Response to Anonymous Referee #3 (https://doi.org/10.5194/gmd-2024-169-RC4)

This is the review report for "FLAML version 2.3.3 model-based assessment of gross primary productivity at forest, grassland, and cropland ecosystem sites". The authors developed a FLAML modeling framework to predict vegetation gross primary productivity using hydro-meteorological variables and variables related to vegetation types and elevation. They focus on sites in China and provide detailed model performance in reproducing forest, grass, and crop sites. While the manuscript is structured, the authors need to check throughout the manuscript to ensure readability, e.g., the abbreviations. More importantly, I have some concerns about the modeling data input, cross-validation, and the selection of hydro-meteorological variables. Also, the evaluation of the absolute values of GPP can smooth out potentially poor performance during extreme situations. Therefore, a specific test for stress conditions and an evaluation of GPP anomalies are both highly recommended.

We are very grateful for your thoughtful and constructive comments, which have been instrumental in improving our manuscript. In response, we have carefully revised the manuscript by addressing each of the issues you pointed out. This process involved a substantial reorganization of the manuscript's structure and a comprehensive update of its content, resulting in extensive modifications throughout the text.

These revisions, we believe, have significantly enhanced both the clarity and scientific rigor of our work. Below, we provide detailed responses to each of your comments, explaining the corresponding changes made.

Methodology

Q1. GPP and RECO Partitioning: The manuscript should provide a clear description of the method used to partition GPP and RECO from NEE. Additionally, it is recommended to test different partitioning algorithms to assess their impact on the results.

Thank you for your insightful comment. Due to data upload inconsistencies, ER data were missing at several sites (DLG, LCA, XLG). To address this issue and ensure data consistency across all sites, we estimated ecosystem respiration (ER) using the Lloyd & Taylor equation (Reichstein et al., 2005; Lloyd and Taylor, 1994), which is a widely adopted method in flux data processing.

This approach distinguishes daytime and nighttime periods using shortwave radiation (Rg), with a threshold of 10 W/m^2 . The temperature – response function derived from nighttime ER observations was then extrapolated to estimate daytime ER. This method is commonly used across many flux tower networks for separating Reco into GPP and ER components, and thus was adopted in our study to maintain methodological consistency. This has been clarified in **Section 2.2.1** of the revised manuscript.

We fully agree with your point that evaluating the impact of different partitioning algorithms on GPP estimation is valuable. However, in the context of this study, flux partitioning serves as a preprocessing step rather than a primary research focus. A detailed comparison of flux partitioning methods would be more appropriate for a dedicated study, and we will consider exploring this direction in future work.

Q2. Train-Test Split Strategy: The procedure for splitting the dataset into training and validation sets needs to be described in greater detail. It is important to test whether the model maintains robustness during stress periods (e.g., droughts or heatwaves). Moreover, model performance should be evaluated not only in terms of seasonal GPP dynamics but also in reproducing GPP anomalies, which are crucial for capturing ecosystem responses beyond typical seasonal cycles.

Thank you for your valuable comments. We have addressed both of the concerns you raised through revisions and clarifications in the manuscript.

First, regarding the dataset split strategy, we have clearly described the methodology in **Section 2.3.1** of the revised manuscript. Specifically, the pre-processed dataset was divided into training and testing sets using the Blocked Time Series Split strategy. Given the temporal dependency of the data, standard cross-validation is not suitable for time series analysis (Reichstein et al., 2019). Instead, a block-based and non-continuous split is applied to preserve the temporal structure. In this approach, the time series is partitioned into several non-overlapping continuous training blocks (e.g., 2003-2005, 2007-2009, 2011-2013, 2015-2017, 2019-2021), with independent years reserved as the validation set following each training block (e.g., 2006, 2010, 2014, 2018, 2022). This strategy ensures that the temporal order is maintained, preventing future data from leaking into the training process and thus avoiding invalid predictions. Additionally, the method incorporates validation over multiple periods, enabling the assessment of model generalization across different climate conditions, which is crucial for evaluating the model's robustness under varying environmental scenarios.

Second, regarding the evaluation of model performance under extreme environmental conditions, we have added corresponding analyses in **Section 3.2** of the revised manuscript. Numerous studies have shown that climate extremes—such as heatwaves, droughts, and high atmospheric vapor pressure deficit (VPD)—can significantly alter ecosystem functioning and reduce carbon uptake capacity (Frank et al., 2015; Reichstein et al., 2013). These events can suppress photosynthetic activity, increase respiration rates, and disrupt the carbon exchange balance between vegetation and the atmosphere. To evaluate the robustness and reliability of the FLAML-LUE model under such stress conditions, we examined model performance in simulating GPP during three types of climate extremes: high temperature, high VPD, and drought. By analyzing model accuracy and bias under these extreme scenarios, we aim to assess its applicability and limitations in challenging environmental settings.

Additionally, we acknowledge that the impacts of other extreme weather events and the ability of the model to reproduce GPP anomalies deserve further exploration, which we plan to address in future studies.

Thank you once again for your constructive feedback, which has helped us to improve the rigor and comprehensiveness of our study.

Q3. Choice of Environmental Drivers: The exclusion of key hydrometeorological drivers such as precipitation, vapor pressure deficit (VPD), and soil moisture raise concerns. While LSWI and PDSI are included, they are indirect proxies and not physically direct controls of vegetation water uptake and stomatal regulation. The authors should justify this choice or consider incorporating more directly linked variables.
Thank you for your valuable suggestion. We fully agree that accurately representing hydrometeorological drivers is critical for modeling GPP and that variables such as precipitation, vapor pressure deficit (VPD), and soil moisture play important roles in regulating vegetation water uptake and stomatal conductance.
In our revised analysis, we have removed the PDSI dataset due to its coarse temporal resolution (monthly), which is inconsistent with the finer-scale (8-day or daily) datasets used in this study. Instead, we incorporated new variables that more directly and comprehensively capture vegetation moisture limitations from multiple ecological dimensions, based on both theoretical considerations and prior research (Chang et al., 2023):

- Atmospheric moisture limitation: Relative humidity and precipitation
- Vegetation-level moisture stress: LSWI and evaporative fraction (EF)
- Soil moisture limitation: Soil water content (SW)

We have updated the manuscript accordingly to clarify our variable selection rationale and better align with your suggestion.

Specific points

Q1. Line 101: The abbreviation "RFR" should be defined upon its first use for clarity.
Thank you for your comment. We have revised the manuscript to define "RFR" (Random Forest Regressor) upon its first appearance to ensure clarity for the readers. Additionally, we have carefully reviewed the entire manuscript to identify and address any similar issues, and have made the necessary changes throughout the text.

Q2. Line 134: To promote transparency and reproducibility, the authors should provide a persistent identifier (e.g., DOI) for the datasets used, rather than referencing a general data repository that hosts multiple sources.
Thank you for your helpful suggestion. We have uploaded all datasets used in this study to Zenodo and provided a persistent identifier (DOI) for transparency and reproducibility. The data and code availability statement at the end of the manuscript has been updated accordingly: https://doi.org/10.5281/zenodo.14542880 (Laijie, 2024).

Q3. Line 178: ERA5-Land should not be categorized as remote sensing data. It is a reanalysis product based on assimilation of observations into a numerical model.
Thank you for your valuable comments.
We have made corresponding revisions in the updated manuscript. Specifically,

Section 2.2.3 (Line 181 - 192) now reads as follows: "ERA5-Land (Hersbach et al., 2020) is a global high-resolution reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus Climate Change Service (C3S). It provides hourly land surface variables at a spatial resolution of 0.1°, generated using a dedicated land surface model driven by the ERA5 climate reanalysis. The dataset integrates advanced land surface modeling and data assimilation techniques, offering a wide range of variables such as air temperature, soil moisture, precipitation, and snow depth. In this study, site-specific variables including air temperature (T), soil water content (SW), precipitation (Pre), and leaf area index (LAI) were extracted from ERA5-Land. In addition, photosynthetically active radiation (PAR), evapotranspiration fraction (EF), VPD and relative humidity (RH) were calculated and derived from available ERA5-Land variables using GEE. "

Q4. Line 195: The acronym “LSWI” should be spelled out in full the first time it appears.

Thank you for your valuable comments. We have revised our manuscript as follows: Vegetation and water indices derived from MODIS data included the enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), and land surface water index (LSWI), which were calculated using the formulas presented in Table 2 **(Line 178)**.

Additionally, we have carefully reviewed the entire manuscript to identify and address any similar issues, and have made the necessary changes throughout the text.

Q5. Table 2: All abbreviations should be clearly defined either in the table caption or as a footnote to enhance readability.

Thank you for your valuable suggestion. We have revised Table 2 (**Line 248**) to include clear definitions of all abbreviations, which are now provided as footnotes to enhance clarity and readability.

Q6. Figure 3 (III): In model evaluation scatter plots, it is more intuitive to place observations on the x-axis and simulations on the y-axis, as this mirrors standard regression analysis practice.

Thank you for your valuable suggestion. We agree that placing observations on the x-axis and simulations on the y-axis provides a more intuitive interpretation and aligns with standard regression analysis practices. Following your recommendation, we have revised the scatter plot accordingly. In the updated version (now presented as **Figure 4**), we have also combined the three ecosystem types into a single figure to facilitate direct comparison across ecosystems.

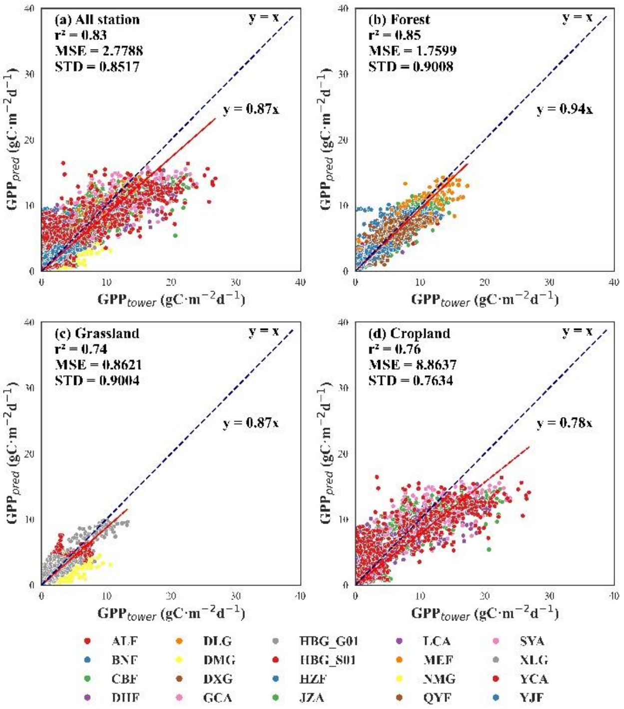

Figure 4. Scatterplot of observed GPP vs. simulated GPP. Different colored dots represent different sites. Note: The simulated GPP values represent the mean of FLAML00 to FLAML25.

Q6. Figures 5/9/13: Do the reported biases account for seasonal differences in GPP variability (i.e., high variability in summer vs. low variability in winter)? Clarifying this would improve interpretation of model performance across seasons.

Thank you for your insightful comment. In our analysis, Figures 5/9/13 show the actual biases between the GPP simulations and observations across different sites and months. We acknowledge that the manuscript does not explicitly consider the seasonal variability in GPP. GPP tends to exhibit higher variability in summer and lower variability in winter, which may lead to higher GPP in summer and lower GPP in winter. In the revised manuscript, although we have included model evaluations under extreme climatic conditions, we have not specifically addressed the seasonal biases in the GPP simulations. Instead, we chose to use the PBias metric to provide an overall assessment of the model's performance across different land surface types. The PBias metric reflects the magnitude of simulation biases between sites, offering a more comprehensive evaluation of the model (Line 311 and Line 455).

Q7. Figure 7: There is a noticeable underestimation of GPP in DLG (typical grasslands) and overestimation in DXG (alpine meadows). Can the authors explain potential causes for these systematic biases?

Thank you for your valuable comment. In the revised manuscript (**Line 441- 453**), we used the PBias (%) metric to evaluate the simulation biases of different vegetation types. As shown in **Figure 5**, there is an underestimation of GPP at DLG (typical grasslands) and an overestimation at DXG (alpine meadows). These systematic biases can be attributed to differences in the biophysical characteristics and climatic conditions of the two ecosystems.

For DLG, the grassland ecosystem typically exhibits high productivity under sufficient water availability, especially during the spring and summer growing seasons. If the model does not accurately represent the seasonal dynamics of water supply and demand, or the interaction between water availability and temperature, it may underestimate the actual GPP.

In contrast, GPP in alpine meadows like DXG is primarily constrained by low temperature and a short growing season. If the model does not fully capture these limitations—particularly under relatively cold conditions—it may overestimate the photosynthetic potential, resulting in an overestimation of GPP.

Q8. Figure 14: Are the farm ecosystems considered in the analysis purely rainfed, or do they include irrigated systems? This distinction is important for interpreting model results under water-limited conditions.

We sincerely thank the reviewer for the insightful comment regarding irrigation regimes at the cropland flux sites. Based on previous studies (Liu et al., 2023; Zhou et al., 2023; Zhang et al., 2023; Zhao et al., 2021), , the cropland ecosystems included in this study encompass both rainfed and irrigated systems. Specifically, SYA and JZA are rainfed single-cropping systems, where agricultural production primarily depends on natural precipitation. In contrast, GCA, LCA, and YCA are high-input, double-cropping systems located in intensively managed irrigated regions, where supplemental watering is essential during critical crop growth stages.

As stated in Section 4.1 of the revised manuscript (**Lines 681 – 686**), the current version of our model does not explicitly differentiate between the irrigation regimes of each site. Although we have identified the irrigation type for each location, this distinction has not yet been incorporated into the modeling framework. We fully recognize the pivotal role that irrigation plays in regulating GPP dynamics, particularly under water-limited conditions, and acknowledge that its exclusion may influence model performance and the scientific interpretation of results.

To address this limitation, we plan to integrate satellite-derived irrigation indicators in future studies—specifically, soil moisture anomalies from the Soil Moisture Active Passive (SMAP) mission and temporal patterns of the Normalized Difference Water Index (NDWI). Incorporating these indicators will enhance the model's ability to represent irrigation effects and more accurately capture the dynamic variability of carbon fluxes in agricultural ecosystems.

Reference

Chang, X., Xing, Y., Gong, W., Yang, C., Guo, Z., Wang, D., Wang, J., Yang, H., Xue, G., Yang, S., 2023. Evaluating gross primary productivity over 9 ChinaFlux sites based on random forest regression models, remote sensing, and eddy covariance data. Sci. Total Environ. 875, 162601. https://doi.org/10.1016/j.scitotenv.2023.162601

Frank, Dorothea, Reichstein, M., Bahn, M., Thonicke, K., Frank, David, Mahecha, M.D., Smith, P., van der Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J.G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Rammig, A., Seneviratne, S.I., Walz, A., Wattenbach, M., Zavala, M.A., Zscheischler, J., 2015. Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. Glob. Change Biol. 21, 2861–2880. https://doi.org/10.1111/gcb.12916

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803

Lloyd, J., Taylor, J.A., 1994. On the Temperature Dependence of Soil Respiration. Funct. Ecol. 8, 315–323. https://doi.org/10.2307/2389824

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M.D., Seneviratne, S.I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D.C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca, S., Walz, A., Wattenbach, M., 2013. Climate extremes and the carbon cycle. Nature 500, 287–295. https://doi.org/10.1038/nature12350

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob. Change Biol. 11, 1424–1439. https://doi.org/10.1111/j.1365-2486.2005.001002.x

Liu, F., Shen, Y.J., Cao, J.S., Zhang, Y.C., 2023. A dataset of water, heat, and carbon fluxes over the winter wheatsummer maize croplands in Luancheng during 2013 – 2017. China Scientific Data (in Chinese), 8, 36–45.

Zhou, L., Geng, J.J., Zhou, G.S., Zhang, S., Wu, Y.X., 2023. A dataset of carbon and water fluxes in the winter wheat and summer maize rotation ecosystem at Gucheng Station (2020–2022). China Scientific Data (in Chinese), 8, 68–77.

Zhang, S, Zhou, L., Zhou, G.S., Jia, Q.Y., Li, R.P., Wang, Y., 2023. A dataset of carbon and water flux observations in the agricultural ecosystem of spring maize in Jinzhou (2005–2014). China Scientific Data (in Chinese), 8, 148–159.

Zhao, F.H., Li, F.D., Zhan, C.S., Zhang, L.M., Chen, Z., 2021. A carbon and water fluxes dataset of the farmland ecosystem of winter wheat and summer maize in Yucheng (2003–2010). China Scientific Data (in Chinese), 6, 222–228.