Reviewer 1:

We thank the reviewer for the detailed comments regarding the use of SPEI and MODIS GPP.

1. SPEI is a widely used to characterize drought conditions particularly because it is standardized and as such always relates to the local climatic context (see Slette et al. 2019). However, this also poses some issues as outlined in Zang et al., (2020). The main argument of Zang et al., is that negative SPEI (e.g. -1 to -2; often characterized as moderate drought) does not necessarily always coincide with "actual" water shortage as determined by the amount of evapotranspiration subtracted from the precipitation. This seems to be the case especially in wet regions of the world such as the tropics or the boreal forest. The consequence is that SPEI may indicate drought conditions when in reality ample water is available (as determined for example by the Maximum Climatic Water Deficit commonly used in tropical regions).

We thank the reviewer for highlighting this important point regarding SPEI and its interpretation in wet regions. In our study, we use the term “drought” to refer to periods of below-average water availability relative to the local climate. This reflects a relative deviation from normal conditions. To address the reviewer’s point, we examined whether soil moisture is low during the multi-year drought periods identified in this study (L413-418 and Supplements S2). We found that soil moisture was consistently lower during drought periods than during normal period, even in energy-limited regions. This confirms that negative SPEI values also reflect reductions in water availability.

2. The second issue is the use of MODIS GPP as a stand-in for "observed GPP". Satellite-derived GPP is as much a model as it is based on satellite observations. In the case of MODIS GPP, GPP is calculated by an algorithm using remotely sensed FPAR, a land cover classification, a parameter for the conversion efficiency of PAR, and some climate inputs (Tmean, VPD). Consequently, it is not really accurate to consider satellite-derived GPP from MODIS as "observed". This obviously does not invalidate the comparison with simulated GPP from LPJmL-5, however, it does warrant some further discussion and potentially the use of an alternative product (e.g. SIF) to strengthen the message of this study.

For the use of MODIS GPP, we agree that it is a model-derived product based on satellite observations and not a direct observation of GPP. We clarified this in the manuscript and expand the discussion on the limitations of satellite-derived GPP (section 2.3.1, L425-444 and Supplements S3).

Regarding land cover, LPJmL-5 was run with prescribed land cover, including crop distributions, which ensures that the land-cover types used by LPJmL-5 match those of MODIS. Further details on the differences between prescribed and dynamic land cover in LPJmL-5 are provided in Supplement Section S4.

Finally, we have already explored the consistency across MODIS vegetation indices in our comparison of GPP and EVI in Supplement Section S2. Also, Ruijsch et al. (2025) analysed the correlation between GPP, EVI, and LAI during multi-year droughts. Both analyses show broadly similar patterns across the different vegetation indices. To further strengthen the study, we extended this comparison by including SIF (Supplements S3) and found a strong correspondense between EVI, GPP and SIF.

Reviewer 2:

This study presents a comprehensive set of analyses; however, its scientific motivation and added value remain unclear. The work mainly consists of descriptive comparisons between LPJmL-5–simulated GPP and MODIS-derived GPP in terms of mean states, variability, drought anomalies, and responses to SPEI across multiple timescales. Such comparisons, by themselves, do not address a clearly defined scientific question, nor do they provide new insight into vegetation responses to drought beyond documenting discrepancies between two fundamentally different GPP products.

Importantly, MODIS GPP is itself a model-based, remotely sensed product rather than a direct observation. As a result, differences between LPJmL-5 and MODIS GPP cannot be straightforwardly interpreted as model deficiencies or ecological mechanisms. As currently presented, the results primarily show that the two products differ substantially, but it remains unclear how these differences advance understanding of ecosystem drought responses or meaningfully address the stated limitation of short satellite records.

Although the authors argue that DGVMs can overcome the temporal limitations of satellite observations, the study does not demonstrate that LPJmL-5 provides additional or improved insight into vegetation responses to long-term or multi-year droughts. No independent evaluation at longer timescales is provided, nor is it shown that the model improves the representation of drought processes or stress regulation. Consequently, the claimed advantage of using LPJmL-5 to assess long-term ecosystem vulnerability is not sufficiently supported.

Overall, the study lacks a clear hypothesis, mechanistic advancement, or demonstrated improvement in modeling drought impacts. Without evidence that LPJmL-5 adds robust, independent information beyond existing GPP products, the analysis remains largely diagnostic and does not justify the broader conclusions drawn.

We thank the reviewer for review and raising these important concerns. We acknowledge that the initial study did not fully clarify its scientific motivation and intended contribution. The primary aim of this work is not to derive new ecological mechanisms of drought response, but to evaluate whether a widely used DGVM (LPJmL-5) reproduces satellite based drought response patterns well enough to support its application in multi-year drought studies beyond the satellite era. We revised the manuscript to include this motivation better (L43-45), and avoid interpretations that imply ecological mechanisms or ecosystem traits. Below we respond to the specific comments in detail.

1- I question whether the comparison between LPJmL-5 GPP and MODIS GPP is meaningful in the first place, given that the two products differ substantially in spatial resolution, temporal aggregation, and conceptual formulation. LPJmL-5 operates at a coarse grid scale and simulates GPP through process-based parameterizations, whereas MODIS GPP is a remote sensing–based product with its own empirical assumptions and limitations. Under these circumstances, differences between the two cannot be unambiguously attributed to vegetation response characteristics.

Moreover, the choice of MODIS GPP as the sole benchmark appears arbitrary. Using an alternative satellite-based product with different resolution and algorithms (e.g., VPM GPP at 500 m and 8-day resolution) would likely lead to different comparison outcomes. This raises a fundamental question that the study does not address: what exactly do these inter-product differences represent, and on what basis can one conclude that one product provides a more correct or ecologically meaningful representation of drought response than another?

We thank the reviewer for raising this important concern. We agree that LPJmL-5 and MODIS GPP differ in spatial resolution, temporal aggregation, and conceptual formulation, and that these differences come from how the model and observations are made, not from the ecosystem itself.

The purpose of this study is not to identify a "correct" product or interpret differences as ecosystem properties, but to investigate model behavior relative to a satellite reference. MODIS GPP was chosen because it provides global coverage, temporal overlap with other satellite vegetation indices, and a common variable (GPP) that is directly comparable to DGVM output.

To address concerns regarding the model-based nature of MODIS GPP, we explicitly evaluated its consistency with other satellite products (EVI and LAI), finding strong correlations and similar spatial drought anomaly patterns (Supplement S2; Ruijsch et al., 2025). This supports the use of MODIS GPP as a benchmark, while we will still fully acknowledging its limitations.

We included this rationale in a new Methods subsection (sections 2.2.1 and 2.3.1) and supplements section (S3), explaining the choice of MODIS GPP, how it is calculated, and how it is used to evaluate LPJmL-5 simulated GPP.

2-Although the use of SPEI-12 is common in studies of multi-year droughts (MYDs), its role as an optimal or sufficient drought indicator has not been systematically validated. Importantly, different ecosystems exhibit markedly different sensitivities to drought at various temporal scales. For instance, croplands and grassland ecosystems typically respond more strongly to short-term moisture anomalies (e.g., SPEI-1 to SPEI-6), whereas the strong smoothing inherent in SPEI-12 may obscure critical short-duration drought signals, potentially leading to biased interpretations of ecological responses. In addition, longer aggregation periods can delay the detection of drought onset and termination, which may affect the characterization of drought event dynamics.

Another concern is a drought event identified from an index such as SPEI only captures the potential drought stress on vegetation relative to the long-term normal period and may not necessarily reflect the actual water deficit of an event.

We thank the reviewer for the comment on ecosystem sensitivities to different SPEI aggregation periods. To capture this, we used the extreme-based method that evaluates SPEI timescales from 1–24 months and identifies the dominant timescale with the strongest impact on vegetation (Figure 5). This allows us to assess both short-term and long-term vegetation responses and ensures that faster-responding ecosystems are not overlooked.

For the definition of multi-year droughts (MYDs), which is the primary focus of this study, we specifically use SPEI-12. SPEI-12 captures long-term moisture deficits, removes seasonal effects, and ensures that brief wet periods do not split up longer drought events. Together, these analyses allow us to evaluate how well LPJmL-5 reproduces both rapid vegetation responses and long-term multi-year drought dynamics.

3- The comparison presented in Figures 4 and 5 does not provide a valid basis for any inference about ecosystem drought resistance. The observed differences between LPJmL-5 and MODIS GPP primarily reflect systematic discrepancies in GPP construction, temporal sensitivity, and model structure, rather than differences in vegetation response or regulation in an ecological sense. In particular, LPJmL-5 does not include any explicit parameterization or calibration targeting ecosystem resistance, buffering capacity, or drought regulation, making it conceptually unjustified to interpret shorter response timescales as reduced drought resistance.

Figure 4 explicitly demonstrates that the model produces sharper and more immediate GPP declines during dry periods, which is a direct consequence of the model's tight coupling between soil moisture and photosynthesis. This behavior reflects model design choices rather than ecosystem properties, and therefore cannot be interpreted as evidence of altered resistance or response dynamics. Without explicitly separating response speed, response magnitude or demonstrating that these metrics are meaningfully constrained by observations, the presented analysis does not support any ecological conclusions regarding drought resistance. As currently formulated, this comparison is fundamentally methodological and cannot substantiate claims about ecosystem response mechanisms.

We thank the reviewer for this important clarification. We agree that LPJmL-5 does not explicitly parameterize ecosystem drought resistance, buffering capacity, or stress regulation, and that shorter response timescales cannot be interpreted as evidence of altered ecosystem resistance.

Our intent in Figures 4 and 5 was to compare model behavior to MODIS, not to infer ecosystem traits. In the revised manuscript, we have removed or rephrased language such as "underestimate drought resistance" and instead describe these patterns as model limitations in reproducing MODIS drought dynamics.

To address the reviewer's concerns, we have revised the manuscript as follows:

1. **Introduction:** We clarified this study's goal and motivation, explicitly stating that the focus is on assessing model performance relative to satellite-derived GPP.
2. **Methods:**
   - **LPJmL-5 drought mechanisms (2.2.1):** We included a section about processes in LPJmL-5 that govern drought responses, such as water stress functions, phenology, and mortality, to clarify what drives model behavior during dry periods.
   - **MODIS GPP (section 2.3.1):** We added a new subsection explaining why MODIS GPP was chosen and how it serves as a benchmark for LPJmL-5. We will also add a comparison to other observational products (EVI and LAI) and acknowledge its limitations.
3. **Results:**
   - We described differences between LPJmL-5 and satellite-derived GPP without referring to ecosystem properties or ecological traits.
4. **Discussion:**
   - We included a paragraph noting that MODIS GPP itself is a model-based product, and that inter-product differences partly reflect differences in construction and assumptions (L425 – 444).