Dear authors,

I want to congratulate for a very well written, thought-through, and structured manuscript covering a wide range of topics and providing a very extensive analysis. Applying a continental model to assess water storage variations can be a key contribution for a better understanding of future threads to endangered ecosystems such as the Amazon River basin.

Before acceptance is possible, however, I have a few critical comments which are provided hereafter separated as general and specific comments on the submitted work.

General comments

- The work contains a lot of modelling work with subsequent extensive validation and comparison of model results with a range of observations. In this sense, it presents a "classical" technical modelling study with potential to be scaled up to other basins. Since droughts can have pronounced societal impact, I would like to read more about how the presented work can help to not only "anticipate future hydrological conditions" but also how this knowledge could be used as leverage to tackle present and future challenges of water management in the Amazon. Both in the Introduction and in the Conclusion adding the societal dimension and possible added value of your work would be of great benefit to the manuscript.
- Throughout the manuscript, there are quite some adverbs such as "relatively", "extremely" and so on. Please ensure that you use those words only when absolutely appropriate.
- I found quite some instances where you describe the figures in the actual text (eg. The first seven lines of chapter 4.1). To improve readability and shorten the text, I advise to limit the descriptions to the figure captions and only refer to the figures in the text.
- You compare interannual and interdecadal results. While I do see the added value of analyzing interannual variation, I am wondering why you decided to compare decades as well? In my opinion this time interval is just not long enough to assess long-term changes. Why not assessing long-term trends over the entire climatology instead?

Specific comments

- Page 2/ line 17: Is it possible to associate the climate and human-induced changes more exactly with the results? What are the driving factors, is it rather the change in climate or the increased human activities that alter the (hydrologic) system in the Amazon?
- Page 2 / lines 23-26: This sentence is in my opinion a repetition of the information presented in the paragraph before (eg. With respect to streamflow reduction). It would be worthwhile considering removing any repetitive statements here and throughout the remaining manuscript.
- Page 4/ line 6: From the manuscript it does not become clear whether the "opposite trend" between model output and GRACE is only a thing in the Amazon River basin or whether it is issue also on global scale. Please provide this information so readers can get a better idea of the severity of this problem. Also, why is there no explanation available? Maybe provide a brief sentence (based on Scanlon 2018?) since it's bit unsatisfying to read at the moment.
- Page 5/ line 5 "The LHF model":
    - Even though the paper was already published, adding a flow chart would help the reader to better understand the LHF model and the modelling steps required.
    - What is the temporal resolution of model output? Please add.
    - Was the model calibrated? If so, how and using which data and parameters? If not, why not? Please add this information.
- Page 5/ line 19: What was the reasoning to use the 2 km version of LHF and not the probably faster 5 km version?
- Page 5/ line 26: It is unclear to me why you perform an extensive spin-up of 150 years but then discard the first year of the modelling period too? Were there issues with model stability or initial conditions? Please explain your choices clearly to avoid any misunderstanding.
- Page 6/ line 1 "Atmospheric forcing":
    - Why did you decide to use the WATCH data as model forcing? Why not the more recent ERA-5 data?
    - The accuracy of forcing plays an important role, also in the Amazon. The work by Towner et al. (2019, HESSD; https://doi.org/10.5194/hess-2019-44) compares several forcing data sets (ERA-I, ERA-5, and re-forecasts) with respect to their relative impact of model accuracy for the Amazon basin. I think that this could provide a good starting point for a brief discussion about the role of forcing in modelling studies and to explain your decision to use WATCH data.
- Page 6/ line 10: I understand you use annual input for the land cover? Are you confident that this is acceptable given the model runs at a different temporal resolution I assume? Please add a brief statement why you are convinced your choice made is appropriate.
- Page 6/ line 19: how did you decide to use LAI value 5 as threshold for transitioning into forest? Is this based on expert knowledge, scientific literature or just an arbitrary decision? I can image that in vegetation-rich areas (ie. With generally high LAI values) such as the Amazon this threshold can have a marked influence on the total area eventually specified as forest.
- Page 7/ line 11: did you use monthly averages? Please add this information.
- Page 7/ line 14: Do all stations cover the same period or not? And what about missing values in the time series – are there any and if so how did you treat them?
- Page 9/ line 29: Why is it that GRACE shows little agreement in small basins? Please add a brief explanation. Also, I am wondering whether using GRACE is then the right approach for

those basins – would it not be an alternative to skip the GRACE comparison for small basins where it's known a priori that agreement will be small?

- Page 11/ line 7: I have my doubts whether the manuscript presents a "state-of-the-art" framework. After all, the model used is already available for quite a while and the main novelty of the presented work is the extensive analysis with GRACE and streamflow data for the Amazon and its sub-basins for a long period of time (which is an important contribution to current process understanding!). Another example would be the WATCH forcing which could be updated with more recent data sets. If you're convince the framework is nevertheless state-of-the-art, I would like to see an elaboration in the model description section why that is the case.
- Page 14/ line 3: What was your motivation to employ HydroSHEDS basins for this specific analysis? Do these "sub-catchments" match the geographical extent of the sub-basins you are referring to in the remainder of the manuscript? If not, this choice somehow complicates the analysis by introducing another geographical unit; in this case, I would advise to stick to the sub-basin definition used for the other analyses.
- Page 15/ line 3: You are mentioning "important insights" but I don't see any further elaboration what those insights could be. Please append this information!
- Page 15/ line 4 "Intensification of the Amazonian Dry Season": This chapter could profit from discussing findings from other literature to put your results into perspective. Please add where applicable!