Response to review comments 1

We thank reviewer 1 for their detailed comments on the manuscript "Observation-inferred resilience loss of the Amazon rain forest possibly due to internal climate variability". In this document we will discuss and respond in detail to the submitted comments.

General comments

- *To my understanding, the authors reproduce the methodology from Boulton et al. (2022). Nevertheless, the observational trends plotted in this paper and in Boulton et al. (2022) look markedly different. The trend in Boulton et al. (2022) is nearly monotonic from 2003 to 2016 (tau = 0.913) whereas the AR(1) trend in the manuscript under review shows a decrease from ~ 2012 and 2016. Together with the inclusion of the period prior to 2003, this seems to contribute to a smaller (tau = 0.62) compared to Boulton et al. (2022). I ask the authors to clarify why the observational curve in the current manuscript is different from the one in Boulton et al. (2022). In addition, it would strengthen the manuscript strongly if the authors would additionally present results for shorter periods (e.g., 2003-2014) given that the results in Boulton et al. (2022) are for the period 2003 - 2016.*

  The observational curve presented in our manuscript matches the curve presented by Boulton et al. (2022) in the supplementary material (Figure S5). In this case, the record from 1995 – 2017 shows even a greater AR(1) increase (tau = 0.646) compared to the period
  2003 – 2017 (tau = 0.562). For that reason, and to include the longest observational record as possible in order to capture decadal variability, we decided to include the full available record. The difference of the reported tau by Boulton et al. (2022) of tau = 0.646 and our
  tau = 0.62 can be explained by the grid cell selection procedure (see Boulton et al. (2022) methods). Further, the difference between our reported observational curve and the main text curve by Boulton et al. (2022) originates from the seasonal-trend-decomposition of the signal. As reported, we use the input parameters trend = 19, season = 13 and period = 12 months with the Python statsmodels STL function, gaining the same result as Boulton et al. (2022) in the supplementary material. Robustness tests for the choice of parameters were done by the same authors.

- *The authors put their work in the context of attribution studies for weather and climate extremes. Using models in attribution studies is only justified if the models adequately simulate the studied phenomenon. I understand that the observation period of the AR(1) time series is too short to evaluate the models and assess the significance of recent trends. Nevertheless, it does not become clear if the models are adequately simulating Amazon rain forest dynamics. A skillful simulation of leaf area index across the Amazon rain forest would substantially increase the confidence in the simulated AR(1) trends. Therefore, I ask the authors to (1) provide a more extensive description of the used models (how complex are they compared to state-of-the-art vegetation models? do all of them simulate vegetation dynamically?), and (2) either refer to previous studies in which the ability of the models to simulate the Amazon rain forest is analyzed or include some analysis, e.g., on the similarity of the simulated mean state and spatial patterns with observations. In addition, the authors do only present anomaly time series from the time mean AR(1) coefficients. Are the temporal mean AR(1) coefficients comparable to the time mean AR(1) coefficients of the observations (in the spatial mean and the spatial patterns)?*

(1) Information on dynamic plant functional type coverage is added to Table 1. Details on land surface model components are added in line 85 – 94.

(2) Reference of estimation of LAI representation in models is added. Changes made to line 95 – 97.

We only present AR(1) anomalies because only the change of AR(1) is of interest analysing forest resilience.

- *The absence of a more extended description of the models also precludes the interpretation of differences between models. In particular, it would be interesting to explore potential reasons for why one model shows a significant difference between the control simulations and the historical period whereas the other models do not show such a change.*
As our study is based on a multi-model approach, it is beyond its scope to understand in detail the behaviour of single models.

- *tau measures the monotonicity of AR(1) trends but it does not quantify the rate of increases/decreases. It would be very interesting to also compare the simulated and observed rate of the increase to assess how exceptional the observed 2003 - 2016 trend is.*
Our study builds upon the observational study of Boulton et al. (2022) and therefore we use the same statistical method as they did.

It is of course interesting to test other methods. For example we noted the recent publication by Wang et al. (2023) uses the Theil-Sen estimator. This method is used to fit a line to data with an underlying trend, but it contains some outliers that may contaminate a normal fit. The figure below shows an application of this method to the MPI-ESM1.2-LR ensemble and the observations. Indeed we see somewhat larger trends, in particular for the observational record. The larger trend in observations is probably achieved because the method filters out data at the ends of the observational record. Observations now have the largest positive trend, but there is still a segment of the historical experiment with a larger in magnitude negative trend. Therefore we anticipate that running more experiments will eventually yield a run with a larger positive trend. Furthermore, it is questionable whether the ends of the observational data record should be treated as outliers, or whether they are part of a slow natural fluctuation.
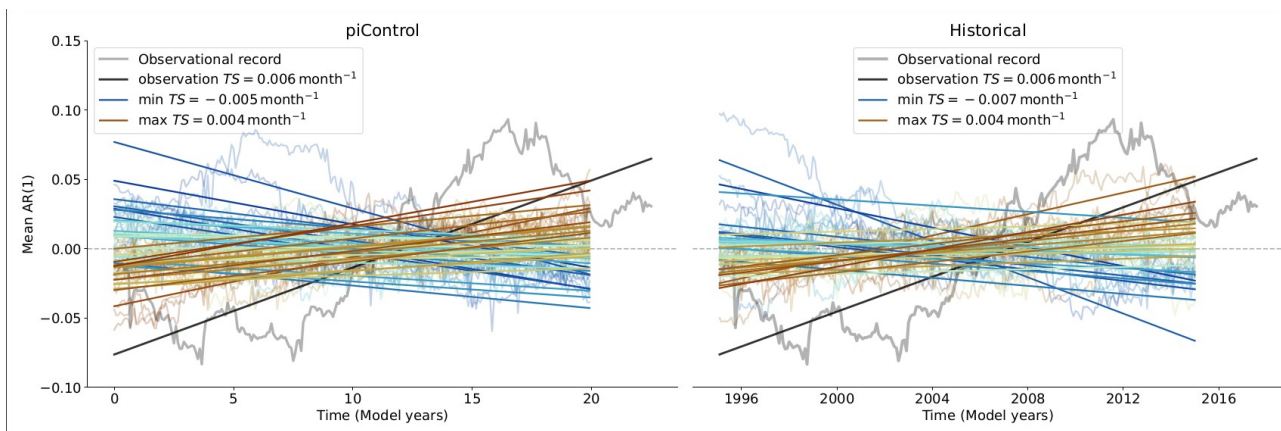

Figure 1: Theil-Sen slopes fitted to the AR(1) trends of observations and MPI-ESM1-2-LR for historical and control runs. TS denotes the slope of the fitted line, color scheme accordingly to Fig. 1 in the manuscript.

Specific comments

- *l. 13: After reading the manuscript, this strong statement does not seem to be supported by the presented evidence. Where in the manuscript is the role of local actors studied?*
  Changes made to line 13.

- *l. 57: Can you give references or examples for increasing AR(1) due to other physical reasons?*
  Reference added to line 57.

- *Sect. 2.1: The description of the used simulations is very short. More information on the complexity of the employed models and similarities/differences between them would be very helpful. In particular, do they all simulate dynamic vegetation (variable PFT coverage frequencies)? If not, are there systematic differences between models with and without dynamic vegetation?*
  See comment above.

- *l. 78: Can you expand on why VOD and LAI are physically closely related?*
  See comment below.

- *l. 78-79: While the manuscript states that LAI and VOD are strongly correlated, this seems to be not the case in the Amazon catchment area (Sect. 4.4 and Fig. 11 in Moesinger et al., 2020). This absence of a strong positive correlation in the Amazon rain forest should be stated explicitly and the use of LAI in models and VOD in observations should be justified in light of this weak correlation.*
  Moesinger et al. (2020) demonstrates that VOD reflects plausible seasonal and short-term changes in vegetation and vegetation dynamics on top of LAI and other related optical biophysical vegetation products from optical remote sensing. A positive correlation is shown globally and possible reasons for negative correlation in the Amazon are given (e.g. drought events that increase LAI and decrease VOD).
  This issue is more likely due to a problem with the observational products, in a model the LAI and VOD are to the best of knowledge tightly related in parameterisations, and also physically we see no obvious reason why the opposite should be the case. However, VOD is not provided as output, hence we use LAI.

- *l. 104-105: This sentence might be misleading. To my understanding, the employed Kolmogorov-Smirnov test does not explicitly test for similarity of the spatial structure (i.e., spatial correlations) but only compares the distribution of the tau values across all considered grid boxes. For example, if all the tau values would be randomly reshuffled (thereby loosing the spatial information), it would not change the result of the KS test.*
  Changes made to line 104.

- *l. 119: One 'together' too much*
  Changes made to line 119.

- *l. 137-143: What are the implications of the fact that the KS-test is only passed for so few ensemble members?*
  As explained later in the manuscript, it implies that the observations are on the edge of what models are able to reproduce. In case of higher simulated mean

tau, we would see much more members that pass the KS-test, indicating that a resilience loss is more probable according to models. Changes made to lines 142 – 143.

- *l. 149-150: It is not clear to me if such a strong differentiation between IPSL and models, that pass the test for 1 or 2 members, is justified. Can you quantify this difference statistically?*
Changes made to line 150.

- *Table 2: Please explain in the caption how pcml is computed*
Changes made to caption of Tab. 2.

- *Fig. 3: It would be insightful to also plot the KS tests for the other models, for example in a supplement. As a major novelty of the paper is the use of multiple models, it does not seem justified to focus on MPI-ESM in several figures and not give equal exposure to all models (or at least the four models with 20 or more ensemble members)*
The KS-test plots for the models with 20 or more ensemble members are going to be provided in the next version of the manuscript.

- *l. 152: Can you explain how the significance is computed here?*
Changes made to line 153.

- *l. 159-167: Why is only MPI-ESM1-2-LR used here and not the other ESMs with plausible ensemble members? Table 2 suggests to me that MIROC-ES2L should be given a larger role in the analysis compared to MPI-ESM since it possesses the most ensemble members that pass the KS-test under historical forcing.*
We chose MPI-ESM for the analysis of SSP5 because we can link the historical runs to the future scenario runs (highlighted members 10 and 28 in Figure 5). This was unfortunately not possible with e.g. MIROC-ES2L that only provides 10 SSP5 runs.

- *l. 165-166: Is the land use change enforced by the SSP scenario? If yes, it should be easy and insightful to quantify its role.*
Figure A7 added. Changes made to line 166.

- *l. 171 Boulton et al. (2022) instead of Boulton et al. 2020?*
Changes made to line 171.

- *l. 174: "are" instead of "is"*
Changes made to line 174.

- *l. 178: "underperform" instead of "under performs"*
Changes made to line 178.

- *l. 178-179: While similar trends to the observations are found in a few ensemble members, the conclusions should emphasize stronger how rarely they occur in the ensembles. For example, the observation's percentiles are at most 0.02 for all piControl ensembles (Table 2) and at most 0.02 for all historical ensembles except for MIROC-ES2L (which is also the only model with a significant change between PI and historical). While it is true that the results suggest that the trend 'could' occur due to internal variability, the likelihood of this occurrence should be the more important quantity, as it is common in other attribution studies.*

We acknowledge the likelihood of the AR(1) trend occurrence e.g. in line 8, 153 – 154, Fig. 4, and new in changed lines 142 – 143.

- *l. 182-183: Can you expand on why the results suggest that "large scale weather events" are responsible for the anomalies?*
We provide possible reasons for large in magnitude AR(1) trends in models and their spatial distribution. Changes made to line 183.

- *l. 189-191: The presented results only analyze trends in AR(1), interpreted as an indicator of rain forest resilience. Where in the study are the reasons for the decline in rain forest extent studied and where is the role of local actors analyzed?*
Changes made to lines 189 – 191.

References

WANG, Huan et al. Anthropogenic disturbance exacerbates resilience loss in the Amazon rainforests. Global Change Biology, v. 30, n. 1, p. e17006, https://doi.org/10.1111/gcb.17006, 2023.