Dear editor and reviewers,

Here we provide a point-by-point response of the changes we made based on the reviews. Thank you for the valuable feedback. We believe that this greatly benefitted the quality of the manuscript.

Kind regards on behalf of all authors,

Pau Wiersma

Point by point response to Reviewer #1

I have only two main concerns and one specific point for this study:

- *The authors test a widely accepted hypothesis that the physical representation and simulation of hydrological model will be improved if its corresponding parameterization is optimized on a global scale. I am not quite sure that a test of a widely accepted hypothesis is a true innovation (I leave this question to the editor). If the test is done by coupling the global hydrological model and global glacier model physically instead of simply replacing the PCR-GLOBWB 2 runoff by the GloGEM runoff for glacierized areas, the novelty of this study make sense at least from a practical point of view.*

We adjusted the first paragraph of section 3.1 to the following: "Within the context of this study, the term 'coupling' refers to the replacement of the PCR-GLOBWB 2 runoff by the GloGEM runoff for glacierized areas. **We deem this simplification of coupling physically plausible since much of the exchange of water between glaciers and the rest of the catchment occurs at the surface in the form of runoff.** To the best of our knowledge, this **coupling approach** has not been applied before for glacio-hydrological modeling purposes. Several situations can be thought of for which further coupling between a glacier model and a hydrological model could be applied, such as surging glaciers damming upstream rivers (Sevestre and Benn, 2015) or the flow of subglacial groundwater (Vincent et al., 2019), but these are considered irrelevant at the considered scale."

- *In both Abstract and Introduction sections, the authors mentioned that global runoff prediction can be improved through the coupling of GHMs and GGMs. However, only runoff "simulation" was tested in this study rather than "prediction". The authors are suggested showing the results of runoff prediction (not for the calibration/validation periods but for the prediction period) as well.*

We added a sentence at the end of section 3.2: "All model setups are run between the hydrological years 2000-2012. As PCR-GLOBWB 2 is not calibrated, the simulation over these 12 years can be seen as a test for the prediction quality of the coupled model versus the uncoupled model."

A specific point: Paragraph 165 in P7, extra periods.

This has been adjusted.

Point by point response to Reviewer #2

I do have a few but important major comments that relate to both the model coupling and verification of the process representation.

(1) It would be useful to speak more about the coupling approach and its compatibility with BMI (L110-113). In particular, this sentence needs to be expanded on: "Communication with hydrological models is independent of the model language through GRPC4BMI (van den Oord et al., 2019) and BMI (Hutton et al., 2020). Additionally, the ESMValTool (Eyring et al., 2016) implementation in eWaterCycle allows for smooth preprocessing and high compatibility of forcing data."

Firstly, please spell out BMI in its first use. In the US in particular, BMI is rapidly becoming the standard for model coupling. It should be noted as to how the coupling approach here is compliant with the BMI standard or can be adapted to BMI.

If the code is not usable with the BMI standard, please add detail to the text so that a user understands that (at least from what is implied by the above sentence) they can use something other than a Jupyter Notebook (Python, for example) to couple the models - because of the language-independent nature of the coupling used by the different modeling components.

Hopefully, this is the case, as it is certainly an important advancement beyond the potential scientific improvements offered with respect to the process representation.

We made two adjustments to the manuscript to incorporate the reviewer's feedback:

- Swap BMI and GRPC4BMI and spell out BMI in L110, which thus becomes: "Communication with hydrological models is independent of the model language through the Basic Model Interface (BMI) (Hutton et al., 2020) and GRPC4BMI (van den Oord et al., 2019)."
- Adding a new paragraph after L154: "The numerical implementation of the coupling is largely done using standard BMI functionality. As mentioned in section 2.3, the eWaterCycle platform uses BMI for communication with the hydrological models and therefore also allows for requesting and modifying model variables using the *get_value()* and *set_value()* BMI functions. In this case, these functions are used to add the GloGEM glacier runoff to PCR-GLOBWB 2 but other combinations of glacier and hydrological models could be coupled using the same interface. While the adjusted landcover fraction maps need to be created manually, they are passed to the model via the model's configuration file in the BMI *initialize()* function.

(2) In the results and conclusions (and abstract), (see L341 as an example), the statement is made that "The coupled model produces higher runoff across all basins." However, there is not a follow-on statement discussing whether this results in better hydrological modeling of the process in that it matches observations. This type of assessment should be more clearly stated in these sections. Simply producing more runoff through this new model coupling does not necessarily mean the modeling results are better or the process representation is more correct. Using an evidence-based approach with a comparison to observations, and showing how this is an improvement over other modeling approaches is preferred. Otherwise, the manuscript's hydrologic contribution is reduced and more emphasis is placed on the model coupling/software engineering contribution. I would be interested to understand the authors' response to this comment.

We added a paragraph and adjusted the final paragraph in section 5.2 as follows:

“A major limitation of using runoff observations at the basin outlet is that they are not a direct measure of glacier runoff, and therefore we can not fully exclude the possibility that GloGEM overestimates the glacier runoff and simply compensates for other deficits of PCR-GLOBWB 2 at the basin level to reach the higher RRD-scores. While we chose the discharge stations as close to the glacier sink as possible, we excluded in many cases other upstream discharge stations from our analysis. Future studies are encouraged to consider multiple discharge stations per basin to limit this identifiability problem. Nonetheless, several aspects of our study point against the abovementioned possibility. Firstly, since GloGEM has been calibrated and validated with glacier mass balance observations (Gardner et al., 2013) it is unlikely that GloGEM heavily underestimates glacier runoff, at least on a monthly scale. Secondly, an indication that the PCR-GLOBWB 2 underestimation stems from glacierized areas is given by the observation at the Aletsch glacier (see section 2 of the Supplement), where PCR-GLOBWB 2 simulates zero runoff over multiple years. Finally, in section 5.1 we provide evidence that the difference in glacier parameterization between PCR-GLOBWB 2 and GloGEM is responsible for a large part of the difference in runoff.

In conclusion, strongly glacier-influenced basins produce at the same time higher and more significant RRD scores, and we have shown this to be mostly attributable to the difference in glacier representation. The coupling of GloGEM is therefore likely to prevent significant underestimation of glacier runoff in PCR-GLOBWB 2. While in this study the coupling does not lead to better results for weakly glacier-influenced basins, it is probable that the glacier parameterization has in fact improved the resulting runoff in these basins, at least close to the headwaters, but that this is not visible in the results.”

Response to the editor’s final comment

The first reviewer raised a question of novelty in the manuscript, which I believe was answered well by the authors but it would be helpful to add more of the response to the text. I would suggest that the last sentences of Bullet 2, Response to Comment 1 be added somewhere in the introduction so that the text could read, "We hypothesize that this simplification makes the coupling of a glacier model with a hydrological model feasible and straightforward to implement without losing too much physical basis. Such a straightforward implementation of one-way coupling of models could demonstrate that this method can be adopted by other global hydrological models. We further hypothesize that the slight loss in physical basis induced by the coupling method can then be compensated by the gain in glacier representation accuracy of the GGM relative to the GHM."

We added two sentences to the end of the introduction: “To benefit its replicability with other GHMs, we apply a simplified coupling method using standard open source libraries. We expect the gain in glacier representation accuracy of the GGM relative to the GHM to compensate for any loss in physical basis following the simplifications applied in the coupling method.”

Response to external comments by Sarah Hanus

We added several missing basins in Table S1 of the supplementary material. Additionally, we specified in figure S5 that the SWE in the figures includes both accumulated “glacier” and actual snow mass over the spinup period, that no importance should be attributed to the absolute values of SWE and that the SWE values only reflect the glacierized areas of the basins.

Overview of changes

- Introduction

    - Updated the Müller-Schmied et al. reference to refer to the published article instead of the preprint
    - Discussed the expected gain in physical basis
- Models and data
    - Updated the Hut et al. reference to refer to the published article instead of the preprint
    - Swapped the order of the BMI and GRPC4BMI references
- Methods
    - Further discussed the physical basis of the methodology
    - Discussed the nature of the BMI implementation in the coupling
    - Specified the distinction between prediction and simulation in this study
- Discussion
    - Added a paragraph in the section “evaluation against observations” on the identifiability problem and the argumentation on how we believe to have overcome it
- Supplementary material
    - Added missing basins in Table S1
    - Elaborated on the SWE interpretation in figure S5