Answer to comments of Gemma Coxon

The original comments of Referee Gemma Coxon are in black color and indicated by "R:". Replies by the authors ("A") are colored in green. Actions are introduced by "Action:", changes in the manuscript are in italics.

R: This paper describes the global hydrological model WaterGAPv2.2d and evaluates its outputs. Overall, the paper is well written and a comprehensive model description that will be useful to a wide array of environmental scientists. There is a clear description of the model updates from WaterGAP 2.2 and each of the model components is well described. It is great to see all the model outputs being made available and there is some interesting model evaluation. Overall my comments are relatively minor revisions, however, I have a couple of broader comments on calibration and I do think the paper is too long and that section 7and 8 need to be significantly shortened or removed.

A: Thank you for the overall positive feedback and the comments to which we will answer below.

General Comments

R: *Summary Table.* The description of each of the model components is generally very clear and comprehensive. However, it would be useful to have a summary table of parameters, stores and fluxes for each of the models. See for example Table 1 in Brauer et al (2014) (doi:10.5194/gmd-7-2313-2014). This would be useful for a reader to refer to, provide a central place where a lot of the acronyms in the paper can be found (there are a lot of acronyms!) and be very helpful as a summary when comparing models in model intercomparisons. If it cannot fit in the main paper, this would still be useful in an appendix or supplementary info.

A: Thank you for the very good suggestion which we are following.

Action: We added a table with describing the symbols used in the equations and one with explaining the acronyms, both in the Supplement. While compiling the list of equations, we discovered 3 symbols which have been used twice but with different meaning and minor inconsistencies in the equations and units which we have corrected.

R: *Calibration.* There are a lot of correction factors applied to the outputs for each basin before a suitable result (defined here as+/-10% of the long term mean annual flow) is obtained. Figure 4 is interesting but the results are not analysed in section 4.9.3 –where do the authors think the major errors are coming from (input data, model process representation) and how does this vary spatially? Given the significant correction factors for a lot of basins, it warrants a discussion on how appropriate the model outputs are for future conditions when you assume that these correction factors remain stationary over time (when

in reality they could change depending on what errors they are accounting for!) and the fact that you correct to the mean but do not consider extremes (so how appropriate the model results may be for floods/droughts). It would also be useful to make these correction factors available as part of the standard model outputs (if they are not already).

A: Thank you for this suggestion. Indeed, we have not discussed the potential reasons for the usage of those calibration factors as a thorough assessment would require much more assessments (e.g. with other precipitation input data, model representations or calibration setups) which is outside of the scope of the manuscript. We have done such an assessment in Müller Schmied et al., 2014 with a sensitivity study and do not want to repeat it. We also do not want to introduce a lengthy discussion in Sect. 4 which is solely a model description. The dominance of the different error sources is spatially and temporally varying and it is difficult to make concise and general statements.

The discussion regarding the usage of historically-derived calibration parameters for future conditions is touched initially in line 624 and referred to the discussion within Krysanova et al. (2018) which we do not want to repeat in the context of this manuscript. Also, a thorough discussion of the impact of calibrating to the mean with respect to floods / droughts (esp. when considering the calibration status) would be an extra study but this suggestion is well received. The many assessments of WaterGAP in context of the Inter-Sectoral Model Intercomparison Project (ISIMIP) model evaluation shows that the model is for large spatial domains the best performing model also for low / high flows (e.g. Zaherpour et al., 2018, Veldkamp et al, 2018, Krysanova et al., 2020). However, we have not yet analyzed yet whether the well performing extreme flows are a result of the calibration or simply of the model structure. We agree that this is a very interesting floor for a separate study but we do not want to overspeculate in this manuscript.

Besides this, we agree that it is a very good idea to provide the calibration parameters to the public.

Action: 1. We have added to the repository four netcdf files with a) gamma, b) CFA, c) CFS and d) the calibration status. 2. We have extended line 626 with the reference Krysanova et al. (2020) as there an assessment of model performance and credibility of climate change impact was done.

R: *Model Application.* Section 7 and 8 do not add to the paper and both need to be significantly shortened or removed. Currently, the paper is very long and while the material before this is very relevant, the results presented in Section 7 do not provide any new information or significant results to the reader. Section 8 could also be significantly shorter. If these sections were shortened then you could expand a little on the interesting discussion of the future model developments that you discuss in the conclusion and perhaps link this in with broader developments in the GHM community. Furthermore, it would make the paper shorter and more readable for readers.

A: Thank you for the suggestion. We disagree and strongly believe that Section 7 does provide new and relevant information to readers who want to understand the WaterGAP model and what is can be applied for. We think that the GMD manuscript type "Model description paper" should include the presentation of model results beyond those that can be compared to observations or independent data (we do this in Section 6 Model Evaluation). A model description paper should also describe the model output for which no observations are available but which has been the whole purpose of model development as it provides information about the global water situation that cannot be obtained without the model. Therefore, we do not want to remove (or shorten) Section 7, which we kept as concise as possible.

We certainly agree that the paper is very long. To shorten it, we agree with the reviewer to remove Section 8 from the main text. We think it is best to move it to the supplement and refer to it in Section 7; it is worth keeping in the supplement as especially for a new model output user (or code user, see the answer to a later referee comment), it can be beneficial to know in which fields the model has been used.

Action: We kept Sect. 7 as is. We moved Sect. 8 to the supplement and renamed the heading to "*WaterGAP application fields*".

R: *Code and Data Availability.* I understand the difficulties with making the code open source. Do you have a timeline of when it will be made open source and how it will be made open source (for example on a platform like GitHub?).

A: Thank you for your understanding. WaterGAP is developed since the mid 90es and therefore a many people have been involved in model development. Since roughly one year we have collected the written consents of all model developers to grant open access code. However, the formal process of making it open accessible is still in negotiations between the Universities where the main parts of code development has been done. We cannot give an estimate when those issues are being solved but hope this is the case in the next couple of weeks or months. The code itself is already integrated in GitHub and can be made available in short time once the license issues are clarified. Nevertheless, this does not necessarily mean that the model can be executed immediately by an external user. This would need a number of input data sets, configurations, user manuals and a strategy of how the model development might be shared within a community. To be able to provide that information and to setup a community strategy, but also to rewrite the model code to a modern style, we are currently seeking for funding opportunities which is not trivial especially as the code is not yet open source.

Action: None required.

R: *Appendices and Supplementary Information.* It would be worth thinking about whether some of your appendix materials (particularly Appendix D) may be better placed in supplementary material rather than the appendix of the paper.

A: Thank you for the suggestion to reduce the paper length.

Action: We moved Appendix D with its seven figures to the supplementary material.

Minor Comments

R: Abstract L9. I would replace 'can be done' with 'can be achieved'

A: Thanks.

Action: replaced as suggested

R: Introduction L11. This opening sentence is too long –it needs rewriting.

A: Thanks, also pointed out by Referee #1.

Action: We modified the first sentence to "*A globalized world is characterized by large flows of virtual water among river basins (Hoff et al., 2014) and by international responsibilities for the sustainable development of the Earth System and its inhabitants. The foundation of a sustainable management of water, and more broadly the Earth system, are quantitative estimates of water flows and storages as well as of water demand by humans and freshwater biota on all continents of the Earth (Vörösmarty et al., 2015).*"

R: Introduction L28. What do you mean by 'proper simulation'?

A: Proper in the sense of "sufficient" or "reasonable" in terms of high simulation quality.

Action: We modified "proper" to "*high performing*".

R: Introduction L43. It would be good to have some specific references here of where this variant of the model has been used.

A: Thanks, we intended to distinguish the model version family 2 (operating at 0.5 deg resolution) and 3 (operating at 5 arc min) but this can be misunderstood.

Action: We revised the beginning of the section to "*Water – Global Assessment and Prognosis (WaterGAP), which has been developed since 1996, is one of the pioneers in this field. WaterGAP as described here operates with a spatial resolution of 0.5° x 0.5° and is called the model family WaterGAP 2. Key model versions are WaterGAP 2.1d (Alcamo et al. (2003), Döll et al. (2003), Kaspar (2004)), 2.1e (Schulze & Döll 2004), 2.1f (Hunger & Döll (2008), Döll & Fiedler (2008)), 2.1g (Döll et al., 2008), 2.1h (Döll et al., 2012), 2.2 (Müller Schmied et al. (2014)), 2.2a (Döll et al., 2014)), 2.2(ISIMIP2a) (Müller Schmied et al., 2016), 2.2b (Müller Schmied (2017), Döll et al. (2020)), 2.2c (description submitted to this journal) and 2.2d (this manuscript). In addition, a model family with 5'x 5' is named WaterGAP 3 (Eisner 2015). While the model family 3 has similar algorithms than the model family 2, this paper only refers to the recent model version WaterGAP 2.2d.*"

R: Section 3.2.2 L180. Can you provide a reference or website for the Environmental Data Explorer?

A: Thank you. Actually this was a reference to a website which is listed in the references but due to missing entries in the bib.tex file, the year disappeared in the text. We apologize for any inconvenience this might have raised.

Action: We modified this and any other references to websites where appropriate. The sentence reads now as follows: "*Additionally, population numbers beyond 2005 as well as information on the ratio of rural to urban population of each grid cell come from UNEP (2015).*"

R: Section 4.6 L415. Can you add the specific version of the GRanD database you are using?

A: This is sadly not possible as it was a preliminary and unpublished version of the one published in Lehner et al. (2011).

Action: none

R: Section 4.6 L416. I think 'Sect. E' should be 'Appendix E'?

A: Thank you, good observation!

Action: modified as suggested.

R: Section 4.9.1 L621. "to avoid that average water resources are misrepresented" –this isn't clear as written, can you be more specific?

A: Good point, the sentence is indeed not very specific.

Action: We modified the beginning of this section to "*The main purpose of WaterGAP is to quantify water resources and water stress for both historical time periods and scenarios of the future. Not only due to very uncertain global climate input data, uncalibrated global hydrological models may compute very biased runoff and streamflow values (e.g. Haddeland et al. 2011). To reduce the bias and simulate at least mean streamflow and thus renewable water resources with a reasonable reliability, WGHM has been calibrated to match observed long-term average annual streamflow at gauging stations on all continents (Döll et al., 2003, Kaspar, 2004). Calibration is required…*"

R: Section 4.9.1 L634. One of the key outcomes from Coxon et al (2015) was that the discharge uncertainty varied significantly between gauging stations and over the flow range. It may be worth adding a sentence somewhere in the paper stating that you recognise that the discharge uncertainty is unlikely to be stationary in space and time but there are no further data to better constrain the uncertainties at these gauging stations so a representative value of +/-10% is used.

A: Thank you for this very good advice.

Action: We added the following sentence to line 635: "*It is noteworthy that the discharge uncertainty (approximated here with +/- 10%) is unlikely to be stationary in space and time (Coxon et al, 2015) but there are no further data available to better constrain the specific uncertainty of each gauging station.*"

R: Section 6.4.1. Missing ‘h’ from ‘withdrawals’ in title

A: Thank you!

Action: typo solved

R: Figure 5. I wasn’t sure whether the NSE and logNSE values presented in Figure 5 were calculated based on all the monthly country data or were a median value for each country? It is notclear how it was calculated for these variables.

A: Thank you for pointing out this potential source of misunderstanding. NSE (and log NSE) in Figure 5 was calculated using each single data point (yearly) of FAO AQUASTAT and the corresponding simulated value.

Action: We added the sentence “*The evaluation metrics (Sect. 6.3.1) are calculated using each single data point of AQUASTAT, without any temporal aggregation by country.*” at the end of Sect. 6.2.1.

R: Section 6.4.2 L785. “is rather satisfying” – I would remove this and just present the results

A: Good point.

Action: We modified the sentence to: “*The performance of WaterGAP 2.2d in terms of monthly streamflow time series at 1319 gauging station (Fig. 8) reaches a median NSE(KGE) of 0.52 (0.61).*”

R: Section 6.5. Can you attribute some of these improvements in model performance to specific changes made to the model?

A: Without an extensive experiment setup which would include simulations (and calibrations) of each modification in a stepwise manner, a scientifically based answer is not possible, and we decided not to speculate.

Action: none

R: Conclusions. An additional development here could be improving the lakes/wetlands and reservoir regulation as you noted this being a limiting factor to good model performance in North America/Canada?

A: Thank you for the suggestion. Indeed, there are many areas for future development such as the improvements mentioned. However, all of these potential improvements require funding to do research and implementation, hence we have listed only those lines that are currently under development (see also Line 1018).

Action: none

References

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., & Siebert, S. (2003). Development and testing of the WaterGAP 2 global model of water use and availability.

Hydrological Sciences Journal, *48*(3), 317–337. https://doi.org/10.1623/hysj.48.3.317.45290

Döll, P., & Fiedler, K. (2008). Global-scale modeling of groundwater recharge. *Hydrology and Earth System Sciences*, *12*(3), 863–885. https://doi.org/10.5194/hess-12-863-2008

Döll, P., Fiedler, K., & Zhang, J. (2009). Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrology and Earth System Sciences*, *13*(12), 2413–2432. https://doi.org/10.5194/hess-13-2413-2009

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G., & Scanlon, B. R. (2012). Impact of water withdrawals from groundwater and surface water on continental water storage variations. *Journal of Geodynamics*, *59–60*, 143–156. https://doi.org/10.1016/j.jog.2011.05.001

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology*, *270*(1–2), 105–134. https://doi.org/10.1016/S0022-1694(02)00283-4

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., & Eicker, A. (2014). Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, *50*(7), 5698–5720. https://doi.org/10.1002/2014WR015595

Eisner, S. (2015). *Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients* [PhD thesis]. Kassel University.

Hunger, M., & Döll, P. (2008). Value of river discharge data for global-scale hydrological modeling. *Hydrology and Earth System Sciences*, *12*(3), 841–861. https://doi.org/10.5194/hess-12-841-2008

Krysanova, V., Zaherpour, J., Didovets, I., Gosling, S. N., Gerten, D., Hanasaki, N., Müller Schmied, H., Pokhrel, Y., Satoh, Y., Tang, Q., & Wada, Y. (2020). How evaluation of global hydrological models can help to improve credibility of river discharge projections under climate change. *Climatic Change*. https://doi.org/10.1007/s10584-020-02840-0

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., & Wisser, D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502. https://doi.org/10.1890/100125

Müller Schmied, H. (2017). *Evaluation, modification and application of a global hydrological model* [Goethe-University Frankfurt]. http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/year/2017/docId/44073

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J., & Döll, P. (2016). Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. *Hydrology and Earth System Sciences*, *20*(7), 2877–2898. https://doi.org/10.5194/hess-20-2877-2016

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

Schulze, K., & Döll, P. (2004). Neue Ansätze zur Modellierung von Schneeakkumulation und -schmelze im globalen Wassermodell WaterGAP. In R. Ludwig, D. Reichert, & W. Mauser (Eds.), *Tagungsband zum 7. Workshop zur großskaligen Modellierung in der Hydrologie* (Issue November 2003, pp. 145–154). Kassel University Press. https://www.upress.uni-kassel.de/katalog/abstract_en.php?978-3-89958-072-3

UNEP. (2015). The Environmental Data Explorer, as compiled from United Nations Population Division. http:/ede.grid.unep.ch

Veldkamp, T. I. E., Zhao, F., Ward, P. J., De Moel, H., Aerts, J. C. J. H., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., & Wada, Y. (2018). Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: A multi-model validation study. *Environmental Research Letters*, *13*(5), 055008. https://doi.org/10.1088/1748-9326/aab96f

Zaherpour, J., Gosling, S. N., Mount, N., Schmied, H. M., Veldkamp, T. I. E., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanasaki, N., Kim, H., Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Schewe, J., & Wada, Y. (2018). Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environmental Research Letters*, *13*(6), 065015. https://doi.org/10.1088/1748-9326/aac547