We thank the reviewer for the valuable comments. We have replied the comments one by one below, and revised the manuscript accordingly. The replies are highlighted in blue color and the modified texts (in the revised manuscript) are shown italic.

Reviewer #2

The paper describes a multimodel assessment of the relative impacts of human activities and climate on mean annual streamflow over the past 4 decades in China. This study shows that unlike previous assessments, the climate impact signal is much more pronounced than the human impact signal in 88% of river segments in China. The study also quantifies the impact of humans across basins and discusses regional differences. In general the paper is publishable after some moderate revisions.

Reply: Thanks for the positive comments. We have replied the comments and revised the manuscript accordingly.

- The use of the term ‘climate change’ in the title and throughout the manuscript is somewhat confusing and misleading because it gives the impression that the paper will be forward looking in time and over the coming several decades (e.g., 2050, 2100). A more appropriate term is ‘climate impacts’

Reply: We used “climate impact” in the title. We have replaced “climate change” by “climate variability” in most cases, and by “climate impact” in some cases where appropriate. “climate change” is still used when refer to future climate change and results from some specific references.

- The 3rd paragraph of the introduction makes the argument that “This is the first study to perform such a quantitative assessment for all rivers of China with comparable modeling experiments.” Being aware of the ISIMIP publications (https://www.isimip.org/outcomes/publications/) in this space with global assessments including many of the authors on this paper, I find this argument to be an exaggeration. I think the last sentence of that paragraph is a key novelty of this work, and as such linking back to the content of the second paragraph in the introduction to make the case would be my suggestion. I do agree that focusing on China is somewhat unique about this study. So one suggestion is to tweak the noted sentence as follow “This is the first study to focus on performing such a quantitative assessment for all rivers of China with comparable modeling experiments.”

o Schewe et al.: Multimodel assessment of water scarcity under climate change. PNAS, 2014.

o Haddeland et al.: Global water resources affected by human interventions and climate change, P. Natl. Acad. Sci. USA, 111, 3251–3256, https://doi.org/10.1073/pnas.1222475110, 2014.

o Veldkamp et al.: Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century, Nature Commun., 8, 15697, https://doi.org/10.1038/ncomms15697, 2017.

o Wada et al.: Human–water interface in hydrological modelling: current status and future directions, Hydrol. Earth Syst. Sci., 21, 4169-4193.

Reply: Thanks for the suggestion. We have revised the sentence following the suggestion.

- P5, L2: I would suggest omitting ‘preliminary’

Reply: Changed.

- P7, L14-17: Showing the individual models in figure S2 makes the figure too busy to read. Why not use the same format as in figure 2 by showing a band around the median. Also, it would be useful to show the same type of figure as figure 2 but for streamflow.

Reply: We have redrawn Figure S2 and added a figure (Figure S3 in the revised manuscript and below) for the simulated and observed annual streamflow following the suggestion.

A brief description has been added.

Revision in the manuscript (Subsection 3.1, the second paragraph (new added)):

The model spreads in the ensembles of seasonal streamflow and the deviations between observation and simulations are relatively larger in the northern basins than those in the southern basins (see Figure S2 for each basin). Comparison between the simulated and observed annual streamflow (Figure S3) shows similar patterns as the seasonal streamflow with respect the discrepancies between northern and southern basins. The Nash-Sutcliffe coefficient was calculated for the multimodel median and observed monthly streamflow at each station (see Table S2), which shows that the multimodel medians have better performance in the southern basins. This evaluation indicates that the multimodel simulations have relatively poor performance in northern basins and most stations with relatively smaller streamflow (e.g., in dry areas or upper reaches). The large spreads between models underline the necessity of using ensemble medians rather than individual models for the attribution of streamflow changes.

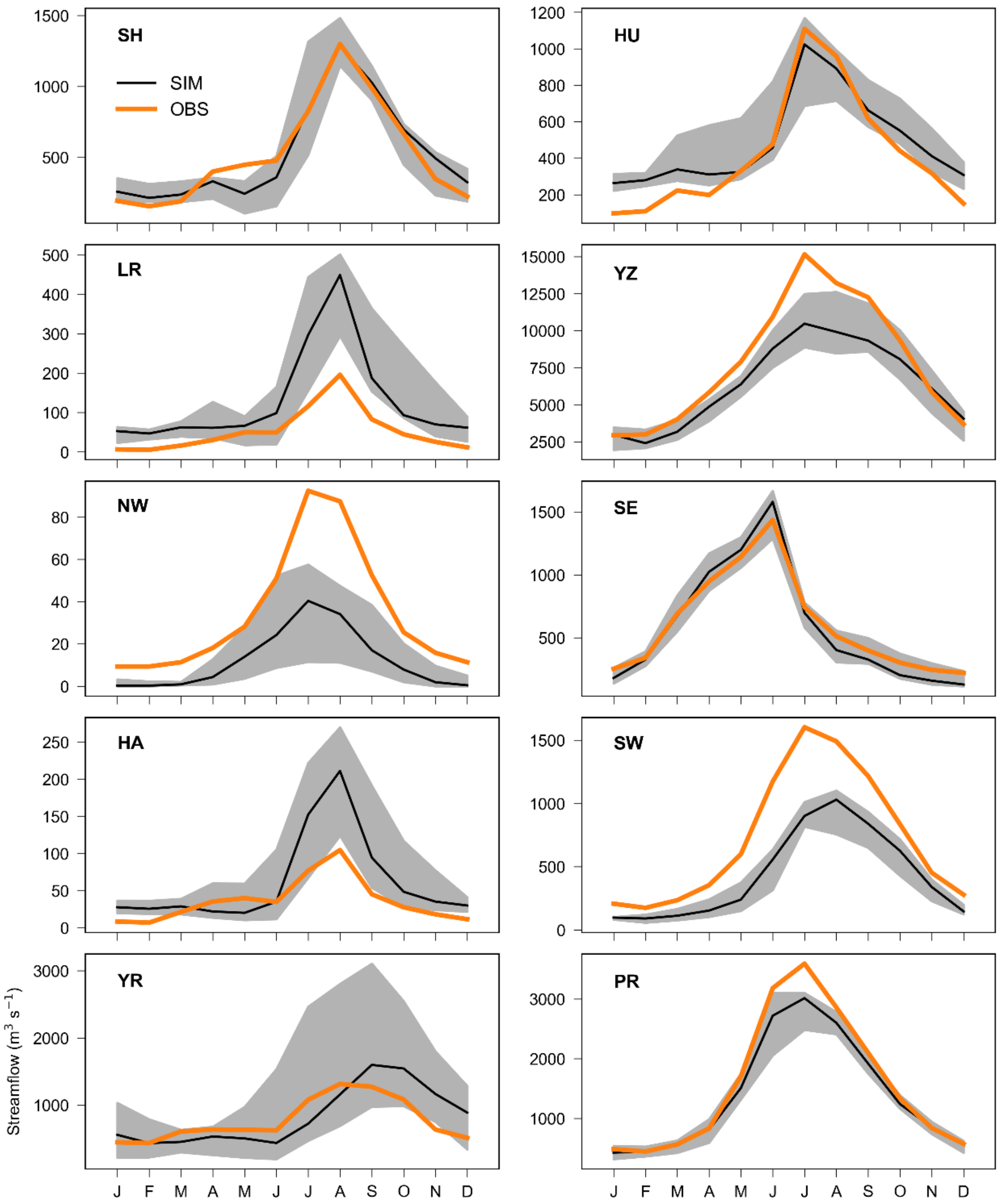

Figure S2. Seasonal cycle of streamflow from observations (orange) and multimodel medians (black). The observations are the average values of the hydrological stations, while the simulations are averaged values over the grid cells identified by the location of stations. The gray areas show the 25th and 75th percentiles of the multimodel simulations. Northern basins: Songhua River (SH), Liao River (LR), Northwest Rivers (NW), Hai River (HA), Yellow River (YR), Huai River (HU); Southern basins: Yangtze River (YZ), Southeast Rivers (SE), Southwest Rivers (SW), Pearl River (PR).

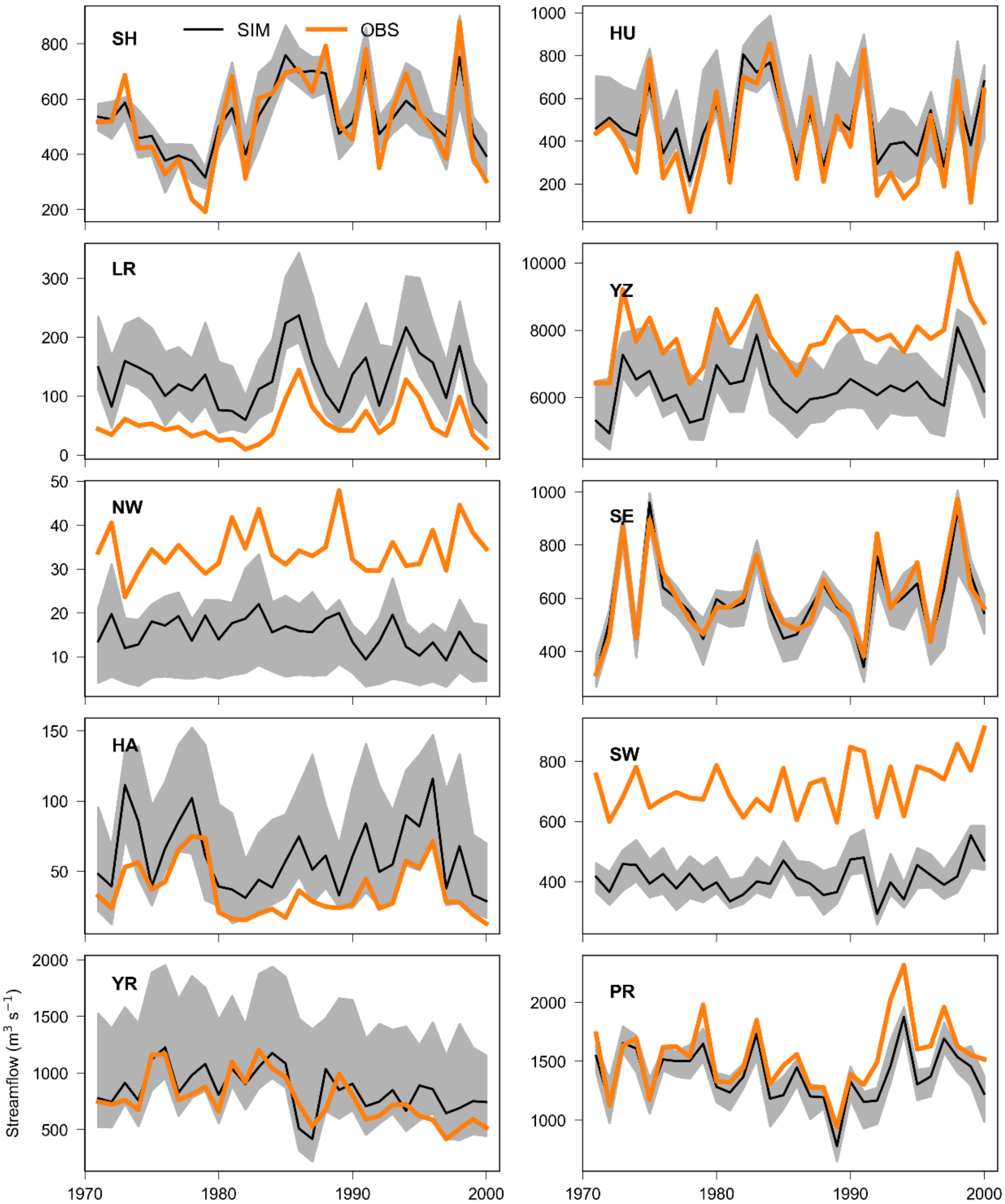

Figure S3. Simulated (black) and observed (orange) mean annual streamflow at the hydrological stations in each basin. The observations are the average values of the hydrological stations, while the simulations are averaged values over the grid cells identified by the location of stations. The gray areas show the 25th and 75th percentiles of the multimodel simulations.

- P7, L18-24: I realize that given the large departures in water withdrawal estimates, matching streamflow gauge observations might be a challenge, unless the authors believe that simulated water withdrawals might be equally or even more reliable than the statistically collected data, which have their own challenges.

Reply: We would not state that the simulations of water withdrawals are equally or more reliable than the statistically collected data. In fact, they are largely based on statistically collected data (e.g., Flörke et al. 2013; Hanasaki et al., 2008). We agree that the large spreads in the multimodel simulations of water withdrawals should be one uncertainty source to the streamflow simulations, but the effect is not superimposing (e.g., Fig. 9 in Müller Schmied et al 2014 for one GHM). On the other hand, Veldkamp et al. (2018) showed that inclusion of human impacts such as water withdrawals leads to better model performances. The simulations of water withdrawals remain a challenge, though great efforts have been made by the community. We have added a statement to address this concern at the end of this paragraph in the revised manuscript, but would not further elaborate it since it is not the focus of this study.

Revision in the manuscript (Subsection 3.1, the last paragraph)
The large deviations in the multimodel simulations of water withdrawals could make the modeling of streamflow more challenging (Döll et al., 2016; Wada et al., 2017).

- P7, L18-24: Are water withdrawals taken from surface water sources or also groundwater sources? What about return flows? Also, I am assuming that glacier melting, which contributes to streamflow, is simulated in these models, but that region is not included in the analysis. I realize that some of these were mentioned in the results, but incorporating some of these details briefly when discussing the method or the results from the evaluation exercise would suffice.

Reply: The sources of water withdrawals are shown in Table S1, depending on models, which may include river channel, reservoirs, groundwater and lakes. Return flows were considered in different ways for different water uses (e.g., Müller Schmied et al., 2014; Wada et al., 2014; Pokhrel et al., 2015). Glacier melting was not simulated in most GHMs (except PCR-GLOBWB) in this study. We have added the absence of glacier melting in the models as a reason of excluding the Tibet plateau region, and described the sources of water withdrawals in the Method section (please see the reply to the 3rd comment of reviewer #1) in the revised manuscript.

Revision in the manuscript (Subsection 2.1, the first paragraph)
The simulations may have large uncertainties over the Tibetan Plateau because long-term meteorological and streamflow observations are sparse in this region (Zhang et al., 2017) *and the modeling of glacier melting is absence in most of the models*.

- P10, L23-29: how does the model specify how much water is taken from surface water vs groundwater sources? Are the small pockets of increased MAF due to human impacts (Fig 5c) attributed to technological change (e.g., irrigation efficiency), or return flow from groundwater pumping, or something else?

Reply: Generally, groundwater is withdrawn when the water use requirement is not met due to limited accessibility to or insufficient quality of surface water. Groundwater withdrawal was considered in most models (see Table S1), but the pumping rate may vary substantially between models (Wada et al., 2016). It is difficult to determine the groundwater pumping rate since groundwater storage is usually unknown. The fraction of groundwater for water use is determined from reported statistics data (Siebert et al., 2010, used in WaterGAP) or estimated in the model (e.g., PCR-GLOBWB, see Wada et al., 2014). We have briefly clarified it in the Method section (also see the reply to the 3rd comment of reviewer #1) in the revised manuscript.

The increased MAF should be mainly due to return flow, but we cannot identify it from which source because of lacking related model output currently. Technological development may improve water use efficiency and reduce the amount of withdrawals. However, it may be not the reason for the slight MAF increase induced by DHI change, because water withdrawal increased over the study period (see Figure 2). We have clarified it in the revised manuscript.

Revision in the manuscript (Subsection 3.4, the first paragraph)

Compared to the first sub-period, in the second sub-period MAF increased by more than 30% in many river segments of the Northwest Rivers and increased by more than 5% in large parts of the Huai River, *which may be due to the return flow from water withdrawals.*

- P11, L7: I would suggest omitting the sentence about the US and Canada. It breaks the flow of the paragraph which is talking specifically about China.

Reply: Removed.

- P13, L14-30: To me this, this is a key contribution of this study. Yes, I agree that the results are not necessarily comparable in term magnitudes due to the highlighted reasons by the authors. But a missing discussion point is to why they fundamentally differ in their findings. I don't agree that either one of these two approaches (small scale using statistical approaches vs large scale modeling similar to this study) is necessarily superior. Each approach has its own pros and cons. So articulating why this approach differs from earlier findings is critical.

Reply: We agree that both the methods has its own pros and cons. At the end of the discussion, we have emphasized the importance of using multiple approach to obtain more reliable assessment. In the revised manuscript, we have rewritten the last paragraph of the discussion, wherein we clearly stated the key difference between the method in previous studies and this study.

Revision in the manuscript (Subsection 4.6, the last paragraph)

One major difference between previous studies (e.g., Li et al., 2007; Bao et al., 2012) and this study is that the former estimates DHI contribution by comparing simulations with observations while we compare two simulation experiments. The former may be subject to uncertainty in comparing the data from two systems (i.e., the model and the real world). In this study, the two simulation experiments favor the estimation of DHI contribution in a consistent manner that is largely free of uncertainty in the data from different systems. The multimodel simulations also allow profiling the uncertainties among models and input forcings, which is difficult for a single *model assessment. However, the deficiency of this approach is that DHI is not real. Therefore, the assessment is inevitably influenced by the extent to which the models can reproduce the real DHI.* Considering the complexity of DHI on streamflow and the ability of current hydrological models in reproducing historical hydrological changes, multimodel simulations and different attribution approaches are well worth obtaining more robust assessments (Liu et al., 2017; Yuan et al., 2018).

Reference:

Bao, Z., J. Zhang, G. Wang, G. Fu, R. He, X. Yan, J. Jin, Y. Liu, and A. Zhang (2012), Attribution for decreasing streamflow of the Haihe River basin, northern China: Climate variability or human activities?, J. Hydrol., 460-461, 117-129, doi: 10.1016/j.jhydrol.2012.06.054.

Döll, P., H. Douville, A. Güntner, H. Müller Schmied, and Y. Wada (2016), Modelling Freshwater Resources at the Global Scale: Challenges and Prospects, Surv. Geophys., 37(2), 195-221, doi: 10.1007/s10712-015-9343-1.

Flörke, M., E. Kynast, I. Bärlund, S. Eisner, F. Wimmer, and J. Alcamo (2013), Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, Global Environ. Change, 23(1), 144-156, doi: 10.1016/j.gloenvcha.2012.10.018.

Hanasaki, N., S. Kanae, T. Oki, K. Masuda, K. Motoya, N. Shirakawa, Y. Shen, and K. Tanaka (2008), An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing, Hydrol. Earth Syst. Sci., 12(4), 1007-1025, doi: 10.5194/hess-12-1007-2008.

Li, L., L. Zhang, H. Wang, J. Wang, J. Yang, D. Jiang, J. Li, and D. Qin (2007), Assessing the impact of climate variability and human activities on streamflow from the Wuding River basin in China, Hydrol. Processes, 21(25), 3485-3491, doi: doi:10.1002/hyp.6485.

Liu, X., Q. Tang, H. Cui, M. Mengfei, D. Gerten, S. Gosling, Y. Masaki, Y. Satoh, and Y. Wada (2017), Multimodel uncertainty changes in simulated river flows induced by human impact parameterizations, Environ. Res. Lett., 12(2), 025009, doi: 10.1088/1748-9326/aa5a3a.

Müller Schmied, H., S. Eisner, D. Franz, M. Wattenbach, F. T. Portmann, M. Flörke, and P. Döll (2014), Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18(9), 3511-3538, doi: 10.5194/hess-18-3511-2014.

Pokhrel, Y. N., S. Koirala, P. J. F. Yeh, N. Hanasaki, L. Longuevergne, S. Kanae, and T. Oki (2015), Incorporation of groundwater pumping in a global Land Surface Model with the representation of human impacts, Water Resour. Res., 51(1), 78-96, doi: 10.1002/2014WR015602.

Siebert, S., J. Burke, J. M. Faures, K. Frenken, J. Hoogeveen, P. Döll, and F. T. Portmann (2010), Groundwater use for irrigation – a global inventory, Hydrol. Earth Syst. Sci., 14(10), 1863-1880, doi: 10.5194/hess-14-1863-2010.

Veldkamp, T. I. E., et al. (2018), Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model

validation study, Environ. Res. Lett., 13(5), 055008, doi: 10.1088/1748-9326/aab96f.
Wada, Y., D. Wisser, and M. F. P. Bierkens (2014), Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dynam., 5(1), 15-40, doi: 10.5194/esd-5-15-2014.
Wada, Y. (2016), Modeling Groundwater Depletion at Regional and Global Scales: Present State and Future Prospects, Surv. Geophys., 37(2), 419-451, doi: 10.1007/s10712-015-9347-x.
Wada, Y., et al. (2017), Human–water interface in hydrological modelling: current status and future directions, Hydrol. Earth Syst. Sci., 21(8), 4169-4193, doi: 10.5194/hess-21-4169-2017.
Yuan, X., Y. Jiao, D. Yang, and H. Lei (2018), Reconciling the attribution of changes in streamflow extremes from a hydroclimate perspective, Water Resour. Res., doi: doi:10.1029/2018WR022714.
Zhang, C., Q. Tang, and D. Chen (2017), Recent Changes in the Moisture Source of Precipitation over the Tibetan Plateau, J. Climate, 30(5), 1807-1819, doi: 10.1175/jcli-d-15-0842.1.