Review of Moritz Adam et al., manuscript number egusphere-2026-626
Title: “Earth system models might overestimate the local plant productivity response to temperature–moisture extremes”

This manuscript evaluates the co-occurrence of temperature-moisture extremes and gross primary productivity (GPP) extremes simulated by Earth system models (ESMs) against the FLUXCOM product. The authors calculate co-occurrence rate using a binary approach, combining with a significant test. They present an overstimated coupling between temperature-moisture and GPP extremes, especially in low latigudes.

I particularly like the binary method and the significant test presented in this manuscript. The binary solution is simple but can effectively capture the link between two types of extremes, nicely bridgeing the modelled and observation-based results. In addition to soil moisture extremes, they also include precipitation as moisture metrics, which offers another interesting perspective of moisture propagation.

In addition, I however find the manuscript needs major revision to be clearer and more informative, as outlined below:

General comments:

1. The key metric of this paper, i.e., the co-occurrence of temperature-moisture and GPP extremes, is somewhat not clearly defined from the beginning of the manuscript. The authors use several statistical labels to indicate this co-occurrence, such as statistical links, coupling strength, etc. It would be beneficial to clearly address the key research question in the introduction and use the principal metric (currently there are statistical links/relationships, coupling strength, etc.) consistently.
2. One of the main findings is that the co-occurrence of both extremes is overestimated by models in low latitude regions. However, this conclusion is drawn by comparing the modelled result with FLUXCOM, which is a product using upscaling methods. Given the limited data availability, FLUXCOM GPP is considered highly uncertain over the tropical regions (Nelson et al., 2024). It is probably true that ESMs also fail to represent the co-occurrence of both extremes in low latitudes, but if there is already large uncertainty in the reference dataset, this conclusion seems to be less convincing. Although there is no directly observed GPP dataset, maybe showing complementary analysis by using another independent GPP product will show robustness of the main results.
   Additionally, I wonder why the authors choose to use FLUXCOM GPP instead of FLUXCOM-X (Nelson et al., 2024) as the reference data. FLUXCOM-X provides daily GPP that matches the temporal resolution of meteorological variables analyzed in this paper. Also, avoid calling FLUXCOM “observations” as FLUXCOM GPP is not observed.
3. When comparing different types of temperature-moisture and GPP extremes, I found it is challenging to put the results into perspectives. It would be helpful to 1) provide statistics. How much is the overestimation? Can you give a number, e.g., in percentage? 2) summarize key findings. Which type of temperature-moisture and GPP extremes contribute to the most uncertainties?
4. Overall scientific writing needs improvement. This is listed as a general comment because many times the structure and phrasings are misleading. For instance, the

last paragraph of introduction reads more like a summary. The end of introduction usually briefly summarizes the aim and the approach, and clearly define the research gap or hypothesis. What the study found is usually not part of it, rather belonging to the results and conclusion.
Several phrasings are informal and ambiguous, for instance, “bracket” (L103), “attractive” (L107), “propagate” (L119; this is often called “upscale”); Other cases are listed in the specific comments. Please revise the manuscript throughout.

Specific comments:

1. Terminology: be consistent - statistical relationship (L10), coupling strength (L11), statistical links (L71), statistical relation (L79), statistical coupling (L116).
2. L18-19: This sentence seems to be disconnected from the main context of this study. Please consider revising.
3. L44: Do you mean “co-occurrence of temperature-moisture and vegetation extremes”?
4. L51-60: the paragraph on LSM simplification/cascading effects is not well integrated with the study objective; tighten or better link to the research aim.
5. L75: What does “mitigating effects” imply? Do you mean increased GPP after temperature-moisture extremes?
6. L136: The recent released FLUXCOM-X product provides GPP at daily resolution (Nelson et al., 2024). This can resolve the mismatch of time scales between different datasets. I think the authors at least need to discuss to which extent the choice of reference data might influence the results.
7. L144: Unclear phrasing. Please clarify what “characteristic response time” and “demanding and beneficial conditions” mean.
8. L146: How did you remove the seasonal trend? Please clarify the calculation processes.
9. L149: “Preliminary tests…”: If you make the claim without directing readers to tests or supporting figures, ensure the methods or supplement documents enough detail so reviewers can judge its validity.
10. Figure 1: This could be a nice figure to understand the method. Currently, the fonts are very small and affects the readability. Panel (c) is not clear to illustrate the calculation of the co-occurrence rate. Is the overlap measured by the product of the two series above? What do the numbers next to the arrows mean? How is the time lag considered when calculating the product?
11. L165: “reduces effective integration timescale” is not clear. Similar phrase “moisture integrating time” appears again at L365. I suppose you mean precipitation show higher variability thus influences hydrological conditions with shorter memory, compared to soil moisture that is accumulated over longer-term. Please clarify.
12. L206: The comparison metric between model and observation-based data is mentioned only briefly. As this is a major part of the results, it is important to prepare the authors which kind of metrics are used to evaluate model performances. For instance, how is the correlation in Figure 3 and Figure S2 calculated? Why do you choose correlation instead other metrics, e.g., RMSE to evaluate model performances?
13. L216: keep wording consistent—use “temperature-moisture” unless specifying precipitation explicitly.

14. Figure 2: Are non-vegetated areas masked out from the analysis? For example, it does make sense that there are co-occurrences between compound and GPP extremes in the Sahara Desert in MPI-ESM.
15. L255: specify where the increase is shown.
16. L272-273 & L277-278: direct readers to specific figures that support these statements.
17. L299: Unclear statements. Please explain how seasonality can influence the response times. What does “stimulate supplies” mean?
18. L380: What does “dipoles” mean?
19. L466: be specific with weaknesses. Data quality? Limited observations?
20. Sections 5.1 and 5.2 both discuss similarities and biases; clarify why they are separate or merge.
21. L470-474 and L508-514: these read as limitations-consider merging them into one coherent limitations paragraph.
22. Figure 6: The calculation of soil moisture rate should be included in the method section.

Technical comments:

1. L244: Do you mean “lagged time” instead of “lead time”?
2. L246-247: The two sentences are repetitive.
3. L251: “Spatially resolved” -> “gridded”
4. L423-424: Unclear sentence

References:

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Dušek, J., Zhang, W., … Jung, M. (2024). X-BASE: The first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X. Biogeosciences, 21(22), 5079–5115. https://doi.org/10.5194/bg-21-5079-2024