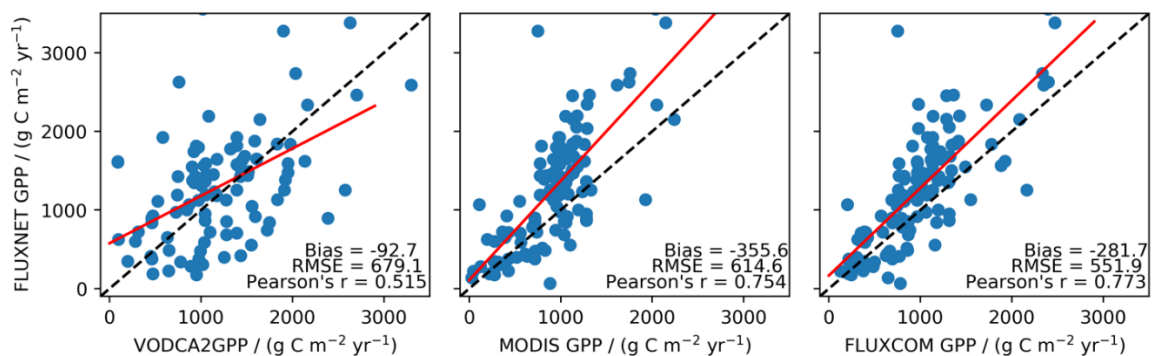


**Figure 2: a) Mean annual GPP as derived from VODCA2GPP for the period 1988-2019. b) Standard deviation of mean annual GPP (1988-2019) as obtained by the uncertainty analysis based on site-level cross-validation.**

13 In figure 3, please check the unit in both x and y axis. Please use dashed line for 1:1 line. Here, you presented the site-level evaluation for mean annual GPP. What about the annual, seasonal (or monthly) evaluation? In fact, Pearson value of around 0.5 seems to be relatively lower for GPP estimation. Did you try 10-fold cross validation for model performance?

VODCA2GPP underperforms mainly for sparsely vegetated biomes, while performance is good for densely vegetated regions (Fig A3). In these regions VODCA2GPP’s performance metrics (Pearson’s r and RMSE) are comparable or even better than those from MODIS and FLUXCOM. Therefore, we suggest that VODCA2GPP can be used as a reliable source of GPP data as long as the users are aware of the performance differences depending on the landcover. We understand that this can be seen as drawback but considering the fact that VODCA2GPP performs well for most regions and is largely independent from current state-of-the-art products, we believe that VODCA2GPP can provide very valuable input for the scientific community. Due to the independence of VODCA2GPP, it also allows users to gain insights on global GPP trends and anomalies that are partly hidden in optical remote sensing based GPP datasets.

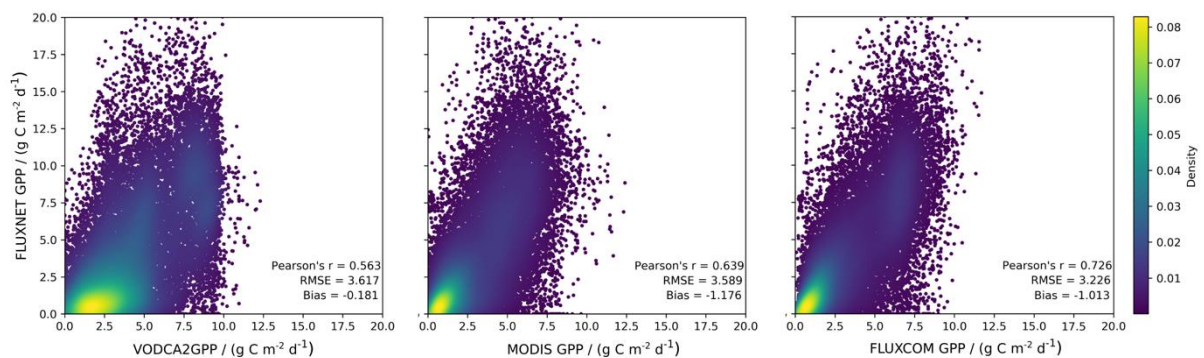
In Fig A1 (previously Fig. 3), we adapted the units, replaced the 1:1 line and added the bias in the performance metrics:



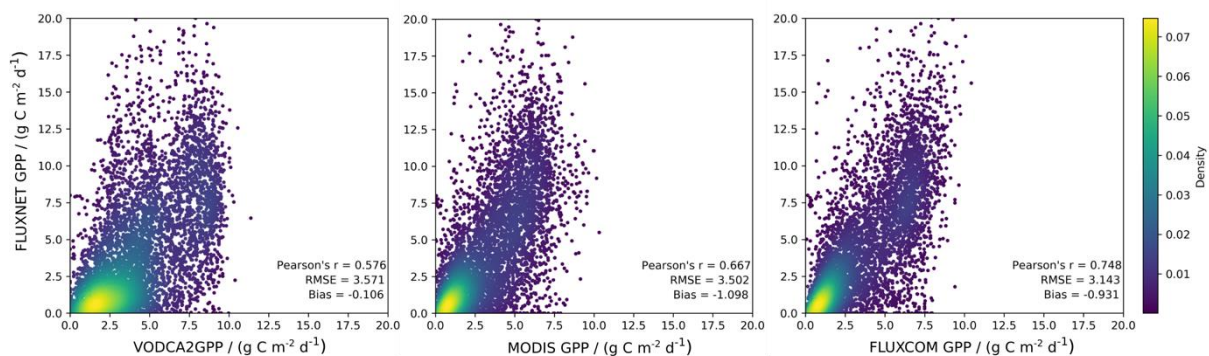
**Figure A1:** Mean annual in-situ GPP (FLUXNET) plotted against mean annual GPP from VODCA2GPP, FLUXCOM and MODIS for the respective grid cells. Mean annual GPP was computed from all available overlapping years and thus each station is represented by one dot. Red lines indicate the best linear fits determined by ordinary linear regression and the black lines represent the 1:1 lines.

In order to give a more thorough overview on the performance of VODCA2GPP on different time-scales we added several validations focusing on the comparison with in-situ GPP estimates from FLUXNET on different time-scales in the revised manuscript. Specifically, we added the following validations:

- Site-level comparisons with eddy covariance measurements for 8-daily (Fig 3) and monthly (Fig A2) sampling (in addition to the mean yearly values that were already included (Fig A1)). In order to allow comparability with the current state-of-the art we conducted these analysis not only for VODCA2GPP but also for MODIS and FLUXCOM. The



**Figure 4:** GPP from FLUXNET plotted against GPP from VODCA2GPP, MODIS and FLUXCOM for the period 2002-2016 with 8-daily sampling.

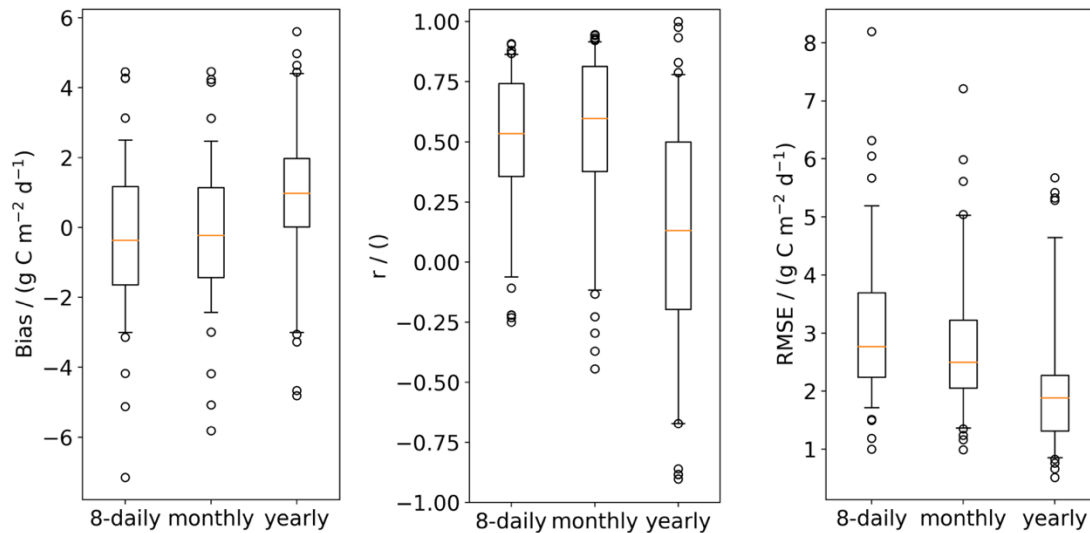


**Figure A2:** GPP from FLUXNET plotted against GPP from VODCA2GPP, MODIS and FLUXCOM for the period 2002-2016 with monthly sampling.

- A site-based cross-validation from which we computed the following error metrics: Root mean square error, bias and Pearson's r. This analysis was conducted for different time-scales: 8-daily, monthly and yearly (Fig 5). The adapted methods section (3.5 Site-level evaluation and uncertainty assessment) reads as follows:

Page 8, Line 243: "The robustness of the model was evaluated based on a site-based cross-validation analysis during which the influence of the selection of available in-situ stations on the GPP model was investigated. For the cross-validation 10 VODCA2GPP models were trained. Each of the 10 models was trained with 90% of the available FLUXNET stations, while

the remaining 10% of the stations were retained for validation (Teubner et al., 2019). Every station was excluded exactly once which is why this approach is classified as pseudo-random. Model performance was assessed at all sites that were omitted in the respective model run by computing Root Mean Square Errors (RMSE), Bias, and Pearson's  $r$  for different time scales (8-daily, monthly and yearly). [...]

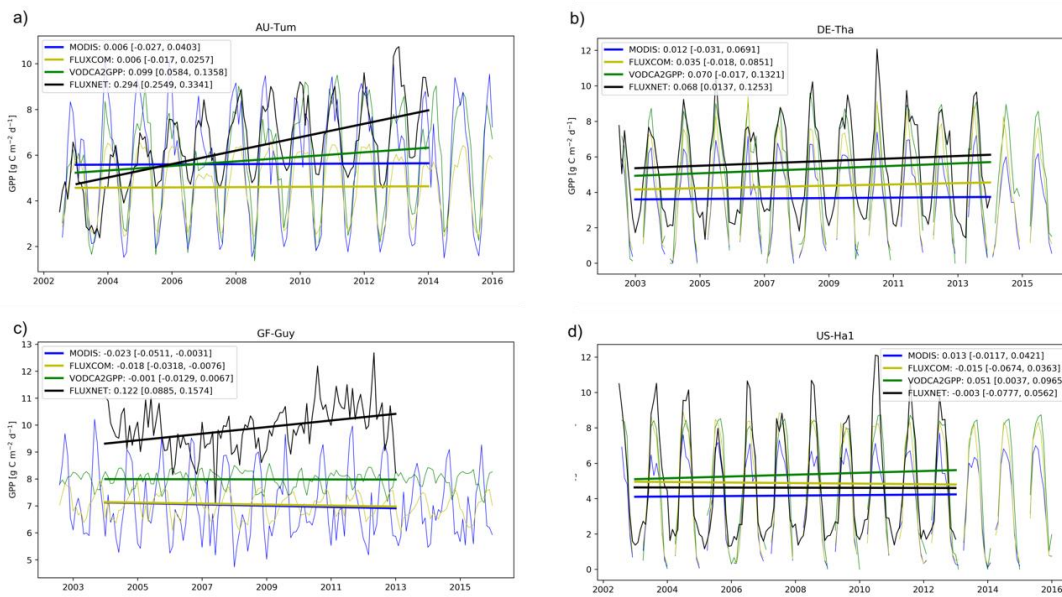


**Figure 5: Site-based cross-validation for 8-daily, monthly and yearly sampling of VODCA2GPP and FLUXNET GPP. RMSE, Bias and Pearson's  $r$  were computed at each of the 10% of FLUXNET sites that were omitted during the respective training run. The whiskers of the boxplots extend to the 0.05/0.95 percentiles.**

We added the following text concerning the cross-validation in the result section:

Page 13, Line 320: *“The site-based spatial cross-validation also exhibits only a small (negative) bias of VODCA2GPP (Fig. 5) for monthly and 8-daily GPP values while the bias for annual variations is positive and slightly higher. High median Pearson's  $r$  for 8-daily and monthly values indicates good model performance for interannual variations. Pearson's  $r$  drops substantially at annual scale which is partly explained by the short period covered by most FLUXNET time series (mean observation time span: 7.27 years), and thus may show little interannual variability. The RMSE decreases with increasing observation length scales.”*

- Structure-wise we merged Sect. 4.2 (Comparison of temporal dynamics) with Sect. 4.1 (Spatial patterns in global annual GPP) into “4.1 Spatiotemporal patterns in global GPP”. We added Sect. 4.2 (Site-level evaluation) which includes the above mentioned analysis.
- We updated Fig A6 and included the exact magnitude (incl. upper and lower confidence intervals) of the derived long-term trends:



**Figure A6: Exemplary collection of time series of in-situ FLUXNET GPP together with extracted time series from MODIS, FLUXCOM and VODCA2GPP. The stations were selected because of their high data availability for the respective landcover. The lines indicate the regression lines as obtained from the Theil-Sen slope estimation for yearly median GPP. The trends are computed for the common observation period with FLUXNET. The slope [g C m<sup>-2</sup> yr<sup>-1</sup>] is depicted in the legend together with the respective 90% lower/upper confidence intervals. The depicted stations are:**

a) AU-Tum: Tumarumba, Australia; Lat: -35.65 °N, Lon: 148.15 °E; Landcover: EBF  
b) DE-Tha: Tharandt, Germany; Lat: 50.96 °N, Lon: 13.57 °E; Landcover: ENF  
c) GF-Guy: Guyaflux, French Guiana; Lat: 5.28 °N, Lon: -52.93 °E; Landcover: EBF  
d) US-Ha1: Harvard Forest EMS Tower, United States; Lat: 42.54 °N, Lon: -72.17 °E; Landcover: DBF

14 line 308, when you mention ‘good temporal agreement’, you need to specify the region, since the correlation is even negative in tropical America.

We agree that this needed to be reformulated more precisely:

Page 12, Line 298: *“VODCA2GPP shows good temporal agreement with MODIS and FLUXCOM for most regions outside the tropics (Fig. 3).”*

15 line 308-309, it should be ‘the highest’ and ‘the lowest’.

Thank you for this suggestion. We have revised this in the latest version.

Page 12, Line 299: *“Pearson’s r is the highest in regions with distinct interannual variability such as sub-arctic, temperate, and semi-arid regions and the lowest for dense tropical forests where even negative correlations occur.”*

16 line 311, please rephrase it.

Thank you for this suggestion. We have revised this in the latest version.

Page 12, Line 300: *Median Pearson’s r between VODCA2GPP and GPP from MODIS and FLUXCOM is 0.77 and 0.75, respectively.*

17 line 313, why monthly GPP anomalies are compared? Why you did not use monthly GPP instead?

Here, we wanted to evaluate VODCA2GPP’s ability to capture extreme events and anomalies in general.

18 line 334, here you used 2002-2016. Why did you use 2003-2015 in next paragraph?

Thank you for bringing this to our attention. The period 2003-2015 is correct. We have revised this.

In order to avoid confusions regarding the time period used in the trend analysis of all products we added the following text:

Page 16, Line 372: *“Table 3: [...] The analysed periods are 2003-2015 which corresponds to the fully overlapping period for all datasets (for the period 2002-2016 there were some data gaps in the used MODIS and FLUXCOM data at the very beginning/very end. Since these data gaps could potentially impact the slope estimation, the slightly shorter period 2003-2015 was used) [...]”*

19 line 335, there are only three sentences in this paragraph. It would be better to merge it with next paragraph.

Thank you for this suggestion. We merged these paragraphs. (Page 15, Line 348)

20 In figure 6, at global scale, the VODCA2GPP is two times of the MODIS and FLUXCOM. So the global GPP can be more than 200 PgC yr<sup>-1</sup>, which is the highest estimation as far as I know. You should give reasonable evidence for such value, or it won't be convincing.

It is true that VODCA2GPP estimates are substantially higher than MODIS and FLUXCOM estimates but as already outlined above, optical products are known to significantly underestimate GPP, especially in very productive regions (Fig 4, Fig A1/A2/Fig A3; Turner et al., 2006; Wang et al., 2017). Thus, differences in global annual GPP between VODCA2GPP and MODIS/FLUXCOM need to be interpreted with caution. In fact, the comparison of VODCA2GPP with in-situ GPP does not suggest a systematic overestimation. VODCA2GPP only shows strong overestimation for (woody) savannas and open shrublands (Fig A3). For all FLUXNET sites located in other landcovers our analysis suggests very mild overestimation or even underestimation of VODCA2GPP. Furthermore, the comparison of VODCA2GPP with the mean of all TRENDY models (Fig. 7) indicates that the overestimation is by far not as severe as implied by MODIS and FLUXCOM.

We added more validations with in situ GPP from FLUXNET which provide a thorough overview on the performance of VODCA2GPP on different time-scales and prove its overall reliability. Please see (13) for details.

21 In figure 7, it should be 2002-2016 or 2003-2015?

2003-2015 is correct. Please see (18) for more details.

22 In section 5.2, TRENDY ensemble is indeed independent from VODCA2GPP. But I couldn't get the meaning of this paragraph. The process-based estimation without accounting for nutrient constraints is also reported to overestimate GPP.

The objective of this section is to emphasize that the validation of global GPP datasets generally needs to be interpreted with caution due to a) the generally low availability and uneven distribution of FLUXNET stations and b) due to the often shared in-situ data-source (i.e. FLUXNET). MODIS, FLUXCOM and VODCA2GPP use the same reference measurements for calibration and validation. As a result MODIS and FLUXCOM cannot be viewed as completely independent from VODCA2GPP although they are using fundamentally differing modelling approaches. Thus, this paragraph specifically aims to transparently inform potential users about these inevitable validation problems.

We removed this section in the revised manuscript and instead rewrote this paragraph and merged it with Sect. 5.3.

23 In section 5.6, in terms of the VODCA2GPP datasets, more similarities with TRENDY than that of other observational products were mentioned. Is it really a strong point for this GPP dataset?

As already mentioned in response 13, one strong argument for VODCA2GPP is its large independence from other (optical) observational products. This independence is evident especially when analyzing long-term trends in GPP where process-based and optical remote sensing products only exhibit very weak correspondence. VODCA2GPP on the other hand exhibits patterns that are visible in at least one other dataset which led us to the conclusion that VODCA2GPP is able to largely bridge the obvious gap between process-based and optical remote-sensing based products. Thus, not the good correspondence with TRENDY alone is a strong point for VODCA2GPP but more importantly its ability to complement existing products, both process-based and observational, is advantageous.

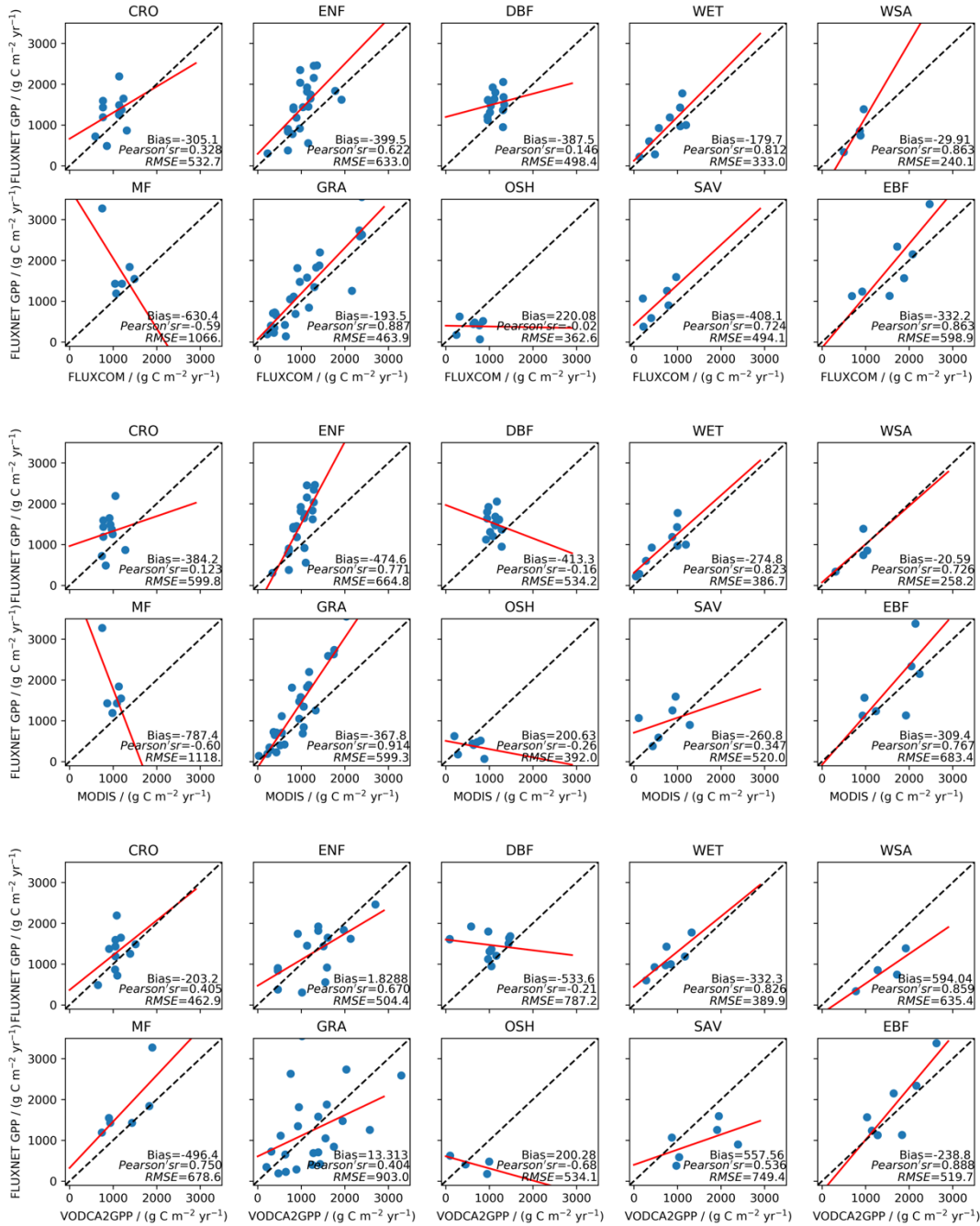
We removed Sect. 5.6 in the revised manuscript and added the following text in the Conclusion:

Page 21, Line 500: *"[...] Furthermore, trends derived from VODCA2GPP contain several plausible patterns that match those derived from the TRENDY-v7 simulations but are not visible in both observational products and vice versa. This suggests that the novel microwave-based approach in VODCA2GPP has the potential to reveal novel findings about temporal dynamics in GPP at large scales that are not yet captured by other GPP products."*

24 In figure A1, please check the unit in both x and y axis. You can give annual, seasonal (or monthly) evaluation as well.

We adapted the units in the axis labelling and added the bias in the performance metrics:





**Figure A3: Scatterplots of mean annual GPP for FLUXCOM, MODIS and VODCA2GPP for the period 2002-2016 per vegetation type. Vegetation types indicate the pre-dominant 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**

25 The code for producing the data and figures should also be publicly available in a repository for transparency reason.

We generally embrace the idea of open access science which is why we made the VODCA2GPP dataset publicly and freely available. However, at the moment we do not consider our work finished. We still strive for improvements of the VODCA2GPP dataset which we will then of course also share with the scientific (and non-scientific) community. Therefore, we are currently not planning to make the VODCA2GPP code publicly available but we are happy to share the code with anyone who approaches us with ideas for collaboration. Regarding the transparency, we described the model

algorithms and the data that is depicted in the figures in detail which is why we do not see a necessity to explicitly share the underlying code.

#### References

Anav, A., et al. (2015), Spatiotemporal patterns of terrestrial gross primary production: A review, *Rev. Geophys.*, 53, 785– 818, doi:10.1002/2015RG000483.

Koffi, E. N., Rayner, P. J., Scholze, M., and Beer, C. (2012), Atmospheric constraints on gross primary productivity and net ecosystem productivity: Results from a carbon-cycle data assimilation system, *Global Biogeochem. Cycles*, 26, GB1024, doi:10.1029/2010GB003900.

Turner, D. P., Ritts, W. D., Cohen, W. B., Gower, S. T., Running, S. W., Zhao M., Costa, M. H., Kirschbaum, A. A., Ham, J. M., Saleska, S. R., and Ahl, D. E.: Evaluation of MODIS NPP and GPP products across multiple biomes, *Remote Sens. Environ.*, 102, 282–292, <https://doi.org/10.1016/j.rse.2006.02.017>, 2006.

Wang, L., Zhu, H., Lin, A., Zou, L., Qin, W., and Du, Q.: Evaluation of the Latest MODIS GPP Products across Multiple 770 Biomes Using Global Eddy Covariance Flux Data, *Remote Sensing*, 9, 5, <https://doi.org/10.3390/rs9050418>, 2017.

Welp, L. R., Keeling, R. F., Meijer, H. A., Bollenbacher, A. F., Piper, S. C., Yoshimura, K., Francey, R. J., Allison, C. E., and Wahlen, M.: Interannual variability in the oxygen isotopes of atmospheric CO<sub>2</sub> driven by El Niño. *Nature*, 477(7366):579–582, <https://doi.org/10.1038/nature10421>, 2011.