Responses to the comments from Dr. Yu Chaoqing:

Thank you for your constructive comments on our manuscript. We have addressed your comments one by one below and revised our paper based on your suggestions.

Comments

Import and export WF in the Heihe River Basin. The WF values in the paper are calculated from the statistical data of local crop and livestock productions, but it does not show a comprehensive footprint chains of the overall water cycling based on the land use of the basin. The exchange values between local and external water footprints are not well considered in their calculation either. Mapping the space-time changes of water footprints within the basin can increase our understanding of the entire processes of virtual water movement, and facilitate risk analysis of water use within the basin. A practical approach is to use models. While the paper mentioned that the annual and monthly blue water resources were derived from the SWAT modeling results, from the description in the paper, it seems that the research only used the output data of the modeling, rather than applying the WF approach to trace and measure the water cycling processes in the distributed hydrological modeling. The import and export goods attached with virtual water are also an important part of the WF, but are missing in the paper. Measuring the inter-basin exchange of WF still remains challenging because of the diversity of the goods and the scarcity of data.

Responses

We agree with the reviewer's comments for many shortcomings of the paper. We understand the above comments in three aspects. First, there is no separation of internal and external *WF*, or the virtual water trade (import and export) is not addressed. Second, the spatio-temporal distribution of *WF* is not addressed. Third, water cycling processes are not traced. We also agree with the reviewer that to address all the above issues still remains a challenging work.

For the first point, we have calculated internal and external *WF* based on input-output models for Gansu province, which covers most part (43%) of the HRB. The results show that the virtual water export of the agricultural products accounted for 10% of the total water resource and 25% of the total water use in the province (Cai et al., 2012). We did not provide a comprehensive calculation of internal and external *WF* in this paper for the HRB because (1) previous research on virtual water trade is based on input-output models, but our approach in this paper is based on the Water Footprint Network method; (2) for the Water Footprint Network method, either the food trade data or the food consumption data should be used to estimate virtual water trade. Unfortunately, both the datasets have not yet collected successfully. We mentioned the above findings as well as the difficulties in partitioning of internal and external *WF* in the manuscript. [Page14, Line 10-26]

For the second point, the spatio-temporal distribution of *WF* assessments remains a very big challenge for the HRB largely due to the scarcity of data, as mentioned by the reviewer. Spatial heterogeneity of climate conditions and land use/cover are very sharp in the HRB with high precipitation and glaciers upstream and low precipitation and desert downstream. There is a need to compare *WF* with water availability at the sub-basin levels. This is out of the scope of this paper, but it is what will be further investigated in the next step. [Page15, Line 6-11]

For the third point, how detailed the calculation of *WF* should go to depends on the objective of the research. To study product *WF*, it is often necessary to trace the supply chain of the product, and add up all the water needed in each chain. However, *WF* assessments at a river basin level are often based on the product *WF* results without going to track trace and measure the water cycling processes in the distributed hydrological modeling. [Page15, Line 12-17]

Comments

Validation is essential in evaluating the WF results, but is not well addressed in the paper. Although the paper compared the WF values with the results from other studies, it is not a correctness proof of the conclusions. The indices of virtual water contents (VWC) are fundamental in WF calculation, but it is not well explained how such indices are derived, or if the numbers of these indices are feasible to the Heihe River Basin in terms of the local species, climate, soil, and management. Validation of the WF calculation is still a difficult task, but the conclusions can be much stronger if there are some sporting evidences from scientific experimental data, such as local farming practice, soil moisture changes, biomass, and irrigation experiments. Some other technologies, such as stable isotope analysis, can also be helpful to trace the water cycling processes and provide solid evidences to prove the results.

Responses

We agree with the reviewer that validation is very important for the research results. But our research is the first study of *WF* assessment at the HRB, it is very hard to validation the *WF* results from some scientific experimental data like local farming practice and soil moisture changes. But we have compared our results with other studies, which have reported *VWC* values.

For different crops, the *VWC* of crops estimated in this paper is slightly higher than China's average values from Liu et al. (2007). One exception is cotton, and its *VWC* value estimated here is about twice the national average value. The climatic condition is one important reason for the higher *VWC* values in the HRB. The HRB is located in arid and semi-arid regions with high potential evaporating capacity. We also find that the *VWC* values of livestock products in HRB are generally higher than those reported in Chapagain and Hoekstra (2004) and Liu and Savenije (2008). Especially for beef, its *VWC* value is 1.6 times the value calculated by Chapagain and Hoekstra (2004). The feed eaten by animals has higher *VWC* values in the HRB due to the dry climate conditions, leading to higher *VWC* of animal meats.

Zhang (2003) have calculated *VWC* of crops and livestock in the Zhangye located in the west of HRB. Except for starchy roots and oil crops, the *VWC* of all other crops and livestock reported by Zhang (2003) are very close to our results. The *VWC* of starchy roots and oil crops calculated by Zhang (2003) are much larger than ours. Mainly because rainfall in the Zhangye region is lower (157-103 mm y^{-1}) than the HRB's average level. These two types of crops mostly depend on green water rather than blue water. Low precipitation leads to high *VWC* of these two crops in the Zhangye region.

We have put the above comparison in the discussion part. [Page 11, Line 6-21]

Comments

Sustainable water use is an ultimate objective of conducting such type researches, and it is addressed in this paper. However how to measure sustainability is still questionable in this paper. The authors used the value of EFR (environmental flow requirements) as an index for WF sustainability assessment, according to the suggestion in Hoekstra et al. (2011, 2012b). In this paper, the EFR value being used is 80% of the total blue water resources of the basin, but what Hoekstra et al. suggested was 80% of the total natural runoff. It is necessary to discuss why the total blue water resources, instead of the total natural runoff, are applied. The number of 80% is a generic value, but is it feasible for the Heihe River Basin at all? Answering this question needs to setup a baseline of a "normal" water status, and then evaluate the actual water requirements, especially from the local ecological systems. The indices of blue water scarcity values are also a rather arbitrary standard, which is worthy of a further study to evaluate if it is suitable for this particular basin.

Responses

Thank you for your good points. One thing we need to clarify here is that the blue water resources estimated in Zang et al. (2012) were for a natural condition without human intervention. Hence the concept of blue water resources used in the paper is equivalent to natural runoff. In the revised version, we have changed blue water resources to natural runoff in many places to reduce the confusion.

As to the EFR value, we choose 80% suggested by Hoekstra et al. (2011; 2012b). Certainly such a value is still questionable. We did not find a more suitable threshold for the environmental flows of the HRB. We also agree that it is worthy of a further study to evaluate whether 80% can be used at the HRB. The best way is to setup a baseline of a “normal” water status, and evaluate the actual water requirements, especially from the local ecological systems. We have discussed this issue in the revised version as follows.

“Second, for the EFR value, we choose 80% as a threshold based on Hoekstra et al. (2011) and Hoekstra et al. (2012b). It is still questionable whether such a threshold can be used for river basins in arid and semi-arid regions such as the HRB. To address this issue, further efforts are still needed to study the environment flows that are required to sustain freshwater ecosystems and human livelihoods and wellbeing that depend on these ecosystems. One effective way is to setup a baseline of a “normal” water status, and evaluate the actual water requirements, especially from the local ecological systems.” [Page14, Line 3-10]

Comments

Overall, this paper demonstrates a good example of calculating WF values at a basin level. It is obviously more advantageous than using the traditional withdrawal index, because the WF approaches take more water cycling processes into account. Accurate assessment of WF values, however, is still challenging because of complexity of the water processes. Beyond the research presented in this paper, there needs a lot of extra efforts, including developing new methodologies, standards, and technologies to improve the WF approaches.

Responses

We agree that there is still many challenging work for the *WF* assessment. We have explicitly mentioned this in our discussion session in the paper as follows.

“Overall, accurate assessments of *WF* still remain a challenging task due to the complex processes of water cycles and human activities, and the lack of many important input data at a river basin level. It is worth extra efforts to collect more detailed information to increase the accuracy of *WF* assessment at river basin scale.” [Page15, Line 20-23]