The responses to the reviewer comments are shown below in blue. Text in italic and between quotation marks indicates sentences from the manuscript, and underlined sentences indicate parts that are newly added in this revision.

Reviewer 1

General comments:

This manuscript titled "Higher tree diversity reduces critical slowing down in the Amazon forest" presents an interesting study investigating the response of the globally most important forest ecosystem to climate change. To this end, the authors simulate changes in tree species diversity and composition under different scenarios varying in spatial (alpha, beta, gamma) scale. Their findings highlight the critical link between biodiversity and ecosystem stability and thus should have important implications for Amazon forest biodiversity conservation.

Generally, I believe that this is an excellent study based on proper statistical analysis given the complexity of the stability and diversity data being investigated across multiple spatial scales. However, I must admit that when reading through the manuscript at times I was a bit confused about the jargon with regard to the hypothesis investigated in this study. For instance, in L55 you refer to the insurance hypothesis by specifically stating "diversity may increase stability by reducing critical slowing down or the likelihood of tipping points." Later on, in L71 you expect "a positive correlation between diversity and stability, or a negative correlation between diversity and critical slowing down". Hence, while it is clear what relationship (negative/positive) you are referring to it might still be worth to improve the clarity of a few sentences by following the logic of arguments. Having said this, I would further recommend rephrasing the title to something more intelligible like "Higher tree diversity reduces the likelihood of Amazon tipping points".

We thank the reviewer for this feedback. We agree that the jargon throughout the text might be confusing for readers, so we will follow the recommendation of changing the title and most of the 'critical slowing down' terminology to 'tipping likelihood' throughout the text. However, since critical slowing down is an important term in Amazon forest research, we will include critical slowing down as an extra keyword and still mention it once in the introduction.

In the abstract:

Page 1, Lines 20-27: "<u>Slower recovery to short-term disturbances, hereafter named reduced stability, is considered an early warning indicator of such tipping points</u>. However, the role of tree species diversity in regulating this vulnerability remains poorly understood, especially across spatial scales. <u>To examine how tree species diversity impacts tipping point likelihoods</u> across multiple spatial scales, we used modelled tree species diversity data at the alpha (local), beta (asynchrony across local communities), and gamma (regional) scales. <u>We quantified tipping likelihood on the same scales using temporal autocorrelation trends</u> in monthly satellite-derived vegetation productivity time series over 2001-2019. <u>Our findings reveal higher tipping likelihoods</u> at the alpha level (25 km^2) compared to the gamma level (209,903 km^2), indicating that Amazonian tipping points are more likely to occur locally than regionally or basin-wide."

In the introduction:

Page 2, Lines 38-43: "*TAC measures the correlation between successive time points in the time series, with an increase indicating a greater similarity between the current and previous ecosystem states over time (Dakos et al., 2012). According to mathematical theory, when dynamical systems such as the Amazon lose stability and approach a tipping point, they are expected to show a slowing recovery to short-term disturbances (translated as an increase in TAC) – a phenomenon known as critical slowing down (Boulton et al., 2022; Scheffer et al., 2009). For easier interpretation, we will rather use the term 'tipping likelihood' hereafter.*"

Page 2, Lines 55-56: *"Consequently, diversity may increase stability by reducing the likelihood of tipping points."*

Page 3, Lines 70-72: *"Based on the insurance hypothesis, we expect higher diversity to be associated with higher stability, or equivalently, a lower likelihood of tipping, at the alpha scale (Liu et al., 2022)."*

Page 3, Lines 75-76: *"In this research, we use changes in TAC of a satellite-derived proxy of canopy productivity as an indicator of tipping likelihood, which we combine with modelled tree species richness across multiple spatial scales for the Amazon forest."*

In the materials and methods:

Page 4, Lines 98-99: *"In this study, we define high Amazon forest stability as a low likelihood of reaching a tipping point, using time series of satellite data across the Amazon."*

In the results:

Page 8, Lines 230-232: *"When including only the satellite pixels overlapping with the regularly distributed alpha diversity plots, we observed negative stability values, indicating increased tipping likelihood, for 40% of the included EVI pixels at the alpha scale and for 20% of the regions at the gamma scale (Fig. 2 and Fig. 3)."*

In the discussion:

Page 11, Lines 274-277: *"We integrated changes in satellite-derived TAC of canopy productivity with modelled tree species richness to understand the impact of tree species diversity on the likelihood of Amazonian tipping points. By incorporating stability and diversity data across multiple spatial scales, we showed that forest areas characterized by higher alpha tree species diversity exhibit lower tipping likelihoods."*

Page 11, Lines 284-285: *"Our analysis reveals that approximately one-third of the analyzed pixels within the Amazon forest exhibited increased tipping likelihoods over the 20-year study period, which decreases to one-fifth of the Amazon at the regional scale."*

While this might represent a minor issue, I would emphasize the need of further clarifying some details with regard to the methodology applied in this study. Specifically, I am not fully convinced about the following points (listed by line numbers): (i) application of temporal autocorrelation (TAC-1), seasonality index (L117-121) and cumulative water deficit (L166-172); (ii) derivation of alpha stability as the negative value of the TAC(+0.1) trend in the EVI time series (135-145); (iii) representation of causal relationships (alpha, beta, gamma) in the conceptual structural equation model (L195 and L259-263).

Our responses on the TAC-1, derivation of alpha stability, and representation of causal relationships were added to the specific comments below. In the following response, we only focused on the seasonality index and cumulative water deficit by including extra information about both in the text:

About the seasonality index:

Page 6, Lines 163-166: "*Lastly, we included seasonality to describe the intra-annual precipitation variability in the Amazon forest, derived from monthly TerraClimate precipitation data from 1980 to 2019 (Abatzoglou et al., 2018) and calculated using the Seasonality Index (SI; Walsh & Lawler, 1981). Low SI values reflect evenly distributed rainfall throughout the year, while high values indicate that all rainfall is concentrated in a single month*."

About the cumulative water deficit:

Page 6, Lines 175-187: "*CWD was calculated for all pixels within the Amazon using monthly TerraClimate precipitation (P) time series from 1980 to 2019 and a fixed evapotranspiration (E) of 100 mm per month, using the following rule (Aragão et al., 2007):*

If CWDn-1 – E + Pn < 0;

Then CWDn = CWDn-1 – E + Pn;

Else CWDn = 0

Where n represents each month in the time series. Months with negative CWD values represent periods when monthly precipitation is insufficient to offset both the evaporation occurring that month and any precipitation shortfall carried over from the previous month, indicating a water deficit. The CWD dataset was then used to calculate the mean and standard deviation of CWD per month. Standardized anomalies were determined per 0.05° pixel by subtracting the monthly mean from the pixel value and dividing the result by the monthly standard deviation. Pixels with CWD anomalies below -1.96 had significantly larger water deficits than average (with p < 0.05). Extreme drought periods were defined as starting with at least two months of significantly dry CWD anomalies and ending when the anomaly became positive."

Specific comments:

First, if I understand correctly you have extracted monthly EVI images from 2001 to 2019 from the daily Moderate Resolution Imaging Spectrometer (MODIS) and used the decomposed remainder (of the seasonal and trend decomposition) to calculate the lag-1 autocorrelation with a moving window length of five years. While I have applied this method myself and do appreciated its value for investigating temporal correlations among biophysical parameters with their environmental drivers, I do wonder why you did not investigate (or at least mention) any other lags (1/3/6/12 months) in the time-series. Usually, one would expect recurring seasonal (e.g., 6 months or annual) pattern between vegetation parameters and climatic drivers (e.g. Hofhansl et al., 2012), which could be investigated by further computing partial autocorrelations and cross-correlation functions (i.e., ACF and CCF, sensu Venables & Ripley 2002).

This is indeed correct. While we do understand the comment, the use of a one-month-lag (lag-1) rather than lag-3 (3 months) or lag-6 (6 months) temporal autocorrelation as a proxy for tipping likelihood is based on the mathematical equations underlying the theory of critical slowing down (Scheffer et al., 2009). Mathematical theory assumes that a dynamical system such as the Amazon forest, characterized by a state variable *x* (e.g., canopy vitality), has the capacity to recover from perturbations (e.g., drought events) back to its reference state before the perturbation (Lenton et al., 2022; Scheffer et al., 2009). The potential function *U(x)* describes the equilibrium states or stable fixed points around which the variable gravitates, as well as the boundaries beyond which recovery will not occur. According to the critical slowing down theory, when *x* approaches a bifurcation point, the potential function loses depth, which causes the dominant eigenvalue characterizing the rate of change around the equilibrium to converge to zero. This shallowing of the basin results in larger excursions away from the equilibrium and elongates the recovery process, meaning that the system recovers increasingly slower from small perturbations. Consequently, this phenomenon leads to an increase in the lag-1 temporal autocorrelation of *x* (see Scheffer et al. (2009) for the mathematical derivations). We decided to follow this theoretical lag-1 value to be in line with other critical slowing down research (Boulton et al., 2022; Forzieri et al., 2022; Runge et al., 2025).

Hofhansl et al. (2014) performed lagged correlation analysis to investigate maximum correlation coefficients between seasonally detrended climate and productivity variables at different time lags to indicate the most significant drivers of tropical lowland net primary production, and they found significant lags between 0 and 36 months. This is different from our goal, as we were interested in capturing early warning signals of tipping points, rather than in specific temporal correlations.

We added this in the methods:

Page 4, Lines 117-121: "*The decomposed EVI remainder was then used to calculate the lag-1 autocorrelation (one-month lag) with a moving window length of five years. The use of longer time lags could highlight recurring seasonal or annual patterns between vegetation and climatic drivers (Hofhansl et al., 2014). However, we used lag-1 temporal autocorrelation rather than longer time lags to be in line with previous critical slowing down research (Boulton et al., 2022; Forzieri et al., 2022; Runge et al., 2025).*"

Second, I can see why you want to avoid a negative number in the denominator of equation (L135) and you mention the additional analysis being based on a subsection of the dataset (and results presented in the appendix A) but I am still not convinced that the resulting values are representative of shifts in ecosystem function (L393). Hence, while I do concur with the assumption that the pattern detected in the analysis is related to signals in remotely sensed EVI patterns and associated shifts in phenology (L390-393) I would expect that this might be related to short-term fluctuations (in mainly canopy characteristics) rather than long-term dynamics (representing shifts in functional species composition) due to tree mortality.

We agree that we would need longer time series to distinguish between medium-term canopy or long-term forest dynamics. This is already partly included in the discussion: *"Finally, while the 20-year timescale might be too short to capture long-term stability trends in the Amazon, the inclusion of three widespread drought years within this period enabled us to quantify how diversity influences medium-term stability responses."*

We changed the term 'ecosystem functioning' to 'canopy characteristics' to make our claim less strong:

Page 15, Lines 407-408: *"However, because EVI does not capture forest structure, changes in EVI TAC do not necessarily reflect increased tree mortality but rather shifts in canopy characteristics."*

Third, you indicate that the assumptions for linear regression, such as the linearity of the data, normality of the residuals, homogeneity of residuals variance, and independence of the error terms, have been checked, and that none of them were violated (L204-207) but I am still not fully convinced it is feasible to throw in seemingly related variables, such as alpha, beta, gamma diversity into one and the same structural equation model (even when using piecewise SEM and accounting for random effects, such as "region ID" (L212). Given that tree species richness in this study is being based on model predictions rather than plot measurements wouldn't be alpha, beta, and gamma diversity therefore be auto correlated and thus violate the assumptions of independence anyway?

We agree that α-, β-, and γ-diversity are not independent variables; their scale-dependence and interdependence are precisely the motivation for our cross-scale analysis. The SEM is not constructed under an assumption that these components are unrelated; rather, it explicitly represents hypothesised pathways linking predictors to biodiversity across spatial grains. In piecewise SEM, distributional assumptions (linearity, homoscedasticity, residual normality, and independence) apply to the residuals within each component model, not to independence among all variables included in the network. To address potential non-independence among observations, we accounted for spatial structure by (i) including region-level random effects in all relevant component models and (ii) incorporating spatial correlation structures, which removed residual spatial autocorrelation.

Furthermore, using model-predicted values as a response variable is not inherently problematic, as prediction per se does not create non-independence among observations. Statistical independence concerns the correlation structure of residuals among observational units (e.g., sites), not whether a variable was measured directly or estimated from a prior model. Predicted values may still be independent across sites unless the prediction model itself imposes residual correlation (e.g., via ignoring spatial autocorrelation or spatial grouping). In the diversity database used here, such dependence structures were taken into account (Keil & Chase, 2019) and are therefore not ignored in our analysis. The primary implication of using predicted responses is the presence of additional estimation uncertainty, which is a measurement-error issue rather than a violation of the independence assumption.

Having said that, several foregoing studies have highlighted that the Amazon is more likely to experience local transitions to a degraded state in response to multiple and linked environmental and anthropogenic factors (Franco et al., 2025) rather than to result in a systematic break-down (Rammig et al., 2010).

We added these papers as references in the discussion:

Page 11, Lines 295-297: *"Furthermore, this pattern also suggests that the Amazon is more likely to experience local transitions to a degraded state than to reach a regional or system-wide critical threshold (Franco et al., 2025; Rammig et al., 2010)."*

Nevertheless, I do believe that this analysis represents a valuable contribution to the scientific debate about the fate of the Amazon forest in response to both natural and anthropogenic disturbance, but as indicated above, I would recommend adding clarifying details on the applied methodology (see the specific points raised above) and to further stress the underlying aspects regarding the presentation of the study findings so that the reader could follow the line of arguments presented in the main text in a clear and concise manner.

References (to be considered):

Franco, M.A., Rizzo, L.V., Teixeira, M.J. *et al.* How climate change and deforestation interact in the transformation of the Amazon rainforest. *Nat Commun* **16**, 7944 (2025). https://doi.org/10.1038/s41467-025-63156-0

Hofhansl, F., J. Kobler, J. Ofner, S. Drage, E.-M. Pölz, and W. Wanek (2014), Sensitivity of tropical forest aboveground productivity to climate anomalies in SW Costa Rica, *Global Biogeochem. Cycles*, 28, 1437–1454, doi:10.1002/2014GB004934

Rammig, A., Jupp, T., Thonicke, K., Tietjen, B., Heinke, J., Ostberg, S., Lucht, W., Cramer, W. and Cox, P. (2010), Estimating the risk of Amazonian forest dieback. New Phytologist, 187: 694-706. https://doi.org/10.1111/j.1469-8137.2010.03318.x

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*, series Statistics and Computing. Springer, New York, NY. doi:10.1007/978-0-387-21706-2

We would like to thank the reviewer once more, for the valuable comments and suggestions.

Reviewer 2

This manuscript provides a valuable multi-scale analysis of the relationship between tree species diversity and ecosystem stability in the Amazon using satellite-based indicators of critical slowing down. The study is generally well structured and clearly written, with high-quality figures and a commendable effort in data integration and analysis. However, in its current form, several issues in the Methods and Results sections require further clarification. My primary concerns are outlined below.

We thank the reviewer for this feedback. We have responded to the comments below.

(1) Limitations of EVI-based indicators

As EVI primarily reflects canopy greenness and phenological dynamics, increases in temporal autocorrelation (TAC) should be interpreted as changes in ecosystem functioning rather than as direct evidence of structural degradation or imminent forest collapse. It is unclear why the analysis relies on a single vegetation index. The authors are encouraged to justify the exclusive use of EVI and to discuss whether incorporating additional indicators (e.g., NDVI, SIF) could improve the robustness of the results.

It is indeed true that that EVI reflects changes in greenness so EVI TAC only gives information about top of canopy dynamics rather than whole tree dynamics. This is mentioned in the text: "*EVI measures the canopy greenness, rather than the woody growth response. Therefore, using*

EVI as a proxy for ecosystem stability assumes that reductions in stem growth coincide with canopy browning (Janssen et al., 2021)." There are multiple other indicators that could be included in this study to make it more robust, but none of them are ideal either: NDVI saturates more quickly in tropical forests than EVI (Huete et al., 2002), there are considerable inconsistencies among diverse SIF datasets and contradictory findings in applying them (Sun et al., 2023), and VOD has coarser spatial resolution (0.25°) than EVI (in this manuscript 0.05°) (Moesinger et al., 2020). However, it is true that using a multitude of satellite products could improve the robustness of the results. We chose to focus on one indicator (EVI) to simplify the manuscript and analyses, and rather go deeper into the different early warning signals (both TAC and standard deviation) and different buffer zone sizes to integrate the diversity and stability data (3x3, 5x5, and 7x7 MODIS pixels).

We added this in the discussion:

Page 15, Lines 407-410: "*However, because EVI does not capture forest structure, changes in EVI TAC do not necessarily reflect increased tree mortality but rather shifts in canopy characteristics. Incorporating additional satellite products that capture other forest characteristics, such as solar-induced chlorophyll fluorescence (SIF) or vegetation optical depth (VOD), would improve the robustness of the results.*"

(2) Uncertainty in climate datasets

The manuscript lacks sufficient description of the climate datasets used in the analysis. Please specify which climate products were adopted and provide justification for their selection. Given that different climate datasets may yield differing causal relationships, an assessment or discussion of climate data uncertainty and its potential influence on the causal analysis would strengthen the study.

We agree with this comment and added more detail in the methods section:

Page 6, Lines 172-177: "*We used precipitation time series from TerraClimate, which merges the high-spatial resolution climatological normals from the WorldClim dataset with time-varying data with a coarser spatial resolution from the Climatic Research Unit time series data version 4.0 (CRU TS4.0) and the Japanese 55-year Reanalysis (JRA-55) (Abatzoglou et al., 2018). CWD was calculated for all pixels within the Amazon using monthly TerraClimate precipitation (P) time series from 1980 to 2019 and a fixed evapotranspiration (E) of 100 mm per month.*"

We also added a discussion on the uncertainty:

Page 13, Lines 354-359: "*Furthermore, as the choice of precipitation dataset influences the variability in droughts stress across the Amazon (Papastefanou et al., 2022), using another dataset may yield different causal relationships. TerraClimate showed slightly better validation performance than CRU when compared to weather station data, although its long-term precipitation trends are inherited from CRU (Abatzoglou et al., 2018). A comparison of multiple precipitation datasets over the Amazon found CRU-based drought estimates to be broadly consistent with other products (Papastefanou et al., 2022). Nevertheless, reliance on a single dataset can still lead to regional over- or underestimation of drought occurrences.*"

(3) Weak explanatory power in Figure 3

Although the relationships shown in Figure 3 are statistically significant at the alpha and beta scales, their explanatory power is very low (marginal R^2 = 0.01–0.02). This raises the question of whether such weak correlations are sufficient to support the conclusions regarding the stabilizing role of tree diversity.

We agree that the explanatory power of the relationships is very weak. However, this is already mentioned in the abstract: *"We also observe significant but weak positive linear relationships between tree species diversity and stability at both alpha and beta scales. This emphasizes both the importance of biodiversity conservation at multiple spatial scales and the complexity of understanding the stability of the Amazon forest."*

And in multiple parts within the discussion:

"However, the low R^2 values of the positive relationships also show the complexity of understanding the drivers of ecosystem stability in the Amazon forest."

"The significant diversity-stability relationships on the alpha and beta scales highlight the role of tree species diversity and its spatial heterogeneity in protecting the Amazon forest from reaching a tipping point. However, the low explanatory power of these relationships could potentially be due to the high functional redundancy in Amazonian forests. At small scales, additional species can still make a difference in terms of niche complementarity, whereas, at larger spatial scales, they may lead to functional redundancy (Poorter et al., 2015)."

We added this extra sentence to suggest a potential next step:

Page 12, Lines 303-304: *"Addressing the low R^2 values and the unexplained variability of the TAC trends will require mechanistic modelling to disentangle the underlying causal processes."*