**Reply to Referee #1**

Dear Referee,

We are grateful for your time and constructive comments and suggestions on our manuscript. The provided comments have contributed substantially to improving the paper. Below, we provide a point-by-point response (in blue) to the comments (in black) and explain how we will address each of the reviewers' comments in the revised manuscript.

Sincerely,

Nariman Mahmoodi, Jens Kiesel, Paul D. Wagner, Nicola Fohrer

**GENERAL COMMENTS:**

The present paper is on "Sustainable use of water resources…". The topic/paper is interesting but needs a number of modifications before it can be accepted.

"Sustainable/ sustainability" is mentioned several times in the text (and the title). In my understanding sustainable water resources management is not simply supplying an ever growing demand (that's a really old, outdated approach from the last century) but an important part is water demand management. In the paper there is nothing about water demand management, but in one of their scenarios the authors simply interpolate past/observed water use/demand into the future.

We agree that "sustainability of water resources" is a complex field and requires proper definition. In this study, we consider water related sustainability within two different aspects: groundwater sustainability (groundwater recharge vs groundwater withdrawals) and surface water sustainability (the range of variability approach). Although we did comprehensive analyses on these two sustainability aspects under different water use systems, water consumption-, and climate change scenarios (sections: 3.1 Groundwater sustainability and 3.2 Hydrologic alteration), we acknowledge that a clear definition is lacking. We will change the subtitle of section 3.2 to "streamflow sustainability" (as suggested in the specific comments) and add a clear definition of sustainability in the revised version.

Overall, the wording (or definitions) needs to be reconsidered. The authors define"sustainable/ sustainability" as state where "GWR>GWW"; i.e. "groundwater recharge" is higher than "groundwater withdrawal". A state where GWW/GWR ratio is higher than "1" is considered "unsustainable". The authors consider GWR only; there is no information (to my understanding) on the groundwater level or volume. The authors assume that the whole water demand can be withdrawn, e.g. stating that "The rate of GWW to GWR is greater than 2..." (p. 7, lines 195/196). Without information on the groundwater level or volume, how can they say that the water volume demanded can be withdrawn? In my opinion the whole text needs to be changed (I will define demand on groundwater here as "GWD"): the authors can only calculate what volume can be "sustainably" withdrawn from groundwater. In cases where "GWD>GWR" water withdrawals will become unsustainable, i.e. water demand on groundwater resources is larger than groundwater recharge, meaning the groundwater level or volume will decline. The (or one version for) correct wording would be that "Only 50% of the water demand can be sustainably withdrawn…" instead of "The rate of GWW to GWR is greater than 2..." (p. 7, lines 195/196). I think the whole text (and Tables/Figures) needs to be changed in this direction.

Thank you very much for your comment. Indeed, there is no data on groundwater level or volume. We believe that our definition of sustainability is in agreement with the view of the reviewer: Withdrawing the amount of groundwater recharge would be sustainable —going beyond this amount is unsustainable. As we implemented groundwater withdrawal in the model, we used it for our calculation, but we agree with the reviewer that the proposed way of relating it to percentage of groundwater demand that can be sustainably withdrawn is more precise. Beyond considering 'groundwater sustainability' only, please note that our definition of 'sustainability' includes the IHA and RVA analysis and the impact of the scenarios (including groundwater withdrawals) on the ecological flow regime (streamflow sustainability). We will thoroughly revise the manuscript in this regard.

It is not clear if the results (e.g. "GWW/GWR") represented (e.g. Fig. 2) a calculated on an annual basis and then averaged over the entire (30 years) period or if it is calculated for the entire (30 years) period. If it is calculated on an annual basis and averaged afterwards, how is groundwater storage, e.g. in a very wet year GWR may be (much) higher than GWD and increase availability (increased groundwater level or volume) in the next year, considered? Also, to my knowledge, there is no groundwater flow between sub-basins in SWAT. In reality a sub-basin (region) with high groundwater withdrawals, resulting in declining groundwater levels, may receive groundwater inflow from neighbouring sub-basins (regions) without groundwater withdrawals, i.e. high(er) groundwater levels.

In our assessment groundwater recharge was calculated for the entire 30-year periods. We will clarify this in the revised manuscript. Not being able to consider groundwater connection between neighboring sub-basins is not a significant disadvantage, because we want to explicitly show the exact location (sub-basin/regions) of imbalances between GWR and GWD in the study area. However, we discussed the connectivity of groundwater in our analysis when the whole basin was treated as an integrated system and the total groundwater recharge was compared to the total water consumptions/demands. The results are shown in Table 6 and 7.

The paper builds on work of Mahmoodi et al. (2020a) (referenced as Mahmoodi, N., Kiesel, J., Wagner, D.P., and Fohrer, N.: Water use systems and soil and water conservation methods in a hydrological model of an Iranian Wadi system. J. Arid Land, 12, 545–560, https://doi.org/10.1007/s40333-020-0125-3, 2020a.), describing the set-up and calibration/validation of a SWAT model to the Halilrood river basin. In a simplified way the water balance can be given as:

SW(t) = SW(t-1) + PRECIP(t) - SR(t) - ET(t) - PERC(t) – SSF(t)

where SW(t) is the soil water content at time step t, PRECIP is precipitation, SR is surface runoff, ET is evapotranspiration, PERC is percolation, and SSF is subsurface flow (with the last two variables describing groundwater-/subsurface part of the system, this part may be differentiated differently depending on the complexity of the approach/model). The important point is that there are some variables that are observed usually, i.e. PRECIP and river flow (combining SR and SSF), while for other variables seldom or only for very few locations observed values are available, e.g. ET and PERC. Assuming (for simplicity) constant SW and known PRECIP and river flow, potential and actual evapo(-transpi-)ration and PERC are unknown. Depending on the parametrisation of the model potential and actual evapo(-transpi-)ration and PERC can differ strongly. In a model calibrated/validated on stream flow only there is high uncertainty if potential and actual evapo(-transpi-)ration and PERC are simulated correctly - here either observed values, e.g. lysimeter measurements (point information), or from other sources, e.g. LAI or actual evapotranspiration from GLEAMS or MODIS (areal information) should be used to validate the model results. In the paper of Mahmoodi et al. (2020a) such a validation is not carried out and therefore I wonder how the authors can prove the reliability of their SWAT parametrisation. For instance in Table 4 (Selected parameters for calibration in the SWAT model) Mahmoodi et al. (2020a) give "EVRCH Reach evaporation adjustment factor from 0.5 to 0.8", to my understanding this means that the potential evaporation is reduce strongly by 20 to 50% (i.e. more water may percolate, become groundwater recharge)! As given in Mahmoodi et al. (2020a, p. 551):

“WYLD=SURFQ+LATQ+GWQ–TLOSS, (1)

where WYLD is the water yield (mm); SURFQ is the surface runoff (mm); LATQ is the lateral flow contribution to stream (mm); GWQ is the groundwater contribution to stream flow (mm); and TLOSS is the transmission losses (mm), i.e., water loss via transmission through the bed of the channels.” Transmission losses in semi-arid environments/rivers can be very high, introducing another uncertain variable in the simulations.

We agree with the reviewer that the model was -as most hydrologic models- not validated for evapotranspiration. However, going beyond other parameterization approaches it was calibrated using a multi-metric approach that includes the comparison to flow duration curves. The study area is a data scarce region. We used the SWAT model as it has proven its ability to perform well under these conditions (Wagner et al. 2012) and even in ungauged catchments (Srinivasan et al., 2010). The spatially distributed parameters of SWAT derived from land use, soil, and slope maps provide an initial parameterization that has proven its use in various regions worldwide. To address the reviewers concerns we will compare the model to potential and actual ET in the catchment and to a mean recharge rate for Iran. As we focus on the long-term water balance, we believe that these aggregated values for the water balance equation along with the validation of the dynamics of streamflow and the proven ability of SWAT to model water fluxes provide confidence in our model results.

In SWAT2012, a high portion of streamflow volume in low flow periods could be lost due to overestimation of channel evaporation since the default reach evaporation adjustment factor is equal to 1 (Arnold et al., 2013; Nguyen et al., 2018). According to the SWAT manual (Arnold et al., 2013) and other studies (Nguyen et al., 2018; Muleta and Nicklow 2005), the EVRCH default setting of 1 represents the upper threshold of reach evaporation and hence may lead to overestimation of evaporation from reaches. The parameter was introduced specifically for arid regions to reduce reach evaporation (Arnold et al., 2013, see chapter 4, page 112). We will further address this issue in the water balance comparison (see first paragraph of this response).

Wadi bed infiltration has an important impact on streamflow and recharge rate, but depends on local conditions and the quantification of bed infiltration is challenging since many factors i.e., evaporation, sediment and deposit properties, volume and velocity of surface discharge, slope and widths of the reaches are involved (Weather et al., 2008; Neitsch et al., 2011). Over/underestimation of reach evaporation and transmission losses, in combination, can affect the SWAT simulated stream flow. From a good/satisfactory performance of our hydrological model in simulating streamflow and potential evapotranspiration we can only infer that bed infiltration is represented sufficiently well.

<u>References:</u>

Neitsch, S.L., Arnold, J.G., Kiniry, J.R. and Williams, J.R., 2011. Soil and water assessment tool theoretical documentation version 2009. Texas Water Resources Institute.

Arnold, J.G., Kiniry, J.R., Srinivasan, R., Williams, J.R., Haney, E.B., Neitsch, S.L. 2013. SWAT 2012 input/output documentation. Texas Water Resources Institute.

Nguyen, V.T., Dietrich, J., Uniyal, B., Tran, D.A. 2018. Verification and correction of the hydrologic routing in the soil and water assessment tool. Water, 10(10), p.1419.

Wagner, P.D., Fiener, P., Wilken, F., Kumar, S. and Schneider, K., 2012. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. Journal of Hydrology, 464, pp.388-400.

Srinivasan, R., Zhang, X., Arnold, J.G., 2010. SWAT ungauged: hydrological budget and crop yield predictions in the upper Mississippi river basin. Trans. ASABE 53 (5), pp.1533–1546.

Muleta, M.K. and Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. Journal of hydrology, 306(1-4), pp.127-145.

Furthermore, the paper builds on work of Mahmoodi et al. (2020b) (referenced as “Mahmoodi, N., Wagner, D.P., Kiesel, J., and Fohrer, N.: modeling the impact of climate change on streamflow and major hydrological components of an Iranian Wadi system. Water Clim. Change, https://doi.org/10.2166/wcc.2020.098, 2020b.”): "The G-RCM CSIRO-SMHI was chosen since it represented the median model of the major hydrological components (Mahmoodi et al., 2020b)." The authors write that they follow the argumentation of Tebaldi and Knutti (2007) and Thober and Samaniego (2014). However, reading the paper of Tebaldi and Knutti (2007) I only find that they discuss using (weighted) averages of a number of climate models, based on the idea that the performance can be improved by averaging or combining results from multiple models.But this is that first a number of climate models are run and then the results are averaged (in the case of impact models the input from a number of climate models are used and then results of the impact models are averaged). In the study of Thober and Samaniego (2014) a number of (meteorological) indices are used to select regional climate models, i.e. reduce the number of impact model runs. In case such an analysis (to select G-RCM CSIRO-SMHI) was carried out in the present study this analysis and its results need to be described (e.g. in the Suppl. Material), otherwise the results are not replicable (I also need to mention that Thober and Samaniego (2014) reduce the number but they do not propose to use results of only ONE climate model). In the paper of Mahmoodi et al. (2020b) I don*t find any information that “G-RCM CSIRO-SMHI … represented the median model of the major hydrological components”. The use of only one climate scenario restricts the value of the paper – for a more wet or dry (climate) future results could be very different. Such results would be needed to enable more robust decisions on future water resources management, especially as the authors state that effects of climate change on surface/groundwater resources are much more significant than future water demand/use.

We believe that the focus on one climate model is instructive to discuss possible future impacts and the interaction of climate change impacts with the impacts of growing groundwater demand. The procedure of analysing the impacts of all climate models in an ensemble on the target indicator (here: streamflow) and then selecting the median model is common in climate change impact studies (ensemble of opportunity, model democracy). We therefore determined a median model using the following procedure: Among the ensemble climate models, one climate model that is representing the median for most of the simulated hydrological components is selected for our current research. So, indeed, we did “run a number of climate models and then the results are averaged (in the case of impact models the input from a number of climate models are used and then results of the impact models are averaged)”. This description will be added to the revised manuscript and we will correct the references to the cited studies as suggested by the reviewer.

However, we acknowledge that we did not propagate the impact of the different streamflow changes through the RVA analysis for being able to properly present the combinations of scenarios. To nevertheless address the raised concerns of uncertainty, we will additionally analyse and discuss the impacts of the min and max climate models —,which lead to the driest and wettest climate condition based on the simulated hydrologic parameters— from the full ensemble on groundwater recharge and the hydrological regime under different water consumptions and WUS scenarios.

When reading the paper Mahmoodi et al. (2020b) I also found that minimum and maximum elevation for the river basin shown in Figures 1 (Mahmoodi et al. (2020b) and the present study) are different - please explain.

Thank you for bringing this to our attention. This happened due to a wrong reference layer during figure preparation. We will correct this in the revised version.

In the whole text please use "streamflow" instead of "flow" when referring to river flow, otherwise readers could be confused as the paper is on surface water and groundwater. The results given for streamflow (IHA, VRA) are all for the outlet of the basin?

We will use streamflow and harmonize it throughout the manuscript and will explain that IHA and VRA analysis were carried out for the outlet of the basin.

## SPECIFIC COMMENTS:

Minor comments regarding typos, rephrasing of sentences or adding of additional information on the model performance will all be addressed in the revised manuscript.

### Introduction:

Page 1; Lines 23: "…alteration caused by natural or anthropogenic activities…" I am not sure what is "natural … activities"; rethink formulation.

We will revised it.

Page 2; Lines 40/41: "…precipitation... rainfall..."; please use either "precipitation" or "rainfall" in the whole text

We will use "precipitation" and harmonize it throughout the text.

Page 3; Line 68: "Further aggravation will put increasing pressure on the..."; "aggravation" of what (rethink formulation)?

We mean "further aggravation of climate change".

### Materials and methods:

Page 3; Lines 80/81: "…increased over the last years at the outlet of the basin during the past 33 years..."; reformulate "…increased over the last years … during the past 33 years..."

Thank you. We will modify it accordingly.

Page 4; Lines 95/96: "Good performance for modeling daily streamflow values was achieved judged by a multi-metric approach including NSE (0.76 and 0.54)..."; according to Moriasi et al. (2007: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3): 885–900) an NSE of 0.54 is considered as "satisfactory".

Thank you, our statement is misleading. The NSE for the calibration period was 0.76 and for the validation period 0.54 at the daily time step. Since the time step in Moriasi et al. (2007) was monthly, our model performance can be rated as very good for the calibration period with NSE=0.81 and satisfactory for the validation period with NSE=0.6. We will add the performance rating separately for each calibration and validation periods in the revised version.

Page 4; Lines 114 and 116 (also line 459 "Statistics Center of Iran"): on page 114 it is "Statistics Cerner of Iran"; on page 116 it is "Statistical Center of Iran" - The homepage (https://www.amar.org.ir/english) gives the translation/name "Statistical Centre of Iran", please use the official translation/name given on the homepage.

Thank you for pointing this out. We will add the official name given on the homepage.

Page 4; Lines 123/124: "To meet the future domestic, agricultural and industrial water demand, increases in the number of wells and qanats are linearly extrapolated...", so you assume that all wells and qanats have the same (water) yield? Please justify this assumption.

The wells and qanats have specific water yields that have been derived from the available measurements provided by Iran Water & Power Resources Development Company (IWPCO) in 2001 and 2006 for the baseline scenario. There is no data on possible future extractions from wells and qanats and we therefore needed to define the most sensible extraction rates. By extrapolating the current rates linearly with population growth, we believe to have chosen a conservative scenario.

**Results:**

Page 7; Line 199: "...unsustainable subbasins (GWW<GWR)...", unsustainable is "(GWW>GWR)"; however, I suggest to change the wording (see "GENERAL COMMENTS") and this part needs to be rewritten anyway…

We will modify this in revised version. Please see our answer for the GERNAL COMMENTS section.

Page 7; Lines 207/208: "... Among these 56 unsustainable sub-basins, GWW/GWR ratio is higher than 5 in 42 sub-basins....", this means "only 20% of the water demand can be supplied sustainably"?; change the wording (see "GENERAL COMMENTS") - all the results presented need to be changed accordingly.

We will thoroughly revise the manuscript in this regard. Please see our answer for the GERNAL COMMENTS section.

Page 8; Lines 211/212: "...it drops from 385 (106 m3 yr-1) in model setup period to 172 (106 m3 yr-1)...", what is the unit of the first numbers (385; 172)?

The unit is "mill. $m^3$ $yr^{-1}$. The number "$10^6$" is changed to "mill." in the revised version of the manuscript.

Page 8; Lines 217: "3.1" is "Groundwater sustainability"; therefore I suggest to call "3.2" "Streamflow alteration",… give a subtitle that's clearly points to Streamflow/Surface water…

Thank you for the suggestion. We will modify it in the revised version.

**Discussion:**

Page 10; Line 282: "…substantial deficits in discharge during..."; is this "groundwater recharge" or "river discharge"?

It is river discharge.

Page 11; Line 285: "…predicted unsustainability of groundwater could..."; correct "…predicted unsustainability of groundwater use could..."?

Thank you. We will correct it.

Page 11; Line 287: "…could lead to a higher groundwater withdrawal in summer season..."; correct "…could lead to a higher demand on groundwater in summer season..."?

Thank you for the suggestion. We modified this to "could lead to higher demand on groundwater withdrawal in summer season when the surface water does not meet the rising demand.

**Conclusion:**

Needs to be adapted according to "GENERAL COMMENTS".

**Supplementary material:**

- in Figure S1 units (e.g. for "Group 1" and "Group 2" the for x-axis "year" and for the y-axis "m3 s-1") should be given

Thank you. We will add those to the revised version.

- Figure S2 needs a better description to understand what is shown (e.g. the title states "Distribution of annual values…" but shown are monthly values; or is it - the single dots - the 30 monthly values of the 30-years per period, e.g. for January?);

Each single dot representing the value of a specific month e.g. January, in a year. We will change the caption in the revised version.

also here units are missing. The results shown are scattered (not ordered) – is this because the results are given according to their temporal occurrence (i.e. the year)? If yes, why a different x-axis when compared to Fig. S1?

We used a different x-axis compared to Fig. S1 to provide a better overview of the changes in each IHA. The ordered values were already shown in Fig. S1. The shape of the violin plots illustrates the changes explicitly.