Answer to comments of Anonymous Referee #1

The original comments of Referee #1 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.

General comments:

R: The authors provide a detailed model description of the latest WaterGAP global hydrological model and specification and validation of its standard output data. The model description part covers the entire model but puts extra weights on the improvements and advances since Müller-Schmied et al. (2014) which reported the last model updates. The standard output data part concisely compiles related information for the potential users.

WaterGAP is a great model that has lead the field of global water resources research for two decades. I believe this paper provides a foundation and a benchmark to the research community.

It is noteworthy that the model description includes the detailed procedure of hydrological parameter tuning. One of the distinct characteristics of WaterGAP is that the developers have conducted painstaking manual hydrological parameter tuning at more than one thousand basins. This feature brings a distinct performance compared to other global hydrological models (mostly untuned), but the procedure was virtually unseeable for non-developers since available descriptions were quite old (Döll et al., 2003; Hunger and Döll, 2008). The clear and detailed description in this paper will be helpful for understanding the outputs of WaterGAP, in particular, those who are interested in intercomparing models.

The disclosure of standard output should be highly appreciated. Although numerous model intercomparison projects have been conducted (e.g. WaterMIP, ISIMIP/globalwater), the performance of models tends to fall short of that of under the original (model-optimum) condition. The broad community will be benefited from the provided data.

The manuscript is certainly long, but well structured and written. The major contents are, as noted earlier, the description of the latest model and outputs which is essentially a summary of past six years of peer reviewed papers. Due to the nature of this manuscript, I haven't rigorously examined the methods and results one by one. Rather I have read and commented this manuscript from the viewpoint of a learner of the model and a user of the outputs. Hope the specific comments below are useful for further revision.

A: Thank you for the overall very positive comments and encouragement. Regarding the comment on the parameter tuning please allow us a clarification. The parameter tuning is not done manually but using a – specifically developed for WaterGAP - automatic calibration framework on multiple nodes of a computation cluster.

Specific comments

R: Line 11: This sentence is too long. Better to split into two or three.

A: We agree.

Action: We modified the introduction sentence as follows: "*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: Line 20 'Environmental Performance Index': This term needs a reference.

A: Thanks. Indeed the EPI in parentheses is a citation (and referred to in the literature) but we agree that it is ambiguous as it can be understood as simple abbreviation. But while cross-checking this indicator (https://epi.yale.edu/), it came up that the EPI methodology does not consider the "water scarcity" indicator (where global models are being used) after the 2010 version of the indicator.

Action: We deleted this indicator from the list.

R: Line 86: 'Hyungjun' reads 'Kim'.

A: Thanks for the hint and sorry for this mistake.

Action: we corrected the citation.

R: Line 137 'Cropping patterns and growing periods are generated for every year': A bit confusing. The authors wrote that growing period is fixed at 150 days. What does this part mean?

A: The principle is described in Section 3.1.1. The individual growing period in a given grid cell have a **fixed length** of 150-days, whereas the **start (day of the year)** is not fixed, but depends on climate data of each grid cell: 30-year-average monthly temperature, precipitation (30-year-average monthly sums and number of rain days) and potential evapotranspiration (30-year-average monthly shortwave radiation and fixed mask of arid/humid grid cells, for Priestley-Taylor approach, see Eq. (7) in line 318).

As described in line 120, using the ranking criteria explained in Döll and Siebert (2002), "The most highly ranked 150-day period(s) is/are defined as growing season(s)"

In the current model version, for the calculation of cropping patterns and growing periods, each year the climate averages are calculated externally to WaterGAP as running means with the current year as the central year (exception: at the beginning or ending part/15 years of the time series).

This is consistent with the assumption that farmers' choice of cropping patterns and growing periods are rather based on long-term experience than on short-term weather.

Action: We have added a reference to section 3.1.1 in line 137.

R: Line 138 'the respective 30-year climate averages': A bit hard to read. Which climate variables are year-specific and which are 30-year mean?

A: As mentioned in the previous comment, for the calculation of cropping patterns and growing periods 30-year averages are used (monthly variables: temperature, precipitation, number of rain days, shortwave radiation).

Action: As mentioned above, we added a reference to section 3.1.1 in line 137.

R: Line 265 'Increases in soil water storage in irrigated areas are not taken into account': I am wondering how evapotranspiration from irrigated area is estimated in this model. I guess abstracted water for irrigation is directly added to evapotranspiration of a gridcell, but this should be clearly elaborated.

A: The model calculates soil evapotranspiration Es (Eq. 17), sublimation Esn (Eq. 14), evaporation from canopy Ec (Eq. 6), evaporation from water bodies (Eq. 22) which in sum can be seen as actual evapotranspiration. For the output described in Table 2 as actual evapotranspiration Ea we add the actual consumptive water use WCa (which is the sum of NAs and NAg (Section 3.3)) to consider the "lost" water to the atmosphere as additional part of evapotranspiration. Hence, consumed water (for all water uses) are included in the model output provided as Ea. We, however agree that this note on Table 2 might not be prominent enough.

Action: We added a sentence after the mentioned line 265. "*To consider anthropogenic consumptive water use in the output variable of actual evapotranspiration Ea (Table 2), we sum up all evapo(transpi)ration components and actual consumptive water use WCa (see note 5 in Table 2).*"

R: Line 318 Equation 7: Seems LAI was used only for canopy storage calculation. Is this really the case? I wish to see the list of variables which are directly affected by the daily dynamics of LAI. This point must be important to understand/interpret the outputs of WaterGAP model.

A: The referee is right. The LAI model is only effective for canopy storage calculation (Eq. 3-6). Hence, it is Ec and Sc directly affected. LAI development in terms of plant growth for irrigation is done specifically in the submodel of GIM (Sect. 3.1). Transpiration is not simulated separately, only jointly with the soil evaporation (Sect. 4.4).

Action: We added the following two sentences before Eq. 6 for clarification: "*It is noteworthy that in WaterGAP L only affects the calculation of the canopy water balance. L is not taken into account in computing consumptive water use of irrigated crops (Sect. 3.1) and evapotranspiration from land (Sect. 4.4).*"

R: Line 611 'Unsatisfied water use is added to NAs of the next day until the end of the calendar year': It sounds that this treatment can result in a quite unnatural hydrograph. For example, for the rivers affected by the monsoon system, the increase in wet season's discharge must be substantially delayed, because the initial increase in runoff is used for the 'repayment of water loan' accumulated in the preceding dry season. Please consider adding a note on the consequences of this assumption/treatment which would be helpful for the readers. Finally, as a hydrologist living in an Asian country, I need to write here that this assumption/treatment is quite odd. The drastic seasonal change in water availability is the heart of the water scarcity problem in our region, which seems largely (if not completely) unaccounted by this model (see discussion in Hanasaki etal. 2008, HESS).

A: Thank you for rising this important issues. With the delayed option we are aiming at compensating that WaterGAP likely underestimates storage of water e.g. by small tanks and dams, and because of the generic reservoir operation scheme. The delayed satisfaction scheme may, however, overestimate satisfaction of surface water demand in particular in highly seasonal flow regimes. With regards to the effect of adding unsatisfied use to NAs of the next day, we have done an additional simulation with this feature disabled. We took the Nash-Sutcliffe-Efficiency as indicator for substantial deviations of the hydrograph. From the 1319 river basins assessed, there are only 20 river basins where this indicator deviates by more than +-0.1 (for 3 stations NSE improved by more than 0.1; the median NSE slightly decreases from 0.5226 in 2.2d to 0.5225 in the variant without delayed satisfaction scheme) in the two model variants, all outside of monsoon regions. For monsoon regions (as the Yangtze river, see Fig. 1 below) the effect is – even though there are large potential NAs values calculated in this basin - not visible in the hydrograph as the seasonal change in the hydrograph is much stronger than the effect of delayed satisfaction of NAs. However, in non-monsoon-regions (Fig. 2 & 3), there are certain effects visible, especial in (or better after) dry phases. Furthermore, we have assessed the per cent satisfaction of actual NAs to potential NAs with and without delayed satisfaction of NAs. With the delayed satisfaction of potential NAs as computed in GWSWUSE, 92.5% of global potential NAs during 1981-2010 is satisfied, but only 82.2% in case of the alternative option that surface water demand needs to be satisfied by available water on the same day. Fig. 4 below provides information of the spatial distribution of these differences. Overall, switching off the delayed reduces the satisfaction of NAs especially in the dry regions. We believe that especially there, local storage systems are installed which might be represented by the delayed satisfaction of NAs.

Action: We have added the three hydrographs to the supplement, added the following paragraph after line 614 and hope that with this additional text we have covered the concerns of the referee including Hanasaki et al. (2008): "*Delayed satisfaction aims at compensating that WaterGAP likely underestimate storage of water e.g. by small tanks and dams, and because of the generic reservoir operation scheme. Without delayed satisfaction, less than 50% of potential NAs could be satisfied in many semi-arid regions (Fig. S8). The delayed satisfaction scheme may overestimate satisfaction of surface water demand in particular in highly seasonal flow regimes. However, this effect is hardly visible in the hydrograph of the monsoonal Yangtze river (Fig. S9) but more visible in semi-arid regions (Fig. S10, S11). With delayed satisfaction of potential NAs, 92.5% of global potential NAs during 1981-2010 is satisfied, but only 82.2% in case of the alternative option that surface water demand needs to be satisfied by available surface water on the same day.*"

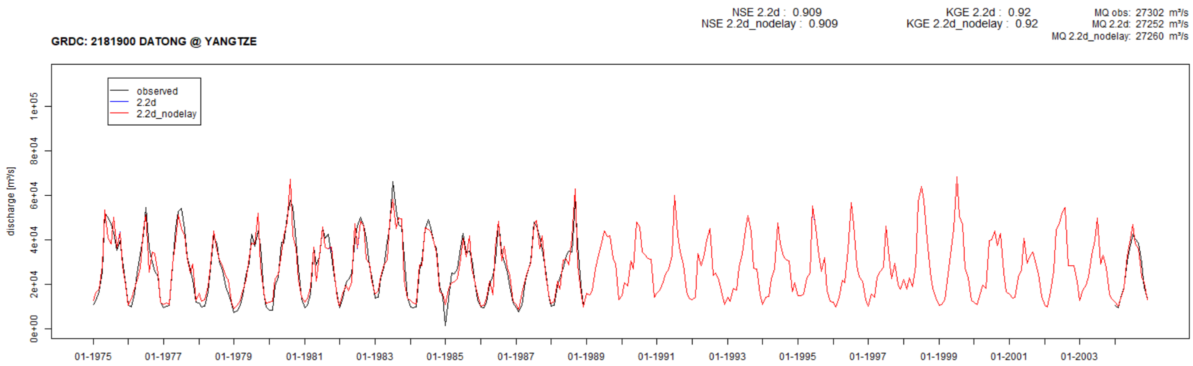

Figure 1: Hydrograph of Yangtze river at Datong station with standard 2.2d and a variant without delayed satisfaction of water use as well as with the GRDC data included.

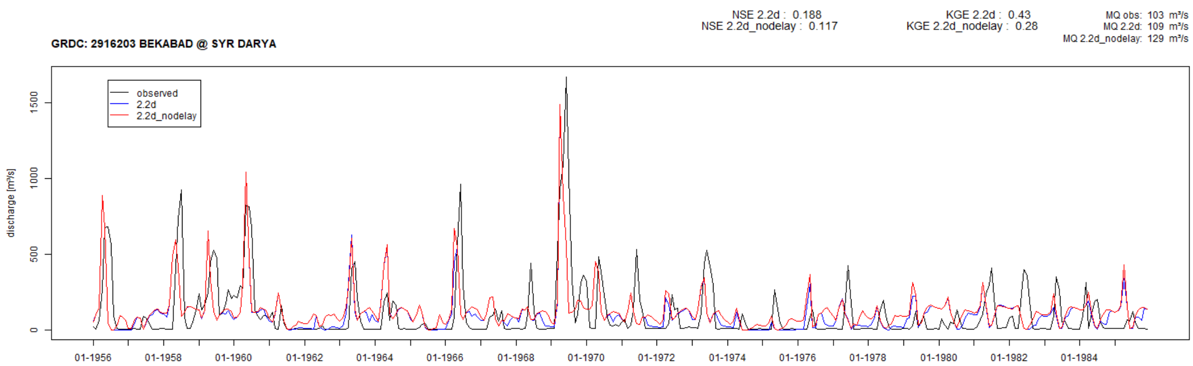

Figure 2 Hydrograph of Syr Darya river at Bekabad station with standard 2.2d and a variant without delayed satisfaction of water use as well as with the GRDC data included.

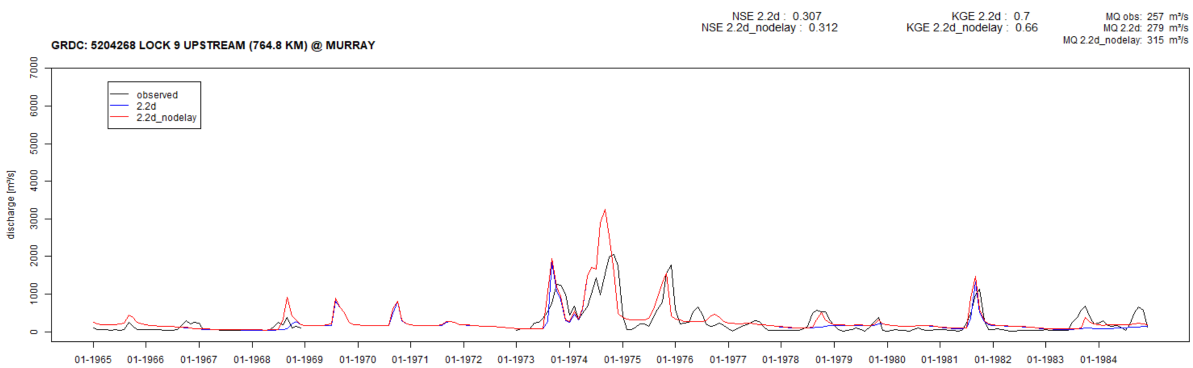

Figure 3 Hydrograph of Murray river at Lock 9 station with standard 2.2d and a variant without delayed satisfaction of water use as well as with the GRDC data included.

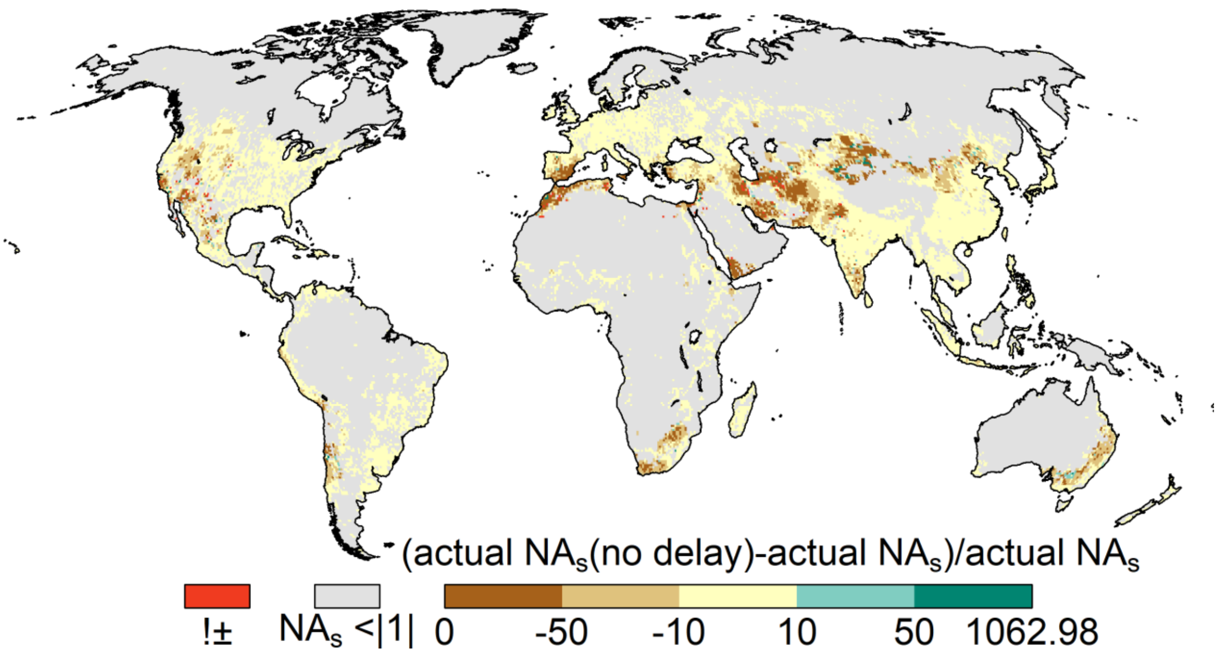

Figure 4: The spatial impact of delayed satisfaction of NAs, showing a lower satisfaction especially in dry regions compared to the standard variant. Values are expressed in percent.

R: Line 629 'areal correction factor (CFA)': Why is this term called 'areal'?

A: In contrast of the station correction factor CFS which is effective only in the outflow cell of the calibration basin, the areal correction factor (CFA) is spatially distributed in the river basin. In line 641 we have briefly explained the calculation of CFA.

Action: We referred to the calibration status CS3 and CS4 and to Hunger & Döll 2008 in line 629.

R: Line 647 'For global water balance assessment the mass balance is kept by the actual evapotranspiration component': Does this mean that the actual evapotranspiration is simply calculated by P – Q?

A: With applying CFS we destroy the water balance in the grid cell where this factor is applied. For example, a streamflow Qsim, of let's assume 1000 m3/s, needs to be multiplied by a factor of 0.5 (Qmod 500 m3/s) to match to the observed streamflow, the water balance lacks of 500 m3/s. One may choose to add this amount to evapotranspiration. But this might lead to an overestimation of evapotranspiration if the reason for the CFS is that precipitation is overestimated by the climate forcing. If, with a CFS > 1; evapotranspiration would need to be reduced and could get therefore to negative values, the water balance stays unclosed. However, for water balance calculations as in Table 6, we have considered the CFS effect by adding (removing) the CFS-adapted streamflow to (from) the actual evapotranspiration. As the second referee pointed out that it would be of benefit to provide the CFS (and gamma and CFA) values for clarity, we follow this advice.

Action: We added gamma, CFA and CFS and calibration status to the model output at Pangaea and added the sentence after line 675: "*Additionally, the calibration factors γ, CFA,* *CFS and the calibration status (Sect. 4.9) are provided.*" and we modified the sentence in line 647 by adding "*by the amount CFS modified streamflow*".

R: Line 779 "NSE and logarithmic NSE": NSE is usually calculated between two time-series at a single location (e.g monthly simulated and observed discharge). How NSE in Figure 5 was calculated? Seems nation-wise NSE was calculated using five-year interval time series (i.e. the typical interval of FAO AQUASTAT is five-year), then averaged globally, but is this really the case?

A: NSE (and log NSE) in Figure 5 was calculated using each single data point of FAO AQUASTAT and the corresponding simulated value. We want to show the general skill of the water use assessment. The logarithmic NSE is used to elaborate differences in small numbers.

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: Line 785 "However, NSE Values below 0 for 259 stations show the complete failure of WaterGAP2.2d to simulate streamflow dynamics in one fifth of the evaluated basins": I feel that this sentence is a bit unclear. What I understood is that monthly and annual variations were not properly reproduced for these 259 stations, although the simulated mean annual discharge agrees well with that of observation due to the calibration.

A: Thank you for this issue with wording and suggestion to revise.

Action: We modified this sentence to: "*However, NSE values below 0 for 259 stations show that WaterGAP2.2d cannot reproduce monthly and annual streamflow dynamics in one fifth of the evaluated basins, although the simulated mean annual streamflow fits to the observations due to the calibration.*"

R: Line 775 "reasonable quality": I don't know what this phrase exactly indicates. The log-log scatter plot (Figure 5) is not very helpful for judging model performance. At least some additional notes are needed for the results of industrial sector (Figure 5e) which indicates frequent occurrence of two orders of magnitude overestimation between simulation and observation.

A: We refer with this phrase to the NSE and logNSE values which indicates quite good values (between 0.67 and 0.92, Fig. 5 d-f). The log-log plot was chosen to avoid distortion due to few high values. The alternative would be to show no-log axes as shown in Figure 5 below which does in our perspective not allow a better assessment. Nevertheless, we see the value of adding this Figure to the supplement. With respect to the mismatch of WaterGAP estimates and FAO AQUASTAT for some data points of the industrial sector, we provide additional notes as suggested by the referee. In general, a mismatch between WaterGAP outputs and data from FAO AQUASTAT can occur through the use of different sources because WaterGAP does not build on AQUASTAT data rather on national statistics (Flörke et al, 2013).

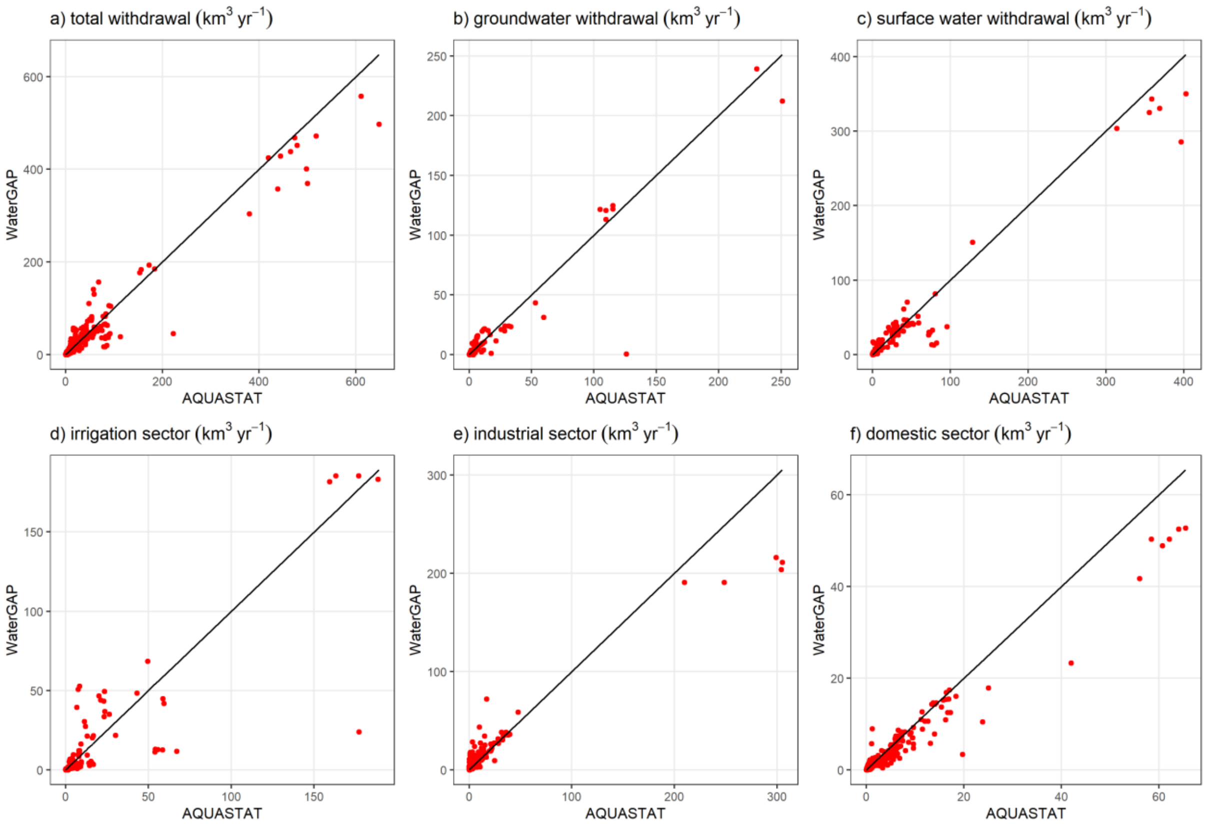

Figure 5: Same as Figure 5 of the manuscript but not with log-log axes.

Action: 1) We added Fig. 5 to the supplement and referred to it in the text.

2) We added additional text after Line 781 to further explain the model performance with regard to industrial water uses. "*In terms of overestimated values, values for India and Germany dominate the differences in the time intervals 2008-2012 and 2013-2016, respectively. Water withdrawals of 56 km³ for the industry sector (including thermoelectric) was assessed by India's National Commission on Integrated Water Resources Development for 2010 (Bhat, 2014). Here AQUASTAT reports 17 km³ yr⁻¹ and WaterGAP simulates 72 km³ yr⁻¹. In case of Germany, AQUASTATs reports only the water use of manufacturing sector but omits the water abstractions of cooling water for thermal electricity production that is included in the WaterGAP results. (i). (ii) The underestimation of industrial water uses >200 km³ yr⁻¹ (Fig. S12e) is particularly biased by the reported numbers from the US statistics. While AQUASTAT data includes both freshwater and saline water abstractions from manufacturing, thermoelectric and mining, WaterGAP only accounts for the freshwater part of the manufacturing and thermoelectric abstractions.*"

R: Line 938-945 "In case of negative NAs…": Highly technical and hard to read. It would be helpful for readers if the authors add links to the directly relevant model description parts (e.g. subsection or equation).

A: Thank you for pointing out that the description should be better phrased.

Action: We modified the section starting in line 936 to: "*As noted in Sect. 4.8, the actual net abstractions can differ from its potential values. The ratio of actual to potential net surface water abstractions NAs (Fig. 15c) shows a heterogeneous pattern, with adjacent grid cells with values below 0.9 and above 1.1. This is explained by the option to satisfy water demand from a neighboring grid cell. In case of negative NAs, potential and actual values are always the same as it is assumed in the model that NAg can always be fulfilled so that return flows to surface water are not changed. There are only a few longer river stretches where actual NAs is smaller than the potential value.*

Actual NAg is equal to potential NAg except in a few grid cells where potential NAs cannot be fulfilled and there is irrigation with surface water (Fig. 15d). In these cells, return flows to groundwater decrease and actual values of NAg increase compared to their potential values. For example, in case of a positive (negative) potential NAg, a ratio of 1.1 (0.9) means that the difference between actual and potential NAg is 10% of the absolute value of potential NAg. In most grid cells, actual NAg is equal to the potential value."

R: Line 949 Table 6: What does negative values for 'actual net abstraction from groundwater' indicate? Does it mean that groundwater recharge has been increased by humanities? I am quite confused because Table 7 indicates that the groundwater is being depleted globally. Similarly, add some extra notes for the negative values for 'change of total water storage' which look constantly increasing by time. What are the key reasons for this?

A: Thank you for the good question. We feel that the description in section 7.3.2 (lines 956 to 967) are not carefully expressed, hence we have rewritten this section, and added an explanation for positive and negative values of NAs and NAg.

Action: 1) In line 936, we added an explanation for positive and negative values of NAs and NAg. "*Positive values of NAs and NAg indicate that human water use results in a net subtraction of water from surface water bodies and groundwater while negative values indicate a man-made addition of water to these water storage compartments.*"

2) We have revised L 958 to L 962 by "*The negative value of actual net abstraction from groundwater in Table 6 indicates that globally aggregated, the groundwater compartment is recharged by return flows from irrigation with surface water (addition of the positive and negative values of NAg in Fig. 5b). A globally averaged anthropogenic increase in groundwater recharge is consistent with a decrease of groundwater storage that is mainly caused the net groundwater abstractions. The global groundwater storage, however, has decreased (Table 7) mainly due to groundwater depletion in those grid cells where (positive) NAg is higher than groundwater recharge (Döll et al. 2014). The anthropogenic net recharge of groundwater in the grid cells with negative NAg in Fig. 5b does not lead to a substantial increase in groundwater storage but mainly increases groundwater discharge to surface water bodies. The decreasing trend of total water storage is dominated by increasing water storage losses that were balanced in earlier periods by increased water storage in newly*

constructed reservoirs while dam construction became less during the last three decades (Table 7, Cáceres et al., 2020)."

References

Bhat, T. A. (2014). An analysis of demand and supply of water in India. Journal of Environment and Earth Science, 4(11), 67–72.

Cáceres, D., Marzeion, B., Malles, J. H., Gutknecht, B. D., Schmied, H. M., & Döll, P. (2020). Assessing global water mass transfers from continents to oceans over the period 1948–2016. Hydrology and Earth System Sciences, 24(10). https://doi.org/10.5194/hess-24-4831-2020

Hunger, M. and Döll, P.: Value of river discharge data for global-scale hydrological modeling, Hydrology and Earth System Sciences, 12, 1305 841–861, https://doi.org/10.5194/hess-12-841-2008, 2008.