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 threats 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. Response: Thank you for your positive evaluation of the
manuscript and the constructive comments that helped tremendously in improving the quality of the paper. We provide a detailed point-by-point response to all comments in the following. We note that we have addressed all your comments and have made necessary changes in the revised manuscript. We have also provided detailed responses to each comment to avoid any confusion along with references directed towards the revised manuscript. Please note that following the other reviewer’s comment, we have made slight changes in the paper outline of the revised manuscript.

General comments:

GC1) 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.

Response: Thank you for the thoughtful comment. We agree that the outlook based on societal aspects will greatly benefit the manuscript, hence giving a more complete picture of the Amazonian droughts. We have provided additional discussion regarding droughts with respect to the societal dimension both in the Introduction and Conclusion. However, we have attempted to keep the discussion on the societal dimension short as this issue is not completely within the overall scope of the present study. To summarize, we discuss the broad impact of droughts on the livelihood of the local population through the disruption in fish yield, navigation, drinking water supply, etc. We also present an overview of the impact on the financial condition of the riverine population caused by incessant droughts. Further, we have provided a discussion on the application of the results presented in this study towards prediction and mitigation of
future drought conditions in the Conclusion section of the revised manuscript.

GC2) 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.

Response: Thank you for the suggestion. We have revised the entire manuscript to avoid the use of adverbs. However, in some locations the adverbs are retained to maintain the statement’s inference and to keep them concise, avoiding excessive increase in words because the manuscript is already a little long.

GC3) 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.

Response: Thank you for the suggestion. We have limited the figure descriptions to figure captions (for example, Figure 1 and 10). General changes to the text are also made to ensure a smooth flow between text and paragraphs.

GC4) 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?

Response: We agree that the time interval of this study is not long enough for a complete interdecadal analysis. However, given that we simulated the Amazon hydrology for 36 years, it was worthwhile to take a first look at the hydrological changes occurring at a decadal timescale. Realizing the shorter time frame for interdecadal trend analysis, we kept the discussion limited to decadal differences rather than specific trends (section 3.4). Our simulation period also encompasses several ENSO episodes, so it is
worthwhile utilizing the simulations to examine how the hydrology is changing decade by decade. Moreover, studies have shown that precipitation exhibit opposite trends over the northern and southern Amazon on interdecadal scales (Lee et al., 2011; Marengo, 2004) making it important to examine how these changes propagate into TWS variations over the region. The long-term trends over the entire climatology have been discussed in the third paragraph of Section 3.4 and in Figure 6.

Specific comments:

SC1) 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?

Response: Hydrological changes in Amazon have been suggested to be a result of a combined impact from climate change and human activities (Cook et al., 2012; Cook and Vizy, 2008; Lee et al., 2011; Malhi et al., 2008; Shukla et al., 1990). Due to the overall scope of the study (focus on droughts) and to keep the manuscript concise, we aimed to keep the Introduction section highly focused on the objective. All the above-mentioned studies, although focus on either climate change or human activities or both, none of them explicitly quantify the causes of the hydrological change, rather it is difficult to do so due to the complex interaction between climate and human activities over a region. Hence, we believe that the overall impact of climate change and human induced changes on the hydrologic system of the Amazon region can be better inferred qualitatively rather than quantitatively. However, we have added the drought related results from these individual studies in the Introduction section.

SC2) 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.
Response: Thank you for pointing this out. We have removed the sentence and have made some editorial changes on our own for better clarity and flow, ensuring the new paragraphs are logically connected.

SC3) 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.

Response: Thank you for this important comment. Scanlon et al. (2018) has shown that most of the global river basins show a different trend behavior in TWS compared to GRACE. In case of the Amazon River basin, models show an opposite trend behavior in TWS. Scanlon et al., (2018) attributes the discrepancy to model shortcomings, such as poor representation of water stores and some hydrological processes, and uncertainty in forcing datasets. Although, their study quantifies the impact of the above-mentioned causes to some extent, the results vary greatly among models and forcing datasets; hence, giving no clear explanation/quantification of the causes of model-GRACE discrepancy. We have added this information about the “opposite trend” between GRACE and model output in the same line. Further, we have also included a reference directing to section 3.3, where we have a detailed discussion regarding the model-GRACE discrepancy and results from Scanlon et al., (2018).

SC4) Page 5/ line 5 “The LHF model”: â€¢ 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. â€¢ What is the temporal resolution of model output? Please add. â€¢ Was the model calibrated? If so, how and using which data and parameters? If not, why not? Please add this information.

Response: Thank you for the suggestion. We have added more information about
the LHF model and its simulation setup in section 2.8. We have prepared a diagram showing the LHF model setup employed in the study; however, we have included this flow diagram in the revised supplementary information because after adding more information about the LHF model and its setup in Section 2.1 and 2.2, we found this to be less important in the main manuscript. LHF model time steps is of 4 minutes but we save the model output at daily intervals. Regarding the third question, we have not calibrated the model using any observed datasets as the land surface, hydrologic and groundwater processes in the model framework are physically based. As such, the entire temporal extent of the model is utilized as validation period rather than dividing it into calibration and validation periods. Although, there are some parameters (e.g., manning’s co-efficient) in the physical equations which can be tuned to have a better correlation with observed datasets, we have not performed any tuning as the model represents these processes rather well with the predestined values. We have added this information in Section 2.8 of the revised manuscript.

SC5) Page 5/ line 19: What was the reasoning to use the 2 km version of LHF and not the probably faster 5 km version?

Response: Given the large areas of Amazonian floodplains and its large-scale interaction with the sub-surface water store, it is crucial to simulate these interactions with higher accuracy, to get a relatively complete picture of the hydrology in the region. A finer grid allows the model to retain the spatial details essential for an accurate simulation of floodplain dynamics and the groundwater processes (e.g., convergence along valleys and lateral flow) which are mainly controlled by local topography (Miguez-Macho and Fan, 2012). Moreover, the computational resources which we had at our disposal allowed us to conduct a finer resolution simulation with enough to spare for further analysis. In the past several years, the hydrologic modeling community has put concerted efforts to increase model resolution and move toward hyper-resolution global modeling (grid sizes of 1km or smaller) (Fan et al., 2019; Wood et al., 2011). To contribute to these community efforts LHF has been further refined to 1km grids (Fan
et al., 2017). Thus, we feel that going back to a coarse resolution will be unjust as long as computational facilities permit such high-resolution simulations. We have been using LHF at 5km grids for the continental US where computational cost constraints the grid resolution (Shin et al., 2018). In case of our 5km version, the model domain consisted of Continental United States (CONUS), which is ∼3.5 times larger than the domain used in this study, hence forcing us to employ a lower resolution (∼5km) over CONUS.

SC6) 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.

Response: Thank you for the thoughtful comment. The spin-up was performed for 150 years with the repeated use of forcing from 1979. All the simulations for 1979 were counted towards spin-up and the main simulation was started from 1980-2015. We did not find any issues with model stability or initial conditions, however, as the manuscript intends to mainly focus on the interdecadal changes, the simulation for year 1979 would anyway have to be discarded, as a single year cannot represent a complete decade. Further, discarding the first year of simulation is also a general methodology adopted in other hydrologic modelling studies. We have added this information in Section 2.8 of the revised manuscript to avoid any misinterpretation.

SC7) Page 6/ line 1 “Atmospheric forcing”: Ê Why did you decide to use the WATCH data as model forcing? Why not the more recent ERA-5 data? Ê 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.
Response: Thank you for pointing this out. We also agree that the accuracy of forcing plays an important role in simulating Amazonian hydrology. We decided to use WATCH Forcing Data methodology applied to ERA-Interim reanalysis data (WFDEI) because it has been suggested to well represent the observations and is more suitable for hydrological modelling of Brazilian water resources in the past literature (Monteiro et al., 2016) compared to other datasets considered in the studies. Several bias and atmospheric corrections were also applied in deriving the WFDEI dataset (Weedon et al., 2014), hence making it a widely used forcing dataset for regional and global studies. On the contrary, ERA5 dataset is fairly new and limited studies exist in the literature showing its suitability for hydrological modelling over Amazon. We could have used ERA5 in our analysis and checked its effectiveness over Amazon, however this approach would be vastly different than the overall objective of this study. To further explore the model-GRACE discrepancy, one of the main objectives of this study, we are conducting multiple LHF simulation sets with different forcing datasets; this analysis and the model accuracy with ERA5 dataset will be addressed in our forthcoming paper. Moreover, to better explain our decision of using the WFDEI forcing in our analysis, we have provided additional discussion in Section 2.2 which includes your suggested article of Towner et al., (2019).

SC8) 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.

Response: We use the European Space Agency Climate Change Initiative’s Land Cover project (ESA-CCI; http://maps.elie.ucl.ac.be/CCI/) land cover maps to represent the land use land cover (LULC) dynamics in our model framework. Land cover dataset from ESA-CCI is the only dataset currently available which satisfies the LHF model requirements in terms of spatial extent, temporal extent and spatial resolution. Although, one can generate land cover maps at higher temporal resolution using Land-
sat imagery, the resulting process will greatly differ from the overall scope of this study. Moreover, we believe that the impacts of LULC change on the regional hydrology dominate on a longer time scale and its seasonal dynamics are well captured through LAI changes. Both of these dynamics are incorporated in the framework with an annual input of land cover and monthly LAI input. Further, usage of annual LULC input is also the general practice in hydrologic impact studies. To facilitate a better understanding of the model setup, we have added these additional details regarding the use of annual land cover input in Section 2.3 of the revised manuscript.

SC9) 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.

Response: Thank you for pointing out this issue. The threshold of LAI=5 for the forest transition is based on scientific literature. The threshold value was mainly based on the study conducted by Asner et al., (2003) which presents a synthesis of global LAI values for different land cover types. Asner et al., (2003) showed that the evergreen broadleaf and needleleaf forests, which are the major forest types in Amazon, have average LAI values greater than 5 (5.8 and 6.7, respectively). Other studies also classify the evergreen forests in the same LAI range (Myneni et al., 2007; Xu et al., 2018); for example, Myneni et al., (2007) studied the seasonal swings in LAI values and showed that the mean annual LAI is ∼5 over the entire Amazonian rainforest (Figure 1A of the citation). Hence, we used the threshold of LAI=5 to get a first-hand approximation of the past forest cover in Amazon. To elaborate a bit more on the method we used to back extrapolate land cover, we have added more information in Section 2.3.

SC10) Page 7/ line 11: did you use monthly averages? Please add this information.

Response: We used the observed monthly averaged streamflow data from Agên-
cia Nacional de Águas (ANA) in Brazil for model validation. This additional information about the observed streamflow data is added in Section 2.4.1 of the revised manuscript.

SC11) 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?

Response: The streamflow stations were selected based on their data length and data gaps. Data periods varied among streamflow stations, with some spanning over the entire modelling period (i.e. 36 years) and others for more than 30 years. As the observed streamflow values are solely used for model validation, the months with missing data were skipped from the statistical analysis. More information about the observed streamflow data and their usage in the analysis is added in Section 2.4.1 of the revised manuscript.

SC12) 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?

Response: GRACE error mentioned in this manuscript refers to bias and leakage correction errors (Landerer and Swenson, 2012; Longuevergne et al., 2010). This type of GRACE error is mainly dependent on the basin size, which increases with a decrease in basin size (Longuevergne et al., 2010). We use the GRACE data to ensure that the simulated TWS variations are within the plausible limits for the Amazon basin and its sub-basins. Also, one of the key discussions made in the manuscript is the discrepancy between model and GRACE, hence skipping GRACE comparison for smaller basins would result into an incomplete analysis. Moreover, to tackle the leakage errors associated with GRACE spherical harmonics products, we also employ the GRACE mascon products in our analysis, which are known to better capture the TWS signal by reducing the error from leakage (Save et al., 2016; Scanlon et al., 2016). Even though,
some of the Amazonian sub-basins are fairly small compared to their neighbors, the smallest river basin (i.e. Japura, \(\sim 256,000 \text{ km}^2\)) under consideration still has a basin area higher than the GRACE footprint (\(\sim 200,000 \text{ km}^2\)) (Longuevergne et al., 2010). Hence, skipping the model-GRACE comparison for these basins will not be wise, as it will not only create a void in the validation but also introduce inconsistency (or even doubt to the reader on why they are excluded) with other analyses done in this study. We have thus decided to include those smaller basins in the analyses.

SC13) 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.

Response: Thank you for pointing this out. We have removed the term “state-of-the-art” from the specified line. We agree that the LHF model has been available for quite a while now, but we cannot stress less on the role of the simulation setup for making the overall framework “state-of-the-art”. The version of LHF model we use in this study was last updated in 2012. Novelty of the LHF model lies in that incorporates sophisticated land surface process such as a prognostic groundwater store, dynamic water table and its interaction with surface water stores, lateral groundwater exchange, sea-level influence on coastal drainage and river-floodplain routing by resolving the full momentum equation of open channel flow. In this study, we further incorporated hydrological interactions with human activities, such as LULC change, through dynamic land cover input at annual scale (see response for SC8) and leaf area index input at a monthly scale, hence creating a comprehensive framework for assessing long-term hydrological changes. Although, many other comparable hydrological models also rep-
resent some of the above-mentioned hydrological processes in their modelling framework, most of them lack the explicit groundwater scheme (e.g., NOAH, VIC) (Ek et al., 2003; Liang et al., 1994) and the flood dynamics in their framework are fairly simplified, essentially making the surface-subsurface interaction mostly linear reservoir based (e.g., WGHM, PCR-GLOBWB) (Alcamo et al., 2003; van Beek and Bierkens, 2009). Similar level of detail as LHF, is also found in the CaMa-Flood (Yamazaki et al., 2011) model which incorporates a river-floodplain routing system; however, the scheme has not integrated the Saint-Venant equation with other land-surface hydrological processes till date, hence demoting it in the hierarchy. Therefore, the evident contribution from the long-term simulation setup incorporating the dynamic human role in impacting the hydrological cycle in addition to the original novelty of the LHF model are the key factors behind highlighting our framework as “state-of-the-art”. To demonstrate this to our readers, we have added more information in respective sections of the revised manuscript. For details regarding the usage of WATCH forcing data in this study, please refer the response for SC7.

SC14) 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.

Response: Yes, the “sub-catchments” exactly match the geographical extent of the sub-basins referred in the rest of the manuscript. The LHF model also utilize the topography information from the HydroSHEDS dataset. We decided to use the “sub-catchments”, which are essentially the sub-basins of the Amazonian sub-basins, to study the hydrological drought propagation at higher spatial resolution. This allows us to take advantage of the wide number of the streamflow estimates obtained from the model, to conduct an in-depth analysis of the complex interaction between LULC changes and streamflow. Analyzing the hydrological drought trends at a sub-basin level
will constrain the analysis at merely 8 points, further making it difficult to understand while severely underestimating the role of LULC in governing streamflow generated from the region. Therefore, even though we agree with the reviewer that we could use sub-basins for the sake of consistency but going down to smaller catchments would provide finer details of the role of LULC in regional hydrology.

SC15) 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!

Response: Thank you for the suggestion. We have moved the statement at a location which more appropriately justifies the term “important insights”. The statement now is situated before the start of the discussion in Section 3.6 trailed by the detailed description of individual insight it infers.

SC16) 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!

Response: Thank you for the suggestion. We have added a more exhaustive discussion regarding the “Intensification of Amazonian Dry Season” by combining the results from previous literature and this study in Section 3.7.

References


Save, H., Bettadpur, S. and Tapley, B. D.: High-resolution CSR GRACE RL05 mascons,


Xu, B., Park, T., Yan, K., Chen, C., Zeng, Y., Song, W., Yin, G., Li, J., Liu, Q.,
