Hydrol. Earth Syst. Sci. Discuss., author comment AC2
https://doi.org/10.5194/hess-2022-160-AC2, 2022

Reply on RC2

Amanda Triplett and Laura E. Condon

Author comment on "Climate warming-driven changes in the cryosphere and their impact on groundwater-surface water interactions in the Heihe River Basin" by Amanda Triplett and Laura E. Condon, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-160-AC2, 2022

We thank the reviewer for their detailed comments to improve this manuscript and have replied to all major concerns below. We hope that the explanations and planned revisions to address questions and concerns are satisfactory.

1. Commonly, specific yield is smaller than porosity. In the manuscript, the authors used specific yield as an analogy for porosity. In addition, the authors simplified specific yield values in the model. Specific yield data from 17 unique values calibrated in Tian et al. (2015a) were simplified to three intervals of 0.1, 0.2, and 0.3. The authors stated that this simplification was used to lessen computational demand. The reviewer doubts the reasonability of such simplification. This oversimplification may cause the model far from real condition and the results may be inaccurate. As supercomputer and parallel computation is so common, computation demand may not be a problem.

These are great points of clarification that should be made in the updated manuscript. First, the authors acknowledge that specific yield is not the same as porosity. In this study, we used the same source data as that for the previously published HEIFLOW model. The HEIFLOW model only contains data for specific yield, while the ParFlow model requires porosity as an input. We decided the specific yield values could be used as reasonable estimates of porosity for two reasons. First, the middle Heihe is mostly an unconfined aquifer, with discontinuous aquitard sections (Yao et al., 2015a). This was confirmed with the input data. In unconfined aquifers, specific yield can more feasibly be used as an analog for porosity, although there is certainly some difference between the two values. Second, we note that the underlying data used to build the specific yield estimates was sparse and uncertain. We make the assumption here that the differences between porosity and specific yield are smaller than the uncertainty of the specific yield values themselves which is explained in more detail below.

As stated above, we initialized our model with the same input data used in the HEIFLOW model (Tian et al., 2015b) and with data provided from the Heihe Program Data Management Center. However, we found that using many of the values directly caused unrealistic behavior in our model, not only for specific yield but for hydraulic conductivity. Furthermore, we also found some of the values from the raw data to be non-physically realistic (i.e. anisotropy values of 8000). Given these complications we determined that we would need to conduct a separate model calibration exercise for our simulations. In the initial testing process, we found that our results were more sensitive to hydraulic conductivity than porosity, so we focused our efforts on calibrating that variable. By taking a lumped approach, we acknowledge how uncertain these variables are, and focus on more impactful hydrogeologic variables like K in our system.

Furthermore, we would like to note that when we say to lessen computational demand, we mean the computational demand to perform satisfactory calibration on all possible input variables. This will be clarified in the future manuscript. We absolutely do take advantage of parallel computing in our simulations, and ParFlow has been demonstrated in previous work to have excellent parallel scaling performance. However, it should be noted, that compute resources are still limited by compute allocations and must be used responsibly. One year of simulation with ParFlow alone (i.e. for our spinup runs) requires roughly 250 core hours (running on 972 cores for about 15 minutes). One year of ParFlow-coupled to CLM requires roughly 4800 core hours (running on 972 cores for about 5 hours). Considering the number of years for the groundwater system to return to equilibrium, especially after changes to hydrogeologic variables, this is a significant computing constraint.

2. The vertical thickness of the model is 472 m, which may be too shallow for a groundwater model. Many regional groundwater models have vertical thickness of 3-5 km. Can the 472 m thickness capture the main groundwater flow system in the study area?

The previous modeling studies of the Heihe are our best source for the hydrogeology of the basin. As explained in our previous answer, we did have to calibrate our model and change the individual values in the domain, but we tried to maintain the general lithology that they developed. The maximum depth of hydrogeologic data used in the previous modeling efforts was 2094m (Yao et al., 2015a; Tian et al., 2015b) with a no-flow boundary imposed below this. After analyzing the data, only a very small percentage of the domain had data below 1072m (about 8%). Additionally, only about 50% of the MHRB domain has data at depth greater than 472m and a K greater than .005 m/h. Originally, we had an additional bottom layer that was 600m thick in our model to cover the depth of the previous models. Note that the HEIFLOW model has variable thickness while our ParFlow model has constant thickness.

Our study is focused on shallow groundwater and groundwater surface water interactions, and we expected that the deep regional flow paths extending below this depth would have a limited impact on our results especially given the other uncertainties involved. We decided to test without the last large thickness layer. After comparing our results before and after this change, we did not see large enough differences in water table depth and streamflow processes on the time order of our simulations (11 years) to warrant the additional computational costs of keeping it.

In response to this comment, we will make our reasoning for the selection of this depth clear in the manuscript in Section 2.4 Model Configuration and Initialization.

3. For a regional groundwater model with relatively large thickness, the decrease of hydraulic conductivity K and specific storage Ss with depth should be taken into account. Did the authors consider the decrease of K and Ss with depth in their model?

We used hydraulic conductivity and specific storage input data as referenced in Tian et al. 2015b and used in additional studies with the HEIFLOW model (Tian et al., 2018; Sun et al., 2018). The creation of these datasets based on observations is described in Yao et al. (2015a). In this dataset, K is shown to decrease with depth. Additionally, any calibration we did was scalar, so the depth relationship we observe in the underlying hydraulic conductivity data has been maintained. The specific storage data, which we used unaltered from the source, was uniform with depth. We believe that some of these discrepancies in the data, for instance only two specific storage values that do not vary with depth, versus 92 values for K which do, highlights the limitations and uncertainty in comprehensive data for regional scale models. We will add text to highlight where these depth relationship trends exist in the input data.

4. The constant flux boundary condition along the border between the Upper and Middle Heihe may not be reasonable. The flux from the Upper Heihe is variable, and the flux is different in different seasons and among different years. The authors should clarify why a constant flux boundary is reasonable.

The reviewer is correct that we would expect groundwater flux to change over the course of the year, however our ability to represent this is limited by observations and we don't have any data to support seasonal variations on the boundary. Tian et al. (2015b) also used a constant flux value based on average annual data, literature estimates and model calibration (table 1) which we translated to our model and further details can be found in that publication. Furthermore, we do think the constant flux boundary is defensible for the following reasons: First, the groundwater flux represents only a small fraction of total water input to the domain, about 5% on average. Thus, the expected range of seasonal flux around this value is not expected to significantly impact the total amount of water entering the domain. We also performed calibration on the groundwater boundary condition (Line 223). The values we tested varied between +/- 75% of the original data. The change showed a minimal impact on groundwater and surface water. Seasonal flux changes seem unlikely to fall outside of that range.

Further, there is a large elevation gradient between the upper and middle basins in the Heihe, so a large fraction of groundwater discharges as streamflow at this boundary which is the same boundary the constant flux boundary condition is applied. So, seasonal changes in baseflow variability will be captured by the surface water inputs to the model.

5. The authors stated that "about 75% of water coming into the middle basin domain is from streamflow, 20% from precipitation, and 5% from the groundwater boundary condition." Where are the percentages from? Are there any evidence for these percentages?

We calculated the annual average volume of streamflow input into the domain from historic time series (Table 1, Heihe Program Data Management Center). We then calculated average long-term recharge or the balance of precipitation (Xiong and Yan 2013) with ET (PML V2 ET product, National Tibetan Plateau Third Pole Environment Data Center - http://data.tpdc.ac.cn/en/data/48c16a8d-d307-4973-abab-972e9449627c/?q=PML_V2), to obtain the rough contribution of precipitation to the water budget. Last, we obtained the magnitude of water entering through the groundwater boundary condition as calibrated in Tian et al. (2015b). The resulting breakdown of these water sources is approximately 75%, 20% and 5%. We will clarify the data source and calculations that result in these percentages in the edited manuscript.

6. A uniform water table depth of 20 m was used in the model. Why the authors use a uniform water table depth? Is there any support material, publications, or evidence for such a water table depth? For such an area, water table depth should be different in different regions.

We did not use a uniform water table depth for simulation. We started from a uniform water table depth of 20m before spin-up in order to have a point to initialize from. We then proceeded to run for 115 years to get a new water table to use in the calibration process. At this point, we have a spatially variably water table depth and we deemed the spin-up to be complete because storage as a change of percent recharge was less than 1%. After this, we calibrated the model and after final parameters were selected, we ran for an additional 73 years, again until storage change as a percent of recharge was less than 1%. At this point, we have the water table that we initialized our simulations from. This is described on lines 214-220. The authors will clarify the language in this section to make sure it is clear to the reader what water table the simulations start from.

7. For the Combined scenario, the authors used 15% reduction in thawing season flow and 50% reduction in baseflow. Are these reduction percentages have any supporting data, references, or other evidences? Or are they chosen arbitrarily?

We understand from this comment and others that the source of the values for the simulations were not sufficiently clear in the manuscript. We give a brief explanation here of where the values come from and will improve the clarity of this explanation in the manuscript. Additionally, we refer the reviewer to our response to reviewer 1 in comment 3 where we address a similar point.

We did consult the literature to assess these values. An overview of the literature consulted for the Glacier scenario and a brief explanation of the value selection is given on lines 244 to 248. Estimates in the literature of glacial contribution to streamflow are based on historic melt rates and cryosphere interactions and cannot account for unforeseen nonlinearity under future climate change. Additionally, many studies are for the mainstem of the Heihe River, and while it is the largest contributor of water to the downstream basin, other tributaries do contribute to the total basin water balance and may have smaller or larger fractions of glacial contribution, such as 32% as predicted by Li et al., (2014). Thus, we decided to reduce the glacial fraction by the largest amount we considered possible which is higher than the largest literature estimates for the Heihe River of around 10% (Chen, 2014; He et al., 2008; Yang, 1991) to account for this. We could then conclude that any other reasonable fraction used would result in a smaller impact and use that to guide the conclusions at which we arrived. Again, because the goal of the study was to look at possible directions and magnitudes of trends and not give exact predictions of what the basin will look like in 2050 for example, we considered this value selection the one that would give the most information.

For the permafrost scenario, we used a study by Gao et al. (2018) which presented data for the increase in winter baseflow in the upper Heihe. Gao et al., (2018) state that since there are no (or few) other contributions to streamflow other than groundwater discharge in this period that any increase to winter flows would be directly caused by increases in baseflow from permafrost degradation. By linear interpolation of their data, the increase over a 30-year period of winter baseflow was 50% for an 8% loss in permafrost area. Assuming a similar loss in permafrost area in an additional 30 years, we increased baseflow by another 50%. Even though the impact of permafrost degradation to streamflow can most easily be assessed in winter, the changes to hydraulic conductivity caused by permafrost degradation are permanent. Thus, we assume the baseflow impact applies year-round.

8. For comparison of the observed and modeled flow at the HRB2 gage, it can be seen from Figure 3 that the modeled values are significantly smaller than the observed values. The fit between observed and modeled data should be greatly improved.

The authors are in agreement that the modeled flows (dark blue, figure 3) are significantly higher than the observed flows (red, figure 3). The authors were modeling a natural flow state, while the data we had to calibrate the model to were subject to operations. Operations include reservoir and canal diversions as well as groundwater pumping which are significant in this region. For this reason, we do not expect to be able to match the observed flows. In fact, if we are modeling the natural system correctly, it is quite impossible that we would match the observed data in such a heavily managed system. For this reason, we chose to focus on matching winter baseflows when there are few water management operations occurring and also chose to attempt to "naturalize" the data to better assess our fit. After those adjustments, our streamflow comparisons are more reasonable (note that we also discuss this point in our response to reviewer 1 in comment 5). We state our plan to address comments 8 and 9 together after comment 9 below.

9. The authors use Spearman's rho as the standard to determine correlation. Why not use the Nash–Sutcliffe efficiency coefficient to evaluate the fit between observed and modeled data? The Nash–Sutcliffe efficiency coefficient is a widely accepted standard for this purpose.

To also address the reviewers concerns about using Spearman's Rho instead of Nash-Sutcliffe efficiency, we wanted to be able to assess improvements in our fit from calibration that were not so heavily weighted by missing high magnitude peak flows in Spring and Summer as we expected to not fit those well due to modeling a natural flow state. Spearman's Rho tells us if we are getting directional changes right, with less emphasis on magnitude. Thus, it is a more appropriate metric for us to compare our natural flow state to observations subject to operations.

This all being said, we understand how this seemingly very poor fit would erode confidence in the findings of the study. So, we will add a section highlighting (1) the fact that we are modeling a natural flow state, (2) justifying that decision, (3) better highlight our comparisons to the baseflow and naturalized streamflow which are a more reasonable metric for comparison here, and (4) speaking to why that still allows this study to make the conclusions it does.

Minor comments:
All minor comments will be fixed as specified below in the final manuscript. Extra care will be taken to review all references and assure that they have the correct family and given name both in the manuscript and reference section so that credit is properly attributed.