General Comments

Q1. First, the choice of simple vegetation indices as dependent variables for the model seem to me dated, especially due to the current availability of Solar Induced Fluorescence (SIF) products, which are more suited as proxies of photosynthesis than EVI, NDVI, etc. Although the authors mention the possible future use of SIF, I would like to know further details to why it was not used in this study, or extra analysis where SIF is included.

Thank you for your insightful comment. We acknowledge that Solar Induced Fluorescence (SIF) is a promising proxy for photosynthesis and has been increasingly used in recent studies. Compared to traditional vegetation indices (e.g., EVI, NDVI), SIF directly reflects chlorophyll fluorescence emissions, providing a more direct link to gross primary production (GPP).

However, in this study, we did not incorporate SIF due to the following reasons:

Data Availability: Solar-induced fluorescence (SIF) observations have significantly advanced in recent years, yet the availability of long-term, continuous SIF datasets with fine spatial resolution remains a challenge. In comparison to well-established vegetation indices such as the Enhanced Vegetation Index (EVI) and the Normalized Difference Vegetation Index (NDVI), which have been monitored for decades using sensors like MODIS, SIF datasets are relatively recent. The SIF data listed in Table 1 highlight various datasets with different temporal coverage, spatial resolutions, and geographic extents. While some datasets, such as GOME-2 and OCO-2, provide global coverage and span several years, none of the available datasets fully meet the temporal coverage requirements for all FLUX station periods. Additionally, combining SIF products from different sources could introduce inconsistencies, leading to potential errors. These inconsistencies pose a significant risk to the reliability and accuracy of analyses, which is why we chose not to use these SIF products for generating a long-term time series.

Table 1 Summary of Satellite Datasets for Solar-Induced Fluorescence (SIF) Observations.

Dataset	Temporal coverage	Spatial resolutions	Time resolutions	Coverage
GOME-2	2007 to present	40 km × 40 km	1-2 days	Global
OCO-2	2014 to present	1.3 km × 2.25 km	16 days	Global
TROPOMI	2018 to present	7 km × 7 km	1 day	Global
GOSAT	2009 to present	10 km × 10 km	3 days	Global
SCIAMACHY	2002 – 2012	30 km × 60 km	35 days	Global
TanSat	2016 to present	1 km × 2 km	16 days	Global
OCO-3	2019 to present	1.6 km × 2.2 km	16 days	Global
CFIS	2016 – 2018	30 m × 30 m	Irregular	Local
TANSO-FTS	2009 to present	10 km × 10 km	3 days	Global

Resolution Limitations: The current global SIF products, such as those from OCO-2 and TROPOMI, often have spatial resolutions that are relatively coarse, typically greater than 1 km. While suitable for large-scale or global studies, this level of resolution is insufficient for capturing fine-scale ecological variations, particularly in heterogeneous or fragmented landscapes. For instance, OCO-2's spatial resolution of 1.3 km × 2.25 km and TROPOMI's 7 km × 7 km resolution may not be ideal for studies requiring detailed local information or the monitoring of small-scale ecosystem dynamics. Some datasets like CFIS, with a resolution of 30 m × 30 m, offer much finer spatial detail while their spatial coverage of datasets is usually incomplete, which cannot meet our continuous and full flux sites coverage needs in a large area.

For these reasons, we did not incorporate SIF datasets in our current study. That being said, we acknowledge the potential benefits of incorporating SIF and are considering its integration in future research. We plan to explore whether SIF-based models can further improve GPP estimations, either as a standalone predictor or in combination with traditional vegetation indices. Once again, we appreciate your valuable suggestion and will take this into account in our future work.

Q2. Second, the resolution of the remote sensing products used (500 meters) does not seem to be compatible with the eddy flux data. At this scale, microclimatic or topographic factors may cause significant divergences in relation to a 500 m size pixel, and lead to inconsistencies. I suggest that if possible data with higher resolution are used (LANDSAT or SENTINEL-2) or arguments are given for the use of the lower resolution product.

Thank you for your thoughtful suggestion regarding the spatial resolution of the remote sensing products used in our study.

First, we understand your concern that the 500 m spatial resolution of MODIS data might not be ideal for capturing fine-scale variations relevant to eddy covariance measurements. However, it is important to note, as described by Schmid (2002), that the footprint of an eddy covariance tower is not fixed but varies with meteorological conditions, typically ranging from 100 m to 1 km. Additionally, Zhang et al. (2021) found that different footprints, such as 500, 1000, and 1500 meters, showed almost no difference in the study area. Given this, we believe that the 500 m resolution of MODIS is appropriate for representing the footprint of the flux tower and is well-suited for our study.

We did consider the use of higher-resolution products, such as LANDSAT and SENTINEL-2, but there are a few important limitations associated with these datasets.

Regarding LANDSAT data, although it offers finer spatial resolution, there are known issues with data quality. Several Landsat satellites, including Landsat 7, suffered from technical failures that resulted in data gaps and missing information. These issues compromise the consistency and reliability of the dataset, particularly for long-term monitoring studies. As a result, the data quality and temporal consistency of LANDSAT may not be suitable for this study.

As for SENTINEL-2, although it provides high-resolution imagery (10 m), its temporal coverage is limited compared to MODIS. SENTINEL-2 data is available since 2015, which means it doesn't fully cover the historical periods needed for our analysis, especially for longer-term studies. Furthermore, while SENTINEL-2 offers good spatial resolution, it may not always be available due to cloud cover and other environmental factors, further complicating its use for continuous monitoring.

Considering these limitations, we chose to use MODIS data with 500 m resolution because it offers a good balance between spatial resolution, temporal coverage, and global availability, making it more suitable for our study's long-term monitoring needs.

We hope this clarifies our choice of data and addresses your concerns. Thank you again for your valuable input, which will help us refine our approach.

Q3. Finally, I would be very interested in the production of a GPP map of China using the FLAML framework, and how it compares with other GPP maps. I think this would greatly increase the manuscript's appeal.

Thank you for your valuable suggestion. Your input has provided us with very useful inspiration. Using the FLAML framework to create a GPP (Gross Primary Productivity) map for China is indeed a meaningful and interesting task. As we have mentioned in the text, the FLAML-LUE models have "the potential to be applied in predicting GPP for different vegetation types at a regional scale". However, these models are only driven from data of 20 stations, which is not enough to cover the entire ecosystem types in China. Therefore, using them for the production of a China GPP map is still not competent enough. This is not related to the limitations of the method, it's just that we need more site data support.

We plan to further develop this aspect in our future research and will provide a detailed discussion of it in the manuscript. We will consider using the FLAML framework to build a GPP prediction model for China and compare it with existing GPP maps to assess its accuracy and applicability. This will not only help us better understand the spatial distribution of GPP in China but also provide valuable insights for global GPP research.

Once again, thank you for your insightful feedback. Your suggestion will undoubtedly enrich the depth and scope of our research. We will continue to explore this direction in our future work and present the results more comprehensively in the manuscript.

Specific comments

Q1. L90 - I would not say ML is "fundamentally different" from regression models, but that they offer advantages in relation to.

Thank you for your insightful comment. You are absolutely right that machine learning is not fundamentally different from regression models but rather offers advantages in certain aspects. We have revised the text accordingly to better reflect this distinction. The revised sentence now reads: "ML is a modeling approach that differs from simple regression models and complex simulation models in its methodology."

Q2. L94 - I would also point out limitations on ML techniques, such as dependence on large training datasets and not being able to link results to real-world processes.

Q3. L96 - ...Which is an advantage when the focus is solely on spatial predictions

Response to Q2 (L94) and Q3 (L96): Thank you for your valuable comments. We acknowledge that machine learning techniques have certain limitations, including their dependence on large training datasets and the challenge of directly linking results to real-world processes. These constraints are important considerations when applying ML models. However, as you pointed out, when the primary focus is on spatial predictions, the ability of ML models to capture complex patterns without requiring explicit process-based formulations can be an advantage. We have revised the manuscript to reflect these points more clearly.

We have revised our manuscript as follows:

"These data-driven models are particularly suited for capturing nonlinear ecosystem dynamics but often require large training datasets and may lack explicit links to real-world processes. However, their ability to uncover spatial patterns without process-based constraints makes them valuable for spatial predictions. Consequently, ML-based approaches have gained popularity in recent years. For example, Kong et al. (2023) developed a hybrid model that combines ML and LUE model to estimate GPP. This hybrid model improves the LUE model by integrating a machine learning approach (MLP, multi-layer perceptron), and estimates GPP using the MLP-based LUE framework along with additional required inputs."

Q4. Fig. 1 - The mini-map on the bottom right corner does not include any sites, or any extra information, maybe remove it? Otherwise, I believe the editors should label these areas in the South China Sea as "under dispute", as stated in the "maps and aerials" section of the submission guidelines.

We sincerely thank the reviewers for their valuable suggestions regarding the mini-map in Figure 1. However, we would like to clarify that the map reflects the distribution of flux sites within China's territory. As required, we have ensured that the map accurately represents China's territorial boundaries. This representation is consistent with the practices in previous publications in Geoscientific Model Development (GMD). For example, in the article by Ren et al. (Ren et al., 2024), Figure 1, and in Figure 1 of the article by Wang et al. (Wang et al., 2022) and Figure 2 of the article by Wu et al. (Wu et al., 2021) , the South China Sea is similarly depicted as part of China' s territory without any additional labels indicating disputes.

We understand the sensitivity of territorial issues and the importance of adhering to journal guidelines. However, given the scientific context of our study and the precedent set by other publications in GMD, we believe that the current representation of the map is appropriate. We hope this explanation addresses the reviewer's concern.

Q5. Table 2 – In contrast to other vegetation indexes, LAI satellite data is based on empirical models, such as previous GPP estimating methods. It would be interesting to check if field LAI data from the sites are available to see if direct LAI measurements improve the ML model.

Thank you for your insightful comment. You are absolutely right that LAI satellite data, unlike other vegetation indices, is often derived from empirical models, similar to GPP estimation methods. We appreciate your suggestion to explore the availability of field LAI data from the study sites. We also believe that incorporating direct field LAI measurements could potentially enhance the performance of the ML model by providing more accurate and site-specific information. Unfortunately, at this stage, field LAI data in most of the 20 sites were not available. However, we plan to explore this avenue in future research and will certainly consider integrating field measurements of LAI if they become available, as they may provide valuable improvements to the model.

Q6. L686 - I would argue then that in the future hyperspectral data + ML would provide much better estimates too, this could be discussed with references.

Thank you for your valuable suggestion. We agree that hyperspectral data, when combined with machine learning (ML) techniques, could provide more accurate and robust estimates in the future. Hyperspectral data offer a rich spectrum of information across many wavelengths, which can capture subtle variations in vegetation properties that other remote sensing datasets might miss. This could indeed improve model predictions by providing more detailed spectral features.

We have revised our manuscript as follows:

"Recent research indicates that satellite observations of solar-induced chlorophyll fluorescence (SIF) provide a more accurate picture of plant photosynthesis dynamics and serve as a more effective indicator for modeling subtropical evergreen vegetation (Sun et al., 2017; Frankenberg et al., 2011). In the future, integrating hyperspectral data with machine learning could lead to more accurate GPP estimates, as hyperspectral data offer finer spectral resolution, enabling better capture of vegetation traits and environmental conditions (Gessner et al., 2015; Zarco-Tejada et al., 2013). This integration could further enhance model performance, particularly for evergreen forests. For example, Zhang et al. (2021) used hyperspectral data (EO-1 Hyperion) to estimate GPP in the temperate forests of Changbai Mountain. Future research should consider incorporating both hyperspectral data and SIF into models to assess their potential for improving GPP estimations across various ecosystems."

We appreciate your input and will explore the literature on this topic to strengthen our discussion.

Reference

Frankenberg, C., Fisher, J.B., Worden, J., Badgley, G., Saatchi, S.S., Lee, J.-E., Toon, G.C., Butz, A., Jung, M., Kuze, A., Yokota, T., 2011. New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. Geophys. Res. Lett. 38. https://doi.org/10.1029/2011GL048738

Gessner, U., Machwitz, M., Esch, T., Tillack, A., Naeimi, V., Kuenzer, C., Dech, S., 2015. Multi-sensor mapping of West African land cover using MODIS, ASAR and TanDEM-X/TerraSAR-X data. Remote Sens. Environ. 164, 282–297. https://doi.org/10.1016/j.rse.2015.03.029

Kong, D., Yuan, D., Li, H., Zhang, J., Yang, S., Li, Y., Bai, Y., Zhang, S., 2023. Improving the

Estimation of Gross Primary Productivity across Global Biomes by Modeling Light Use Efficiency through Machine Learning. Remote Sens. 15, 2086. https://doi.org/10.3390/rs15082086

Ren, F., Lin, J., Xu, C., Adeniran, J.A., Wang, J., Martin, R.V., van Donkelaar, A., Hammer, M.S., Horowitz, L.W., Turnock, S.T., Oshima, N., Zhang, J., Bauer, S., Tsigaridis, K., Seland, Ø., Nabat, P., Neubauer, D., Strand, G., van Noije, T., Le Sager, P., Takemura, T., 2024. Evaluation of CMIP6 model simulations of $PM_{2.5}$ and its components over China. Geosci. Model Dev. 17, 4821–4836. https://doi.org/10.5194/gmd-17-4821-2024

Schmid, H.P., 2002. Footprint modeling for vegetation atmosphere exchange studies: a review and perspective. Agric. For. Meteorol., FLUXNET 2000 Synthesis 113, 159–183. https://doi.org/10.1016/S0168-1923(02)00107-7

Sun, Y., Frankenberg, C., Wood, J.D., Schimel, D.S., Jung, M., Guanter, L., Drewry, D.T., Verma, M., Porcar-Castell, A., Griffis, T.J., Gu, L., Magney, T.S., Köhler, P., Evans, B., Yuen, K., 2017. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. Science 358, eaam5747. https://doi.org/10.1126/science.aam5747

Wang, P., Mao, K., Meng, F., Qin, Z., Fang, S., Bateni, S.M., 2022. A daily highest air temperature estimation method and spatial–temporal changes analysis of high temperature in China from 1979 to 2018. Geosci. Model Dev. 15, 6059–6083. https://doi.org/10.5194/gmd-15-6059-2022

Wu, R., Tessum, C.W., Zhang, Y., Hong, C., Zheng, Y., Qin, X., Liu, S., Zhang, Q., 2021. Reduced-complexity air quality intervention modeling over China: the development of InMAPv1.6.1-China and a comparison with CMAQv5.2. Geosci. Model Dev. 14, 7621–7638. https://doi.org/10.5194/gmd-14-7621-2021

Zarco-Tejada, P.J., Guillén-Climent, M.L., Hernández-Clemente, R., Catalina, A., González, M.R., Martín, P., 2013. Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV). Agric. For. Meteorol. 171–172, 281–294. https://doi.org/10.1016/j.agrformet.2012.12.013

Zhang, Y., Wang, A., Yuan, F., Guan, D., Wu, J., 2021. The application of EO-1 Hyperion hyperspectral data to estimate the GPP of temperate forest in Changbai Mountain, Northeast China. Environ. Earth Sci. 80, 353. https://doi.org/10.1007/s12665-021-09639-x