General comments

The study used a well-known framework to analyze the water scarcity in some large basins in China. Although the method is not new, the topic is interesting. However, some details about the method should be added (please see the following point-to point remarks), and the presentation of the results should be improved. In the results part, I found that the analysis was not complete for each basin, the results were not well organized, and the figures are hard to follow. These limitations made me a little bit hard to understand the results and conclusions (some are due to a lack of quantitative analysis, and some are due to a lack of complete summary and necessary discussions; particularly, the result about water scarcity was not well interpreted). Finally, the authors had three objectives, but the imbalance between upstream and downstream regions was not well quantified, and the third one was only discussed in a very simple way.

Response: We thank you for your recognition of our work and appreciate the constructive comments and insightful suggestions that will help improve our paper. We will address your concerns in the revisions. The detailed responses for your point-to-point remarks are listed below.

Specific comments:

P4L1: how did you do the model calibration to show that theta is most sensitive to topography? The details about the model calibration were missing. The theta value was constant for all the basins?

Response: The parameter theta of Fu-Budyko framework does change from one basin to another and from upstream to downstream. Here three basins, Hei River, Yellow River and Yangtze River, were taken as examples to show the change of theta in the following table.

River	P_{up}	PET_{up}	$theta_{up}$	P_{down}	PET_{down}	$theta_{down}$
Yangtze River	705	1083	1.7	1106	996	2.0
Yellow River	498	866	1.7	433	1027	2.3
Hei River	375	916	1.3	215	1027	2.0

The climate conditions of the three basins are different, from humid Yangtze River basin, to semi-arid Yellow River basin and arid Hei River basin. Some have more arid upstream regions while others have more arid downstream regions. The similarity, however, was identified that the theta of upstream is lower when compared to its downstream counterpart. Given the fact that upstream regions have steeper terrains, the lower theta is probably related to the topography. The result is consistent with study from Sun et al. (2007), indicating that three factors - infiltration rate, water storage capacity and average slope - had impact on the parameter theta of Fu-Budyko framework.

Reference:
Sun, F., Yang, D., Liu, Z., and Cong, Z., 2007. Study on coupled water-energy balance

in Yellow River basin based on Budyko Hypothesis (in Chinese). Journal of Hydraulic Engineering, 38(4), 409-416

P4L2: the uncertainty of the model should be evaluated more completely. 6.9% was only the average. However, how about the spatial distribution of the uncertainty? Which basins had the largest uncertainty?
Response: This is a very good point. Generally, the biases are smaller in humid basins while larger in arid basins. The southeastern basins (Min and Qiantang) have the lowest biases around 3%. Followed are the southern basins (Pearl and Yangtze) and northeastern basins (Songhua) with biases around 6%. Yellow basin and the northwestern basins have approximately 10% biases. Hai river basin has the highest biases of ~20%.

P4L3: please give references to show this framework can be suitable for annual studies. In my experience, this frame is only suitable for mean annual studies.
Response: Zhang et al. (2008) has tested the Budyko model over 265 Australian catchments at different time scales, including mean annual, annual, monthly and daily. They found at annual scale, the model works well for most of the catchments with 90% of them having values of the coefficient of efficiency greater than 0.5 and less than 3% of the catchments have bias values greater than 10%. Meanwhile, there are some catchments where the model performed poorly. A reference will be added.

Reference:
Zhang, L., Potter, N., Hickel, K., Zhang, Y., and Shao, Q., 2008. Water balance modeling over variable time scales based on the Budyko framework - Model development and testing. Journal of Hydrology, 360, 117-131.

P4L17: here, why was ET0 calculated by the Hargreaves equation rather than the Penman equation? The gridded meteorological data can be also obtained by interpolating the station-based data to grids.
Response: Both equations were used in the study. The Hargreaves method was chosen because only temperature and precipitation were available in the gridded meteorological dataset. And the PM-based potential ET from pointed dataset was used to corrected the Hargreaves-based potential ET, which will be greatly improve the accuracy especially in the eastern regions.
Both the gridded and pointed meteorological data have their advantages and disadvantages: the pointed dataset contains more meteorological variables but is sparse in the western regions, while the gridded dataset is denser in the western regions but contains only temperature and precipitation. Considering the complex terrain and dry climate in western regions, we think the distribution of meteorological gauges is more influential factor for the accuracy of interpolation of potential ET. Thus the combination of gridded data and pointed data was chosen in this study to reduce the errors of potential ET in western regions. Additional notes will be added in revision for clarification.

P5L2, please give the reference for the classification method of AI.

Response: The classification of AI is following the method of Ponce et al. (2000) with arid, semi-arid, semi-humid, and humid regions ranging from 12~5, 5~2, 2~0.75, and 0.75~0.375. In this manuscript, there were mistakes to label the limits of AI and these will be corrected in the revised manuscript. The following references will be added:

Arora, V.K., 2002. The use of the aridity index to assess climate change effect on annual runoff. Journal of Hydrology, 265(1-4), 164-177.

Ponce, V.M., Pandey, R..P., and Ercan, S., 2000. Characterization of drought across climate spectrum. Journal of Hydrologic Engineering, ASCE, 5(2), 222-224.

P5L5, according to Figs. 4 to 9, I think you focused more on the changes, so maybe the trend was less important. Please consider to delete the trend analysis contents to make the results more coherent.

Response: Thank you for your suggestion. The reason for keeping the trend analysis in the manuscript was in case someone is interested in the trends. After careful consideration of your suggestion, we decide to delete Fig. 3 which demonstrates the P and potential ET trends. Meanwhile, the Fig. 2 is still kept because we think the trends of natural and observed runoff are helpful for readers to understand the changes of water scarcity.

P5L9, the definition of water scarcity is expressed by two indicators, but this is not very easy to follow, especially in Fig. 9. I suggest define a new indicator, e.g., WS=WTA/Shortage? Maybe it is easier to compare this indicator among different decades, basins, and reaches.

Response: Thank you for your thoughts. In fact, "Shortage" refers to the available surface water resources per capita and it is related to the demographic-driven change of water scarcity. We will change the description of "Shortage" and use a new abbreviation instead of "Shortage".

P6L15, the correlation coefficient of natural and observed runoff means what? As defined by the authors, natural runoff and observed runoff could be totally unrelated, so I don't know what R means. 1961-1970 was the period for model calibration, so why did you show the degree of suitability of the model during 1961-2010? If the authors assumed that period from 1961-1970 was nearly natural, you should divide the period into two sub-periods: one for calibration and the rest one for validation. I noticed that the model's performance in some basins listed in the right most column of Figure 1 was very poor during 1961-1970. Is the framework suitable for these basins?

Response: At the beginning, we thought the correlation between observed and natural runoff might reflect the human interventions on runoff. We will reevaluate our earlier conclusion. We will use correlation coefficient in calibration period to show the model performance here. And we worried about if five-years would be long enough to calibrate the model. We will try the suggestion to divide 1961-1970 into calibrate and validate period.

For the arid basins, we divided the arid basins into upper mountainous sub-basins and downstream sub-basins, and applied Budyko framework in the former sub-basins. For the downstream sub-basins, we use the observed runoff and evapotranspiration to calculate runoff. Please see the answer to P6L29 for the detailed explanation.

P6L24, it is very difficult to see which gauges are in the upstream and which gauges are in the downstream. The authors should think about how to present the locations of the gauges clearly.
Response: Thank you for the comments. We will label these hydrological gauges in Figure 1 to make these visible.

P6L25, can you explain why a gauge with a positive trend in rainfall can have a negative change (Fig. 3)?
Response: Here the rainfall's change was not calculated as the trend (mm/year) * year, instead it was calculated as the differences between two periods - 2000s and 1960s. Thus the change of a gauge was only dependent on the differences between 2000s and 1960s but not the trend. For those gauges with fluctuations and no significant trend, it is possible that a gauge with a positive trend has a negative change.

P6L29, in northwest of China, such as Heihe, Tarim, river runoff is mostly contributed by snow melt. Is the framework suitable for these basin?
Response: Budyko framework performs bad in arid and snow basins, which has been proved in the previous studies. Here we found that Du et al. (2016) successfully applied a Budyko framework in arid Heihe River Basin by dividing it into six sub-basins according to basin characteristics. They calibrated the model separately in different sub-basins and found the model performed quite well in the upper mountainous regions with little interference of human activities. So we also divided the arid basins into upper mountainous sub-basins and downstream sub-basins, and applied Budyko framework in the former sub-basins. For the downstream sub-basins, we use observed runoff and evapotranspiration to calculate runoff. Detailed explanation will be added in method section.

Reference:
Du, C., Sun, F., Yu, J., Liu, X., and Chen, Y., 2016. New interpretation of the role of water balance in an extended Budyko hypothesis in arid regions. Hydrology and Earth System Sciences, 20, 393-409.

P6L25, P71, the authors gave subjectively the reasons for the trend (a significant increase in rainfall, recent global warming), I don't see any supporting analysis.
Response: There are two reasons for the increase of observed runoff, one for increasing precipitation and the other for decreasing evapotranspiration. Given the insignificant change of potential evapotranspiration, we think the main driver for increasing runoff is the increase of precipitation.

P7L11-15, from Figure 4, I can't see these interesting analyzes. And, please add the AI in this figure.

Response: Thank you for the suggestion on Figure 4. We think the statement of the paragraph (P7L11-15) is too subjective. Thus we will use numbers to describe the change of water stress and include additional discussions here.

P7L14-15, this is also too subjective.

Response: As mentioned above, we are considering to describe the change of water stress using numbers here.

P7L20, in Figure 5a, I suggest add an average of 1970s~2000s for each basin. Here, how did you define "continuously"? Obviously, WTA in the Yangtze, Pearl, Min River and Songhua did not increase monotonously.

Response: Thank you for the suggestion on Figure 5. And we are sorry for the improper "continuously". Because the fluctuations of WTA are small in these humid basins within a range less than 2%, so we think the fluctuations might be caused by the modelling errors and can be ignored. We will fix the improper word and replace it with a new one.

P7L17-25, these results should be discussed to give the possible reasons.

Response: Thank you for the helpful suggestion. We will further discuss the turning point of the decreasing surface water stress for different basins in the revised manuscript, which is supposed to be related to the economic development.

P8L5, Figure 7 is about water shortage, so I don't know why the authors were talking about surface water availability.

Response: "Shortage" is defined as surface water availability per capita, so both water availability and population can influence water shortage. In this paragraph, we aimed to explain the converse phenomenon of water shortage between northern and southern basins, which was mainly related to the surface water availability.
We will add an explanation before the paragraph and change the description of "Shortage" in the revised manuscript.

P8L11, water availability is determined by natural runoff, so I can't understand why population can affect water availability.

Response: We feel sorry about the typing mistakes in this paragraph. It should be "water shortage" or "water availability per capita" but not "water availability". These mistakes will be corrected in the coming revised manuscript.

P8L19, from Figure 9a, I can't see the aggravation of water scarcity in China. This figure is not visual to show this aggravation trend.

Response: When the dots of water scarcity move to up-right direction, the aggravation of water scarcity happened because of higher WTA and lower Shortage. We will try to seek a solution by either changing the figure or adding explanation to make the figure

more understandable.

P8L25, water scarcity is defined with water stress and water shortage, here, why is it related to surface water availability?
Response: We agree that it is an improper expression here. It should be "dramatic increase of surface water withdrawal and little change of water availability per capita, suggesting it is demand-driven water scarcity in semi-humid/arid basins". We will change the expression here.

P8L28, fig. 9a and 9c cannot show this competition (at least I don't know how to interpret). And this paragraph was about water scarcity, but the authors were talking about water withdrawal. So it is very hard to understand these sentences.
Response: We are sorry for the bad exhibition of Figure 9. We will try to revise the figure or add more statement to explain the figure.
And water withdrawal was mentioned here because it is the most influential factor on water scarcity for northern basins. We will reorganize the paragraph to make it less confused.

In Figure 8: in the Liao, Huai, and Qiantang, why were there no upstream, middle, and downstream?
Response: Because the hydrological data of some gauges is not available in the three basins. For example, the record of hydrological data in Liao's upstream gauge started from 1984, which was too short to conduct the analysis; the hydrological data of Huai's downstream gauge was missing; and there was only hydrological data in tributary gauges for Qiantang basin. A short explanation will be added in the data section.

P9L4-5, no analysis supporting the statement here.
Response: We are sorry for the improper statement here. Stricter expression will replace the old one as "This study showed that climate change was the major driver of natural runoff."

P9L16, the possible impacts of the policies on water scarcity in all the basins were not fully discussed.
Response: Thank you for the advise. More discussion about the polices and their impacts on relieving water scarcity in China will be added in the discussion section.