Response Letter

Dear Editor,

Thank you very much for considering our manuscript titled “CEDAR-GPP: spatiotemporally upscaled estimates of gross primary productivity incorporating CO2 fertilization” for publication in Earth System Science Data. We deeply appreciate the time and effort you and the three reviewers have taken to review our manuscript. The reviewers’ thorough and constructive feedback has provided valuable insights into our study. We have carefully addressed their comments and made substantial revisions that we believe have enhanced the quality of our manuscript.

Here I would like to highlight several key modifications:

1) We have performed additional analysis to evaluate the performance of our models in predicting long-term GPP trends at global eddy covariance sites across different plant functional types and climate zones.
2) We have introduced a comprehensive guideline for users to assist with the dataset selection from our product for specific research and application goals.
3) We have provided detailed clarifications and discussions regarding several methodological strategies, particularly on the incorporation of the direct CO_2 fertilization effects and the model evaluation.
4) We have improved the description of the machine learning upscaling approach and more clearly distinguished between direct and indirect CO_2 fertilization effects.

Below, we provide a point-to-point response to the reviewers’ specific comments and detail the corresponding revisions made to the manuscript. We hope that our responses and revisions adequately address the reviewers’ concerns. We eagerly look forward to any further insights and suggestions from you or the reviewers!

Once again, thank you and the reviewers for your time and consideration our manuscript!

Sincerely,

Yanghui Kang
On behalf of all co-authors

Response Letter to Reviewer #1

CEDAR-GPP: spatiotemporally upscaled estimates of gross primary productivity incorporating CO_2 fertilization

Dear reviewer,

Thank you very much for your thorough review of our manuscript. We greatly appreciate your constructive feedback and comments, which have significantly improved our paper. Below, we outline our responses (in blue) to your comments (in black) and detail the corresponding revisions made to the manuscript (in red). We included line numbers from the "track-change" version of the revised manuscript for your reference.

Thank you once again for the time and efforts you dedicated to reviewing our manuscript. We hope that our revisions have adequately addressed your comments and concerns. We eagerly look forward to any further feedback and suggestions you may have!

Sincerely,

Yanghui Kang
On behalf of all co-authors

Kang et al. developed a new machine learning based global GPP dataset (CEDAR-GPP), which incorporated CO2 fertilization effect (CFE). Since the CFE significantly affected the trend of global GPP, the new dataset can answer the question of how the CFE benefits GPP over the last 40 years. This dataset overcomes the previous dataset without consider the direct CFE to GPP (i.e. FLUXCOM, FLUXSAT), so it could have the potential for evaluating the GPP increasing trend caused by the CFE. However, I found serval main issues in the current manuscript and dataset.

Main point#1 Since there are ten different setups for CEDAR-GPP, so which one is best among them? Or can the authors provide a guidance for the readers use this product to address the specific question? This is a particularly important point for the dataset availability. As the results mostly showing the NT data, why not move all the DT results to the SI?

Response: Thanks for the suggestion! We agree that guidance on dataset selection would be greatly helpful. We have incorporated detailed guidelines on dataset selection in the "Data availability and usage" section. We also added the details in the user guide accompanying the dataset in Zenodo. Furthermore, we moved the DT results in Section 3.1 – 3.3.to SI for clarity and added a few more displays in SI to illustrate the performance of DT models.

Revisions:
Line 845 – 868 (Section 5): *The CEDAR GPP product offers GPP estimates derived from ten different models. Models are characterized by 1) temporal coverage, 2) configuration of CO2 fertilization, and 3) GPP partitioning approach (Table 2). We provide a structured approach to selecting the most appropriate dataset for research or applications.*

1) Study period considerations: the Short-Term (ST) setup is ideal for studies focusing on periods after 2000. These models are constructed using a broader range of explanatory predictors, offering higher precision and smaller random errors. The Long-Term (LT) datasets shall be used for research assessing GPP dynamics over a longer time period (before 2001). It is important to note that trends from the ST and LT datasets are not directly comparable, as they were derived from different satellite remote sensing data.

2) CO2 Fertilization Effect (CFE) configurations: the CFE-Hybrid and CFE-ML setups are preferable when assessing temporal GPP dynamics, especially long-term trends. The CFE-Hybrid setup includes a hypothetical trend for the direct CO2 effect, while CFE-ML is purely data-driven and does not make any specific assumption about the sensitivity of photosynthesis to CO2. Averaging the CFE-Hybrid and CFE-ML estimates is acceptable, with the difference between them reflecting the uncertainty surrounding the direct CO2 effect. Note that the Baseline setup should not be used to study long-term GPP dynamics, especially those induced by elevated CO2. The Baseline setup may be useful to compare with other remote sensing-derived GPP datasets that do not consider the direct CO2 effect. Differences between these setups regarding mean GPP spatial patterns, seasonal and interannual variations are minor.

3) GPP partitioning methods: We recommend using the mean value derived from both the "NT" (Nighttime) and "DT" (Daytime). The difference between these two provides insight into the uncertainties arising from the partitioning approaches used in GPP estimation from eddy covariance measurements.

Figure 3 was revised to include NT results only.
Figure 5 was revised to include NT results only and with an additional display of the validation of estimated trend across different biomes and climate zones.
Figure 6: Remove the DT model results from the display of latitudinal distributions.
Figure S1: newly added to demonstrate DT model performance in predicting monthly GPP and its spatiotemporal dynamics.
Figure S2: newly added to display DT model performance by biomes and climate zones.
Figure S3: newly added to illustrate DT model performance in predicting long-term trends of GPP across eddy covariance sites.
Figure S5: Added a panel showing the latitudinal profiles of ten CEDAR-GPP datasets

Main point#2 The authors claimed that they consider the direct CFE to photosynthesis. If my understanding is right, the direct CFE benefits the GPP over the past 4 decades and the CFE-ML based GPP (consider the direct CFE) will have a higher increasing trend compared to the Baseline based GPP. The baseline GPP did not contain the direct CFE, but it has the trend of VI, which can represent for the in-direct CFE trend. The reason I have this question is, as the machine learning model is black box, the trend is based on statistic relationships, so the output for the direct CFE or indirect CFE may just depend on the model training. This means that the Baseline GPP could also capture the indirect CFE to GPP and the indirect CFE can reflect the GPP trend in the real world. Without the site validation, the GPP trend from CEDAR-GPP is hard to distinguish whether they are right or not.

Response: Thank you for your question. Our study distinguishes between the direct CFE, which influences photosynthetic rate or light use efficiency, and the indirect CFE, which is related to leaf area expansion and enhanced light interception. Since these are separate effects, we note that both need to be considered to fully represent GPP's long-term changes in response to increased CO2.

The effectiveness of a machine learning model in capturing functional relationships depends on the inclusion of relevant controlling factors during training. To this end, our CFE-ML model incorporates atmospheric CO_2 levels and remote sensing greenness indicators (VIs, LAI) as inputs, allowing it to capture both the direct and indirect CFE. In contrast, the Baseline model, lacking CO_2 as an input, reflects only the indirect CFE trends through changes in LAI and VIs. We understand concerns regarding the "black box" nature of the ML models and would like to address that our objective is to accurately estimate GPP trends by incorporating both CFE pathways. Quantitatively disentangling these effects is critical; However, it would require separate investigations.

To verify the CFE-ML and CFE-Hybrid models, we validated long-term GPP trend estimates against eddy covariance site data, as detailed in Section 3.1.3 and Figure 5. Results show that the Baseline models substantially underestimated GPP long-term trends, while the CFE-ML and CFE-Hybrid models showed significant improvements aligning more closely with observed data. Since the only difference between the CFE-ML and Baseline models is the inclusion of CO_2 input in CFE-ML models, these results suggest that the underestimation of the Baseline models is due to the overlooking of the direct effect of CO_2 on GPP.

In response to the feedback, we have revised our introduction to better explain the direct and indirect CFE and highlighted our site-level validation results in the abstract. Additionally, we corrected the legend in Figure 5, which falsely indicated eddy covariance data (EC) as "Model."

Revisions:

Line 27 – 29 (Abstract): "*Incorporation of the direct CO_2 effects substantially enhanced the predicted long-term trend in GPP across global flux towers by up to 51%, aligning much closer to a strong positive trend from eddy covariance data*."

Line 82 – 91 (Section 1): "*Increasing atmospheric CO_2 directly stimulates the biochemical rate or the light use efficiency (LUE) of leaf-level photosynthesis, known as the direct CO_2 fertilization effect (CFE). Enhanced photosynthesis could lead to greater net carbon assimilation, contributing to an increase in total leaf area. This expansion, contributing to a higher light interception, further enhances canopy-level photosynthesis (i.e. GPP), which is referred to as the indirect CFE. The direct CFE has been found to dominate GPP responses to CO_2 compared to the indirect effect, from both theoretical and observational analyses (Haverd et al., 2020; Chen et al., 2022).*"

Main point#3 The validation for the GPP product is inadequate, especially the GPP trend. It should have a specific validation for ten setups at Monthly, Seasonal, Annual trend for different vegetation types or climate zones. Therefore, the readers choose the specific region with high accuracy data they needed. I am not doubting that the global GPP has continuously increased during the past 40 years. However, different regions may have different greening/browning patterns, so what is their contribution to the global GPP trend? The spatial continuously product can answer this, so the author should first provide the GPP trend at sites and proved they are right, and then upscale this result to the global. Therefore, the site validation is very important!

Response: Thank you for the suggestion! We concur that model performance varies by biomes and climate zones. In the revised manuscript, we incorporated the assessment of predicted GPP trends in eddy covariance sites by plant functional types and climate zones in Section 3.1.3 and Figure 5. The results consistently suggest that the CFE-ML and CFE-Hybrid models agreed better with eddy covariance GPP than the Baseline model. We note that the validation results of monthly anomalies and seasonal cycles by biomes and climate zones are presented in Section 3.1.2 and Figure 4. We focus on evaluating long-term trends using annual GPP data instead of monthly/seasonal values to mitigate the impact of seasonality.

Revisions:
Line 467 – 485 (Section 3.1.3): *Aggregated eddy covariance GPP experienced increasing trends of varied magnitudes across different climate zones and plant functional types (Figure 5b,c; Figure S3b,c). While the machine learning models generally did not fully capture the enhancement in GPP for most categories, the CFE-ML and/or CFE-hybrid models consistently outperformed the Baseline models in both ST and LT setups. The CFE-ML setup predicted a higher trend than CFE-hybrid in most cases, suggesting that the data-driven approach captured more dynamics not represented in the theoretical model, which was based on conservative assumptions regarding the CO2 sensitivity of photosynthesis (see Sect. 2.3.2 and Appendix A). The choice of remote sensing data (ST vs. LT configurations) did not lead to substantial differences in the predicted GPP trend. Most long-term flux sites (at least 10 years of records) with a significant trend experienced an increase in GPP, and the CFE-ML and/or CFE-hybrid models aligned closer to eddy covariance data than the Baseline models (Figure S4). Additionally, we found a considerably higher trend in eddy covariance GPP measurements derived from the day-time versus night-time partitioning approach, potentially associated with uncertainties in GPP partitioning methods (Figure S4). Yet, machine learning model predicted trends were not strongly affected by GPP partitioning methods (Figure S3, S4).*

Figure 5: Comparison of observed and predicted GPP (from NT models only) trends across eddy covariance flux towers by plant functional types and koppen climate zones.

Main point #4 The annual trend should be evaluated on the long-term reliable observation data. Although the authors used the site GPP with more than 5 years for validation by the method provided by Chen et al. 2022 PNAS. However, the ICOS and the FLUXNET2015 has the longer-term observation data, so the author can evaluate the long-term trend (>10 yr) from these sites. Sometimes, the earlier year GPP data will have more data missing during the year, which underestimated the annual GPP at sites, so the annual trend is not reliable when only using 5 years of GPP observations.

Response: Thanks for the comment. We aimed to incorporate a substantial number of sites for the robustness of our analysis which evaluates GPP anomalies aggregated over multiple sites. Therefore, we used a threshold of 5 years in site selection following Chen et al. (2022). Regarding missing data, we excluded one site year with more than one month of missing data. We have detailed the approach in Section 2.3.3 of the revised manuscript.

Furthermore, we performed additional analysis to evaluate GPP trends in sites with at least 10 years of valid data. We found that most sites showing a significant (p-value < 0.3) GPP trend exhibited a positive change. Notably, our CFE-ML and CFE-Hybrid models consistently outperformed the Baseline models in capturing trends in these sites. The results were incorporated in Section 3.1.3 and Figure S4.

Revisions:
Line 341 – 342 (Section 2.3.3): *We excluded a site-year if less than 11 months of data was available and used linear interpolation to fill the remaining temporal gaps.*
Line 475 – 478 (Section 3.1.3): *Most long-term flux sites (at least 10 years of records) with a significant trend experienced an increase in GPP, and the CFE-ML and/or CFE-hybrid models again aligned closer to eddy covariance data than the Baseline models (Figure S4).*
Figure S4 was newly added to compare estimated and observed trends in long-term eddy covariance sites.

Main point #5 Cross validation among the GPP products should have more discussions. Although the authors did some comparison between the CEDAR-GPP with existed products, they just showed the results, but did not analyze the results and provide the insight on which product can be improved or what is the inherit reason for the uncertainties. So the discussion section can list some viewpoints for the further GPP product improvement.

Response: Thank you for the suggestion! In the manuscript, we have discussed potential factors that led to differences between GPP products when relevant. For example, we discussed the potential impacts of eddy covariance data selection on the predicted GPP magnitudes in upscaled products (Section 4.1.1 last paragraph). Additionally, we analyzed the discrepancy in the global GPP trends across products and their connection to the representation of CO2 fertilization effect in Section 3.2.4. In the revised manuscript, we have expanded our discussions to provide more in-depth explanations for these differences and insights for future improvement, particularly regarding interannual variabilities (Section 3.2.3) and long-term trends (Section 4.1). Our discussion mainly focuses on machine learning upscaled products (Section 4.1, 4.2).

Revisions:
Line 570 – 574 (Section 3.2.3): *The lack of interannual variability in FLUXCOM-ERA5 is attributable to the use of mean seasonal cycles of remotely sensed vegetation greenness indicators rather than their dynamic time series*.
Line 741 – 743 (Section 4.1.2): *For example, the lack of GPP interannual variabilities in FLUXCOM-ERA5 manifests the importance of incorporating dynamic vegetation signals from remote sensing in the upscaling framework.*
Line 824 – 836 (Section 4.2): *Our results suggested that variations in the estimated GPP long-term trends from different products were largely related to the representation of CO2 fertilization. Products that did not consider the direct CO2 effect, including our Baseline models, FLUXSAT, FLUXCOM, and MODIS, showed minimal long-term changes in tropical GPP, while the CEDAR CFE-ML and CFE-Hybrid models demonstrated significant GPP increases aligning with predictions from the terrestrial biosphere models (Anav et al., 2015). FLUXCOM-ERA5, not accounting for dynamics changes in vegetation structures and CO2, did not capture either the direct or indirect CO2 fertilization resulting in a slight negative GPP trend attributable to shifted climate patterns. Notably, rEC-LUE exhibited contrasting trends before and after circa 2000, primarily attributed to changes in VPD, PAR, and LAI, while the direct CO2 fertilization effect remained consistent (Zheng et al., 2020). Nevertheless, considerable differences between CEDAR-GPP and rEC-LUE, as well as between our CFE-ML and CFE-Hybrid products, warrant more in-depth investigations into long-term GPP responses to changes in atmospheric CO2 and climate patterns.*

Main point #6 The water stress effect should be considered. Line 718 to L720, "Yet the model assumed a fixed ratio of leaf-internal to ambient CO2, and thus did not include any responses to vapor pressure deficit.". So is the CEDAR-GPP consider VPD effect for GPP? Although CFE is one of the most significant effects to the GPP trend. However, the VPD trend cannot be omitted. Li et al. 2023 reported that the VPD significantly affects the GPP at different vegetation types, so the CEDAR-GPP consider this or some setups reproduce such condition? If the CEDAR-GPP cannot reproduce the VPD effect, so under what condition can I use this product? The soil moisture is also coupled with VPD, so is the CEDAR-GPP consider the soil moisture stress, and can it reproduce the stress from soil moisture? These are needed to be mentioned.

Li, S., Wang, G., Zhu, C., Lu, J., Ullah, W., Hagan, D. F. T., ... & Peng, J. (2023). Vegetation growth due to CO2 fertilization is threatened by increasing vapor pressure deficit. Journal of Hydrology, 619, 129292.

Response: Thank you for the comment! We acknowledge the importance of VPD and soil moisture in influencing GPP dynamics. The mentioned sentence was intended to clarify that in our theoretical model within the CFE-Hybrid setup, the internal to ambient CO2 ratio (χ) was held constant to a long-term average value typical for C3 plants, when describing CO2 sensitivity in photosynthetic rate. This assumption does not imply that the estimated GPP from CEDAR ignores VPD effects. Indeed, as we incorporated VPD from ERA5-Land and soil moisture from ESA-CCI as predictors in machine learning models (Section 2.2), CEDAR-GPP predictions are expected to represent both effects. Thus, trends in GPP driven by increasing VPD, as outlined in Li et al. (2023), should be captured by our model. We have revised the text for clarity and expanded our discussions on the uncertainties associated with χ.

Revisions:
Line 792 – 799 (Section 4.2): *For simplicity, we assumed a fixed ratio of leaf internal to ambient CO2 (χ) representing an average long-term value typical for C3 plants in the theoretical CO2 sensitivity function. However, χ varies by environmental conditions, including temperature and vapor pressure deficit, and robustly modeling these dependencies remains challenging (Wang et al., 2017). Future work could incorporate more comprehensive representations of the χ and evaluate how the associated uncertainties affect the quantification of GPP and its temporal variations.*

Specific comments:

L14 sugars and starches->carbohydrate

Response: We have modified it accordingly.

L26 the annual trend validation with statistics should be highlighted since it is the most important outcome for CFE induced GPP trend.

Response: Good point! We have revised the abstract accordingly.
Revisions: Line 27-30: *Incorporation of the direct CO_2 effects substantially improved the predicted long-term trend in GPP across global flux towers by up to 51%, aligning much closer to a strong GPP enhancement based on eddy covariance data*.

L38 references needed

Response: We have provided a reference to the statement.

L90 'yet, this important mechanism is still missing in GPP products upscale from in situ eddy covariance flux measurements.' This is not accurate, since Zheng et al. also parameterize the model by in situ eddy covariance flux measurements. So this sentence should be revised.

Response: Thanks! We have revised the sentence accordingly.
Revisions: Line 101 – 103: "*yet, this important mechanism is still missing in GPP products upscaled from in situ eddy covariance flux measurements based on machine learning models.*"

L125 So totally 233 sites are used but not listed, the authors should list them at the SI.

Response: Thanks for the comment! We have added a table listing all sites and their data citation in SI.
Revisions: Table S1

L137 The C3 and C4 map is constant among the investigated years or not? Will the C3/C4 vegetation fraction change?

Response: We used a static C4 proportion map, but the fraction could potentially vary over time. In the revision, we made the clarification and highlighted potential future improvements from incorporating dynamics C3/C4 distributions.
Revisions: Line 218 – 219: "*The C4 percentage dataset was constant over time*."
Line 754 - 757: "*potential improvement may be achieved by incorporating datasets related to agricultural management practices (crop type, cultivar, irrigation, fertilization) (Xie et al., 2021), plant hydraulic and physiological properties (Liu et al., 2021), dynamic C4 plant distributions (Luo et al., 2024), root and soil characteristics (Stocker et al., 2023), as well as topography (Xie et al., 2023),"*

Table 1 'Surface reflectance b1 – b7, Vegetation indices (NIRv, NDVI, kNDVI, EVI, GCI, NDWI), percent snow'. What is GCI?

Response: Thanks for the note! The correct term should be 'CIgreen', representing Green Chlorophyll Index, as stated in the main text (Line 177 – 180). We have made the correction.

L 158 vegetation indexes or vegetation indices?

Response: We have changed "indexes" to "indices" in the manuscript.

L172 'PKU GIMMS LAI4g consisted of AVHRR-based LAI from 1982 to 2003 (generated using machine learning models trained with Landsat-based LAI data and NDVI4g) and MODIS BNU LAI from 2004 onwards (Yuan et al., 2011).' To my knowledge, PKU GIMMS LAI4g didn't use the Yuan et al 2011 as input data.

Response: Thanks for the comment. Upon reviewing the PKU GIMMS LAI4g paper (Cao et al., 2023), we confirmed that the AVHRR GIMMS LAI4g data was consolidated against the reprocessed MODIS LAI generated based on approaches from Yuan et al. (2011). The final product comprised AVHRR-based GIMMS LAI4g (1982-2003) and reprocessed MODIS LAI (2004-2020) (Sections 3.2 and 2.3 in Cao et al., 2023). We have rephrased our description to align with the details more accurately from the GIMMS LAI4g paper.
Revisions: Line 197 - 199: *PKU GIMMS LAI4g consisted of consolidated AVHRR-based LAI from 1982 to 2003 (generated using machine learning models trained with Landsat-based LAI data and NDVI4g) and reprocessed MODIS LAI (Yuan et al., 2011) from 2004 onwards.*

L210 'GIMMS LAI4g and NDVI4g data were only filled with mean seasonal cycle due to their low temporal resolution (bimonthly).' This is unclear, as the CEDAR-GPP with the minimum temporal resolution is at monthly, why the bimonthly data just filled with mean seasonal cycle?

Response: Thanks for the note. The original description was unclear. The temporal resolution of GIMMS LAI4g and NDVI4g is biweekly. We performed gap-filling at the native temporal resolutions of each dataset before resampling them to a monthly time scale to produce CEDAR-GPP. Since vegetation structure could experience significant changes at bi-weekly (half-month) intervals, gap-filling using temporal medians within moving windows could introduce significant uncertainties and potentially over-smooth the time series. We have revised the text to clarify this aspect.

Revisions: Line 236 – 240: *GIMMS LAI4g and NDVI4g data were only filled with mean seasonal cycle due to their low temporal resolution (half-month). This is because vegetation structure could experience significant changes at half-month intervals, and gap-filling using temporal medians within moving windows could introduce considerable uncertainties and potentially over-smooth the time series.*

Figure 2 The CO2 should be separate at the left. Besides, I cannot understand why the CFE-ML is omit in the long-term product.

Response: Thank you for the suggestion. We consider CO2 concentration an atmospheric condition and thus have included it in the "Climate" category. We linked "Climate" directly to "Biophysical Theory" in the diagram since the theoretical function depends on both air temperature and CO2. Therefore, we did not present CO2 as a separate group for cohesion and simplicity of the schematic. We hope this addressed your concerns, but please feel free to let us know if our interpretation was inaccurate.

Regarding the exclusion of the CFE-ML in the long-term product., we would like to clarify that the model was trained on data spanning 2001 to 2020, during which CO2 levels ranged from 370 to 412 ppm. Applying it to the 1980-1990s, when CO2 levels were much lower (340 to 365 ppm), would likely result in inaccurate GPP responses to CO2. This is because machine learning models typically do not extrapolate outside the range of their training data. Unlike CO2, other climate and environmental conditions did not have dramatic shifts in their magnitudes, despite interannual variations. Therefore, we include only Baseline and CFE-Hybrid setups in the long-term products. In the revised manuscript, we have added a more detailed explanation of this choice.

Revisions: Line 268 – 272: *Due to the limited availability of eddy covariance observations before 2001, we did not apply the CFE-ML approach to the long-term setups. The CFE-ML model, when trained on data from 2001 to 2020 with atmospheric CO2 ranging from 370 to 412 ppm, would not accurately predict GPP response to CO2 for the period 1982 – 2000 when the CO2 levels were markedly lower (roughly 340 – 369 ppm). This is because machine learning models, especially tree-based models, could not extrapolate beyond the range of the training data.*

L236 should be section 2.3.2

Response: Thanks for the note. We have made the correction.

L238- L241 If the CFE-ML model is well trained by the LAI4g and other input data during 2001-2020, it could also be extent 1982-2000. So why missing the flux data before 2000 hinder the CFE-ML processing? The CFE-ML before 2000 should also be compared to other products.

Response: Thanks for the question. We have clarified the rationale for excluding the CFE-ML setup from our long-term products in our response earlier (the comment starting with "Figure 2"). We have revised the text to clarify the underlying reasons for excluding CFE-ML models in the long-term product.

Revisions: Line 268 – 272: *Due to the limited availability of eddy covariance observations before 2001, we did not apply the CFE-ML approach to the long-term setups. The CFE-ML model, when trained on data from 2001 to 2020 with atmospheric CO_2 ranging from 370 to 412 ppm, would not accurately predict GPP response to CO_2 for the period 1982 – 2000 when the CO_2 levels were markedly lower (roughly 340 – 369 ppm). This is because machine learning models, especially tree-based models, could not extrapolate beyond the range of the training data.*

L249 ' do not consider the direct effect of CO_2 on light use efficiency', this is inaccurate, the VI sometimes can capture the LUE change from direct CFE.

Response: We would like to clarify that vegetation indices, such as NDVI and EVI, primarily respond to vegetation structural variations, reflecting the amount of healthy green biomass. They were designed to exploit the strong reflectivity in NIR associated with light scattering within leaf internal mesophyll structures rather than leaf biochemical properties (Gamon et al., 1996). Therefore, these VIs mainly capture changes in vegetation greenness or LAI, indicating canopy light absorption (fAPAR). They do not directly reflect changes in light use efficiency (LUE) associated with the CO_2 impacts on the photosynthetic biochemical processes. Thus, our Baseline model, along with other upscaled datasets including FLUXCOM and FLUXSAT, does not consider the direct CFE on LUE.

It's also noteworthy that the narrowband Photochemical Reflectance Index (PRI), sensitive to xanthophyll pigment changes, reflects leaf to canopy level LUE, particularly under physiological stresses (Garbulsky et al., 2014; Penuelas et al., 2011). PRI can be derived from satellite data, such as ocean bands from MODIS. However, its relationships with LUE are confounded by sun-sensor geometry, vegetation structure, and soil background (Middleton et al., 2016). Therefore, future research is necessary to assess the potential of PRI data derived from satellites in tracking long-term LUE changes.

We have revised the corresponding sentence and improved the description of the direct and indirect CFE in the introduction for clarity. We hope that explanations in our response and revisions throughout the manuscript clarify this important aspect of our work compared to existing upscaled datasets.

Reference:

Gamon, J. A., Field, C. B., Goulden, M. L., Griffin, K. L., Hartley, A. E., Joel, G., Peñuelas, J., and Valentini, R.: Relationships Between NDVI, Canopy Structure, and Photosynthesis in Three Californian Vegetation Types, Ecological Applications, 5, 28–41, https://doi.org/10.2307/1942049, 1995.

Garbulsky, M. F., Filella, I., Verger, A., and Peñuelas, J.: Photosynthetic light use efficiency from satellite sensors: From global to Mediterranean vegetation, Environmental and Experimental Botany, 103, 3–11, https://doi.org/10.1016/j.envexpbot.2013.10.009, 2014.

Middleton, E. M., Huemmrich, K. F., Landis, D. R., Black, T. A., Barr, A. G., and McCaughey, J. H.: Photosynthetic efficiency of northern forest ecosystems using a MODIS-derived Photochemical Reflectance Index (PRI), Remote Sensing of Environment, 187, 345–366, https://doi.org/10.1016/j.rse.2016.10.021, 2016.

Peñuelas, J., Garbulsky, M. F., and Filella, I.: Photochemical reflectance index (PRI) and remote sensing of plant CO_2 uptake, New Phytologist, 191, 596–599, https://doi.org/10.1111/j.1469-8137.2011.03791.x, 2011.

Revisions: Line 82 - 91 (Introduction): *Increasing atmospheric CO_2 directly stimulates the biochemical rate or the light use efficiency (LUE) of leaf-level photosynthesis, known as the*

direct CO2 fertilization effect (CFE). Enhanced photosynthesis could lead to greater net carbon assimilation, contributing to an increase in total leaf area. This expansion, contributing to a higher light interception, further enhances canopy-level photosynthesis (i.e. GPP), which is referred to as the indirect CFE. The direct CFE has been found to dominate GPP responses to CO2 compared to the indirect effect, from both theoretical and observational analyses (Haverd et al., 2020; Chen et al., 2022).
Line 279 – 283 (Section 2.3.2)*: As such, the models only include indirect CO2 effects from the satellite-based proxies of vegetation greenness or structure representing changes in canopy light interception, and they do not consider the direct effect of CO2 on leaf-level photosynthetic rates (or light use efficiency, LUE). Our baseline model is therefore directly comparable to other satellite-derived GPP products that only account for indirect CO2 effects (Jung et al., 2020; Joiner and Yoshida, 2020).*

L257-271 an open question here, I read the EEO theory from Chen et al. 2022 PNAS. The authors may refer to the SI from that paper, you may see the EEO theory cannot reproduce the GPP trend at tropical rain forests (GF-Guy). So whether the CFE-hybrid can capture the actual trend of GPP, more validation is needed.

Response: Thanks for the question. We acknowledge that it remains challenging to accurately model long-term GPP trends, and all existing models present uncertainties of various extent, and some models, including the EEO model by Chen et al., (2022), are underperformed in specific biomes such as evergreen broadleaf forests. In the revised manuscript, we have included direct validation of the CFE-Hybrid and CFE-ML models in predicting long-term trends across global eddy covariance sites, categorized by different plant functional types and climate zones. These results could provide comprehensive information on the uncertainties of our products in representing GPP dynamics.

Revisions: Section 3.2.4, Figure 5, Figure S3, S4

L325 To my knowledge, the FLUXCOM has removed the annual trend which has been pointed out by the data use guideline, so why used the FLUXCOM data?

Response: Thank you for the question. The FLUXCOM product, as described in Jung et al. (2020), "does not account for CO2 fertilization effects", hence presents unrealistic trends in carbon fluxes. (Note that it was not specifically detrended.) The CEDAR-GPP product aimed to address this issue by introducing the direct CO2 effect into upscaled GPP estimates. Comparing our product with FLUXCOM thus presents an informative assessment of the effectiveness of our approach in producing long-term GPP trends. Moreover, benchmarking CEDAR-GPP with FLUXCOM and other products also provides useful evaluation regarding GPP spatiotemporal patterns, including mean annual values, mean seasonal cycles, and interannual variability.

Figure3 it should have systematic validation for ten setups but not only validate the CFE-Hybrid_NT. Besides, the GPP anomaly seems to be underestimated compared to the site observation?

Response: Thank you for the comment! We would like to clarify that Figure 3b, d, f, h shows R2 for all five models based on the NT setup, and Figure S1 displays the results for the other five models (DT). Table S3 presented the performance metrics (RMSE, Bias, and R2) for the estimation of monthly GPP and its mean seasonal cycles and interannual variabilities by all ten model setups. Additionally, thanks for your observation regarding the underperformance of GPP anomaly prediction, which is a long-standing challenge documented by previous investigations.

In the revised manuscript, we have improved the clarity of the text when describing the results of GPP anomalies.

Revisions: Line 399 – 400 (Section 3.1.1): *However, all models underestimated monthly anomalies across the sites, with R2 values below 0.12 (Figure 3e*-f).

Line 750 – 757 (Section 4.1.2): *Nevertheless, accurately capturing interannual anomalies remains challenging for certain biomes, such as evergreen needleleaf forest, cropland, and wetland (Figure 4), as acknowledged by previous studies (Tramontana et al., 2016; Jung et al., 2020). This suggests that vital information on GPP is missing or inadequately represented in existing datasets. To this end, potential improvement may be achieved by incorporating datasets related to agricultural management practices (crop type, cultivar, irrigation, fertilization) (Xie et al., 2021), plant hydraulic and physiological properties (Liu et al., 2021), dynamic C4 plant distributions (Luo et al., 2024), root and soil characteristics (Stocker et al., 2023), as well as topography (Xie et al., 2023).*

Figure5 the trend should be analyzed among PFT and climate zone similar to figure4. More importantly, the sites showed a higher GPP trend than the models. Do you mean the CEDAR-GPP underestimate the GPP trend? Why the trend in CFE-ML-NT is much higher than it in CFE-ML-DT?

Response: Thank you for the feedback. We have incorporated additional evaluation results by PFT and climate zones in Figure 5 and noted the underestimation of the GPP trend by the CEDAR models in section 3.1.3.

Moreover, the observed differences between the NT and DT models for the CFE-ML setup reflect that the GPP trends from the NT partitioning can be more effectively captured by machine learning, based on the set of predictor variables, than those from DT based on GPP. This discrepancy may highlight the underlying uncertainties or systematic biases in the NT and DT approaches (see Section 4.1.1, Line 668 – 671). These results also suggest the considerable uncertainties present in the machine learning quantification of CO2 fertilization and underscore the need for further studies using advanced approaches, such as explainable machine learning and causal inference.

Revisions: Figure 5, Section 3.1.3

Line 468 – 471 (Section 3.1.3) *While the machine learning models generally did not fully capture the enhancement in GPP for most categories, the CFE-ML and/or CFE-hybrid models consistently outperformed the Baseline models in both ST and LT setups.*

Line 816 – 823 (Section 4.2): *Nonetheless, the considerable ensemble spread in the CO2 trends from the CFE-ML model and discrepancies between the CFE setups (Figure 11, Figure 13) underscored a high level of uncertainty in the machine learning quantified CO2 effects. Moreover, disentangling the direct CO2 effects on LUE, water use efficiency, and its indirect effects on fAPAR remains challenging with machine learning models due to the correlations and interactions between CO2 and other climatic or environmental factors. Future work may exploit explainable machine learning and causal inference to unravel the complex mechanisms and distinct pathways of CO2 effects on vegetation carbon uptake.*

Figure7 The standard deviation of GPP at different months should be added to the figure.

Response: Thank you for the suggestion. Since the global and regional variations of GPP mean seasonal cycles are substantial, we are concerned that adding the standard deviation to Figure 7 may reduce its clarity. Instead, we have provided this information in an additional figure (Figure S7) in the Supplementary Information.

Revisions: Figure S7

Figure8 The FLUXCOM data cannot be applied to evaluate the IAV as the product guideline has pointed out. The baseline and CFE-ML product should be listed here.

Response: Thanks for the comment! We recognize the issue of underestimated IAV in FLUXCOM product noted by Jung et al., (2020). However, we used an updated FLUXCOM GPP dataset based on MODIS v006 dataset (as indicated in Section 2.4), which showed substantial improvement in IAV representation (as shown in Figure 8).

Regarding the CEDAR-GPP model setups, we presented the estimated GPP IAV from all CEDAR model setups in Figure S7, where all models showed consistent IAV patterns. Therefore, in Figure 8, which aimed to compare IAV patterns between CEDAR-GPP and other products, we selected the ST_CFE-Hybrid_NT as a representative example to minimize redundancy. We have added a clarification regarding this aspect in the text.

Revisions: Line 572 – 574 (Section 3.2.3): *Ten CEDAR-GPP model setups presented consistent patterns in interannual variability, and differences were minimal (Figure S7).*

Figure11 Can I infer the GPP trend from 2001 to 2018 is double compared to it from 1982 to 2000? Is there any existed GPP product (i.e. machine learning or TRENDY data) can support this result?

Response: Thanks for pointing this out! We would like to clarify that trends from the ST (Figure 11b) and LT setups (Figure 11d) are not directly comparable because the models were driven by different sets of remote sensing input datasets. For trends analysis between 2001 and 2020, we recommend using the short-term (ST) datasets as they were based on a more comprehensive set of remote sensing data and have demonstrated a better performance in capturing trends from the eddy covariance data (Figure 5). To analyze trends from 1982 to 2020 and assess shifts in trends between 1982-2000 and 2001-2020, the long-term (LT) dataset should be used for consistency in underlying satellite data and machine learning models. We have highlighted these distinctions and considerations in the newly added user guideline.

The LT product (Figure 11c) did indicate a small difference in GPP trends between the 1982 - 2000 and the 2001 – 2020 periods (Figure S10). This suggests a potential shift in trends from the underlying GIMMS datasets. Also, note that the significance of this difference can be affected by the specific start and end years chosen for analysis due to significant interannual variabilities in GPP. Thus, future analysis would benefit from careful statistical evaluation and intercomparison with other products to robustly assess the changes in GPP trend over time.

Revisions: Line 850 – 854 (Section 5): *1) Study period considerations: the Short-Term (ST) setup is ideal for studies focusing on periods after 2000. These models are constructed using a broader range of explanatory predictors, offering higher precision and smaller random errors. The Long-Term (LT) datasets shall be used for research assessing GPP dynamics over a longer time period (before 2001). It is important to note that trends from the ST and LT datasets are not directly comparable, as they were derived from different satellite remote sensing data.*

Function A2 A = AC = Aj? I cannot agree with this! There are just one condition Ac = Aj is the photosynthesis transfer from lighted-limited to nutrient limited.

Response: We would like to clarify that our derivation of the CO2 sensitivity is based on the eco-evolution theory, which predicts that light-limited (A_j) and Rubisco-limited (A_c) converges through Vcmax acclimation to optimize resource allocation over the scale of a few weeks (Harrison et al., 2021; Wang et al., 2017). Thus, the assumption of the equality of A_c and A_j is

valid at the monthly time scale in our model. Additionally, given that Aj has a lower response to CO_2 than A_c (Walker et al., 2021), our approach provides conservative estimates of the direct CFE. We described this approach and the underlying assumptions in the main text (Section 2.3.2). In the revised manuscript, we have further elaborated our approach in Appendix A.
Revisions: Line 902 – 906 (Appendix A): *Eco-evolutionary theory predicts that the electron-transport-limited (light-limited) (A_j) and Rubisco-limited (A_c) rates of photosynthesis converge on the time scale of physiological acclimation, which is in the order of a few weeks (Harrison et al., 2021; Wang et al., 2017). Thus, at a monthly time scale, we assume that …A = A_c = A_j (A3).*
Line 908 – 910: *In the following, we derive our sensitivity function based on A_j, which has a smaller response to CO_2 than A_c , thus providing conservative estimates of the direct CO_2 fertilization effect (Walker et al., 2021).*

Response Letter to Reviewer #2

CEDAR-GPP: spatiotemporally upscaled estimates of gross primary productivity incorporating CO_2 fertilization

Dear Dr. Wang,

Thank you very much for your thorough review of and positive comments on our manuscript. We greatly appreciate your constructive feedback, which has significantly improved our paper. Below, we outline our responses (in blue) to your comments (in black) and detail the corresponding revisions made to the manuscript (in red). We included line numbers from the "track-change" version of the revised manuscript for your reference.

Thank you once again for the time and efforts you dedicated to reviewing our manuscript. We hope that our revisions have adequately addressed your comments and concerns. We eagerly look forward to any further feedback and suggestions that you may have!

Sincerely,

Yanghui Kang
On behalf of all co-authors

General comments:

This manuscript presents a new global GPP dataset by using the machine learning method and many various datasets. They mainly considered the direct CO2 fertilization effect on GPP through both the machine learning and the theoretical approaches. They then modelled the short and long-term global GPP by using different periods of earth observations and climate variables. They also validated their new GPP data by comparing to flux sites GPP data and other products.

This study fits well for ESSD and I appreciate the efforts that the authors made. Several minor concerns from my side are listed below, which I think may be helpful for revising the manuscript.

Specific comments:

Line 20: Maybe not the "first", since you also said that the revised EC-LUE GPP have considered this.

Response: Thank you for pointing this out. We have revised this sentence to elaborate our product was the first dataset upscaled from eddy covariance measurements using machine learning models that incorporated the direct CO2 fertilization effect.
Revisions: Line 20 – 21 (Abstract): *Here, we introduce CEDAR-GPP, the first global machine-learning-upscaled GPP product that incorporates the direct CO2 fertilization effect on photosynthesis.*

Line 114: Some flux sites have the data before 2001? Why did include them into the model? And, this means that the information before 2001 from flux sites are not included into your GPP products, right?

Response: Thank you for the comment! It is correct that eddy covariance data before 2001 were not included in our models. While we recognize that sites with data back to the 1990s are valuable for evaluating long-term changes, such sites are very few (e.g. only four sites have data before 1996) and could induce model biases towards a limited range of environmental conditions in these early years. This is particularly critical for modeling the relationship between GPP and CO2. Therefore, we were concerned that applying such models globally could result in unreliable predictions. Moreover, the short-term models were based on MODIS data that was only available after 2000, and thus did not require training data beyond the time range. This approach also ensures the consistency between our short-term (ST) and long-term (LT) models.
Revisions: Line 140 – 143 (Section 2.1): *Note that we did not include eddy covariance data before 2001, since it was limited to only a few sites. This scarcity might introduce biases in the machine learning models, particularly in the relationship between GPP and CO2, leading to unreliable extrapolations across space and time*.

Line 214: Both the MODIS and CISF data are used in the model. Will them from the same observations, i.e., the CSIF is also be trained from MODIS data right. Will the multi-collinearity issue exist in this?

Response: This is a great point. Indeed, CSIF, along with other variables derived from MODIS, are highly correlated. We expect the XGBoost model to effectively manage such multi-collinearity in the feature space, as it does not rely on assumptions about the statistical distributions of the predictor variables. While multi-collinearity may affect model interpretability, repetitive information shared among the different MODIS-derived variables can actually enhance the machine learning model's predictive ability and reduce noise.
Revisions: Line 319 – 321 (Section 2.3.3): *Without relying on prior assumptions about the functional forms or statistical distributions, the model is also robust to multi-collinearity between the predictors in our dataset, particularly for the variables derived from MODIS data*.

Line 252-256: When adding the CO2 into the machine-learning model, the output of your model is GPP, but LUE. That means, the model could not distinguish the direct CO2 fertilization on LUE and the indirect CO2 impact on FAPAR? A simple way to check this is to check the CO2 impacts on GPP, LUE and FAPAR at some FACE sites, and to check that if your model is correct.

Response: Thank you for the feedback and suggestion! We agree that disentangling the direct CO2 effect on LUE and the indirect effects on FAPAR is a major challenge within machine learning frameworks, which primarily focus on optimizing the overall predictive accuracy of GPP.

Indeed, FACE experiments offer unique opportunities to quantify the direct CO2 effect on LUE and indirect effects on LAI, as thoroughly analyzed by previous studies analysis (e.g. Ainsworth et al., 2005). However, applying our machine learning models directly to such experimental conditions is challenging. These experiments involved elevated CO2 levels (e.g., 542 ppm in Duke, 547 ppm in Oak Ridge) significantly above current ambient concentration (420 ppm as of 2020). Given that the machine learning models do not extrapolate over unseen conditions, they cannot reliably produce predictions under the FACE experiment scenarios.

Our objective was to provide comprehensive quantifications of the overall GPP trends, encompassing both the direct and indirect CFE, to address omissions of the direct effects in previous upscaled datasets. By incorporating both CO2 and remote sensing proxies indicating canopy structure (LAI, fAPAR) as predictors, our machine learning models (i.e. CFE-ML) aimed to capture the overall CO2 effects. The enhanced performance of the CFE-ML models compared to the baseline models, in capturing the significant increase of GPP from eddy covariance data, suggests that the CFE-ML model effectively captured the direct CFE absent in the Baseline models, providing an improved representation of the overall trend.

In light of your valuable feedback, we have expanded our discussion (Section 4.2) to emphasize the importance and challenges of isolating the direct and indirect CFE with the machine learning frameworks and highlight future opportunities through advanced machine learning and causal inference techniques. Moreover, we have extended our model validation against eddy covariance data across different plant functional types and climate zones (Section 3.1.3, Figure 5, Figure S3), and included further analysis at sites undergoing long-term GPP changes (Figure S4). These additional analyses aimed to demonstrate our models' capacity to capture the overall GPP trends including both the direct and indirect CFE.

Revisions: Line 818 – 823 (Section 4.2): *Moreover, disentangling the direct CO2 effects on LUE, water use efficiency, and its indirect effects on fAPAR remains challenging with machine learning models due to the correlations and interactions between CO2 and other climatic or environmental factors. Future work may exploit explainable machine learning and causal inference to unravel the complex mechanisms and distinct pathways of CO2 effects on vegetation carbon uptake.*

Line 467 – 485 (Section 3.1.3): *Aggregated eddy covariance GPP experienced increasing trends of varied magnitudes across different climate zones and plant functional types (Figure 5b,c; Figure S3b,c). While the machine learning models generally did not fully capture the enhancement in GPP for most categories, the CFE-ML and/or CFE-hybrid models consistently outperformed the Baseline models in both ST and LT setups. The CFE-ML setup predicted a higher trend than CFE-hybrid in most cases, suggesting that the data-driven approach captured more dynamics not represented in the theoretical model, which was based on conservative assumptions regarding the CO2 sensitivity of photosynthesis (see Sect. 2.3.3 and Appendix A). The choice of remote sensing data (ST vs. LT configurations) did not lead to substantial differences in the predicted GPP trend. Most long-term flux sites (at least 10 years of records) with a significant trend experienced an increase in GPP, and the CFE-ML and/or CFE-hybrid models again aligned closer to eddy covariance data than the Baseline models (Figure S4). Additionally, we found a considerably higher trend in eddy covariance GPP measurements derived from the day-time versus night-time partitioning approach, potentially associated with uncertainties in GPP partitioning methods (Figure S4). Yet, machine learning model predicted trends were not strongly affected by GPP partitioning methods (Figure S3, S4).*

Line 261-262: A constant χ value for global and long-term analysis seems will introduce large uncertainties. Since we know that χ will change with different climate conditions, and also will change largely with the CO2 concentration. Although we know that the modelling of global χ is very hard, but some discussions and limitations on this aspect is at least needed.

Response: Thank you for the feedback! We totally concur that applying a constant χ value in global and long-term analysis can introduce uncertainties. We have expanded our discussion to underscore the dynamics of χ in response to meteorological conditions and associated uncertainties, and provided insights for potential improvements in future work.

Revisions: Line 792 – 799 (Section 4.2): *For simplicity, we assumed a fixed ratio of leaf internal to ambient CO2 (χ) representing an average long-term value typical for C3 plants in the* *theoretical CO_2 sensitivity function. However, χ varies by environmental conditions, including temperature and vapor pressure deficit, and robustly modeling these dependencies remains challenging (Wang et al., 2017). Future work could incorporate more comprehensive representations of the χ and evaluate how the associated uncertainties affect the quantification of GPP and its temporal variations.*

Response Letter to Reviewer #3

CEDAR-GPP: spatiotemporally upscaled estimates of gross primary productivity incorporating CO_2 fertilization

Dear reviewer,

Thank you very much for your thorough review of our manuscript. We greatly appreciate your constructive feedback and comments, which have significantly improved our paper. Below, we outline our responses (in blue) to your comments (in black) and detail the corresponding revisions made to the manuscript (in red). We included line numbers from the “track-change” version of the revised manuscript for your reference.

Thank you once again for the time and efforts you dedicated to reviewing our manuscript. We hope that our revisions have adequately addressed your comments and concerns. We eagerly look forward to any further feedback and suggestions that you may have!

Sincerely,

Yanghui Kang
On behalf of all co-authors

In this manuscript, Kang and colleagues present a novel GPP product that incorporates the CO2 fertilization effect. After several thorough readings over the past two weeks, it has become evident to me that this manuscript is both timely and well-prepared. I have some minor suggestions to enhance its clarity and impact:

1. In the Introduction section, it would be beneficial to provide a more comprehensive explanation of why the authors chose scaling as their approach for global GPP development. Given that there are various other GPP development methods, such as LUE models and SIF retrievals, it's essential to highlight the unique advantages of scaling in comparison to these alternatives.

Response: This is a great point. We have enriched the introduction with a more detailed description of the upscaling approach and its advantages.
Revisions: Line 56 – 64: *This “upscaling” approach provides data-driven and observation-based quantifications without prescribed functional relations between GPP and its climatic or environmental drivers. It offers unique empirical constraints of ecosystem carbon dynamics, complementing those derived from process-based and semi-process-based approaches such as terrestrial biosphere models or the Light Use Efficiency (LUE) models (Beer et al., 2010; Jung et al., 2017; Schwalm et al., 2017; Gampe et al., 2021). In recent years, the growth of global and regional flux networks, coupled with increasing efforts in data standardization, has offered new opportunities for the advancement of upscaling frameworks, enabling comprehensive quantifications of terrestrial photosynthesis (Joiner and Yoshida, 2020; Pastorello et al., 2020).*

2. The concept of "direct CO2 fertilization effect" may not be familiar to all readers. It would be helpful if the authors could provide a clearer explanation of this term and include relevant references in the introduction to ensure a better understanding among readers.

Response: Thank you for the suggestion! We have revised the introduction to provide a clearer description of the direct and indirect CO2 fertilization effect.
Revisions: Line 82 – 19: *Increasing atmospheric CO2 directly stimulates the biochemical rate or the light use efficiency (LUE) of leaf-level photosynthesis, known as the direct CO2 fertilization effect (CFE). Enhanced photosynthesis could lead to greater net carbon assimilation, contributing to an increase in total leaf area. This expansion, contributing to a higher light interception, further enhances canopy-level photosynthesis (i.e. GPP), which is referred to as the indirect CFE. The direct CFE has been found to dominate GPP responses to CO2 compared to the indirect effect, from both theoretical and observational analyses (Haverd et al., 2020; Chen et al., 2022).*

3. Figure 1 effectively illustrates the spatial distribution of flux tower data used in this study. it would be more beneficial to indicate how these flux tower sites represent different biomes. I would suggest adding a Whittaker biome figure within Figure 1 to visually depict the biome types associated with the flux tower locations used in the study.

Response: Thank you for the great suggestion! We have added a Whittaker biome plot to illustrate site distributions in the climate space in Figure 1d. Additionally, we have incorporated a plot showing the site distribution by the actual plant functional types (Figure 1c). We would also like to note that the number of sites in each biome and climate zone was also provided in Figure 4 to assist the interpretation of the results.
Revisions: Line 138 – 140 (Section 2.1): *Despite their uneven geographical distribution, these sites effectively cover a diverse range of climatic conditions and are representative of global biomes (Figure 1c, 1d).*

4. Given the potential wide interest in the various GPP products developed in this work, it would be very useful to include a section that provides guidance on how future researchers and users can effectively utilize these datasets in the future. This could include tips on data access, processing, and interpretation to facilitate broader adoption.

Response: Thank you for the suggestion! In the revised manuscript, we have provided a comprehensive guideline on data usage and selection in the Section “Data availability and usage”. We have also incorporated this information into the user guide in the Zenodo data repository.
Revisions: Line 845 – 868 (Section 5): *The CEDAR GPP product offers GPP estimates derived from ten different models. Models are characterized by 1) temporal coverage, 2) configuration of CO2 fertilization, and 3) GPP partitioning approach (Table 2). We provide a structured approach to selecting the most appropriate dataset for research or applications.*

1) Study period considerations: the Short-Term (ST) setup is ideal for studies focusing on periods after 2000. These models are constructed using a broader range of explanatory predictors, offering higher precision and smaller random errors. The Long-Term (LT) datasets shall be used for research assessing GPP dynamics over a longer time period (before 2001). It is important to note that, due to basis in different satellite remote sensing data, trends from the ST and LT datasets are not directly comparable.

2) CO_2 Fertilization Effect (CFE) configurations: the CFE-Hybrid and CFE-ML setups are preferable when assessing temporal GPP dynamics, especially long-term trends. The CFE-Hybrid setup includes a hypothetical trend for the direct CO_2 effect, while CFE-ML is purely data-driven and does not make any specific assumption about the sensitivity of photosynthesis to CO_2. Averaging the CFE-Hybrid and CFE-ML estimates is acceptable, with the difference between them reflecting the uncertainty surrounding the direct CO_2 effect. Note that the Baseline setup should not be used to study long-term GPP dynamics, especially those induced by elevated CO_2. The Baseline setup may be useful to compare with other remote sensing-derived GPP datasets that do not consider the direct CO_2 effect. Differences between these setups regarding mean GPP spatial patterns, seasonal and interannual variations are minor.

3) GPP partitioning methods: We recommend using the mean value derived from both the "NT" (Nighttime) and "DT" (Daytime). The difference between these two provides insight into the uncertainties arising from the partitioning approaches used in GPP estimation from eddy covariance measurements.

5. To streamline the manuscript, consider moving some supplementary materials, such as Table 1 and Table 3, to the supplementary section. This would help maintain a concise and focused main manuscript while still providing access to important supporting information.

Response: Thank you again for the feedback! To streamline the manuscript, we have moved Table 3 and results related to the DT models (i.e. setups based on day-time partitioned GPP) in Figure 3, Figure 5 to the supplementary information. We have opted to keep Table 1 in the main text as it provides essential information on the input datasets, which are critical to our upscaling approaches.