Review of "Observation-inferred resilience loss of the Amazon rain forest possibly due to internal climate variability"

The manuscript by Grodofzig et al. studies trends in lag-1 auto-correlation, AR(1), of leaf area index in the Amazon rain forest as simulated by an ensemble of Earth system models. The study complements recent observational evidence by providing a modeling perspective on the statistical significance of estimates. As increased lag-1 auto-correlation is an indicator for the loss of forest resilience, which can serve as an early warning signal for the system to reach a tipping point, the understanding of natural and forced trends in AR(1) is highly relevant. The paper is concise and well-written but I have a few general suggestions that would in my opinion provide the reader with a better understanding of the results and strengthen the scientific significance of the findings.

General comments

- Tom 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 τ ($\tau = 0.62$) 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 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)?
- 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.
- τ 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.

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?
- l. 57: Can you give references or examples for increasing AR(1) due to other physical reasons?
- 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?
- l. 78: Can you expand on why VOD and LAI are physically closely related?
- 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.
- 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 τ values across all considered grid boxes. For example, if all the τ values would be randomly reshuffled (thereby loosing the spatial information), it would not change the result of the KS test.
- l. 119: One 'together' too much
- l. 137-143: What are the implications of the fact that the KS-test is only passed for so few ensemble members?
- 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?
- Table 2: Please explain in the caption how p_{cml} is computed
- 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).
- l. 152: Can you explain how the significance is computed here?
- 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.
- 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.
- l. 171 Boulton et al. (2022) instead of Boulton et al. 2020?
- l. 174: "are" instead of "is"
- l. 178: "underperform" instead of "under performs"
- 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.

- l. 182-183: Can you expand on why the results suggest that "large scale weather events" are responsible for the anomalies?
- 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?

References

Boulton, C. A., Lenton, T. M., and Boers, N.: Pronounced loss of Amazon rainforest resilience since the early 2000s, Nat. Clim. Chang., 12, 271–278, https://doi.org/10.1038/s41558-022-01287-8, 2022.

Moesinger, L., Dorigo, W., De Jeu, R., Van Der Schalie, R., Scanlon, T., Teubner, I., and Forkel, M.: The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA), Earth Syst. Sci. Data, 12, 177–196, https://doi.org/10.5194/essd-12-177-2020, 2020.