To Reviewer #1,

Comments:

In this manuscript the authors aim to isolate the effects of vegetation greening on water yield in a large basin that serves as donor basin for a major water diversion project. To do so the authors designed a modelled-based scenario analysis. The overall finding of their analyses suggests that greening has the potential to considerably reduce basin water yield and thus supply for the water diversion project. While the experimental set-up is systematic and, in principle, logical and the manuscript is well written, I nevertheless have a number of serious concerns that need to be addressed and resolved in detail before this manuscript could be considered for publication.

Response: We appreciate the reviewer's insightful and constructive comments, which helped us a lot in improving this manuscript. Please find our specific responses below (in blue).

Comments:

(1) The experiment is designed, the results are interpreted and, as a consequence the manuscript is framed from a purely engineering perspective with focus on water yield available for the diversion project. As a consequence, the non-explicit message that is delivered between the lines here is the following: to secure water supply for the diversion project greening needs to be reduced. Or in other, more explicit words: stop afforestation - chop down the forest! I am not sure that this can and should be the message to be conveyed here for the simple reason that this analysis is not comprehensive enough to draw this conclusion.

Response:

This is a valuable and important point! Indeed, as suggested by the reviewer, we did not provide a balanced message about the positive ecosystem goods and services that forests provide, but we mainly focused on the water supply. It was not our intent to have the readers take "stop afforestation-chop down the forest" as the message between the lines from this work. Rather, our message is that despite the myriad ecosystem goods and services forests provide, there can be tradeoffs to these services. We demonstrated that rapid increase in forest cover (i.e., greening up) can significantly reduce freshwater supply, which is itself an essential societal good.

In the revised version of the manuscript, we highlighted positive ecosystem services forests provide, including significantly reducing sediment in the streamflow, improving water quality, and carbon sequestration to mitigate global warming. At the same time, it is important for policy makers to understand that there can be unintended consequences from afforestation and that programs focused on reforestation/afforestation should be planned accordingly, especially in the context of a warming climate. These tradeoffs in reforestation/afforestation programs have been seen around the world, especially in water shortage regions. For example, a recent report in the Kathmandu Post (https://kathmandupost.com/editorial/2019/12/23/an-unwise-decision-to-plant-conifers-is-parching-the-land) shows that the rapid plantation of pine forests in Nepal jeopardized people's livelihoods due to excessive use of water by the trees. The water yield in most part of the study watershed is sensitive to vegetation change where precipitation is less than 900 mm/yr.

What we intended to stress in this study is the balance between environmental improvement and water resource availability for water diversion projects. Our study was from the perspective of water yield alone. Therefore, what we demonstrated here is a risk of water supply reduction due to afforestation. We advocate for a comprehensive watershed management strategy to deal with water use by forests under a warming future.

Specifically, we made following revisions in the Discussion and Conclusions sections to deliver a more balanced message, highlighting both the goods and services provided by forests and the potential tradeoffs and unintended consequences of afforestation:

a) In "4.1 Reduced water yield and drought amplification from greening", we discussed the greening effects more conservatively. Specifically, we stressed that "However, as the latest round afforestation efforts are about to conclude, implementation of afforestation projects will likely soon slowdown in China, and combined with potential limitations from future water stress, the greening trend will not likely continue at their same levels in the UHRB." And the discussion about the future leaf area increases was removed.

b) The "4.2 Reduced capacity of water supply to SNWDP from greening" was revised as "Trade-offs among ecosystem goods and services induced by greening" and its content was reframed. Specifically, to emphasize the "trade-off" instead of water availability alone, we substantially discussed water quality improvement, as well as other ecosystem goods and services due to greening in the first paragraph of 4.2. Then we shortened the discussion on how greening reduced water availability and moved the discussion about water supply capacity to the water diversion project to section 4.3.

c) In "4.3 Implications for water diversion projects", we provided more discussion on watershed and forest management. Specifically, we moved the discussion about water supply capacity to the water diversion project from 4.2 to combine with the initial first paragraph of 4.3 to discuss the water supply concerns of water diversion projects. Then we offered more specific recommendations for alleviating effects of greening on water supply. For example, we suggest that using natural regeneration with local tree species rather than artificial plantations to control erosion and conduct ecological restoration. Forest management such as stand thinning to reduce water use and fire risk and increase resilience of the ecosystem should be also considered as part of the integrated watershed management.

d) We added "4.4 Limitations" to discuss the limitations of this study in terms of experiments and external impact factors. We first discussed the possible influences of the interaction between vegetation and climate due to limitations of the experimental design. We then substantially discussed the uncertainty from possible moisture recycling. The enhancement of local ET had the potential to increase P. However, the P in the UHRB is primarily controlled by the Monsoon, which comes from the Pacific Ocean in southeast. The significant and widespread greening was also observed in the southeast China and evaporated a larger amount of moisture which might be brought to the UHRB and potentially induce a P increase. If this happens, such precipitation should already been capture by the precipitation data we used. In the last part of 4.4, we discussed the possible effects of increasing atmospheric CO_2 on water cycle.

Accordingly, we revised the conclusion to stress the "trade-off" and suggest that the "improved watershed management (e.g., forest management and reducing water use) is needed to maximum the ecosystem service benefits in watersheds serving as water sources of large water diversion projects."

Comments:

Little explicit consideration is given to potentially relevant feedback effects of greening the water cycle. This includes the potential for increased local precipitation recycling (P), reduced vapour pressure deficits (VPD), temperatures (T) reduced through increased latent heat flux, which in turn affect water partitioning and thus water yield. These points need at least to be discussed in substantial detail and the conclusion needs to explicitly state these limitations.

Response:

We modeled water yield by considering precipitation and ET, which is a function of NDVI, T, VPD, and radiation. The direct effects of those variables on ET should have been captured by the model, though as suggested by the reviewer, the complex feedbacks and interactions (e.g., increased ET/latent heat flux leading to increased P and reduced surface temperatures) may not have been. Therefore, what we did in this study is solely to disentangle the direct effects of climate and land cover (mainly vegetation greening).

The feedback effects mentioned by the reviewer are indeed not explicitly accounted in our model. We expect some of the feedback effect (e.g., T, P, and VPD) should be captured by the meteorological data we used. Investigating these indirect feedback effects, while very important, is beyond the scope of this study and likely not achievable with the models used here. We realized that the omission of these feedback mechanisms may cause additional uncertainty in modeling results of water supply reduction.

We added the discussion in “4.4 Limitation” in the updated manuscript: “Vegetation greening may in turn affect climate itself, such as through evaporative cooling and moistening of the atmosphere. For example, vegetation greening in the local and upwind area may potentially increase P downwind via increasing atmospheric vapor. This water vapor cycling likely offset some negative vegetation effects of increasing ET on WY. We could not, however, explicitly account for the potential feedbacks between vegetation (NDVI and land cover) and climate (e.g., T, P, and VPD). The climate data used as model inputs would implicitly include a certain level of the feedback effects from vegetation greening, but those effects could not be explicitly disentangled in our analysis. Thus, our results represent an attempt to estimate the direct, first-order net effects of climate and vegetation greening on WY.”

Comments:

(2) Linked to (1) is the design of scenarios S2 and S3, which may indeed contain a fallacy. By fixing NDVI and land cover to the values of 2001 (S1) and 2018 (S2), respectively, the authors aim to isolate the “real” and potential effects of greening. This is in principle ok. BUT: it is not clear from the description of the experiment how the related feedbacks in P, T, VPD and even radiation (changes in albedo!) were accounted for. As far as I understood, the assumption for the 2 scenarios was that *only* NDVI and land cover is fixed to the values of the 2 individual years. If this is so, the authors overlook that the observed past P, T, VPD and radiation data already account for changes in NDVI and land cover. What are the effects of that? Do the results then really allow to isolate effects of greening? If I am mistaken here, then I would nevertheless ask the authors to makes this much clearer in the description of their experiment.

Response: Indeed, S2 and S3 are hypothetical scenarios, assuming everything else is the same except the land cover and NDVI. As discussed in the reply above, we did not explicitly account for the feedback effects of T and P as these values are not dynamically coupled in the model, but are driving factors provided externally from existing climate data. The feedback effects on T, P, and VPD should be implicitly included in those input data though they cannot be explicitly decoupled in our model experiments.

To address the reviewer’s comment, we conducted new simulation experiments to separate effects of P, T, VPD and radiation. However, all of P, T, VPD and radiation had insignificant trends during the study period (figure A1). Consequently, their effects on the long-term trends in ET and WY were minor. Therefore, we only showed the combined effects of these climatic variables and did not discuss them separately. In Scenario S2, NDVI and land cover were fixed in 2001, under which changes in WY would be the effects of climate change only. The climate data were derived with in-situ measurements and a climate model, thus they must already contain certain level of feedback effect from a greening world. We cannot remove such effects from the climate data.

Despite these limitations, we still believe that the experiment provided insight on how land cover, greening, and climate each influenced ET and consequently WY. As suggested by the reviewer, we added discussion of the limitations of the modeling in section 4.4:

"……. However, it is impossible to completely decouple the effects of climate change and vegetation greening on ecosystem goods and services due to the tight biophysical interactions and feedbacks between vegetation and climate. The observed vegetation change, as indicated by NDVI, represents the combined effects of changes in climate, human activity (i.e., reforestation, irrigation), and natural vegetation (re)growth. However, because P, T, VPD and radiation had relatively small and insignificant trends during the study period (figure A1), the effects of climate on the vegetation change (NDVI trend and land cover change) were likely relatively minor compared to direct human modification of the landscape and vegetation growth.

Vegetation greening may in turn affect climate itself, such as through evaporative cooling and moistening of the atmosphere. For example, vegetation greening in the local and upwind area may potentially increase P downwind via increasing atmospheric vapor. This water vapor cycling likely offset some negative vegetation effects of increasing ET on WY. We could not, however, explicitly account for the potential feedbacks between vegetation (NDVI and land cover) and climate (e.g., T, P, and VPD). The climate data used as model inputs would implicitly include a certain level of the feedback effects from vegetation greening, but those effects could not be explicitly disentangled in our analysis. Thus, our results represent an attempt to estimate the direct, first-order net effects of climate and vegetation greening on WY."

Comments:

(3) Linked to (2), the description of the models and their actual implementation does not provide sufficient detail to allow the reader to meaningfully assess the results and interpretations. For example, it is completely unclear how NDVI, land cover but also soil data were used. The reader can only assume that these data are somehow used to

estimate the variables PAR and FPAR (Eq.1). Even if this is described elsewhere in detail, it will be necessary to provide this crucial information here as well. In addition, it remains completely opaque, which parameters the two models required and which of those had to be calibrated and which were a priori fixed (e.g. from look-up tables). How were the models calibrated? Was the calibrated model tested on independent data (which should be a rather standard procedure in the year 2021)?

Response:

We added more details and made it clearer in updated version as shown below:

FPAR, R_s, T_s, and W_s were calculated according to Sims et al. (2005), King et al., (2011), Raich et al., (1991), and Landsberg and Waring, (1997), respectively, as:

$$FPAR = 1.24 \times NDVI - 0.168, \tag{3}$$

$$R_s = 1 - K_1 \times R_a / R_{cs}, \tag{4}$$

$$T_s = \frac{(T - T_{min}) \times (T - T_{max})}{(T - T_{min}) \times (T - T_{max}) - (T - T_{opt})^2}, \tag{5}$$

$$W_s = exp\,(-K_2 \times (VPD - VPD_{min})), \tag{6}$$

Where R_a and R_{cs} are respectively actual and clear-sky radiation. The calculation of R_{cs} is based on Raes et al. (2009). T_{min}, T_{max}, and T_{opt} are respectively the minimum, maximum and optimal air temperatures for photosynthetic activity, varying by biome (derived from the land cover data). VPD_{min} is the minimum VPD exceeding which moisture stress starts to take effect, which also varies by biome, and K_1 and K_2 are biome-specific empirical parameters that scale the radiation and VPD effects, respectively. The parameters (ε_{pot}, T_{min}, T_{max}, T_{opt}, VPD_{min}, K_1 and K_2) are calibrated based on global FLUXNET data through a Monte Carlo simulation (Zhang et al., 2016, 2019).

The Sacramento Soil Moisture Accounting Model (SAC-SMA) was used to model WY in the WaSSI model, driven by ET, precipitation (P), and soil parameters. Soil parameters (see following Table) were generated from multi-layer soil particle-size distribution and depth according to Anderson et al. (2006). The algorithm divides the soil layer into lower and upper zones at different depths and estimates the distribution of moisture—including both tension water components (driven by evapotranspiration and diffusion) and free water components (driven by gravitational forces) in each of these two zones. We add the following Table in Appendix A.

Table: Soil parameter for driving model.

Parameters types	Parameters	Abbreviations
Soil Storage Parameters	Upper Layer Tension Water Capacity	UZTWM
	Upper Layer Free Water Capacity	UZFWM
	Lower Layer Tension Water Capacity	LZTWM
	Lower Layer Supplemental Free Water Capacity	LZFSM
	Lower Layer Primary Free Water Capacity	LZFPM
Baseflow Discharge Rate Parameters	Depletion Rate from LZFPM	LZPK
	Depletion Rate from LZFSM	LZSK
	Interflow Depletion Rate from UZFWM	UZK
Upper to Lower Layer Percolation Parameters	Percolation fraction direct to LZFW	PFREE
	Percolation Curve Shape	REXP
	Maximum/Minimum Percolation Rate Ratio	ZPERC

Ref.:

Sims, D. A., Rahman, A. F., Cordova, V. D., Baldocchi, D. D., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson, L., Monson, R. K., Schmid, H. P., Wofsy, S. C., and Xu, L.: Midday values of gross CO2 flux and light use efficiency during satellite overpasses can be used to directly estimate eight-day mean flux, 131, 1–12, https://doi.org/10.1016/j.agrformet.2005.04.006, 2005.

King, D. A., Turner, D. P., and Ritts, W. D.: Parameterization of a diagnostic carbon cycle model for continental scale application, 115, 1653–1664, https://doi.org/10.1016/j.rse.2011.02.024, 2011.

Raich, J. W., Rastetter, E. B., Melillo, J. M., Kicklighter, D. W., Steudler, P. A., Peterson, B. J., Grace, A. L., Moore, B., and Vorosmarty, C. J.: Potential Net Primary Productivity in South America: Application of a Global Model, 1, 399–429, https://doi.org/10.2307/1941899, 1991.

Landsberg, J. J. and Waring, R. H.: A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning, 95, 209–228, https://doi.org/10.1016/S0378-1127(97)00026-1, 1997.

Raes, D., Steduto, P., Hsiao, T. C., and Fereres, E.: Aquacrop-The FAO crop model to simulate yield response to water: II. main algorithms and software description, 101, 438–447, https://doi.org/10.2134/agronj2008.0140s, 2009.

Zhang, Y., Song, C., Sun, G., Band, L. E., McNulty, S., Noormets, A., Zhang, Q. and Zhang, Z.: Development of a coupled carbon and water model for estimating global gross primary productivity and evapotranspiration based on eddy flux and remote sensing data, Agricultural and Forest Meteorology, 223, 116–131, https://doi.org/10.1016/j.agrformet.2016.04.003, 2016.

Zhang, Y., Song, C., Band, L. E. and Sun, G.: No Proportional Increase of Terrestrial Gross Carbon Sequestration From the Greening Earth, Journal of Geophysical Research: Biogeosciences, 124(8), 2540–2553, https://doi.org/10.1029/2018jg004917, 2019.

Anderson, R. M., Koren, V. I., and Reed, S. M.: Using SSURGO data to improve Sacramento Model a priori parameter estimates, in: Journal of Hydrology, 103–116, https://doi.org/10.1016/j.jhydrol.2005.07.020, 2006.

Comments:

(4) Linked to (3), no attempt, whatsoever is made, to estimate the uncertainty around the models, their parameters and the associated results. Not even confidence intervals around the regressions (and the underlying parameters) are given. Quite frankly, I find this very surprising, as this should be part of any meaningful and serious scientific protocol.

Response: This is a good point. We validated the model results at multiple scales and provided R^2 and RMSE as well as the confidence intervals for the model results based on observed streamflow data (section 3.2 and Fig. 4). As suggested by the reviewer, we now also provide confidence intervals on the regression slopes in our figures (Figs. 3-6) and in the main text.

Comments:

Additional specific comments:

2, l.42: this is a completely non-sensical use of the term “drought”. The term “drought” is always refers to a negative anomaly with respect to a specific local reference value, defining a “normal”, typically a median. By convention, conditions below this normal are then defined as “drought”. By extension, there can then be no location with more “frequent” droughts than other locations, as drought is always the deviation from the local/regional normal.

Response: We rephrased the sentence as "the UHRB is quite vulnerable to hydrological drought events". There are many drought indices to identify drought with different criteria (Hayes et al., 2002). In general, hydrological drought refers to a severe lack of water in the hydrological system, manifesting in abnormally low streamflow in rivers and abnormally low water levels in lakes, reservoirs, and groundwater (van Loon, 2015). Although there are, by definition, half of periods with water yield below the average (or median) for all areas, not all incidents with streamflow below the average can be called drought events. Some areas have unstable water yield with greater fluctuations than others. Therefore, these areas generally experience longer and more severe drought events.

Ref.,

Hayes, M. J., Alvord, C. and Lowrey, J.: Drought indices, National Drought Mitigation Center, University of Nebraska., 2002.
van Loon, A. F.: Hydrological drought explained, Wiley Interdisciplinary Reviews: Water, 2(4), 359–392, https://doi.org/10.1002/wat2.1085, 2015.

2, l.50-51: if afforestation was meant to safeguard water availability, then this is in contradiction with l.55-56. Please rephrase.

Response: The statement was rephrased as "safeguard water quality and increase soil water storage".

2, l.56: greater leaf area in itself does of course not increase transpiration. Vegetation metabolic activity increases transpiration. Leaf area is merely an indicator for increased metabolic activity and thus transpiration.

Response: Revised "greater leaf area" as "vegetation greening".

2, l.61: should read as "…are not feasible…"

Response: Changed as suggested. Thanks!

3, l.69: what are "hydrological entities"?

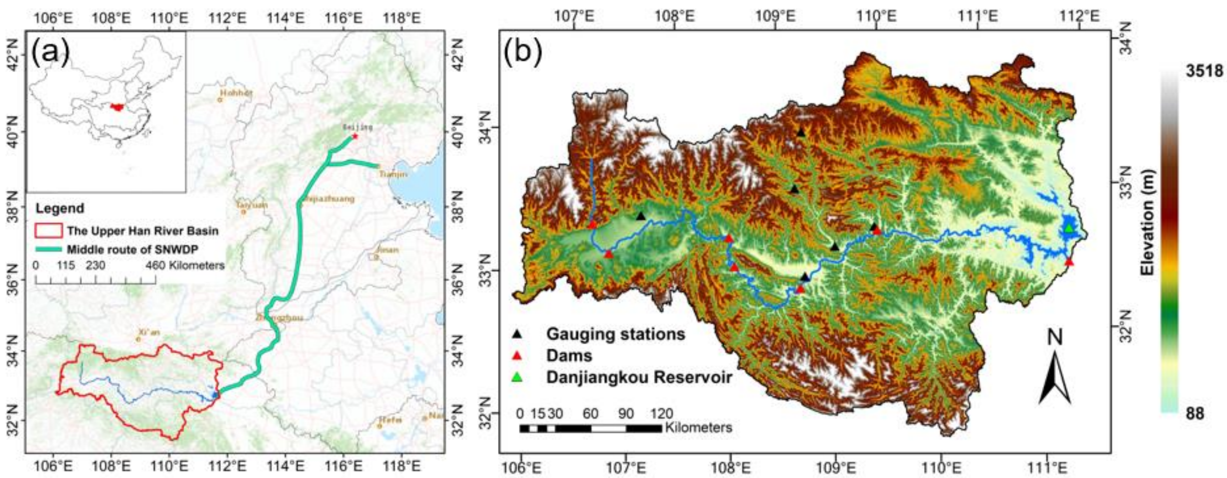
Response: Revised as "hydrological variables".

3, l.76: droughts are low frequency phenomena that require considerable time to develop and to recede. The 18 years of this study are thus likely not enough to make a meaningful statement about changes in drought regimes.

Response: Revised as “hydrological drought risks”.

4, Figure 1: please also show the location of the reservoirs and hydroelectric facilities.

Response: We added a legend to mark the location of the hydroelectric facilities as is shown in following fig. (Fig 1).



6, l.130-131: irrelevant, can be omitted

Response: Removed.

6, l.132-133: not clear what is meant here. Which other models?

Response: We now refer to more specific models in the manuscript: the Penman-Monteith (Penman, 1948; Monteith, 1965), RHESSys (Tague and Band, 2004), ORCHIDEE (Krinner et al., 2005) models etc.

Ref.:

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, https://doi.org/10.1029/2003GB002199, 1 March 2005.

Monteith, J. L.: Evaporation and environment, 1965.

Penman, H. L.: Natural evaporation from open water, bare soil and grass, 193, 120–145, https://doi.org/10.1098/rspa.1948.0037, 1948.

Tague, C. L. and Band, L. E.: RHESSys: Regional Hydro-Ecologic Simulation System—An Object-Oriented Approach to Spatially Distributed Modeling of Carbon, Water, and Nutrient Cycling, 8, https://doi.org/10.1175/1087-3562(2004)8<1:RRHSSO>2.0.CO;2, 2004.

6, l.138: more detail is needed for this choice here. Why 45%? How sensitive is the model to this choice?

Response: The percent of PAR in the total solar radiation indeed varies slightly from place to place. We took the value based on the published value from Running et al., (2000). However, because this parameter is a simple scalar, a change in this parameter would not influence the interannual variation of evapotranspiration or streamflow since the other parameters in the LUE model were optimized to maximize fit between the measured and modeled ET (i.e., any change in the percentage of shortwave radiation that is PAR would be proportionally offset by adjustments in the calibrated light-use efficiency).

Running, S. W., Thornton, P. E., Nemani, R., and Glassy, J. M.: Global Terrestrial Gross and Net Primary Productivity from the Earth Observing System, in: Methods in Ecosystem Science, Springer New York, 44–57, https://doi.org/10.1007/978-1-4612-1224-9_4, 2000.

6, l.139-140: how were PAR and FPAR determined?

Response: PAR was calculated as 45% of the total shortwave radiation; FPAR was calculated based on NDVI (Sims et al. 2005):

$$FPAR = 1.24 \times NDVI - 0.168$$

We have further clarified this in the manuscript.

Ref.

Sims, D. A., Rahman, A. F., Cordova, V. D., Baldocchi, D. D., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson, L., Monson, R. K., Schmid, H. P., Wofsy, S. C., and Xu, L.: Midday values of gross CO2 flux and light use efficiency during satellite overpasses can be used to directly estimate eight-day mean flux, 131, 1–12, https://doi.org/10.1016/j.agrformet.2005.04.006, 2005.

6, l.139-150: Much more detailed is needed on which parameters these models feature and how the parameters were determined, including their prior distributions and the calibration strategy applied.

Response: We have clarified which parameters were tunable and how they were calibrated (see previous responses above).

7, l.163: R2 and NSE have a very similar information content: NSE collapses to R2 in the absence of a bias. Thus, I am not sure of the added value of using R2 as performance metric here.

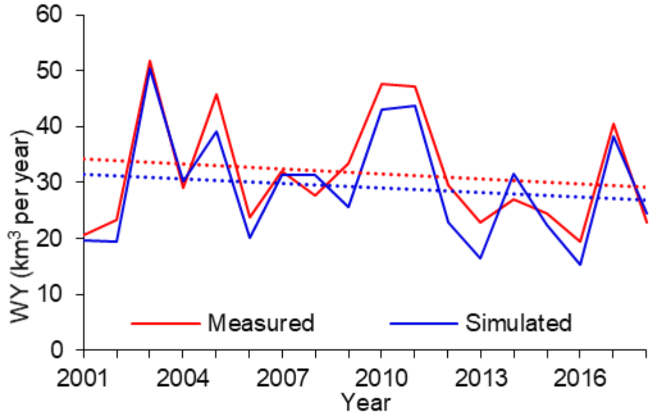
Response: This is a good point, though the initial intention of using both R^2 and NSE was to provide more information to readers. Since bias may indeed be present, we have decided to show both of them, and many readers may not be as familiar with NSE as R^2.

7, l.164: reliable? Many would argue otherwise (e.g. Schaefli and Gupta, 2007, HP). In addition, what does "reliable" actually mean here?

Response: We thank the reviewer for pointing this out. The word "reliable" was removed.

10, Figure 4: given that the model only provides monthly estimates of water yield, the model does not do a particularly good job in reproducing the observed water yield, in particular for the 2012-2014 period. What is the implication of this? What are the uncertainties around that? How does it affect the results and interpretation?

Response: The inflow of the Danjiangkou Reservoir is controlled by dams on the mainstream Han River. Since the flow records of hydropower plants in China is not publicly accessible, to what extent the inflow of the Danjiangkou Reservoir can represent the WY of overall UHRB is unknown. However, we think the effects of bias of WY during 2012-2014 were limited. The measured and simulated WY not only have good correlation but also have nearly identical trends (see the following figure). This situation has been stressed in "2.4 Model Evaluation", and we added statement about the comparison of trend of measured and simulated WY in the model evaluation section.



11, l.225-226: I am concerned that the change point analysis here is really very sensitive to the rather short time period considered and that the points identified here may be mere artefacts (e.g., Zhou et al., 2019; HSJ). I strongly suggest to omit this from the analysis.

Response: As suggested, we have removed this from the analysis.

11, l.220ff: am I right to assume that scenarios S1 and S2 are shown and discussed in this section? Please clarify and make this explicit.

Response: Yes, we have clarified this in the revised manuscript.

12, 237ff: not clear what is considered here. Is it the difference between S1 and S2? If yes, I wonder how much of the correlation is spurious, as NDVI is kept constant, while still using observed T and VPD that are the result of a variable vegetation cover. This needs to be made much clearer.

Response: Yes, it is the difference between S1 and S2. We have made this clearer in the revised manuscript.