Dear Editor and Reviewers,

We have revised the manuscript according to the insightful comments provided by the editor and reviewers. All recommendations have been addressed in the revised manuscript. We would like to thank you for the thorough consideration and critical comments that helped us improve our manuscript.

Editor Comments

Editor Comment: You received two new reviews, one minor revisions but with some critical points and one rejected (with extensive review). Both reviewers agree the manuscript is really too long, the groundwater aspect of Lake Urmia (lake-groundwater relationship) should be discussed much more in depth and the model set-up, assumptions and uncertainty need more attention. This could mean the manuscripts will be even longer. See some suggestions to shorten it below.

Response: *We would like to thank you for your time and suggestions. We have reduced the manuscript size and provided some description of the lake-groundwater relationship, model setup, and uncertainty. We hope the changes have made the manuscript suitable for publication and we look forward to your response.*

Editor Comment: A formal uncertainty analysis is not required for me (as requested by reviewer 2), but some more quantification and discussion as requested by the second reviewer needs to be included.

Response: *We have changed the model set up to consider uncertainty. We have discussed the uncertainty based on different optimal parameter sets which were obtained from the optimization algorithms (GA and NSGA-II).*

Editor Comment: Having gone through the paper I do want to stress I think editorial or only textual changes are not sufficient, I recommend to have a fresh look at the current manuscript and bring in some more focus (choice of results to present and discussion and decide which less important results should be presented in supplement only).

Response: *We have restructured the results and discussion sections and rewriting other parts accordingly. We have put some parts of the result section related to modifying NA in the supplement.*

Also, we have put some parts of the material and methods in the supplement. In the revised version, we have focused more on uncertainty and model setup.

Editor Comment: Could you reduce the introduction, and see whether the method section really needs all info? I could imagine that the well-known WaterGAP model(s) could be summarized in terms of Urmia application only (more, extensive info in supplement). For me, the main aspect (novelty) of your article seems using the comparison of calibrating using RS without and with ground information, so keep that in main text but bring all short references to your input data to the supplement as well.

Response: *We have restructured the manuscript based on your suggestion and put all the detailed information in the supplement. Also, in the revised version, an automatic approach was developed for calibrating WGHM based on evolutionary optimization algorithms. To reduce the length of the manuscript, we have focused on two calibration variants i.e. calibrating using RS with and without ground information as you mentioned in the next comment. We have summarized the descriptions of the WaterGAP model in the manuscript instead we have added some descriptions for the model in the supplement.*

Editor Comment: The results section is very long whereas the discussion is shorter. Typically, one would like to see the reverse. Instead of giving all correction parameters and time-series, you could try to highlight 1-2 results (and present rest in supplement). Furthermore, quite some paragraphs in the results section have discussions. Please re-evaluate whether these parts could not be combined in the discussion.

Response: *We agree that the results section is longer than normal size. We have reduced the size of the results and put some parts in the supplement. Also, we reconsidered the result and discussion sections based on your comment (see the result and discussion sections in the revised version).*

Editor Comment: Note, section 3 consists only of 1 subsection (3.1).

Response: *In the revised version it has two subsections.*

A point-by-point response to the Referee #2

Referee #2: The revised version of the article is much more understandable and better. However, I still think it is too long and there are a lot of assumptions made in the text which can be criticized separately.

But, I think a minor revision would be proper for the final decision considering an additional comment below.

Response: *First of all, we genuinely appreciate your time in reviewing the revised version. We have reduced the manuscript size and revised it according to your comments. We hope that you are satisfied, after the changes have been made.*

Referee #2: I think that you are aware that the role of groundwater on the lake has been discussed by many researchers of this field. Some of these researches e.g. Amiri et al. (2016) think that there is no significant relationship between lake and groundwater in the basin but some like Ashraf et al. (2017), and Vaheddoost and Aksoy (2018) believe that there is strong evidence of a relationship between them. Since your modeling includes analysis of groundwater I think you should compare your results by these articles and try to confirm or reject obtained results by them using similarities and dissimilarities in results.

Response: *We run the models under two scenarios (1) there is a relationship between lake and groundwater and (2) there is no relationship between lake and groundwater. In scenario 1, the seasonality of the groundwater storage was strongly misrepresented. Therefore, we could not calibrate the model when there is a relationship between lake and groundwater. However, WGHM as a hydrological model that does not include a gradient-based groundwater model has some limitations for studying groundwater-lake water flows. We believe that there is an indirect relationship between lake and groundwater i.e. groundwater-river, river-lake as accepted by ULRP (2015). In addition, as you mentioned some studies e.g. Amiri et al. (2016) using isotopic analyses and chemical tracer rejected the significant relationship between lake and groundwater. Also, Danesh-Yazdi and Ataie-Ashtiani (2019) stated that the study by Vaheddoost and Aksoy (2018) is not reliable and there is some doubt in accepting that. Therefore, we have added the following part in the discussion of the revised version.*

"It is worth mentioning that WGHM as a hydrological model that does not include a gradient-based groundwater model has some limitations for studying groundwater-lake water flows. We attempted to calibrate WGHM under the assumption that there are direct water flows between lake and groundwater. Under this assumption, the seasonality of the groundwater storage was strongly misrepresented. Therefore, as accepted by ULRP (2015c), we assumed there is no direct flow between the lake and groundwater. While Vaheddoost and Aksoy (2018) using traditional hydrograph separation methods

claimed that there is a significant relationship between the lake and groundwater, Danesh-Yazdi and Ataie-Ashtiani (2019) rejected their claim. Equally, some studies that applied isotope and chemical tracer analyses (e.g. Amiri et al. 2016) rejected any significant relationship between lake and groundwater. In conclusion, the results of this study support the idea that there are no significant direct interactions between lake and groundwater within the basin."

References:

Danesh-Yazdi, M., and Ataie-Ashtiani, B. (2019). Lake Urmia Crisis and Restoration Plan: Planning without Appropriate Data and Model Is Gambling. Journal of Hydrology, 576, 639-651.

Amiri, V., Nakhaei, M., Lak, R., and Kholghi, M. (2016). Geophysical, isotopic, and hydrogeochemical tools to identify potential impacts on coastal groundwater resources from Urmia hypersaline Lake, NW Iran. Environmental Science and Pollution Research, 23(16), 16738-16760.

Vaheddoost, B., and Aksoy, H. (2018). Interaction of groundwater with Lake Urmia in Iran. Hydrological processes, 32(21), 3283-3295.

ULRP (2015), Urmia Lake - Causes of shrinkage and potential threats. 36 pp (In Persian).

A point-by-point response to the Referee #3:

Referee #3: Overall assessment: The aim of the study essential is to quantify the impact of human activities (mostly in terms of water consumption) vs climatic changes on the Lake Urmia water balance. Even though 3 of the referees provided detailed reviews and pointed out to several shortcomings of the paper mostly on model setup and uncertainty, the authors' revision is minimal and in fact insufficient as many comments are effectively ignored.

Response: *Firstly, we are thankful for your time in reviewing our manuscript. We do not agree with you about the "many comments are effectively ignored" because referee#2 was satisfied with the revised version. In the new revised version, we have addressed the shortcomings of the model setup and uncertainty.*

Referee #3: All reviewers except Chaudhari took issue with the experiment design, particularly the model set up, input data time period, and lack of adequate model calibration and evaluation. Yet, authors have not changed the experiment design, and only added two performance metrics. No uncertainty or sensitivity analysis was conducted whatsoever, which is a common analysis required for any hydrological modeling study. This is even more serious as the manuscript is in fact inconsistent on the issue of uncertainty. While authors discussed the limitations of the work such as parameter uncertainty, no account of the hydrogeology of the lake, assuming constant bathymetry, among others; not only they have not accounted for these uncertainty sources by even a simple uncertainty/sensitivity analysis, they kept pushing that their study is "a holistic and reliable modelling approach", "we are confident that human water use reduced lake inflow that would have occurred without human water use during 2003-2013 by about 41%", and "This study proved that even without human water use Lake Urmia would not have recovered from the significant loss of lake water volume caused by the drought year 2008", among other instances of false overpromises.

Response: *We have developed a new model set up based on an auto-calibrated approach using a genetic algorithm (GA) and Non-dominated sorting genetic algorithm II (NSGA-II). We also have discussed uncertainty arising from the different possible optimal parameters that were obtained from the optimization algorithms. About the limitations, we should mention as you know all modeling studies faced some limitations that are inevitable. About the "holistic modeling", if you review the modeling study on Lake Urmia, all modeling applied a trial and error method based on a single objective calibration. Therefore, when we consider all possible data for our model, we consider it as a holistic one that includes TWSA, inflow, groundwater, lake volume. About the "reliable modeling" we agree with you. We have removed it from the manuscript. We also replaced the "we are confident ..." and "This study proved that" with "we found ..." and "Based on the results can be claimed that…", respectively.*

Referee #3: The manuscript is filled with redundant discussions (either well established in the literature or not relevant to the core research question of the paper), and is written in poor language with several typos. It is a waste of the editor's time and reviewers to resubmit a manuscript that has not been proof-read especially that most reviewers pointed this out. Further, some of the sources are either not peer-reviewed or in Persian. While local knowledge can be useful, the credibility of a non-peer reviewed source is always questionable. In such cases, authors should provide reason and demonstrate clearly instead of just referring to the source.

<u>Response</u>: *We have reformatted the manuscript, results and discussions have made shorter. We have checked the writing of the manuscript. We believe that readability is sufficient now. After acceptance HESS will proof again the manuscript by native speakers. Regarding the Persian references, we agree with you but unfortunately there are no English references, however, all ULRP reports have been reviewed by the scientific committee. Besides, actually we have no other way, for example about the "water withdrawals data" how we can cite it? Or about "irrigated area"? We should either use Persian references or do not provide any sources for these cases. Moreover, referring to ULRP as the main data center for lake Urmia is common in all published papers in international journals.*

Referee #3: In my evaluation, the manuscript is rejected. If authors wish to resubmit the work, they must revise the modeling experiment to sufficiently address the comments by reviewers, and by addressing I specifically mean to change their modeling setup by providing a more robust calibration than a simple and insufficient trial and error, transparently explaining the modeling setup to ensure (somewhat) the reproducibility of the study, and perform some sensitivity or uncertainty analysis. Also, remove all the redundant discussions from the manuscript and focus their discussion on the relevance of the results to the lake given the uncertainties. While most of the comments by reviewers still hold, here I high-lights a few urgent ones.

<u>Response</u>: *As aforementioned, we have revised the model setup and added uncertainty analysis to the revised version. Readers might not just focus on one objective of the study (quantify the impact of human water use and climate variation on the Lake Urmia water balance). As the editor has stated, one novelty of this study is assessing the value of remote sensing data and ground information for calibrating a hydrological model. Therefore, there are not such redundant discussions as you stated. However, we have reduced discussions with more direct focus on two objectives of the study.*

Referee #3: Problem description: The problem description (i.e. drivers of the lake desiccation) particularly in the introduction as well as the later discussion of the results is problematic. It is misleading and inaccurate to stack the climatic and human drivers together (e.g. page 29 lines 1- 10). Khazaei et al. [2019] disentangled these two: they compared the influence of these two sets of drivers for the lake drying and demonstrated the regional human activities (including water management, but not limited to) are the primary drivers compared to climatic changes (including atmospheric droughts). AghaKouchak et al. [2015] also argued that droughts cannot be the primary driver. These two studies, among others, are based on directly analysing the data themselves, without relying on a model of the system which in most cases are inadequate. While these studies have their own shortcomings, as any scientific study has, your modeling results are inadequate to challenge them. As pointed out by the reviewers your modeling setup has several issues. Inadequate models, regardless of the extent of their inputs and their results, are inadequate. Authors said: “This study proved that even without human water use Lake Urmia would not have recovered from the significant loss of lake water volume caused by the drought year 2008”. This conclusion is in direct contradiction with AghaKouchak et al. [2015] conclusion that “a satellite based gauge-adjusted climate record… of Lake Urmia basin's Standardized Precipitation Index… indicates no significant trend in droughts over the past three decades at the 0.05 significance (95% confidence) level... In fact, the region has experienced more severe drought events in the past (e.g., 1997– 2002) that did not lead to a substantial change in the lake's surface area. Thus, we caution against overrating the role of droughts in the disruption of the lake's water balance to the extent that would cause such a massive shrinkage”. Given that AghaKouchak et al. [2015] directly analysed the historic data of the lake without relying on any inadequate model of the lake system, it is reasonable to say this contradiction indicates the shortcoming and (un)reliability of the model set-up in this study. Notwithstanding the unscientific language of this sentence. Science is not in the business of proving anything. In science we demonstrate and approximate. This is more so the case when it comes to hydrological modeling with numerous types and sources of uncertainty including both model structure and data. Also, what do you mean by “climate variations”? Are you referring to only natural climate variability or climatic changes which include both natural variability and human-induced changes?

Response: *We agree with you about the “page 29 lines 1-10”. To clarify we have removed the word “main” in line 6, now the sentence would be accurate. About two mentioned studies, Khazaei et al. [2019] and AghaKouchak et al. [2015], we discussed in “Comparison to human vs. climatic contribution as determined in previous studies” section, why our results are different with some studies and in line with some other studies. Analysis of the monthly data used in calculating SPI cannot take into account the changes in the pattern of daily precipitation. To approve this claim, we refer to two studies that they also* *used directly in-situ data without modeling. Bavil et al. (2018) showed a significant increase in the frequency of daily precipitation of less than 5 mm and a significant decrease in the frequency of daily precipitation of 10-15 mm, suggesting a runoff reduction even in case of constant annual precipitation. Also, Hosseini-Moghari et al. (2018) showed that an increasing frequency of days with less than 5 mm precipitation in combination with decreasing monthly precipitation has led to reductions of inflow into two dams in the Lake Urmia basin that are located downstream of the areas with insignificant human water use. Therefore, we did not rely just on our results and provided some facts from pure in-situ data which has been rarely mentioned in the previous studies. Moreover, Shadkam et al. (2016) who considered only inflow into the lake, reported the same results for the impact of human water use on the reduction of inflow into the lake. About "climate variations" we cannot claim that it is a natural climate variability or climatic change. To speak about climate change must be conscious. It might be considered as any change or variations in climate variables data in our study.*

References:

Bavil, S. S., Zeinalzadeh, K., and Hessari, B.: The changes in the frequency of daily precipitation in Urmia Lake basin, Iran, Theoretical and Applied Climatology, 133(1-2), 205-214, doi:10.1007/s00704-017-2177-7, 2018.

Hosseini-Moghari, S.M., Araghinejad, S. and Ebrahimi, K.: Monthly Precipitation Assessment: a misleading tool for understanding the effects of climate change, 8th Global FRIEND-Water Conference, Beijing, China, 6-9 November 2018.

Shadkam, S., Ludwig, F., van Oel, P., Kirmit, Ç., and Kabat, P.: Impacts of climate change and water resources development on the declining inflow into Iran's Urmia Lake, Journal of Great Lakes Research, 42, 942-952, doi:10.1016/j.jglr.2016.07.033, 2016.

Referee #3: Introduction: The introduction is very long, has redundant sections:

- Figure 3 has already been published and discussed by Aghakouchak and been repeated many times in the literature. (also pointed out by the referee 1)
- Citation to Zarghami (2011) on page 29 is irrelevant.

• Last 2 paragraphs are “method” material and not introduction

Another issue is that some of the sources are not peer-reviewed. Whether right or wrong, it is hard to rely on such sources. While local knowledge and literature may be a valuable source, it has to treated with caution, not to propagate any errors. So, I am hesitant to accept such discussions.

Response: *Figure 3 and Zarghami (2011) have been deleted and the last two paragraphs of introduction have been reconstructed. Almost all references are peer-reviewed except for some sources that reported in-situ data. It should be noted that all ULRP reports were reviewed by scientific committees. Unfortunately, there are no English references in some cases.*

Referee #3: Time period: the time period 2003-2013 is inadequate for modeling the lake dynamics. Before 2000 the lake was not as heavily impacted by over-regulation of the river flows, and also between 2000-2003 there is significant variation in the lake level and annual inflows to the lake. Therefore, it is essential to include these years, for as many variable as possible. Otherwise, the model is biased and not representative of the lake dynamics.

Response: *We have considered this period due to the fact that the observed data was available for this period. We completely agree with you; it was better to consider a longer period for calibration. However, we don't prefer to reconstruct data, that is error-prone. The GRACE data and irrigated areas are not available for the period 2000-2003. Further, we don't want to use the model for out of calibration period, therefore we believe that for using the model in the calibration period there is no concern about the bias.*

Referee #3: Model setup: as pointed out by R1 comment 2 (and a few other comments), the model setup is not transparent and justifiable. Surely, documenting model setup (even as supplement) is more necessary than many redundant discussions. As a test, an adequate model setup should work against the updated observation (2013-present). Can the authors demonstrate this?

Response: *We have added some descriptions for the model setup in the revised version. Unfortunately, there is no information about the irrigated areas after the calibration period. Forcing data also available until the end of 2013. However, we have tested the model against lake volume which is really is more challenging those simulation variables used in the calibration process.*

Referee #3: Standard model: inclusion and discussion of the standard model, as a reviewer pointed out, in unnecessary given that this model is not calibrated for this catchment.

Response: *As you know the outputs of the standard version of WGHM as a global model have been commonly used worldwide. For instance, Tourian et al. (2015) used the standard version of WGHM over lake Urmia. Therefore, it would be valuable to show the validity of using the standard version of WGHM in another study. Hence, we would keep this section in the revised version.*

Reference:

Tourian, M. J., Elmi, O., Chen, Q., Devaraju, B., Roohi, S., and Sneeuw, N.: A spaceborne multisensor approach to monitor the desiccation of Lake Urmia in Iran, Remote Sensing of Environment, 156, 349-360, doi:10.1016/j.rse.2014.10.006, 2015.

Referee #3: Model calibration and evaluation: the model calibration is also questionable. It is only based on trial and error so the identified parameter sets are not reliable. Also, it is not clear how sensitive the model results are to these parameters. There is issue about over-parametrizing the model as in each variant new data is added (issue about correction factors pointed out by reviewers). A single model run for each variant is not sufficient, even if the calibration was done through an automatic parameter space search scheme. This is well-established in the literature and a model ensemble is required to account for the uncertainties, even though partly. If the model is run on a daily basis, authors should be trans-parent and present the daily results too.

Monthly and annual performance metrics are usually high for most models. The devil is in the details though. On table 4, the flow is only calibrated on an annual scale and not evaluated at all. There is significant seasonality in this region. It is quite possible that seasonal errors are just cancelling each other out and give seemingly good annual results. This can be seen on figure 8 where the model exhibits unrealistic seasonality which is not in the observed data, e.g. in Fig 8e, there is generally a negative bias in the first half of the RS_Q_GW_NA simulation, and positive bias in the second half (red line is first below the black line systematically, and then above it).

<u>Response</u>: *We have reconsidered the model calibration by using two optimization algorithms, namely genetic algorithm (GA) and Non-dominated sorting genetic algorithm II (NSGA-II). As editor request, to keep the length of the revised version no more than the previous version, we just focus on two calibration variants using RS with and without ground information. Other variants only considered for some discussions. In the revised version, we used an ensemble of model outputs that were obtained from different GA runs or Pareto front for NSGA-II. Therefore, we have discussed the uncertainties in the revised version. Adding daily outputs making the manuscript longer than the current version and we believe that it is redundant. Because most of the outputs are anomaly based on the monthly mean, therefore daily data could not be beneficial. However, to consider your concern about the inflow into the lake, we have plotted the daily inflow here. About Fig 8e, it should be noted that modeling lake volume is not a simple task for a hydrological model. Therefore, we believe that the performance of the model is quite acceptable, while in the revised version the performance of the model improved in this regard.*

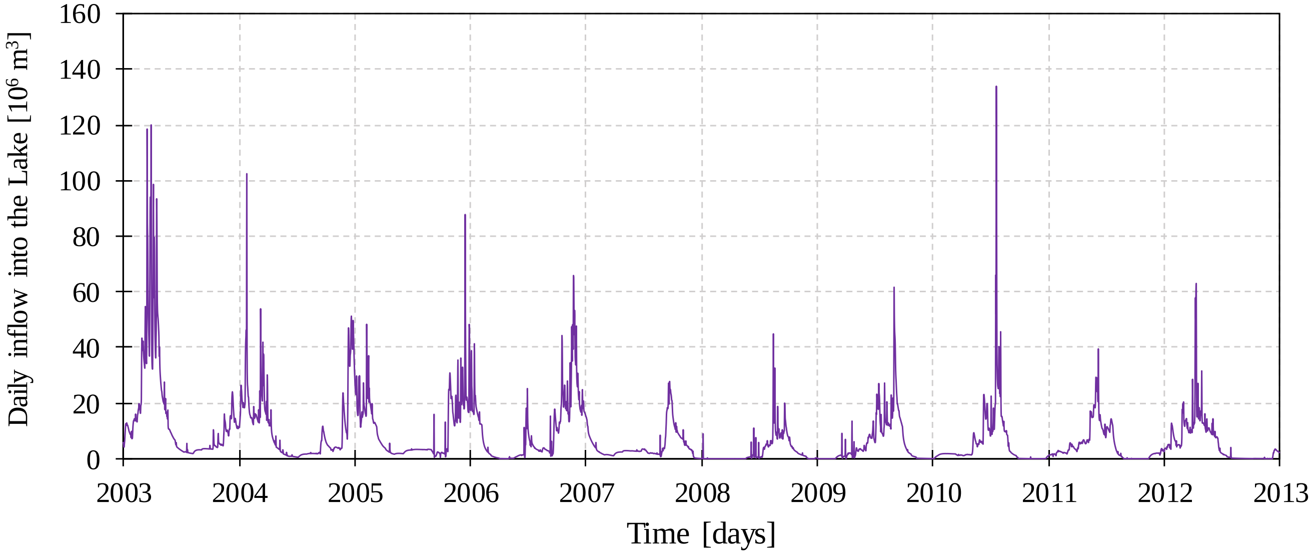


Figure R: Time series of simulated daily inflow into the lake

Referee #3: Uncertainties: as Referee 1 pointed out in their comment 6 and elsewhere, the uncertainties are playing a crucial role here. While authors added a section of uncertainty discussion, it is an ad-hoc discussion. While the discussion obviously undermines the experiment design and the results, they have not revised the experiment design to account for these uncertainties, and also they keep overpromising about the reliability of their results. Further, their discussion of uncertainty shows a lack of understanding about the area of model uncertainty. For instance, they said “Model parameter uncertainty was reduced by the comprehensive multi-observation calibration”. Parameter uncertainty will not be reduced by just adding more data to the model; data uncertainty matters, “garbage in, garbage out” [Kuczera et al., 2010]. The authors must demonstrate how adding input to the model reduced parameter uncertainty, while justifying the credibility of the data themselves. They have not done any uncertainty or sensitive analysis whatsoever.

Response: *We have revised the model set up and added some uncertainty analysis to the manuscript. We believe the parameter uncertainty should be reduced when a multi calibration approach is used. Because the model should satisfy more than one objective therefore change in one parameter should be done with less freedom. We already showed this issue in Table 5 where the model provides changes in total water storage through only by changing lake depth. In addition, use more observed data always help us improve the calibration uncertainty that is arising from our lack of awareness of the observed values of a given variable.*

Thank you very much again for your time and for providing valuable comments.

Quantifying the impacts of human water use and climate variations on recent drying of Lake Urmia basin: the value of different sets of spaceborne and in-situ data for calibrating a hydrological model

Seyed-Mohammad Hosseini-Moghari[1], Shahab Araghinejad[1], Mohammad J. Tourian[2], Kumars Ebrahimi[1], Petra Döll[3,4]

[1]Department of Irrigation and Reclamation Engineering, University of Tehran, Karaj, Iran
[2]University of Stuttgart, Institute of Geodesy (GIS), Stuttgart, Germany
[3]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany
[4]Senckenberg Leibniz Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany

Correspondence to: Seyed-Mohammad Hosseini-Moghari (Hosseini_sm@ut.ac.ir)

Abstract. During the last decades, the endorheic Lake Urmia basin in northwestern Iran has suffered from declining groundwater tables and a very strong reduction in the volume as well as recently in the extent of Lake Urmia. For the case of Lake Urmia basin, this study explores the value of different locally and globally available observation data for adjusting a global hydrological model such that it can be ~~reliably~~ used for distinguishing the impacts of human water use and climate variations. The WaterGAP Global Hydrology Model (WGHM) was for the first time calibrated against multiple in-situ and spaceborne data to analyse the decreasing lake water volume, lake river inflow, loss of groundwater, and total water storage in the entire basin during 2003-2013. The calibration process was done using an automated approach including a genetic algorithm (GA) and Non-dominated sorting genetic algorithm II (NSGA-II). Then the best-performing ~~calibration variant was~~ calibrated models were run with ~~or~~and without considering water use to quantify the impact of human water use. Observations encompass remote-sensing based time series of annual irrigated ~~area~~areas in the basin from MODIS, monthly total water storage anomaly (TWSA) from GRACE satellites, and monthly lake volume anomalies. In-situ observations include time series of annual inflow into the lake and basin averages of groundwater level variations based on 284 wells. In addition, local estimates of sectoral water withdrawals in 2009 and return flow fractions were utilized. ~~Four calibration variants were set up in which the number of considered observation types was increased in a stepwise fashion. The best fit to each and all observations, including the time series of lake volume not used for calibration, was achieved if the maximum amount of observations was used for calibration.~~ Calibration against MODIS and GRACE data alone improved simulated inflow into Lake Urmia but inflow and lake volume loss were still ~~still~~ overestimated, while groundwater loss was ~~understimated~~underestimated and seasonality of groundwater storage was shifted as compared to observations. Lake and groundwater dynamics could only be simulated well if calibration against groundwater levels led to an adjustment of the fractions of human water use from groundwater and surface water. Thus, in some basins, globally available space-born observations may not suffice for improving the simulation of human water use. According to ~~our study~~WGHM simulations with 18 optimal parameter sets, human water use was the reason for ~~50~~52-57% of the total basin water loss of about 10 km^3 during 2003-2013, for ~~40~~39-43% of the Lake Urmia water loss of about 8 km^3 and for up to 87-90% of the groundwater loss.

Lake inflow was ~~40~~39-45% less than it would have been without human water use. ~~This~~The study ~~proved~~shows that even without human water use Lake Urmia would not have recovered from the significant loss of lake water volume caused by the drought year 2008. These findings can support water management in the basin and more specifically Lake Urmia restoration plans.

1 Introduction

Iran is a country with an arid and semi-arid climate where population growth and the government's aim of food self-sufficiency has led to increasing irrigated crop production and exploitation of surface water and groundwater resources. Climate change has resulted in increased temperatures and, in particular the northwest of the country, in decreased precipitation (Tabari and Talaee, 2011a, b) and thus decreased renewable water resources. In the last decades, numerous wetlands and lakes in Iran have dried up, and groundwater levels have strongly declined in most areas (Madani et al., 2016). The most serious disaster has occurred in the Lake Urmia basin, an interior basin in the northwest of Iran located in the three provinces West Azarbaijan, East Azarbaijan, and Kurdistan that covers an area of 52,000 km^2 (Fig. 1). At the downstream of the basin, 17 permanent rivers and 12 seasonal rivers discharge into the largest natural water body in Iran, Lake Urmia. Over the past two decades, climate variations and human activities (Hassanzadeh et al., 2012) have decreased inflow into the lake. Precipitation in the basin shows a decreasing trend over the period 1951-2013, with particularly low values after 1995, and evaporation has increased (Alizadeh-Choobari et al., 2016). Lake water volume is now approximately $30 \cdot 10^9$ m^3 below its historical maximum (ULRP, 2015a).

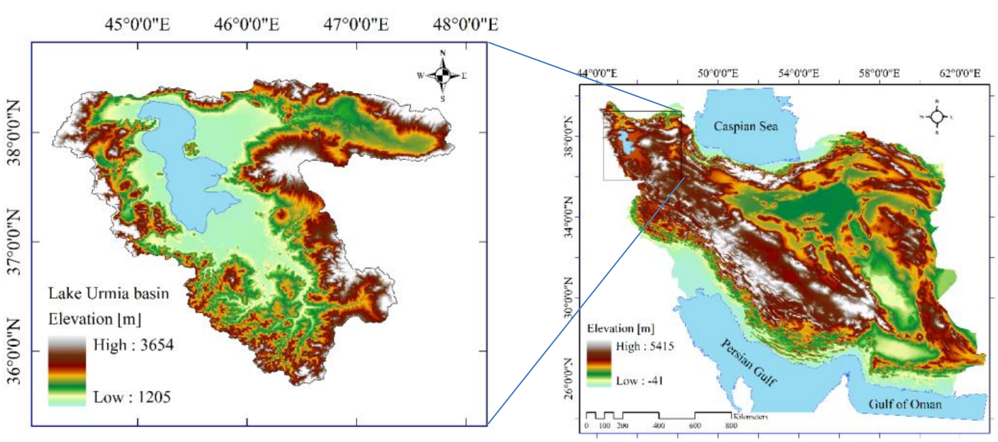


Figure 1: Location of Lake Urmia basin.

Lake Urmia is one of the largest hypersaline lakes in the world, which due to its ecological and natural features is a National Park, a Ramsar Site and a UNESCO Biosphere Reserve (Eimanifar and Mohebbi, 2007). It is a terminal lake that loses water only by evaporation (Hassanzadeh et al., 2012). Abbaspour and Nazaridoust (2007) estimated that inflows of at least $3 \cdot 10^9$ m^3/yr are needed to compensate for lake evaporation, while Alborzi et al. (2018) estimated values between $2.9 \cdot 10^9$

to $5.4 \cdot 10^9$ m^3/yr depending on climatic conditions. According to Alborzi et al. (2018), recovery of the lake could range from 3 to 16 years depending on climatic ~~condition~~conditions, water use reductions, and environmental releases. Inflow from groundwater to the lake was estimated to be less than 3% of total inflow from precipitation, rivers, and groundwater (Hasemi, 2011). In the 1970s and 80s, the water table of Lake Urmia was approximately at 1,276 m above sea level and then increased to more than 1,278 m in 1995 due to a few wet years (Shadkam et al., 2016). ~~Afterwards~~Afterward, the water table dropped to 1,274 m in 2003 ~~specially~~ because of the severe drought in 1999-2001 exacerbated by human water use (Shadkam et al., 2016). From 2003 to 2014, lake extent was approximately halved, and water level declined by another 3 m, while seasonal variability of lake water extent increased (Tourian et al., 2015) (Fig. 2). ~~After 2015, lake extent and storage have stabilized (Fig. 3) due to the relatively high precipitation in 2015 and 2016, increased releases from reservoirs and management activities for decreasing water consumption (ULRP, 2015b).~~

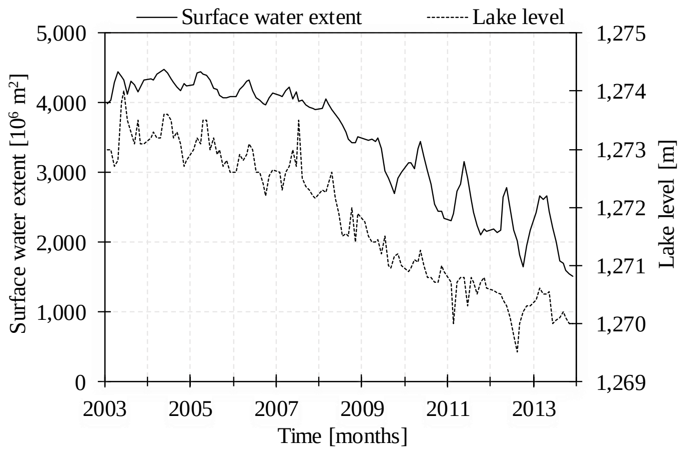


Figure 2: Time series of surface water extent and water table elevation of Lake Urmia (data from Tourian et al., 2015).

Studies on various aspects of the Lake Urmia disaster abound. With decreasing lake water volume, salt concentration has increased, endangering the aquatic biota feeing birds; exposed salt layers may lead to salt storms (Pengra, 2012). Precipitation reduction, temperature increase, agricultural development including construction of man-made dams and building a causeway across the lake have been identified as the ~~main~~ reasons for the degradation of Lake Urmia (Abbaspour and Nazaridoust, 2007; Zeinoddini et al., 2009; Delju et al., 2012; Jalili et al., 2012; Sima and Tajrishy, 2013; Fathian et al., 2014; Farajzadeh et al., 2014; Banihabib et al., 2015; AghaKouchak et al., 2015; Azarnivand ~~et al., 2015~~and Banihabib 2017; Alizadeh-Choobari et al., 2016; Ghale et al., 2018; Khazaei et al., 2019). By using Gravity Recovery And Climate Experiment (GRACE) satellite observations, altimetry data for Lake Urmia and outputs of the Global Land Data Assimilation System (GLDAS), Forootan et al. (2014) estimated the trend of groundwater storage changes in the Lake Urmia basin as -11.2 mm/yr between the years of 2005 to 2011, the largest decrease of the six investigated Iranian basins. ~~Zarghami (2011) examined four routes to transfer the water from Aras basin in the north of Lake Urmia basin to provide an alternative for the water supply for the agricultural and drinking demands in the north of the basin.~~ Ahmadzadeh et al. (2016) investigated the effect of irrigation system changes in the basin from the surface to pressurized systems; they found that such changes would increase water productivity but would have no effect on lake inflow and would reduce groundwater levels by 20%.

~~**Figure 3: Lake Urmia during the time period 2002-2016 (Google Earth Timelapse, last accessed: 28 Apr. 2018).**~~

Three hydrological modelling studies for Lake Urmia basin focused on quantifying the contributions of various factors on lake water volume (Hassanzadeh et al., 2012), lake inflow (Shadkam et al., 2016) or both (Chaudhari et al., 2018). Using a lumped system dynamics modelling approach and observed time series of lake water volume for model calibration, Hassanzadeh et al. (2012), determined that about 65% of lake level decline between 1997 and 2006 was due to reduced river inflow, while four major man-made reservoirs contributed 25% and diminished precipitation on the lake surface 10%. Shadkam et al. (2016) evaluated the impact of climate, irrigation with surface water and reservoirs on inflow into the lake for the period 1960-2010 using a modified version of the macro-scale gridded hydrological model Variable Infiltration Capacity (VIC) model, which was calibrated against time series of river discharge at six observation station at the downstream end of six sub-basins draining into Lake Urmia. While the model was driven by global gridded WFDEI climate data set with a spatial resolution of 0.5°, basin-specific information on 41 reservoirs and on the temporal development of irrigated areas were taken into account. The study found that reservoirs had a very small impact on annual inflows and that climate variations accounted for 60% of lake inflow decrease of 48% over the 50-year period. In the model, all irrigation requirements need to be fulfilled by available surface water. Therefore, reduced availability of surface water during the 2000s due to low precipitation and high temperature resulted in unfulfilled irrigated water demand and a cap on the effect of human water use in the model while in reality, groundwater abstractions occurred and even increased (Delju et al., 2012; Hesami and Amini, 2016). In addition, the modelling study of Shadkam et al. (2016) did not consider the impact of domestic and industrial water use in the basin which can be expected to have increased during the last decades, given a population increase from 4.8 to 5.9 million from 2002 to 2010 (http://ulrp.sharif.ir/en/page/about-urmia-lake-basin, last accessed: 28 Apr. 2018). Chaudhari et al. (2018) used the output of the global HiGW-MAT model, with 1°×1° grid cell size of approx. 10,000 km^2, to distinguish climatic and anthropogenic contributions to the shrinkage of Urmia Lake. By running the model with and without human impacts (surface and groundwater use as well as reservoirs), they estimated that the human-induced river flow decline between 1995-2010 to account for 86% of the observed decrease of lake volume. However, a comparison with GRACE TWSA showed that the model overestimates the decrease in TWSA in the basin between 2003 and 2010. The HiGW-MAT model was not calibrated for the Lake Urmia basin but net irrigation requirements were simulated specifically for this study based on Landsat satellite images for 5 years between 1987 and 2016. The lake water balance is not simulated by the model such that no comparison with observed lake water levels was possible. A comparison with river discharge or groundwater observations was not done either.

~~The aim of our study was twofold. On the one hand, we wanted to quantify, by a holistic and reliable modelling approach,~~In previous hydrological modeling studies of Lake Urmia basin, there either no model calibration or calibration was only done using a single observation type, in particular surface water inflow into the lake. Although streamflow observations are very informative for hydrological modelling as they integrate over processes in the whole upstream basin, a good fit of simulated and observed streamflow may not necessarily lead to an appropriate simulation of other flows and storages (Beven and Freer, 2001). Therefore, additional types of observations have to be added to avoid equifinality (Beven and Freer, 2001; Döll et al., 2016). In this study, a multi-observation calibration approach was used to calibrate a hydrological model which was then applied to quantify the contributions of climate variations and human activities to the decrease of Lake Urmia water volume and river inflows~~ as well as, different from previous studies, to groundwater storage and total water storage in the whole Lake Urmia basin. Such a modelling approach requires the set-up of a model that is able to simulate the impact of surface and groundwater use as well as of climate variations on these water storages and flows. The hypothesis is that if model output for all these variables fit well to observations, then the model can be used to assess the contribution of human water use by comparing the outputs of two model variants, one with human water use and one where human water use is assumed to be zero. To achieve a good fit to observations, hydrological models need to be calibrated by comparison of observations with model output variables. While hydrological models are usually calibrated only against observations of river discharge, it is well known that a good fit of simulated and observed river discharge does not lead necessarily lead to an appropriate simulation of other flows and storages (Beven and Freer, 2001). However, in previous hydrological modeling studies of Lake Urmia basin, model calibration was either not done at all or only using a single observation type. On the other hand~~. In addition, using Lake Urmia basin as a test case, we wanted to explore the value of different types of observation data for adjusting a global hydrological model by multi-observation calibration. Currently, global hydrological models are mostly uncalibrated but globally available space-born observations have increased the opportunity for model calibration at the global scale (Döll et al., 2016).

~~We used the state-of-the-art global hydrological model WaterGAP 2.2c (spatial resolution 0.5°×0.5°) which simulates human water uses from surface water and groundwater and how these affect river discharge, groundwater, lake water, and total water storage. In its standard version, WaterGAP is calibrated against observed mean annual river discharge at 1319 stations worldwide by adjusting 1-3 model parameters related to runoff generation and streamflow (Müller Schmied et al., 2014), but for reasons of data availability not for a station in Lake Urmia basin. A previous WaterGAP version was calibrated, for 22 large basins, against streamflow and total water storage anomalies by adjusting 6-8 parameters (Werth and Güntner, 2010). For this study on the differential impacts of climate and human water use on Lake Urmia basin, WGHM was for the first time calibrated for a specific basin by using multiple types of independent data. Multi-observation calibration included the adjustment of temporally constant model parameters as well as the adjustment of human water use input data. To understand the value of different observations or other regionally available data for understanding dynamics of water flows and storages in a basin, WGHM was calibrated sequentially by considering, in each calibration variant, an additional data type. In the first variant, only remote sensing data were used (variant RS). In-situ river discharge observations were added in variant RS_Q. In the third variant RS, discharge and groundwater level data were used (variant RS_Q_GW), and finally RS, discharge, groundwater levels as well as regional data of basin-wide total withdrawals plus estimated return flow fractions (RS_Q_GW_NA variant). Model evaluation was done by comparison of simulated lake water volume anomalies against observed anomalies. The best-performing model variant RS_Q_GW_NA was then applied to simulate the water flows and~~

~~storages in Lake Urmia basin that would have occurred under naturalized conditions, i.e. without any human water use (and man-made reservoirs). By comparing the output of the naturalized run with the output of the model run with human impacts, we determined the contributions of human water use and climate variation on lake inflow and water storages in the period 2003-2013. In section 2, we describe the utilized data and the simulation setup. The results of the four~~ For this purpose, the WaterGAP global hydrology model (WGHM) was calibrated by means of genetic algorithm (GA) and Non-dominated sorting genetic algorithm II (NSGA-II) for the Lake Urmia basins. Descriptions of the used data and the simulation setup are presented in section 2. The results of the different calibration variants and the impacts of human water use are shown in section 3. Section 4 discusses multi-observation calibration and the analysis of human impact as well as the limitations of the study. Finally, conclusions are drawn.

2 Methods and data

We analyzed the 11-year period from the beginning of 2003 until the end of 2013, as both GRACE data and global climate data to drive WaterGAP ~~where~~were available for this period. In the following sections, WaterGAP, ~~WaterGAP~~its input data and the observational data used for calibration as well as the calibration ~~variants~~approach are described.

2.1 WaterGAP

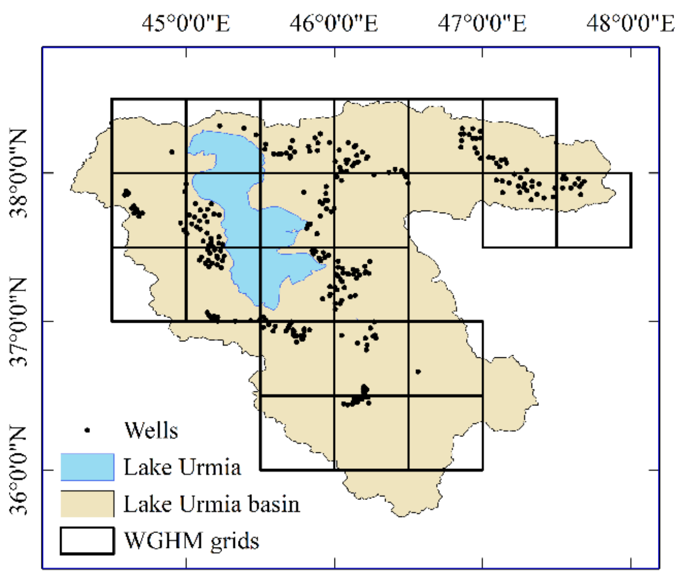
WaterGAP is a global hydrological model for assessing water resources under the influence of humans (Döll et al., 2003; Müller Schmied et al., 2014). With a spatial resolution of 0.5°×0.5°, it simulates water abstractions and consumptive water use (so-called net abstractions, i.e. the amount of water that evapotranspirates during use and does not flow to surface water bodies and groundwater afterwards) in five sectors (irrigation, livestock, domestic, manufacturing and cooling of thermal power plants); then net abstractions from either groundwater (NAg) or surface water bodies (NAs) are computed (Müller Schmied et al., 2014; Döll et al., 2012). ~~Negative values of NAg occur where return flow to groundwater from irrigation with surface water is so high that water is added to groundwater storage by human water use. NA is the sum of NAg and NAs and equal to consumptive water use.~~ Time series of NAg and NAs in each grid cells are then input to the WaterGAP Global Hydrology model WGHM that simulates their effect on water flows and storages. In ~~WGHM, NAg~~its standard version, WaterGAP is calibrated against observed mean annual river discharge at 1319 stations worldwide by adjusting 1-3 model parameters related to runoff generation and ~~NAs are subtracted from either the groundwater or surface water bodies (lakes, reservoirs or rivers)~~streamflow (Müller Schmied et al., 2014~~).~~

~~WGHM simulates daily~~), but due to lack of data not for any station in Lake Urmia basin. A previous WaterGAP version was calibrated, for 22 large basins, against streamflow and total water storage ~~as well as flows like evapotranspiration, groundwater recharge (Döll and Fiedler, 2008), runoff, and river discharge for all continents except Antarctica. Water is transported between grid cells according to the DDM30 drainage direction map (Döll et al., 2003). Water storage compartments encompass snow, canopy, soil, groundwater, rivers, lakes, wetlands, and man-made reservoirs (Eicker et al., 2014). Lake water~~

~~storage is simulated as the difference of precipitation on the lake, evapotranspiration, inflows, and outflows. Outflow is zero for end lakes like Lake Urmia. The temporal variation of lake area, affecting precipitation on and evapotranspiration from the lake, is simulated as a non-linear function of lake water storage. WGHM contains more than 20 parameters that can be potentially be adjusted by calibration~~anomalies by adjusting 6-8 parameters (Werth and Güntner, 2010).

~~WaterGAP includes a multitude of global data sets including information on irrigated areas, the fraction of irrigated areas that is equipped to be irrigated with groundwater (Siebert et al., 2010) and artificial drainage affecting return flows to surface water (Döll et al., 2012). For more information on data and model algorithms used in WaterGAP please refer to Müller Schmied et al. (2014) and Döll et al. (2014a).~~ WGHM can be run globally or for a specific ~~basins only~~basin. In this study, it was run only for the 22 0.5° grid cells that represent the Lake Urmia basin in WGHM (Fig. ~~4).~~3). A more detailed description of WGHM can be found in the supplement.

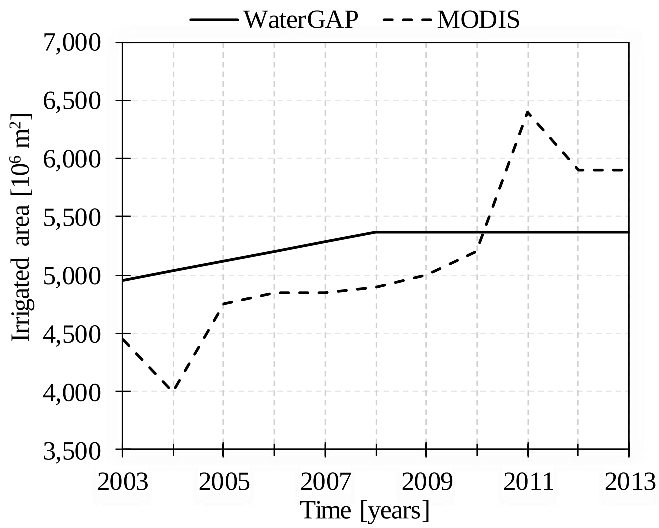
~~WaterGAP outputs were extensively compared to in-situ streamflow observations (e.g., Döll et al., 200; Müller Schmied et al., 2014), to GRACE TWSA (Döll et al., 2012, 2014a, b) and GPS TWSA (Döll et al., 2014b). Results were shown to depend on applied climate input data sets (e.g., Müller Schmied et al., 2014, 2016; Döll et al., 2014b), model structure (Müller Schmied et al., 2014), and assumptions on water use (Döll et al. 2014a, b). Comparison of observed streamflow regime indicators (different streamflow percentiles representing statistical low and high flows) to the values computed by nine (or seven) GHMs showed that WaterGAP is one of the best fitting models (Gudmundsson et al. 2012; Tallaksen and Stahl, 2014). Prudhomme et al. (2011) concluded that "of the three global models considered here, WaterGAP is arguably best suited to reproduce most regional characteristics of large-scale high and low flow events in Europe." Regarding the fit to GRACE and GPS TWS, Döll et al. (2014b) found that WaterGAP underestimates seasonal variations of TWS on most of the land area of the globe and that seasonal maximum TWS occurs one month earlier according to WaterGAP than according to GRACE on most land areas.~~



~~**Figure 4: Grid cells in WGHM corresponds to Lake Urmia basin along with the location of groundwater wells across the basin.**~~

2.2 Data

~~2.2.1~~We used the following observations for calibrating WGHM: (1) Remote sensing data including irrigated area in Lake Urmia basin~~. Based on MODIS images, Kamali~~ and ~~Youneszadeh Jalili (2015) estimated annual time series of irrigated areas in Lake Urmia basin from 2001 to 2012. Considering that water management in the basin aims at preventing any increase of irrigated areas, it is assumed that irrigated area in 2013 remained at the 2012 value (Fig. 5).~~



~~**Figure 5: Irrigated area in Lake Urmia basin assumed in WaterGAP and derived from MODIS (data from Kamali and Youneszadeh Jalili, 2015).**~~

GRACE ~~**total water storage anomalies.** GRACE satellite data allow derivation of monthly time series of total water storage anomalies (~~TWSA~~) over all continents. TWSA describes the total amount of water stored on the continents, including water storage in surface water bodies, groundwater and soil, as compared to the mean value of total water storage over a reference period. In our study CSR GRACE RL05 mascon solutions (Save et al., 2016; , last accessed: 17 Jul. 2018) were used. While it is recommended GRACE data products only for areas with at least 100,000 km^{2} (Watkins et al., 2015; Landerer and Swenson, 2012), studies by Tourian et al. (2015) and Lorenz et al. (2014) showed that signal strength or the so-called gravimetric resolution is determining the applicability of GRACE data. In fact, Lake Urmia basin has experienced an 8·10^{9} m^{3} change in the water volume in the last decade, which allows the use of GRACE for monitoring the changes in water storage in the basin (Tourian et al., 2015). This fact is supported by the very small gain factor of 1.0083 for the Lake Urmia basin based on Community Land Model 4 (CLM4) for spherical harmonic solutions (Landerer and Swenson, 2012), which is the factor with which signal attenuation due to leakage could be balanced. We can assume errors of the applied GRACE monthly time series of TWSA are small compared to the uncertainty of TWSA as computed by WGHM, such that model calibration against GRACE TWSA is meaningful.~~

, (2~~.2.2~~) inflow into Lake Urmia

~~We used total annual observed inflow into the lake during 2003-2013 which was computed by the Urmia Lake Restoration Program (ULRP) based on 19 hydrometric stations around the lake (data available in _(In Persian), last accessed: 12 Nov. 2017). Monthly observations were not available. It was compared to the sum of simulated river discharge of all WGHM grid cells flowing into the grid cell representing Lake Urmia.~~

~~2.2.~~Q, (3) groundwater levels

~~For evaluating the groundwater status in Lake Urmia basin, we used groundwater head data of 284 wells during 2003-2013 (Fig. 4). To obtain a monthly time series of average groundwater level in the basin, first the average of all groundwater level in each 0.5° grid cell was calculated and then the average values of all grid cells (see Strassberg et al., 2009).~~

~~2.2.4~~ from well observations, which were converted into groundwater storage anomalies GWSA (see section S2) and (4) statistical information on water withdrawals and consumptive uses

~~There are no water withdrawals time series data in Lake Urmia basin. However, water withdrawals in the Lake Urmia basin for 2009 was reported to be 4,825·10^6 m^3 (ULRP, 2015c) of which 89% is used for irrigation (Table 1). 57% of the withdrawn water is taken from surface water, the rest from groundwater. According to the report of Mahab Ghodss Consulting Engineering (2013), 16% of the water withdrawn for irrigation returns to groundwater and only 2% to surface water bodies, while the respective values for industrial and domestic water withdrawals are 50% and 10%. In this study, observed consumptive irrigation use was computed by subtracting total return flow from total water withdrawals for irrigation. Thus, it was set to 82% of water withdrawals for irrigation, while observed consumptive use in the domestic/industry sector was set to 40% of sectoral water withdrawals. The sum of consumptive water use in all sectors is the so-called total net abstraction (NA) from either surface water bodies or groundwater.~~

~~Table 1: Water withdrawals in Lake Urmia basin in 2009 [10^6 m^3] (data from URLP, 2015c).~~

~~Source~~	~~Sector~~			~~Total~~
	~~Agricultural~~	~~Domestic~~	~~Industry~~	
~~Surface water~~	~~2424~~	~~276~~	~~33~~	~~2733~~
~~Groundwater~~	~~1867~~	~~190~~	~~35~~	~~2092~~
~~Total~~	~~4291~~	~~466~~	~~68~~	~~4825~~

~~2.2.5 Climate~~

~~The 0.5° gridded EartH2Observe, WFDEI and ERA-Interim Data Merged and Bias-corrected for ISIMIP (EWEMBI) dataset (Lange, 2016) was used as forcing data set. EWEMBI includes daily climate data for 1979 to 2013. For EWEMBI, ERA-Interim Reanalysis Data were bias-corrected with monthly observation data on temperature, precipitation and the number of wet days as well as daily radiation data. We compared, for the period 2003-2013, basin-average monthly precipitation and~~

~~temperature values of EWEMBI dataset with those derived as the mean over monthly values observed at 143 rain gauges and six temperature gauging stations. The correlation coefficient (CC), Nash-Sutcliffe efficiency (NSE), and Willmott's refined index of agreement (Willmott et al., 2012) were 0.985, 0.946, and 0.897, respectively, for precipitation, and 0.996, 0.983, and 0.941 respectively, for temperature.~~

~~2.2.6 Lake volume~~

in the basin. In addition, time series of lake volume based on remote sensing ~~data for lake extent and water table elevation as well as on in-situ bathymetry data, a time series of monthly water volume in Lake Urmia for the period 2003-2013 was generated by Tourian et al. (2015) (their Fig. 9). It was used for evaluation of the model variants.~~ was used for validation. The 0.5° gridded EWEMBI data set was used as climate forcing. Irrigated area and Q are at the annual time scale, TWSA, GWSA and lake volume on the monthly scale and the climate forcing is on a daily scale. All data cover the period 2003-2013 (see section S2 for details).

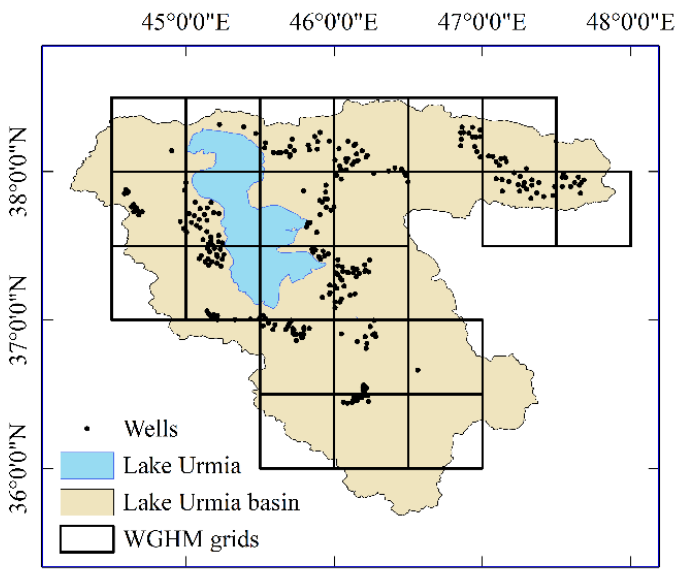
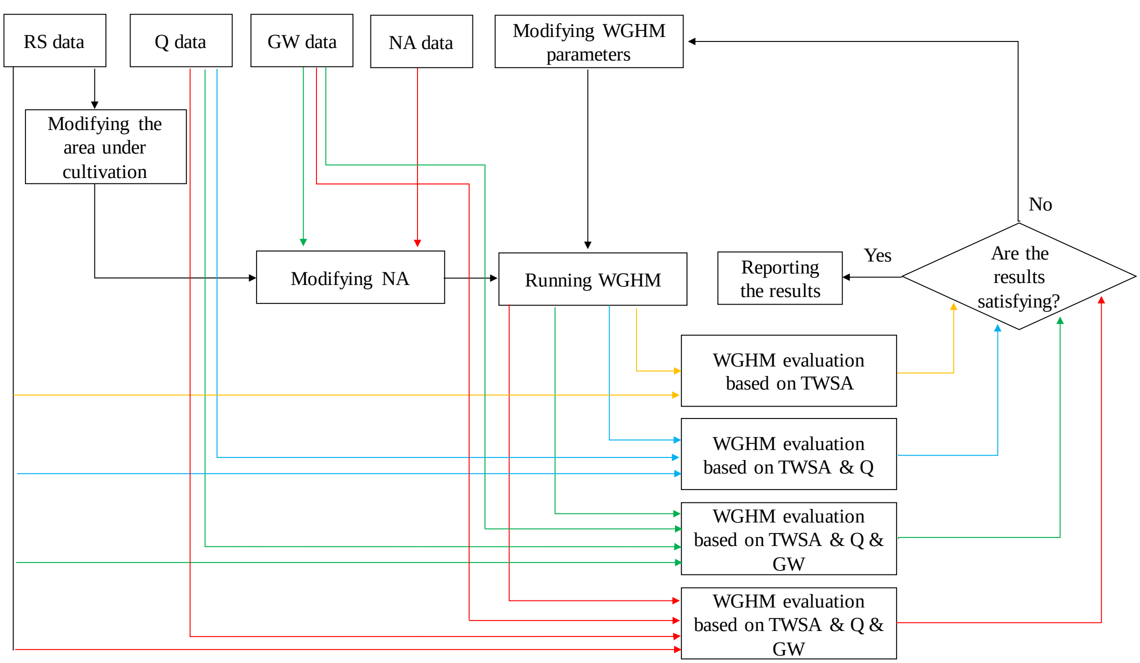


Figure 3: Grid cells in WGHM corresponding to Lake Urmia basin along with the locations of groundwater wells across the basin.

2.3 Calibration ~~variants~~approach

~~Calibration was done by trial-and-error. It included the modification time series of irrigated area~~Two calibration variants were applied. In the RS variant, only the remote sensing information was used for calibration, including irrigated area from MODIS and GRACE TWSA. In the variant RS_Q_GW_NA, ground-based information was used in addition to the remote sensing observations. This included inflow into the lake, groundwater data and statistical information regarding water use. Calibration was done using the genetic algorithm (GA) for variant RS, with just one calibration objective, and the non-dominated sorting genetic algorithm II (NSGA-II), a multi-objective version of GA, for the variant RS_Q_GW_NA. To integrate optimization algorithm with WGHM, we scripted the codes in shell and R environments by modifying *'GA'* (Scrucca, 2013), and *'nsga2R'* (Tsou, 2013) Packages in R. GA and NSGA-II are the most common evolutionary optimization algorithms in hydrological model calibration (e.g. Azarnivand et al. 2019). Both algorithms start with a random population (here WGHM parameters) and after evaluating the objective function(s) (here KGE) the better parameter sets are selected based on the value of the objective function (in GA) and non-domination and crowding distance (in NSGA-II). Then, the crossover and mutation operators are applied and the process will be continued until one stopping criteria met. The details of GA and NSGA-II can be found in Mirjalili (2019) and Deb et al. (2002), respectively. Because of the use of the random generators in GA and NSGA-II, we did five runs for each algorithm to achieve more reliable results. The selected parameters for each algorithm are presented in the supplement (Table S3). Fig. 4 shows the flowchart of these algorithms along with a schematic of the calibration process for the two calibration variants. In short, calibration included the modification time series of irrigated areas, of NAg and NAs, with different multipliers for individual years, as well as the modification of ~~a maximum of~~ seven temporally constant model parameters or, in case of spatially heterogeneous parameters, multipliers~~,~~ (see Table 1). Modifications were done homogeneously for the whole basin. Months with assumed irrigation in Lake Urmia basin according to WaterGAP correspond to the actual irrigation months (Apr. and Oct.) in the basin according to Saemian et al. (2015~~),~~). Thus~~,~~ no correction of ~~the~~ seasonality was needed in the calibration process. ~~Fig. 6 shows a schematic of the calibration process for the four calibration variants. Please note that the identified parameter combinations are not the only ones that would lead to a good fit to observations.~~



~~**Figure 6: Flowchart for the four calibration variants. The black line is common in all variants, the mustard, blue, green and red lines represent calibration based on RS data (RS variant), RS data and inflow data (RS_Q variant), RS, inflow and groundwater level data (RS_Q_GW variant), and RS, inflow, groundwater level and net abstraction data (RS_Q_GW_NA variant), respectively.**~~

~~2.3.1 RS variant: Calibration using remote sensing data~~

~~Irrigated area in Lake Urmia basin used in the standard version of WaterGAP is larger than the MODIS-based irrigated area until 2010, and smaller afterwards (Fig. 4). The largest differences, in 2004 and 2011, exceed 20%, or 1,000 km², and the strongly increasing trend is not represented in WaterGAP. The constant value of irrigated area in WaterGAP is due to the fact that the Food and Agricultural Organization of the UN does not provide more recent estimates of irrigated area in Iran (see , last accessed: 13 Feb. 2018). To utilize the MODIS-based time series, consumptive irrigation water use in the whole basin of WaterGAP in year i was first adjusted by multiplying it by a correction factor CF1(i), with:~~

~~where $Area_{irri}^{MODIS}(i)$ is irrigated area from MODIS in year i and $Area_{irri}^{WG}(i)$ is irrigated area from WaterGAP database. The modified consumptive irrigation use was then added to the consumptive use of WaterGAP for the other sectors to obtain an updated basin-wide NA for each year. Then, modified monthly NAg and NAs in year i were calculated by multiplying, for each grid cell, the standard WaterGAP NAg and NAs values with the ratio of modified over standard basin-wide NA in year i. Then, WGHM was run with the modified NAg and NAs time series, and a small number of WGHM parameters was varied until achieving a good fit to monthly time series of basin-average GRACE TWSA (Fig. 6, yellow lines).~~

~~2.3.2 RS_Q variant: Calibration using remote sensing data and inflow into the lake~~

~~Model parameters of WGHM driven by modified NAs and NAg from the RS variant were adjusted to achieve a good fit for both GRACE TWSA and the time series of annual total inflows to Lake Urmia (Fig. 6, blue lines).~~

~~2.2.3 RS_Q_GW variant: Calibration using remote sensing data, inflow into the lake, and groundwater level~~

~~Since WGHM does not compute groundwater level but only groundwater storage, and there is no good information of basin-wide specific yield that would allow a translation of observed groundwater level variations into storage variations, model calibration in this variant aimed at optimizing the fit between the monthly time series of normalized basin-average observed groundwater levels (calculated by subtracting the mean and dividing by the standard deviation) to the monthly time series of normalized WGHM groundwater storage. To achieve a good fit to groundwater levels, and at the same time to GRACE TWSA and observed inflow into the lake, NAg and NAs as adjusted in variant RS had to be further modified. Keeping total NA(i) constant, correction factors $\alpha(i)$ and $\beta(i)$ were determined , with:~~

~~and new values of temporally constant~~More details are provided in the supplement. During calibration, seven model parameters ~~were identified (Fig. 6, green lines).~~

~~2.3.4 RS_Q_GW_NA variant: Calibration using remote sensing data, inflow into the lake, groundwater level, and net abstractions~~

~~In the most involved calibration variant, statistical data on water withdrawals in 2009~~ (Table 1) ~~was used together with information on return flow to compute a consumptive irrigation water use Cu_{irri}^{Obs} in the basin of 3,520·10^6 m^3. To estimate irrigation use in all other years, with different climatic conditions, the per area consumptive irrigation water use from WaterGAP was used to compute, for each year, a climatic correction factor CF2(i) as~~

~~where $CF2(i)$ is represents the difference in the per area consumptive irrigation use in year i and the year 2009, $Cu_{irri}^{WG}(i)$ is consumptive irrigation use in year i obtained in standard WaterGAP. Finally,~~were adjusted that are known to have an impact on TWSA, Q and GWSA. We used a modified version of the Kling Gupta efficiency (KGE) as the objective function, where the trend of the time series was added as a fourth component to the KGE (see Eq. ~~4 was used for estimating water consumption time series over Urmia basin:~~5 below).

~~where $Cu_{irri}(i)$ is consumptive irrigation water use in year i. Unlike in the RS_Q_GW variant, consumptive use of the other sectors was added based on withdrawal data in Table 1 and a return flow fraction of 60%, resulting in total NA. Then, new values for correction factors $\alpha(i)$ and $\beta(i)$ (Eq. 2) were identified by trial-and-error, and model parameters were modified to obtain a good fit to the data also used in the RS_Q_GW variant (Fig. 6, red lines).~~

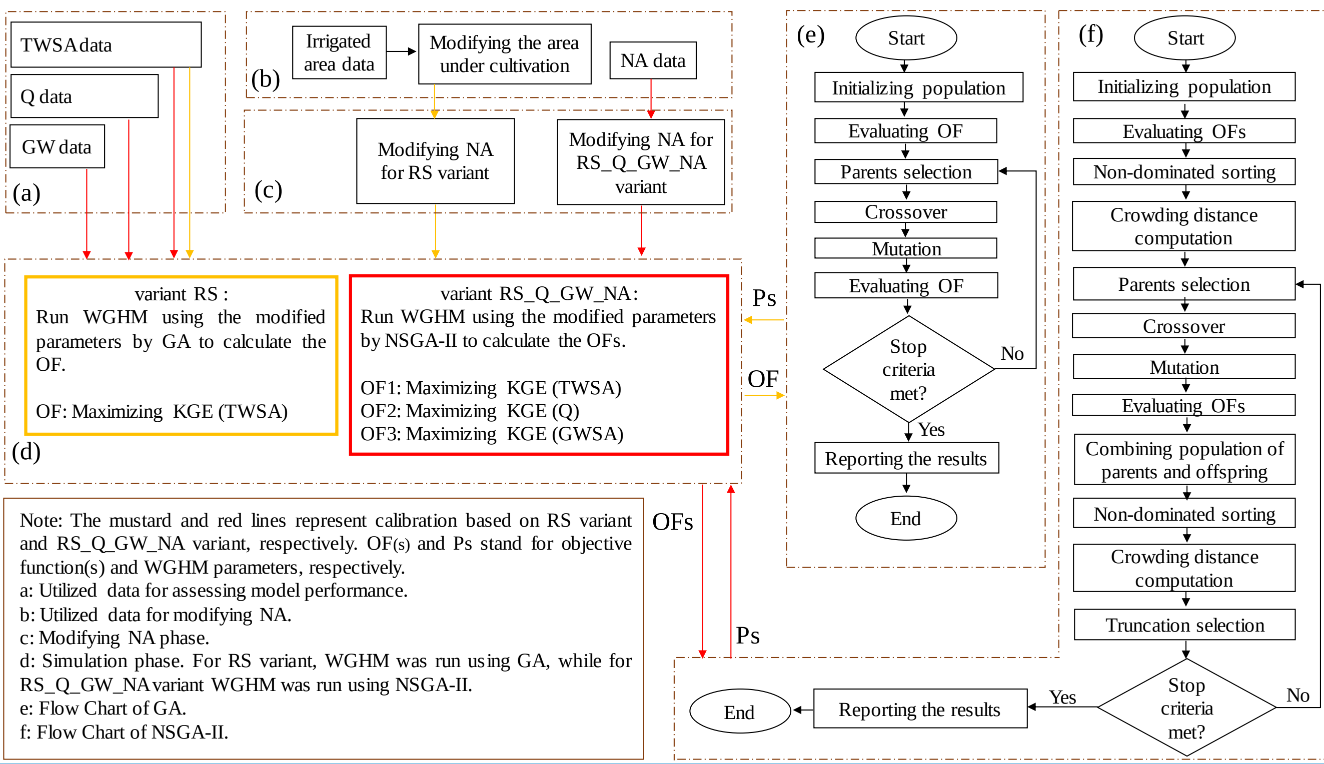


Figure 4: Flowchart of the WGHM calibration approach.

Table 1: WGHM parameters with the most effect on TWSA, inflow into the lake, groundwater storage.

Parameter	Value		
	Default	Minimum	Maximum
P1: Rooting depth multiplier	1	0.5	3
P2: Maximum active lake depth [m]	5	2	12
P3: Runoff coefficient multiplier	1	0.5	1.5
P4: Multiplier for the fraction of total runoff that becomes groundwater recharge	1	0.5	5
P5: Maximum amount of groundwater recharge per day multiplier	1	0.5	5
P6: Minimum amount of daily precipitation necessary in arid/semi-arid areas to get groundwater recharge [mm]	12.5	5	15
P7: Maximum canopy storage [mm]	0.3	0.1	1.4

2.4 Performance indicators

Performance of the ~~calibration variants of~~ WGHM was evaluated using the correlation coefficient (CC~~,~~), Nash-Sutcliffe efficiency (NSE~~,~~),, root mean square error (RMSE), relative absolute error (RAE), and a modified version of the Kling Gupta efficiency (KGE~~, Gupta et al., 2009~~) with

$$CC = \frac{Cov\ (Obs, Sim)}{\sigma_{obs} \times \sigma_{Sim}} \tag{1}$$

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Sim_{(t)} - Obs_{(t)})^2}{\sum_{t=1}^{T}(Obs_{(t)} - \overline{Obs})^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Obs_{(t)} - Sim_{(t)}\right)^2} \tag{3}$$

$$RAE = \frac{\sum_{t=1}^{T} |Obs_{(t)} - Sim_{(t)}|}{\sum_{t=1}^{T} |Obs_{(t)} - \overline{Obs}|} \tag{4}$$

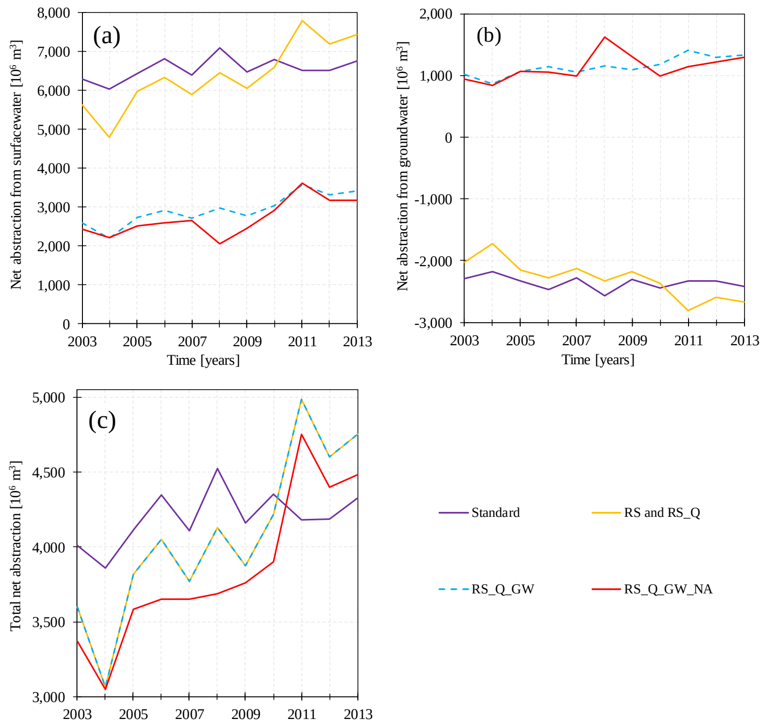
$$KGE = 1 - \sqrt{(CC-1)^2 + \left(\frac{\sigma_{Sim}}{\sigma_{obs}} - 1\right)^2 + (\frac{\overline{Sim}}{\overline{Obs}} - 1)^2 + (\frac{Trend_{Sim}}{Trend_{Obs}} - 1)^2} \tag{5}$$

where *Cov* is covariance function, σ refers to standard division, *Trend* indicates the linear trend of the time series, *Obs* is observed value, *Sim* is simulated value, *t* refers to time counter and *T* is the period length. Optimum values of CC, NSE and KGE are 1, and of RMSE and RE are 0. Trends and overall behaviour of the time series were also analysed.

3 Results

3.1 ~~Multi-observation~~ Model calibration

~~In variants RS and RS_Q, annual time series of irrigated area in Lake Urmia basin derived from MODIS (Fig. 4), which were applied in all four calibration variants, lead to a more strongly increasing trend of NA (consumptive water use) and NAs, as compared to the standard WaterGAP version (Fig. 7). Due to the dominant irrigation with surface water assumed in the standard version of WaterGAP, return flows from irrigation are larger than groundwater withdrawals, and there is a net recharge of groundwater by irrigation, i.e. a negative NAg. Therefore, a more strongly increasing irrigation with surface water in variants RS and RS_Q leads to return flows to groundwater that increase more strongly over time, i.e. NAg becomes increasingly negative with time (Fig. 7). Average NA in 2003-2010 decreased from 4,185·10^6 m^3/yr in the standard version to 3,815·10^6 m^3/yr, and increased from 4,233·10^6 m^3/yr to 4,781·10^6 m^3/yr in 2011-2013. However, increased net recharge of groundwater by return flows was found to be incompatible with decreasing observed groundwater levels (Fig. 8c). Positive NAg values were found to be necessary to simulate the observed lowering of groundwater levels from 2003 to 2013 Therefore, in variant RS_Q_GW, NAg and NAs were adjusted according to Eq. 2 by applying α and β time series presented in Table 2. With these adjustment factors, average NAg changed from -2,294·10^6 m^3/yr in variants RS and RS_Q to 1,147·10^6 m^3/yr in variant RS_Q_GW (Fig. 7b). Keeping annual NA constant, NAs decreased accordingly from 6,373·10^6 m^3/yr to 2,931·10^6 m^3/yr. Total NA slightly decreased in variant RS_Q_GW_NA as compared to the other calibrations variants.~~

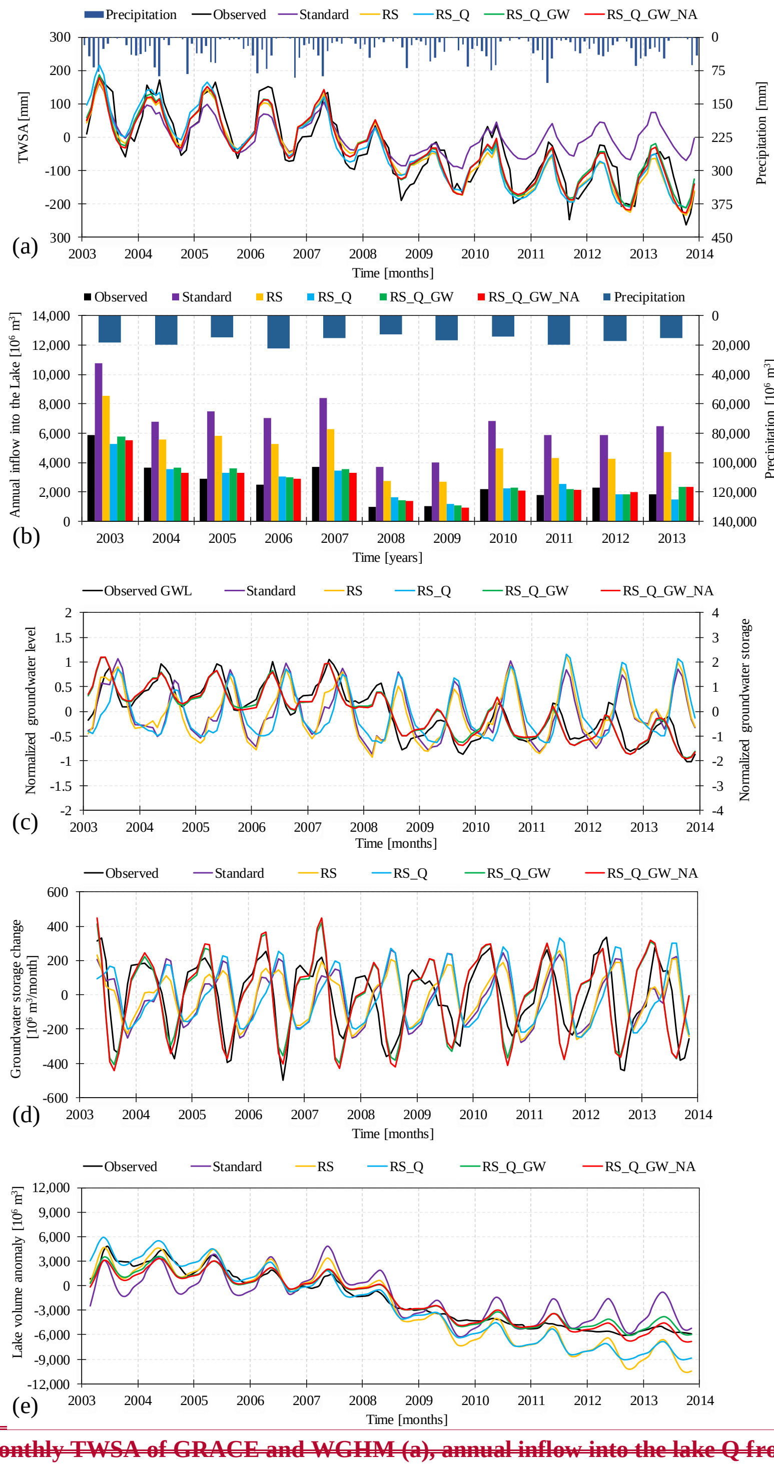


~~Figure 7: Time series of net abstractions from surface water (a) and groundwater (b), as well as total net abstractions (i.e. consumptive use) (c) in Lake Urmia basin in the standard version of WaterGAP as well as the various calibration variants.~~

~~Table 2: Correction factors for modifying NAs and NAg (see Eq. 2).~~

~~Variant~~	~~RS_Q_GW~~	~~RS_Q_GW_NA~~

~~Year~~	~~α~~	~~β~~	~~α~~	~~β~~
~~2003~~	~~0.47~~	~~-0.48~~	~~0.39~~	~~-0.41~~
~~2004~~	~~0.46~~	~~-0.49~~	~~0.37~~	~~-0.39~~
~~2005~~	~~0.46~~	~~-0.50~~	~~0.39~~	~~-0.46~~
~~2006~~	~~0.46~~	~~-0.50~~	~~0.38~~	~~-0.43~~
~~2007~~	~~0.46~~	~~-0.50~~	~~0.42~~	~~-0.43~~
~~2008~~	~~0.45~~	~~-0.52~~	~~0.29~~	~~-0.63~~
~~2009~~	~~0.46~~	~~-0.49~~	~~0.38~~	~~-0.57~~
~~2010~~	~~0.47~~	~~-0.48~~	~~0.43~~	~~-0.41~~
~~2011~~	~~0.47~~	~~-0.47~~	~~0.56~~	~~-0.49~~
~~2012~~	~~0.46~~	~~-0.51~~	~~0.49~~	~~-0.52~~
~~2013~~	~~0.45~~	~~-0.52~~	~~0.47~~	~~-0.54~~



~~**Figure 8: Time series of monthly TWSA of GRACE and WGHM (a), annual inflow into the lake Q from observations and WGHM (b) normalized observed groundwater level and normalized groundwater storage from WGHM (c), groundwater storage change GWSC from month to month from observations and WGHM (d) and the monthly lake volume anomaly (e), for standard WaterGAP and the four calibration variants.**~~

~~Model runs driven by the different NAg and NAs of the four variants lead to the best fit to the variant-specific observational datasets if seven model parameters were re-set to the values listed in Table 3. It is emphasized that the listed parameter sets are not the only possible ones but those requiring the least number of parameters to be changed. In all four calibration variants, the minimum daily precipitation values for which groundwater recharge can occur in semi-arid regions (Döll and Fiedler, 2008) was slightly decreased (increasing groundwater recharge) and the maximum canopy storage was increased (increasing canopy evaporation). When the more observational data types were considered in the calibration process, the number of parameters that needed to be adjusted increased whereas the required parameter changes decreased.~~

~~According to GRACE observations, total water storage in Lake Urmia basin declined by $9.9 \cdot 10^9$ m^3 from its annual average in 2003 to its annual average in 2013, while the standard~~First, NA was adjusted based on either MODIS data only (variant RS) or MODIS data and information of basin water use (variant RS_Q_GW_NA) (section S3). Then, optimal model parameters were identified using GA and NSGA-II for both variants. Figure 5a shows the calibration history of WGHM based on the best performance of GA among five runs for the variant RS. GA started from a KGE value with respect to TWSA near 0.60 and reached to 0.87 after about 5,000 functional evaluations (WGHM runs). Figure 5b illustrates the final Pareto fronts obtained by five runs of NSGA-II for the variant RS_Q_GW_NA. For the variant RS_Q_GW_NA after about 12,000 functional evaluations (for each NSGA-II run), NSGA-II found 18 optimal parameter sets. Figure 6 shows the parameter ranges (5 and 18 values for each parameter for variants RS and RS_Q_GW_NA) obtained by five different runs of GA and NSGA-II in RS and RS_Q_GW_NA variants. Then, an ensemble of WGHM simulations was generated for the variants RS and RS_Q_GW_NA which comprises the model runs with the optimal parameter sets.

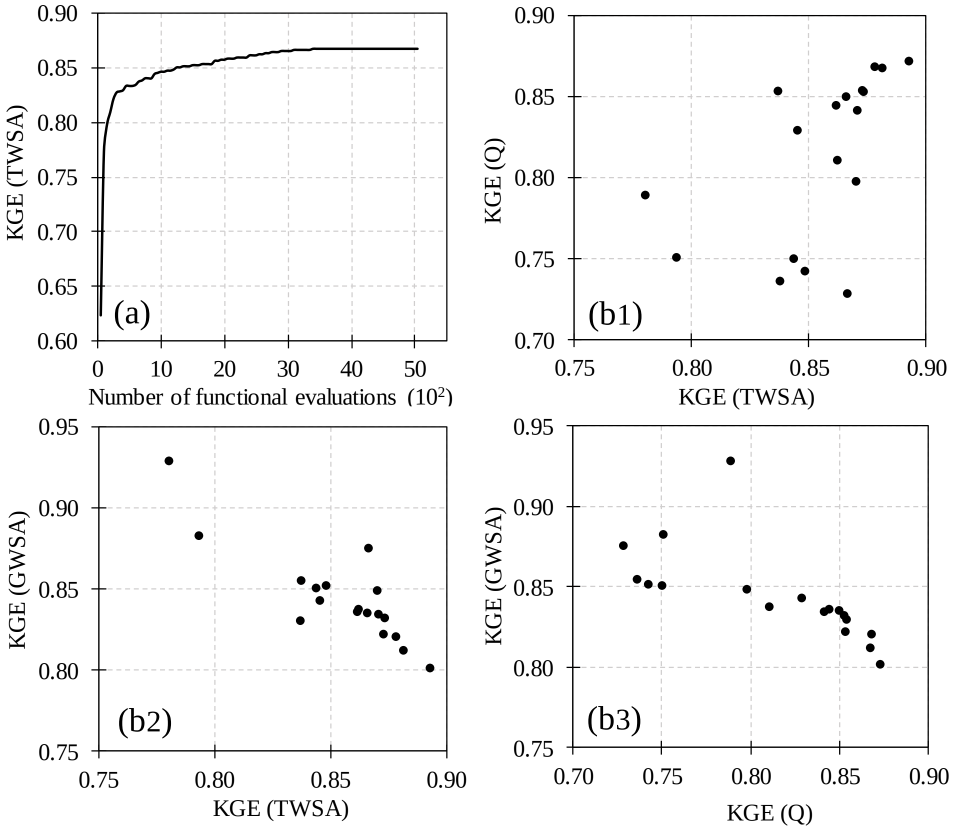


Figure 5: Best convergence history of GA in calibrating WGHM for the variant RS (a) and Pareto fronts for the multi-objective calibrations generated by NSGA-II for the variant RS_Q_GW_NA (b1), (b2) and (b3).

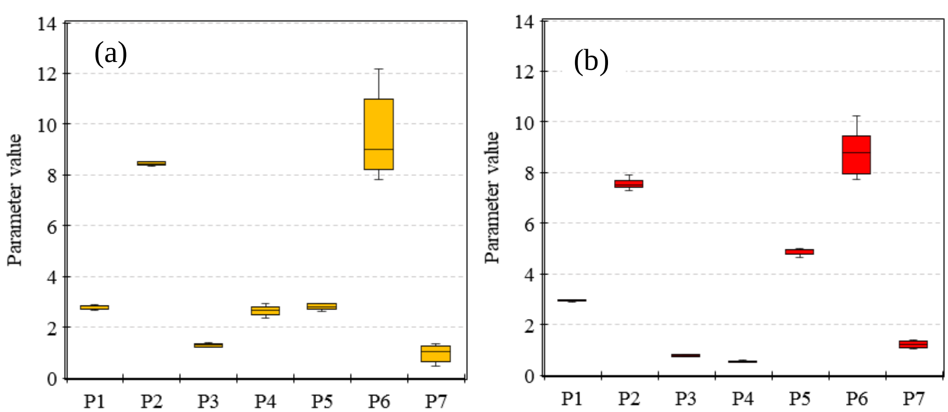


Figure 6: Adjusted WGHM ~~version computes~~ parameter values for variant RS (a~~ much smaller loss. According to~~) and RS_Q_GW_NA (b).

Figure 7 compares the ~~data of Tourian et al. (2015), about 80 %~~output of the calibrated model ensembles (variants RS and RS_Q_GW_NA) with observations and the output of the standard version of WGHM. The minimum and maximum value of each variable in each time period are shown as uncertainty bound of the results in each variant. Standard WGHM underestimates total water ~~loss~~storage decline in the Lake Urmia basin ~~was~~between 2003 and 2013as compared to GRACE observations. A good fit to GRACE results in calibration variant RS, due to ~~the loss of lake water.~~1) a stronger increase of human water abstractions over time ~~(Fig. 7a), doubling~~as indicated by MODIS (Fig. S4), 2) an almost tripling of rooting depth and thus soil water capacity ~~and~~(P1), 3) an increased fraction of runoff that recharges the groundwater (P4-P6) and a 4) a higher maximum canopy storage everywhere in the basin~~, as well as~~ (P7) and 5) an increase of maximum active lake depth of Lake Urmia from 5 m to ~~9~~more than 8 m in variant RS ~~resulted in a good fit of WGHM TWSA to GRACE TWSA (Fig. 8a~~(P2) (Figs. 6a and 7a). With the larger soil and canopy water storage capacities, runoff and thus inflow into Lake Urmia decrease as compared to standard WGHM (Fig. ~~8b). More water could be stored in canopy, soil, and the lake at the beginning of the period such that storages could react to the decline of inflows and decrease after 2007.~~7b). Still, simulated inflows into Lake Urmia computed in variant RS are still much higher than the observed values (Fig. ~~8b~~7b) and seasonality of groundwater ~~levels~~storage is totally misrepresented (~~Figs. 8c, d).~~

~~**Table 3: WGHM parameter values adjusted by calibration in the different model variants.**~~

~~Variant~~	~~Rooting depth multiplier~~	~~Maximum active lake depth~~	~~Runoff coefficient multiplier~~	~~Multiplier for the fraction of total runoff that becomes groundwater recharge~~	~~Maximum amount of groundwater recharge per day multiplier~~	~~Minimum amount of daily precipitation necessary in arid/semi-arid areas to get groundwater recharge [mm]~~	~~Maximum canopy storage [mm]~~
~~Standard~~	~~1~~	~~5~~	~~1~~	~~1~~	~~1~~	~~12.5~~	~~0.3~~
~~RS~~	~~2~~	~~9~~	~~1~~	~~1~~	~~1~~	~~10~~	~~1~~
~~RS_Q~~	~~2.8~~	~~10~~	~~0.9~~	~~1~~	~~1~~	~~10~~	~~1~~
~~RS_Q_GW~~	~~3~~	~~9~~	~~0.8~~	~~0.5~~	~~4~~	~~10~~	~~1~~
~~RS_Q_GW_NA~~	~~3~~	~~8~~	~~0.8~~	~~0.5~~	~~5~~	~~10~~	~~1~~

Fig. 7c). The required reduction of computed lake inflow (Q) can be achieved ~~(Fig. 8b) by further increasing soil water storage capacity~~ in variant RS_Q~~, together with small~~_GW_NA by adjustment of the runoff coefficient and ~~active lake depth (Table 3~~a slight further increase in maximum soil and canopy storage (Fig. 6), while the fit to GRACE TWSA remains good (Fig. ~~8a~~7a). However, the seasonality of groundwater ~~table fluctuations is still not simulated properly. This~~ storage could only be achieved by adjusting the sources of total net abstractions~~. Only if net abstractions from groundwater are multiplied by approximately -0.5 (Table 2),~~ in variant RS_Q_GW~~, does the seasonality of computed groundwater storage variations fit to observations (Fig. 8c).~~ _NA (Fig. 7c). NAg in the standard~~, RS~~ and RS~~_Q~~ variants is negative, which means that there is an artificial groundwater recharge due to irrigation by surface water during the summer irrigation months, leading to an increase in groundwater ~~level and~~ storage. Groundwater ~~level~~storage observations, however, show a decrease during this period, indicating that irrigation causes a net abstraction from groundwater. ~~Multiplication of standard WGHM NAg by a negative~~

~~value leads to a net abstraction of water from the groundwater body, and results in a seasonality of groundwater storage that fits well to the seasonality of the mean groundwater table in the basin. In addition to the NAg and NAs adjustment, two groundwater recharge-related parameters had to be re-set in variant RS_Q_GW (Table 3). The fit to observed TWSA and lake inflow remains good (Figs. 8a, b). Use of local information on water withdrawals and return flows in variant RS_Q_GW_NA barely changed the parameter values (Table 3) and the fit to all observational data (Fig. 8~~Therefore, annual values of NAg as computed by WGHM were multiplied, in variant RS_Q_GW_NA, by a negative correction factors (Table S2).

~~From the results of the RS_Q_GW_NA variant, which was the most comprehensive calibration variant, we estimated the average specific yield of the aquifers in the Lake Urmia Basin, i.e. the change in groundwater storage per unit change of the elevation of the groundwater table. We first divided the standard deviation of the simulated groundwater storage time series by the basin area to obtain groundwater storage variability in terms of equivalent water height and then divided this value by the standard deviation of the observed groundwater levels. This resulted in a specific yield estimate of 0.02, which is equal to the average value derived from pumping tests at 10 locations south of the lake (Hamzekhani and Aghaie, 2015). Estimated specific yield allows to compute an "observed" groundwater water storage anomaly, and thus an observed decline of groundwater storage between the year 2003 and 2013 of $1.8\cdot10^{9}$ m^{3}, accounting for 18% of the observed total water storage loss in the basin. We compared the time series of simulated groundwater storage changes from month to month (GWSC) to those derived from observations of groundwater level changes. Since groundwater level observations were done only once per month and at different days, three-month moving averages were compared (Fig. 8d). Observations and model variants RS_Q_GW and RS_Q_GW_NA agree that the strongest monthly increase in groundwater storage occurs in early spring, and the largest decrease in early autumn.~~

~~The~~ Performance indicators CC, NSE, RMSE, RAE, and KGE with respect to monthly TWSA (Fig. ~~8a~~7a), annual Q (inflow to Lake Urmia, Fig. ~~8b~~7b) and monthly ~~GWSC~~GWSA (Fig. ~~8d~~7c) are presented in Table ~~4~~2 for the standard version and ~~four calibrated~~the ensemble means of the two calibration variants. Regarding the fit to TWSA observations, NSE increased from 0.48 in the standard version to 0.~~84~~86 in the RS variant for which TWSA was the only observation considered, and increased slightly to 0.88 when groundwater observations were taken into account in ~~variants RS_Q_GW and~~ RS_Q_GW_NA ~~variants~~variant. This performance improvement is also reflected by CC, RMSE, RAE, and KGE. Although the performance of WGHM with respect to the observed lake inflow ~~to the lake only improves marginally by~~ was improved in the RS variant, the variant does not yet provide reliable simulations of lake inflow. The calibration against ~~TWSA, in variant RS. Only calibration against~~ inflow observations in variant RS_Q_GW_NA strongly improves ~~model performance~~inflow simulation, with NSE and KGE jumping from negative values for the standard variant to values ~~around~~0.93 and 0.~~9 and RAE from 3.92 to 0.30. Integration of groundwater observations again leads to a small performance improvement (see also RMSE).~~82, respectively. The good performance shown by CC for all model variants indicates that all model variants identify correctly high and low flow years. In ~~the~~ case of ~~GWSC~~GWSA, all performance indicators show that consideration of remote sensing ~~and streamflow observations~~data only ~~do~~does not lead to an acceptable simulation of groundwater storage. Only the ~~two~~

~~variants~~variant for which groundwater observations were taken into account lead to satisfactory performance. ~~With a maximum NSE of 0.59 and KGE of 0.75, the fit to GWSC remains lower than the one to TWSA and~~

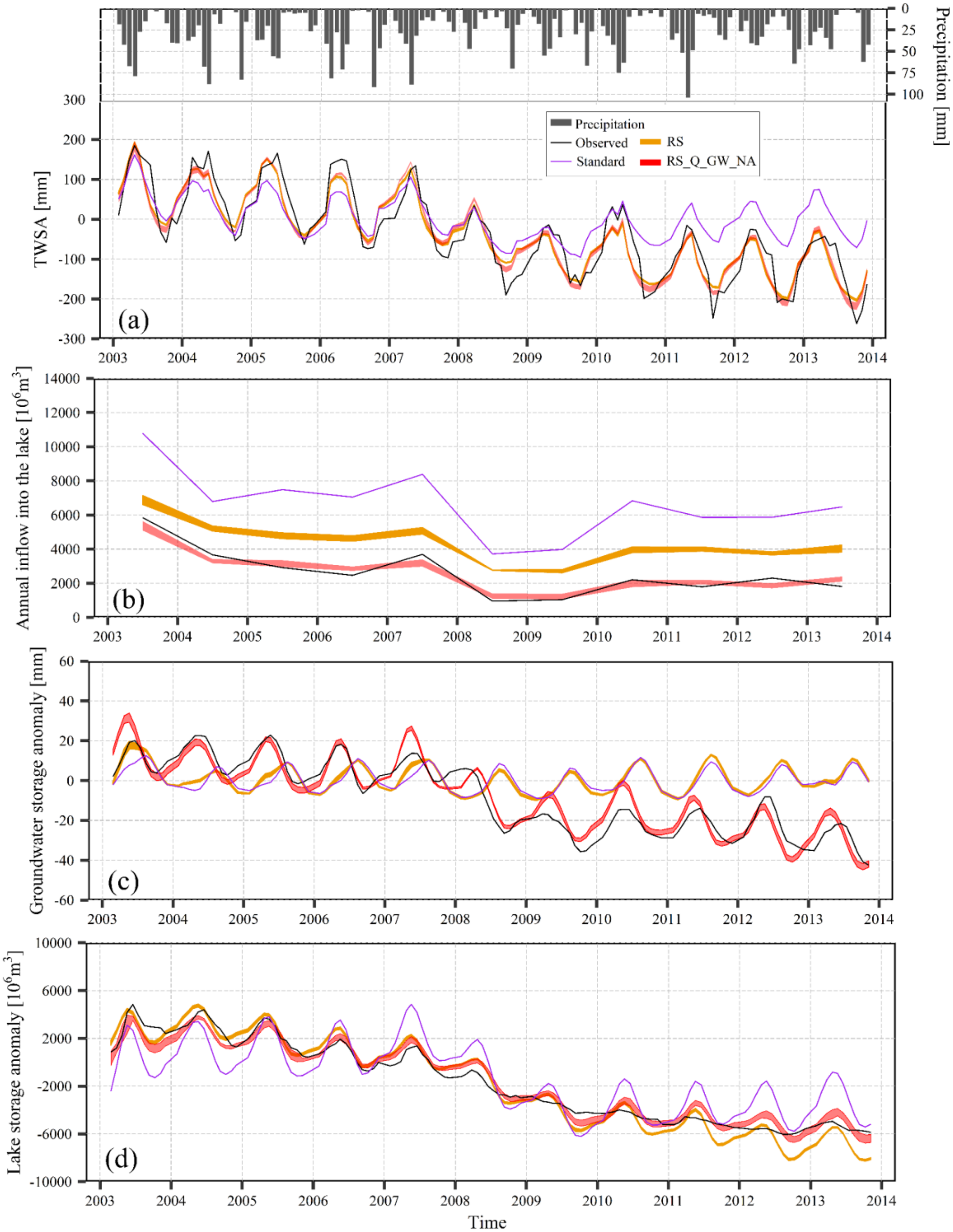


Figure 7: Time series of monthly total water storage anomaly TWSA (a), annual lake inflow~~, which may also be due to the uncertainty in estimating the basin-wide average~~ Q (b), monthly groundwater storage ~~behavior from well observations. The most data-demanding variant RS_Q_GW_NA achieves the best fit to all three observational time series. The fit, however, is only slightly better~~

~~**than the fit of variant RS_Q_GW, and a much more variable time series of NAg and NAs correction coefficients (Table 2) is necessary as compared to variant RS_Q_GW (Table 2).**~~**anomaly GWSA (c) and monthly lake volume anomaly (d), from observations, standard WGHM and the two calibration variants RS and RS_Q_GW_NA.**

For model performance evaluation, we compared the lake volume simulated by WGHM with the observed lake volume of Tourian et al. (2015) (Fig.~~8e~~7d and Table ~~4~~2). The standard model underestimates the decline in both lake water and TWSA, ~~all~~both calibrated variants simulate the TWSA trend correctly, but ~~both~~ variant RS~~ and RS_Q, with worse KGE than the standard version~~, overestimate the decline of lake water storage, thus compensating for not decreasing sufficiently groundwater storage (Fig. ~~8d~~7c) due to assuming a net groundwater recharge due to surface water irrigation. Only ~~variants RS_Q_GW and~~variant RS_Q_GW_NA ~~simulate~~simulates not only the groundwater dynamics but also the decline of lake water volume correctly. ~~NSE~~KGE for the monthly lake volume anomaly is 0.~~68~~52 for the standard WGHM and improves to 0.~~77~~75 for RS~~, where GRACE TWSA could be simulated well by approximately doubling both soil and lake water storage capacity (Table 3).~~. Including groundwater level data further improved the fit to observed lake volume, leading to a very high ~~NSE~~KGE of 0.~~94 or 0.95~~89 (Table ~~4~~2). We conclude that the calibration of WGHM against diverse observations (that do not include lake volume observations) leads to improved simulation of lake volume dynamics.

~~**Table 4: Performance of standard and calibrated WGHM variants with respects to observations of TWSA, inflow to lake, GWSC and lake volume anomaly**~~

~~Phase~~	~~Variables~~	~~Criteria~~	~~Standard~~	~~RS~~	~~RS_Q~~	~~RS_Q_GW~~	~~RS_Q_GW_NA~~
~~Calibration~~	~~Monthly TWSA~~	~~CC~~	~~0.84~~	~~0.93~~	~~0.92~~	~~0.94~~	~~0.94~~
		~~NSE~~	~~0.48~~	~~0.84~~	~~0.83~~	~~0.88~~	~~0.88~~
		~~RMSE [mm]~~	~~77~~	~~42~~	~~44~~	~~38~~	~~37~~
		~~RAE~~	~~0.72~~	~~0.41~~	~~0.42~~	~~0.37~~	~~0.36~~
		~~KGE~~	~~0.64~~	~~0.80~~	~~0.79~~	~~0.82~~	~~0.83~~
	~~Annual Q~~	~~CC~~	~~0.94~~	~~0.96~~	~~0.95~~	~~0.97~~	~~0.97~~
		~~NSE~~	~~-8.51~~	~~-2.33~~	~~0.88~~	~~0.91~~	~~0.93~~
		~~RMSE [10^6 m^3/year]~~	~~4121~~	~~2438~~	~~458~~	~~390~~	~~358~~
		~~RAE~~	~~3.92~~	~~2.32~~	~~0.38~~	~~0.33~~	~~0.30~~
		~~KGE~~	~~-0.60~~	~~0.07~~	~~0.84~~	~~0.88~~	~~0.91~~
	~~Monthly GWSC~~	~~CC~~	~~-0.14~~	~~0.05~~	~~-0.31~~	~~0.80~~	~~0.82~~
		~~NSE~~	~~-0.72~~	~~-0.39~~	~~-1.05~~	~~0.55~~	~~0.59~~
		~~RMSE [10^6 m^3/month]~~	~~271~~	~~244~~	~~296~~	~~109~~	~~103~~
		~~RAE~~	~~1.28~~	~~1.13~~	~~1.42~~	~~0.60~~	~~0.58~~
		~~KGE~~	~~-0.57~~	~~-0.44~~	~~-0.79~~	~~0.71~~	~~0.75~~
~~Evaluation~~	~~Monthly lake volume anomaly~~	~~CC~~	~~0.82~~	~~0.97~~	~~0.99~~	~~0.98~~	~~0.97~~
		~~NSE~~	~~0.68~~	~~0.77~~	~~0.81~~	~~0.94~~	~~0.95~~
		~~RMSE [10^6 m^3]~~	~~1922~~	~~1837~~	~~1611~~	~~757~~	~~739~~
		~~RAE~~	~~0.51~~	~~0.47~~	~~0.42~~	~~0.21~~	~~0.20~~
		~~KGE~~	~~0.70~~	~~0.34~~	~~0.41~~	~~0.88~~	~~0.90~~

Table 2: Performance of standard and calibrated WGHM variants with respects to observations of TWSA, inflow to lake, GWSA and lake volume anomaly.

Phase	Variables	Criteria	Standard	RS	RS_Q_GW_NA
Calibration	Monthly TWSA	CC	0.84	0.93	0.94
		NSE	0.48	0.86	0.88
		RMSE [mm]	77	40	37
		RAE	0.72	0.39	0.36
		KGE	-0.36	0.85	0.86
	Annual Q	CC	0.94	0.97	0.97
		NSE	-8.51	-0.75	0.93
		RMSE [10^6 m^3/yr]	4121	1767	356
		RAE	3.92	1.67	0.33
		KGE	-0.61	0.29	0.82
	Monthly GWSA	CC	0.03	0.16	0.95
		NSE	-0.31	-0.28	0.89
		RMSE [mm]	21	20	6
		RAE	1.07	1.04	0.30
		KGE	-0.87	-0.83	0.85
Evaluation	Monthly lake volume anomaly	CC	0.82	0.98	0.98
		NSE	0.68	0.92	0.96
		RMSE [10^6 m^3]	1922	928	656
		RAE	0.51	0.25	0.18
		KGE	0.52	0.75	0.89

3.2 Differential impacts of human water use and climate variation on Lake Urmia basin

The impact of human water use and man-made reservoirs on water flows and storages was quantified by comparing the output of WGHM in which human water use and man-made reservoirs are considered (this is normally done, now called WGHM-ANT) with the output of a model run for naturalized conditions, where it is assumed that there are no reservoirs and no human water use (WGHM-NAT). We determined that the results of the naturalized run differ by less than 2% from a run with reservoirs but without human water use. Therefore, differences between WGHM-ANT and WGHM-NAT outputs can be considered to be caused by human water use. It should be mentioned that all simulated and observed storages (total, groundwater, lake) are not absolute values but anomalies with respect to the mean water storage during 2004-2009 (baseline period used for the provided GRACE data). Moreover, to quantify the uncertainty in the model calibrations, WGHM-ANT and WGHM-NAT were run based on all 18 optimal parameter sets were obtained from Pareto front for variant RS_Q_GW_NA. All results were presented by min-max ranges.

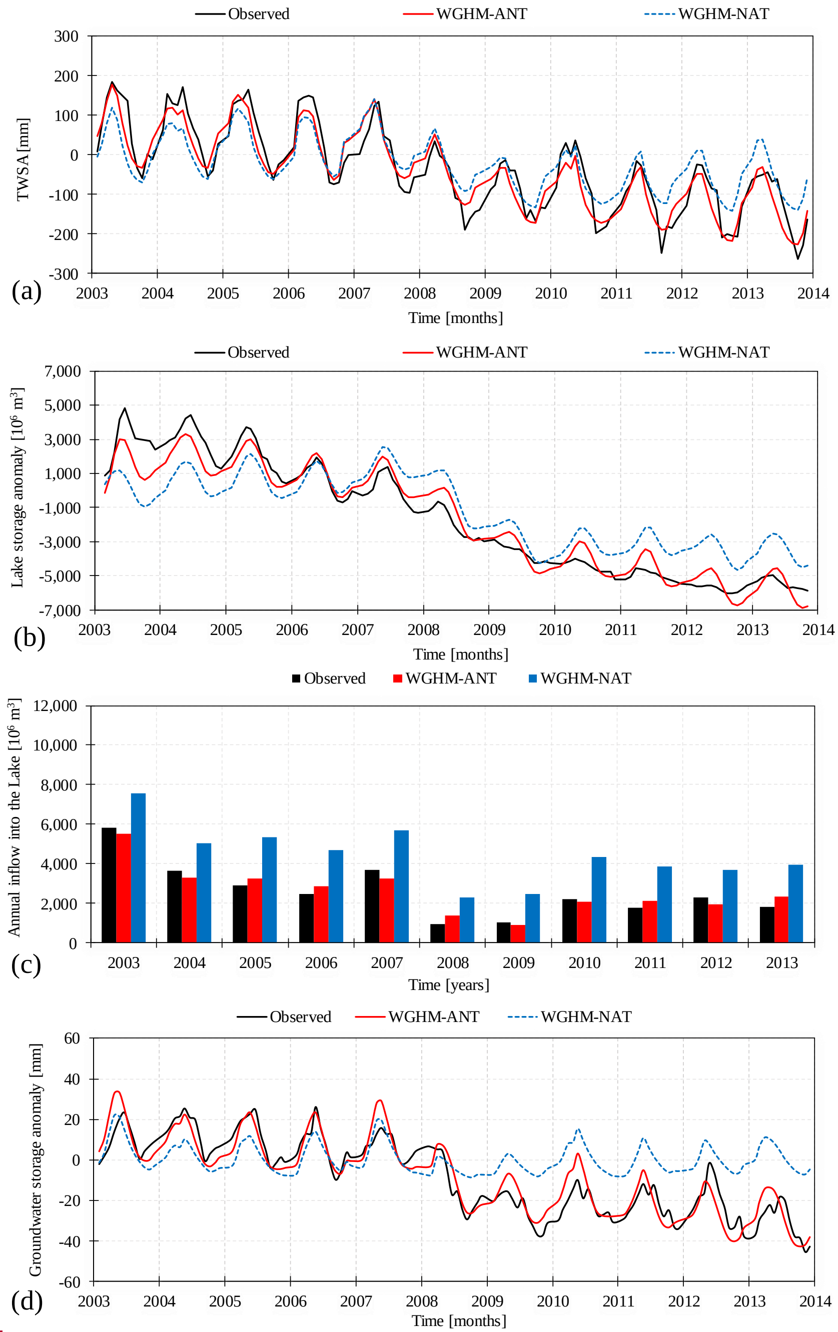
When comparing TWSA under anthropogenic and naturalized conditions in Fig. ~~9a~~8a, remember that TWSA in Lake Urmia basin is dominated by water storage in Lake Urmia. Seasonal TWSA variation of WGHM-ANT and WGHM-NAT do not differ much. Starting after the heavy rain in April 2007 and strongly caused by the lack of spring precipitation in 2008, both WGHM-ANT and WGHM-NAT (as well as GRACE TWSA) show a decreasing trend that is only somewhat more pronounced in WGHM-ANT (Fig. ~~9a~~8a). Thus, this decrease is mainly due to dry climate conditions during the well-known severe drought of 2008, with ~~an~~ annual precipitation of only 241 mm, i.e. 74% of the mean value for 2003-2013~~ (Fig. 8b).~~.

Also in the absence of human water use, total water storage would not have recovered after 2009 but would have stayed 50-100 mm below the values occurring before 2008. However, while in WGHM-NAT the minimum storage in late summer, i.e. the period with high irrigation, remains almost at a constant level after 2009, it decreases each year in WGHM-ANT due to consumptive increasing irrigation water use (see Fig. ~~7c~~S4). The linear ~~trend~~trends of WGHM-ANT and WGHM-NAT TWSA time series for the period 2003-2013 ~~is -24.5~~are between -23.6 and -25.1 mm/yr (GRACE: -24.4 mm/yr) and between -10.1 and -11.~~8 mm/yr~~9, respectively. The TWSA trend for two sub-periods before and after 2008, 2003-2007 and 2009-2013 ~~-14.2~~[-11.7, -18.5] and ~~-~~[-10.6,-16.3] mm/yr, respectively, for WGHM-ANT and only [-1.8,3.3] and [-2.9,-0.~~7 and -3.85~~6] mm/yr, respectively, for WGHM-NAT. The last-mentioned trends are not significant at the 5% confidence level based on Mann-Kendall's test. According to WGHM, the basin lost, on average during 2003-2013, between 1,~~274~~226·10^6 and 1,305·10^6 m^3 water/yr, while in the absence of human water use, it would have lost ~~614~~between 524·10^6 and 618·10^6 m^3 water/yr, i.e. 52-57% less. Of this total water volume~~, 920~~ between 914·10^6 and 975·10^6 m^3/yr of lake water was lost, while only ~~548~~523·10^6 and 598·10^6 m^3/yr would have been lost without human water use (Fig. ~~9b~~8b).

The smaller decreasing trend for lake water volume under naturalized conditions is clearly caused by more inflow into the lake, even though lake evaporation is somewhat higher under naturalized inflow conditions due to the larger lake extent. While mean inflow during 2003-2013 is computed to be ~~4,454~~between 4,323·10^6 and 4,685·10^6 m^3/yr under naturalized conditions, it decreases by ~~41%~~39-45% reached to between 2,~~639~~463·10^6 and 2,742·10^6 m^3/yr under anthropogenically altered conditions (Fig. ~~9c~~8c). The difference is only 50% of NA as only a fraction of (potential) net abstractions from surface water NAs (required to allow optimal irrigation) could be made 1) due to a lack of water availability in the surface water bodies and 2) because a fraction oft of NAg is provided a decrease in groundwater storage. Since 2008 the inflow into the lake has never reached 3,085·10^6 m^3/yr. This is the value estimated to be the minimum environmental water ~~requirements~~requirement that compensates the amount of annual evaporation from ~~of~~ the lake surface (Abbaspour and Nazaridoust, 2007). Therefore, a decrease ~~of~~in lake water storage can be expected for the ~~best estimate of~~estimated inflow by WaterGAP ~~of~~between 2,~~639~~463·10^6 and 2,742·10^6 m^3/yr~~.~~ during 2003-2013. In WGHM-NAT, the inflow was lower than 3,085·10^6 m^3 only in 2008 and 2009. Still, the average inflow into the lake from 2009-2013 of between 3,~~670~~528·10^6 and 3,840·10^6 m^3/yr would have been only enough to keep the lake from further ~~loosing~~losing volume (needed to compensate for lake evaporation). Thus even in the WGHM-NAT, inflow into the lake would not have been enough for a recovery to conditions between 2003 and 2007 (Fig. ~~9b), as during this time period, mean inflow under naturalized conditions would have been 54% larger. The ratio of inflow into the lake over precipitation in the basin varies strongly among the years, reaching a maximum value of 0.30 and 0.41 for anthropogenic and naturalized conditions, respectively, in 2003, and a minimum value of 0.11 and 0.18 in the drought year 2008. For the period 2009-2013, these ratios are, with 0.11 (ANT) and 0.22 (NAT), much smaller than the values for 2003-2007, 0.21 and 0.32. Thus, the drought year 2008 as well as the relatively small ratio of inflow into the lake over precipitation in the last five years of the study period play an equally important role as human water use in the decline of inflow and lake water storage.~~8b).

~~While~~ Groundwater storage is estimated to decline by ~~251~~between 239·10^6 and 267·10^6 m^3/yr during 2003-2013 in WGHM-ANT, the decline is only ~~27~~between 24·10^6 and 35·10^6 m^3/yr in WGHM-NAT (Fig. ~~9d~~8d). Different from lake water storage, groundwater storage would have recovered after 2008/2009 if there had been no (increasing) net groundwater abstractions (Fig. ~~9d~~8d, compare Fig. ~~7b~~S4b), even though mean groundwater recharge ~~was on~~were between 2,~~579~~340·106 and 3,103·106 m3/yr during 2009-2013 as compared to ~~3,310~~between 3,091·10^6 and 4,179·10^6 m^3/yr during ~~2009-2013. In WGHM, the groundwater compartment is modelled using a linear storage model where the change of groundwater storage is the difference between inflows to groundwater and outflow to surface water bodies, supplement by a prescribed outflow due to human groundwater use in case of anthropogenic conditions. Long-term average outflow from groundwater to surface water is proportional to the groundwater storage. Therefore, in case of less groundwater recharge, also the outflow to surface water bodies is decreased, while mean groundwater storage decreases only slightly, in particular in areas with a low average groundwater recharge like the Lake Urmia basin. In the absence of groundwater abstractions, the groundwater level cannot drop below the level of the surface water in WGHM. WGHM cannot simulate the case where groundwater switches from discharging groundwater to surface water bodies to receiving water from rivers and other surface water bodies. In case of groundwater abstractions, however, storage can drop below the level of the surface water, and outflow to surface water bodies ceases in this case.~~

~~In the WGHM-ANT simulations, such a drop below the surface water level, indicated by a negative water storage, value occurs in 7 out of the 22 0.5° grid cells within the basin (Fig. A1a). In 6 of these 7 grid cells, groundwater levels were stable during~~ 2003-~~2007and only declined from 2008-2013, caused by increased NAg and decreased groundwater recharge in the latter part of the study period. It is these 7 cells that cause the basin groundwater decline under anthropogenic condition shown in Fig. 9d. For naturalized conditions, peak seasonal water storages decrease somewhat but minimum water storages cannot drop appreciably given the very low minimum seasonal storage values already during the relatively wet five first years of the growing period (Fig. A1b). Thus, the contribution of human water use to groundwater storage decline might therefore be overestimated as WaterGAP cannot simulate a possible drop of the groundwater table below the surface water level in the absence of groundwater abstractions.~~2007. To summarize, human water use was the reason for 52-57% of the total water loss in the basin, for a maximum of 87-90% of the groundwater loss and for ~~40~~39-43% of the Lake Urmia water loss during 2003-2013, and lake inflow was ~~41~~39-45% less than it would have been without human water use.



~~Figure 9: Time series of simulated (variant RS_Q_GW_NA) and observed monthly TWSA (a), lake water storage anomaly (b), annual inflow into the lake (c), and monthly groundwater storage anomaly (d), under anthropogenic (WGHM-ANT) and naturalized (WGHM-NAT) conditions.~~

4 Discussion

4.1 ~~Multi-observation~~Model calibration

~~The output of hydrological models at all scales is uncertain as these models suffer from uncertain model inputs (e.g., climate variables or soil properties), parameter values and model structure (Döll et al., 2016). To decrease uncertainty, model calibration against independent data (e.g. observations) is performed by adjusting, for example, model parameters. While observations of river discharge are ideally suited for validating hydrological models because the point observation integrates over processes in the whole upstream basin of the gauging station, additional types of observations have to be added to avoid the well-known problem of equifinality (Beven and Freer, 2001; Döll et al., 2016). Without additional data, more than one parameter combination can lead to a good fit to e.g. observed river discharge. While e.g. total groundwater storage dynamics would be simulated very differently by model variants with the parameter sets that simulate river discharge time series equally well.~~

Global hydrological models suffer from a~~ particularly~~ high uncertainty, in particular as model inputs are uncertain. For example, climate input data are based on low-density climate observations and information on water use is often very scarce and outdated. For modelling at the global scale, it is generally not possible to obtain, the same detailed data for a specific region compared to the case that modelling this region only. Still, a global hydrological model includes all data for simulating water flows and storages in specific regions of interest everywhere on the globe, and model calibration against multiple (regional) observations is a means for improving the performance of the global model regionally. In this way,~~ an~~ efficient simulation of regional water flows and storages can be achieved, possibly as an alternative to a costlier setup of a regional model. More importantly, the regional-scale multi-observation calibration done in this study can serve to inform efforts for global-scale but region-specific multi-observation calibration of global hydrological models that would allow to strongly improve the performance of global hydrological models at the scale that they are made for (Döll et al., 2016).

Remote sensing data are the most accessible data for calibration of global hydrological models, including TWSA from GRACE. Therefore, the model variant RS only used globally available RS data, MODIS and GRACE data products. However, MODIS data can only be used to determine the temporally variable extent of irrigated areas in dry regions of the globe such that the important adjustment of temporal dynamics of statistics-based irrigated areas is not possible everywhere. GRACE TWSA quantify the anomalies and changes of water storage aggregated over all land water storage compartments such as snow, soil, groundwater, lakes, wetlands, and rivers. Considering GRACE TWSA improved the simulation of the important water storage compartment Lake Urmia. However, the unsatisfactory simulation of inflow into Lake Urmia and of groundwater dynamics clearly shows that a good fit to observed TWSA does not guarantee a good simulation of river flows or groundwater ~~storages~~storage. Still, calibration against TWSA did, even if only very slightly, improve model performance also with respect to lake inflow and groundwater dynamics.

~~By adding discharge data,~~

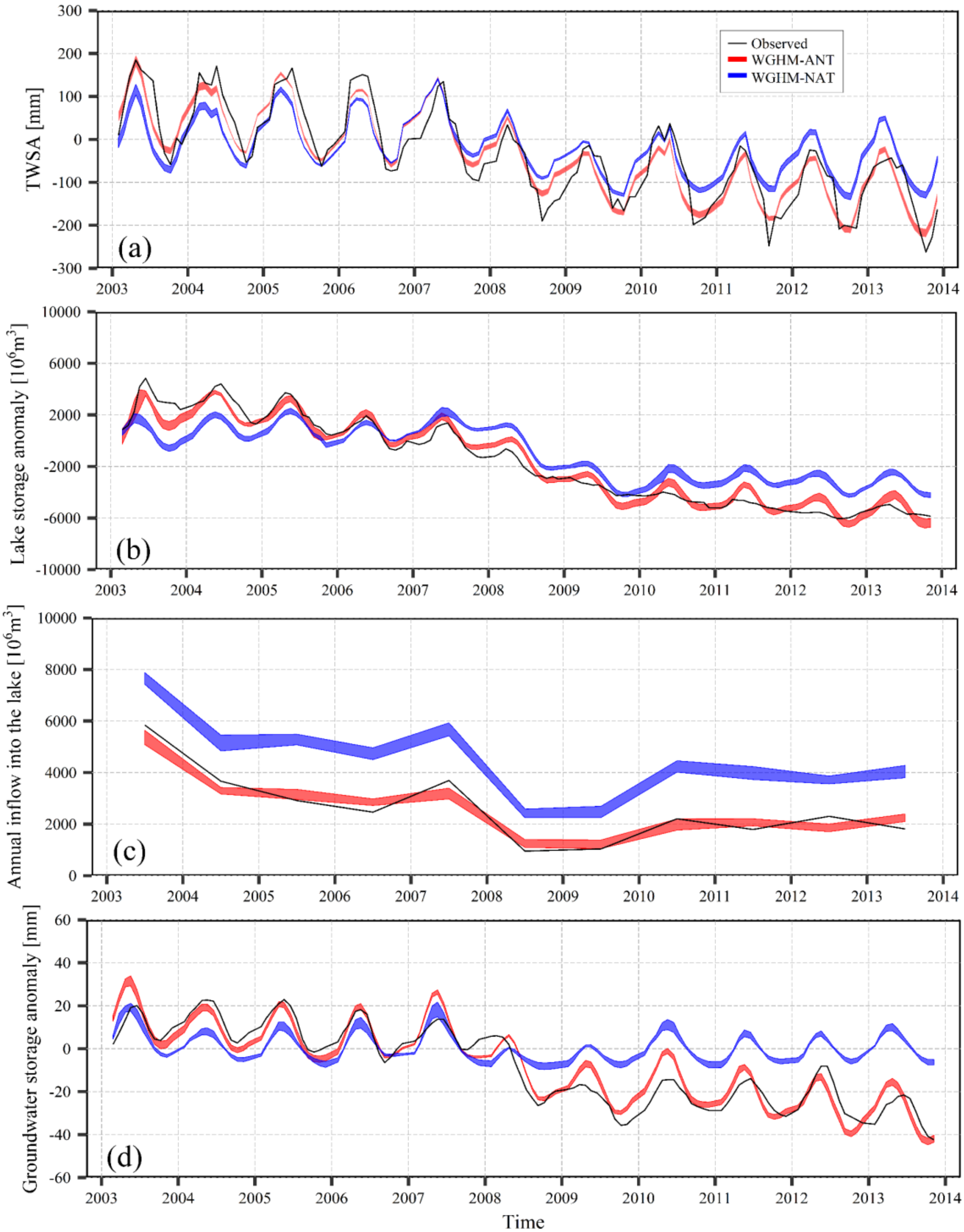


Figure 8: Time series of simulated (variant RS_Q_GW_NA) and observed monthly TWSA (a), lake water storage anomaly (b), annual inflow into the lake Q (c), and monthly groundwater storage anomaly GWSA (d), under anthropogenic (WGHM-ANT) and naturalized (WGHM-NAT) conditions.

To assess the value of using inflow into the lake (Q), groundwater observations (GW) and observed lake volume (LV) time series in model calibration, WGHM was calibrated manually based on some other variants i.e. RS_Q, RS_LV, RS_Q_LV and RS_Q_GW in a step-wise fashion (not shown). Based on the results, by adding discharge data (RS_Q variant), the model was able to simulate TWSA and Q accurately without changing the inputs of the model and only based on modifying the parameters~~, mainly increasing the rooting depth further (Table 3). Interestingly, the significant increase of the rooting depth multiplier from 2.0 to 2.8 strongly increased evapotranspiration but barely affected TWSA (Figs. 7a, b). In the case of the Lake Urmia basin, no trade-off between the fit to TWSA and river discharge exists as the performance indicators with respect to TWSA for variant RS_Q are even slightly higher than for variant RS (Table 4).~~

. Groundwater level data were found (variants RS_Q_GW and RS_Q_GW_NA) to be necessary to identify that different from what is estimated by the standard version of WaterGAP, there is more irrigation with groundwater and less with surface water such that a net abstraction of groundwater and not artificial groundwater recharge occurs due to irrigation. Information on groundwater level dynamics with a suitable spatial density is not readily available for most regions of the globe. To simulate groundwater dynamics properly, it was not sufficient to adjust parameters of the hydrological model (in particular two groundwater recharge related model parameters~~, Table 3~~ (Fig. 6b), but it was necessary to alter the fractions of net water abstractions that come from groundwater and surface water bodies. Only then, groundwater storage decline by net groundwater abstraction was simulated, and lake water storage decline could be correctly simulated instead of being overestimated when only TWSA and lake inflow data are used for calibration. As in the case of adding lake inflow as calibration data type, no trade-off between the fits to the different data types occurred.

Consideration of regional estimates of human water withdrawals in a specific year as well as regional estimates of return flow fractions in variant RS_Q_GW_NA does not improve the fit to observations compared to variant RS_W_GW significantly and only leads to slight parameter adjustments. This indicates a reasonable simulation of per hectare water consumption for irrigation by the WaterGAP model. To summarize, consideration of more and more observations and other independent data results with improved fits to three ~~type~~types of observations, TWSA, lake inflow, and groundwater dynamics, while at the same time more ~~and more~~ parameters need to be adjusted (Tables ~~3~~1 and 2 and ~~4~~Fig. 6). No trade-offs between the fits to the three observational data types occurred in the case of the Lake Urmia basin.

While the introduction of annually varying corrections for NAg and NAs (~~Eq. 2,~~ Table ~~2~~S2) for ~~variants RS_Q_GW and~~variant RS_Q_GW_NA leads to the ~~most suitable~~best fit to multiple observation types, it may be preferable to have instead of 11 free parameters just 1, i.e. a temporally constant β.With a temporally constant β of -0.5~~ in variant RS_Q_GW~~, the fit to TWSA and inflow to the lake does not change at all, and groundwater storage is only slightly increased in the dry ~~year~~years 2008 and 2009. Thus, given the uncertainty of observed groundwater storage variations, a temporally constant NAg correction factor is sufficient for achieving a good fit for all observations.

~~To assess the potential of using observed lake volume time series as calibration target and not only for model evaluation, we also calibrated WGHM against RS observations and lake volume (RS_LV variant) and against RS, lake inflow~~

~~and lake volume (RS_Q_LV variant).~~ In the RS_LV variant, simulation of TWSA and ~~GWSC~~GWSA did not change appreciably but ~~not only~~both simulated lake volume anomaly ~~but also simulated~~and lake inflow ~~into the lake~~ greatly improved as compared to the RS variant. NSE for monthly lake volume anomaly and annual lake inflow reaches 0.95 and 0.44, respectively. Inflow into the lake is much less overestimated than in variant RS. To achieve these fits, the variant RS parameters ~~where~~were adjusted ~~by increasing~~ the rooting depth multiplier to 2.5 and setting the potential evaporation multiplier to 2. Adding lake volume observations on top of lake inflow observations in RS_Q_LV variant leads to an improved fit to lake volume observations, with NSE increasing from 0.81 to 0.95, but the fit of observed inflow into the lake slightly worsens from 0.88 in RS_Q to 0.85 in RS_Q_LV. In this variant, the RS_Q variant parameters were used, except the maximum active lake depth was set to 9 m and the potential evaporation multiplier to 2. We conclude that in the case of the end lake, Lake Urmia, calibration against time series of lake volume anomalies could, in the absence of inflow data, help to improve the simulation of inflow, while calibration against time series of inflow could, in the absence of lake volume observation, improve the simulation of lake volume anomalies. Still, calibration to both observational data types leads to the best simulation of both annual lake inflow and lake volume anomalies. However, the groundwater storage dynamics could not be improved without calibration against groundwater level dynamics.

~~Finally,~~In many hydrological model calibrations, trends are not used as performance criterion. We found that model variants obtained by calibration ~~aimed at optimizing the five criteria CC, NSE, RMSE, RAE~~without a trend criterion, and ~~KGE with respect to monthly time series of observed total, groundwater and lake storages, with almost~~which have a very similar ~~achieved~~ performance ~~values (Table 4), does~~criterion, do not necessarily lead to similar estimates of total and compartmental water losses over the whole time period 2003~~ to ~~-2013. For example, using variants RS_LV and RS_~~LV have the same values for all five performance criteria (expect KGE~~ Q with ~~0.1 difference)~~ similar NSE with respect to monthly time series of TWS ~~(not shown) but~~, TWS loss between 2003 and 2013 is simulated to be ~~11.15·10^9 m^3 and~~ 7.86·10^9 m^3 and 12.20·10^9 m^3, respectively (Table ~~5~~3). TWS loss according to variant RS_Q_GW_NA ~~is, with 10.04~~(based on ensemble mean) is 9.84·10^9 m^3, ~~in between and quite different,~~ even though NSE ~~and KGE are~~is only 0.04 ~~and~~higher, while modified KGE (Eq. 5) for RS_LV, RS_Q, RS_Q_GW_NA is 0.68, 0.~~06 better,~~71, and 0.86 respectively. We conclude that in the case of relevant trends, the calibration criteria should include the minimization of the difference between observed and simulated trends.

~~**Table 5. Water loss in Lake Urmia basin between 2003 and 2013 as observed and simulated by the different calibrated WGHM variants.**~~

	~~Water loss between 2003 and 2013 [10^9 m^3] (mean annual storage in 2003 minus mean annual storage in 2013)~~							
	~~Observed~~	~~Standard~~	~~RS~~	~~RS_LV~~	~~RS_Q~~	~~RS_Q_LV~~	~~RS_Q_GW~~	~~RS_Q_GW_NA~~
~~Total~~	~~9.9~~	~~3.62~~	~~11.15~~	~~7.86~~	~~12.20~~	~~8.24~~	~~9.78~~	~~10.04~~
~~Groundwater~~	~~1.8~~	~~0.17~~	~~0.11~~	~~0.06~~	~~0.02~~	~~0.03~~	~~2.68~~	~~2.52~~
~~Soil water~~	~~N.A.~~	~~0.15~~	~~0.15~~	~~0.20~~	~~0.29~~	~~0.24~~	~~0.25~~	~~0.23~~
~~Lake water~~	~~8.0~~	~~3.16~~	~~10.76~~	~~7.37~~	~~11.83~~	~~7.78~~	~~6.62~~	~~7.02~~

Table 3: Water loss in the storage compartments of Lake Urmia basin between 2003 and 2013 as observed and simulated by the WGHM variants that were calibrated using different observation variables.

	Water loss between 2003 and 2013 [10^9 m^3] (mean annual storage in 2003 minus mean annual storage in 2013)							
Storage compartment	Observed	Standard	RS	RS_LV	RS_Q	RS_Q_LV	RS_Q_GW	RS_Q_GW_NA
TWS Storage	9.90	3.62	10.30	7.86	12.20	8.24	9.78	9.84
GW Storage	1.80	0.17	0.33	0.06	0.02	0.03	2.68	2.26
Soil Water Storage	N.A.	0.15	0.26	0.20	0.29	0.24	0.25	0.25
Lake Storage	8.00	3.16	9.53	7.37	11.83	7.78	6.62	7.24

Based on spaceborne TWSA and lake level observations, total water storage in Lake Urmia basin declined by $9.9 \cdot 10^9$ m^3 from its annual average in 2003 to its annual average in 2013 and about 80% was due to the loss of lake water (Tourian et al. 2015). Observed decline of groundwater storage was $1.8 \cdot 10^9$ m^3, i.e. 18% of the observed total water storage loss in the basin. WGHM overestimates observed loss from groundwater in both calibrations variants that take into account groundwater observations. In WGHM simulations, groundwater decline and depletion below the level of surface water storages occur in only 7 out of the 22 0.5° grid cells within the basin (Fig. S5a). In 5 of these 7 grid cells, groundwater levels were stable during 2003-2007 and only declined from 2008-2013, caused by increased NAg and decreased groundwater recharge in the latter part of the study period. It is these 7 cells that cause the basin groundwater decline under the anthropogenic conditions shown in Fig. 8d. For naturalized conditions, peak seasonal water storages decrease somewhat but minimum water storages cannot drop appreciably given the already very low minimum seasonal storage values during the relatively wet five first years of the investigate period (Fig. S5b) because WaterGAP cannot simulate a possible drop of the groundwater table below the surface water level in the absence of groundwater abstractions. Thus, the contribution of human water use to groundwater storage decline might be overestimated as 1) groundwater storage decline under the impact of human water use is overestimated (Table 3, variant RS_Q_GW_NA as compared to observations and 2) groundwater storage decline under naturalized conditions without human water use may be underestimated.

It is worth mentioning that WGHM as a hydrological model that does not include a gradient-based groundwater model has some limitations for studying groundwater-lake water flows. We attempted to calibrate WGHM under the assumption that there are direct water flows between lake and groundwater. Under this assumption, the seasonality of the groundwater storage was strongly misrepresented. Therefore, as accepted by ULRP (2015c), we assumed there is no direct flow between the lake and groundwater. While Vaheddoost and Aksoy (2018) using traditional hydrograph separation methods claimed that there is a significant relationship between the lake and groundwater, Danesh-Yazdi and Ataie-Ashtiani (2019) rejected their claim. Equally, some studies that applied isotope and chemical tracer analyses (e.g. Amiri et al. 2016) rejected any significant relationship between lake and groundwater. In conclusion, the results of this study support the idea that there are no significant direct interactions between lake and groundwater within the basin.

4.2 Comparison to human vs. ~~climate~~climatic contribution as determined in previous studies

In order to ~~define~~design the Lake Urmia restoration program, it is vital to know which factors contribute how much to the shrinkage of the lake. All previous studies (e.g. Hassanzadeh et al., 2012; AghaKouchak et al., 2015; Ghale et al., 2018; Chaudhari et al., 2018) agreed that shrinkage is caused by both climate variations and human activities, but there is no consensus about the relative contributions. For example, ~~Chaudari~~Chaudhari et al. (2018) concluded that human-induced changes accounted for 86% of the lake volume decline during 1995-2010, while we determined ~~the value~~values of ~~40~~39-43% for ~~2002~~2003-2013. According to our study, human water use was the reason for ~~41~~39-45% inflow reduction into the lake during 2003-2013 which is very similar to the values of Shadkam et al. (2016) for the years 2003-2009 (comp. their ~~Figs~~Fig. 8). Discrepancies are likely due to different analysis methods but different analysis periods~~, as well as different~~ and conceptualizations~~,~~ make a direct comparison of the estimated ~~relative~~ contributions difficult.

While Ghale et al. (2018) seem to support the results of Chaudhari et al. (2018~~),~~) as they state that 80% of drying of Lake Urmia is due to anthropogenic impacts during 1998-2010, ~~there~~their statistical analysis assumes that ~~river~~lake inflow from rivers can be considered to reflect “anthropogenic impacts” while precipitation and evaporation ~~changes~~ reflect climatic ~~variations while river~~variation. However, inflow is in reality also affected by ~~climate~~climatic variations. ~~Also~~ Using a statistical change point analysis and without modelling, Khazaei et al. (2019) stated that given the stable conditions of precipitation and temperature, climatic changes ~~cannot~~could not explain the dramatic decline of the lake level~~.~~; however, they did not use in-situ data (except lake water level data) for their analysis. Based on ~~a~~an analysis of the Standardized Precipitation Index (SPI), a drought index, AghaKouchak et al. (2015) reported there was no significant trend in droughts over the basin during the past three decades and concluded from this that human activities and not climatic variations ~~are~~were the main reason for lake shrinkage. Different from our study and the modelling studies of Shadkam et al. (2016) and Chaudhari et al. (2018), these three studies consider only the dynamics of monthly and annual precipitation~~, not taking into account the~~ and neglect changes in the variability of daily precipitation. During the last three decades, there was a significant increase the frequency of daily precipitation of less than 5 mm and a significant decrease in the frequency of daily precipitation of 10-15 mm, suggesting a runoff reduction even in case of constant annual precipitation (Fig. 2 in Bavil et al., 2018). Hosseini-Moghari et al. (2018) showed that an increasing frequency of days with less than 5 mm precipitation in combination with decreasing monthly precipitation has ~~lead~~led to the observed reduced inflow into two dams in the Lake Urmia basin that are located downstream of areas with insignificant human water use. We conclude that analyses should be done ~~at the daily time scale or smaller.~~ on a daily time scale or smaller. Moreover, we examined the ratio of annual inflow into the lake (based on the ensemble mean) over annual precipitation during the study period. This ratio reached maximum values in 2003 (0.29 and 0.41 for the anthropogenic and naturalized conditions, respectivly) and minimum values in 2009 (0.07 and 0.15). Averaged over the period 2009-2013, these ratios are, with 0.11 (ANT) and 0.22 (NAT), much smaller than the values for 2003-2007, 0.20 and 0.32. Thus, the drought year 2008 as well as the relatively small ratio of inflow into the lake over precipitation in the last five years of the study period play a significant role in the decline of inflow and lake water storage.

~~In addition~~For quantifying human and climatic contributions to observed hydrological changes, a comprehensive modeling approach~~ is preferable~~ that takes into account, for example, the impacts of changing temperatures on runoff and thus river inflow and on evapotranspiration of the lake itself~~.~~ is preferable. Such comprehensive modelling was done by Chaudhari et al. (2018) but their uncalibrated global hydrological model that represented the basin by 5-6 cells only was not able to simulate well the flows and storages in the basin. For example, simulated annual inflow into the lake was estimated to be 3,700·10^6 m^3 in 2003 (their Fig. 8) while observed inflow was much higher, 5,835·10^6 m^3. In 2009, observed inflow, with 1,036·10^6 m^3, was only half of the simulated one. Therefore, the very high human contribution to the lake volume decline of 86% determined by Chaudhari et al. (2018) may arise from the poor performance of the uncalibrated model.

4.3 **Limitations**

Even after multi-objective calibration of a state-of-the-art comprehensive hydrological model, there remain many uncertainties that affect the accuracy of the model results. Like the results of all hydrological models, our results are affected by uncertainties in model input, model parameters, and model structure. Model parameter uncertainty was reduced by the comprehensive multi-observation calibration, albeit conditioned on just one climate input data set and using just one model (instead of the state-of-the-art multi-model ensemble approach, compare www.isimip.org~~, last accessed: 14 Dec. 2018).~~). Given the low spatial model resolution (0.5°×0.5°), the model results are ~~only valid for~~preferably aggregated to the basin as a whole ~~and~~as results for individual grid cells are very uncertain. Also due to a lack of data at the basin scale, the hydrogeology of the basin was not taken into account in the model. Information on the irrigated area in each grid cell was taken from a global data set of areas equipped for irrigation from groundwater and surface water (Siebert et al., 2010), which was ~~adapted~~adopted in this study by scaling it by basin-wide correction factors to better capture the temporal development of irrigation. Calibrated modeling results are also affected by the uncertainties of the observation data. GRACE TWSA data are more reliable for larger (100,000 km^2 ~~(~~according to Landerer and Swenson, 2012~~))~~) areas than the basin area of 52,000 km^2. Estimation of groundwater storage changes based on water level data for unevenly distributed wells is rather uncertain due to the unknown heterogeneities in the subsurface~~. Evaluation results, here the good fit of simulated to~~ and uncertain specific yields. The “observed” lake water volume decline~~, are be affected by a~~ likely ~~underestimation of~~underestimates the actual decline ~~by the “observed” value derived~~as a constant bathymetry was assumed when deriving lake water volume decline from remote sensing of lake water level elevation and lake water area ~~by~~ (Tourian et al. ~~(~~2015~~) assuming a constant bathymetry.~~). However, there was an increase in the elevation of the lake bottom due to sedimentation and salt precipitation (Shadkam et al., ~~2016) so that the “observed” water volume decline was likely lower than the actual one, and our model would underestimate the lake storage decline, too.~~2016).

5 Conclusions

This study investigated the differential impact of human water use and climate variations on ~~water storage (~~total, groundwater~~,~~ and lake~~)~~ water storage in the Lake Urmia basin as well as on inflow into the lake during 2003-2013. This was done by utilizing the information contained in multiple types of observation data to calibrate, specifically for the Lake Urmia basin, the global hydrological model WGHM~~ that~~, which takes into account the impact of human water use and man-made reservoirs on flows and storages. Observations include remote sensing data (for irrigated area, TWSA, and lake volume), in-situ streamflow observations (for of lake inflow), groundwater well data (for deriving groundwater storage anomalies) and statistical data on water use in the basin. A time series of observed lake volume was used for evaluation. Using the ensemble of best-performing models where all available observations were used for model ~~variant~~calibration, the impact of human water use was determined by comparing the output of ~~a~~ naturalized run, ~~where~~with human water use ~~was~~ assumed to be zero, with the ~~run~~runs with the ~~historic~~historical water use. To understand the value of different observational data types for calibration, ~~four calibration~~WGHM was calibrated in six variants ~~were defined where, in a step-wise fashion, basin-wide averages of 1) remote sensing data (for irrigated area and TWSA), 2) in-situ streamflow observations (for of lake inflow), 3) groundwater well data (groundwater level and storage),~~(two auto-calibrated and ~~4) statistical data on water withdrawals in the basin were added. A time series of observed lake volume was used for evaluation.~~four manually calibrated) to different combinations of observational data types.

We found that the time series for water demand by irrigation, as assumed in the standard WGHM version, had to be adjusted using MODIS data such that the modification of ~~four~~seven model parameters could result in a good fit to observed GRACE TWSA. Consideration of these remote sensing data somewhat improved the dynamics of both inflow into Lake Urmia and lake water storage, but lake inflow ~~into the lake~~ was still ~~strongly~~ overestimated by ~~a factor of 0.92%,~~66% and the seasonality of groundwater ~~dynamics should a~~storage was strongly shifted~~ seasonality~~. Additional calibration against observed inflow into the lake did not affect TWSA simulation and slightly improved the simulation of the lake water storage anomaly. Only by using monthly time series of mean groundwater level variations in the basins for calibration, we could adjust the fractions of human water use taken from groundwater and surface water such that seasonality of groundwater storage was simulated correctly. Only then it was possible to simulate the observed groundwater loss, and loss of lake volume was no longer overestimated. Statistical information on sectoral water withdrawals in the basin for one year as well as estimates for sectoral return flow fractions further improved the model, but only slightly. We recommend to include, in case of relevant trends in observations, the difference between observed and simulated trends as one of the calibration criteria, not only differences between time series of daily, monthly or annual values.

The calibration exercise showed that the calibration variant for which the highest number of observational data types were used, WGHM variant RS_Q_GW_NA, showed the best fit to all observations. Certainly, no general conclusions on the worth of specific observation data types for model calibration, including trade-offs among fit to multiple data types, can be derived from this study. Lake Urmia basin is particular with respect to 1) draining into a large end lake that dominates TWSA,

2) the strong impact of human water use and 3) the fact that the standard WGHM version estimates a net recharge to the groundwater due to surface water irrigation, which had to be corrected to a net abstraction. In basins with large lakes, and in particular with end lakes, remotely sensed time series on lake area and the elevation of the lake water table should be used to estimate time series of lake water storage as these observational data can be expected to be of high value for understanding the freshwater system by hydrological model calibration. Groundwater storage cannot be observed from space but relies on in-situ observations on groundwater heads in wells but, as in the case of Lake Urmia basin, such data can be crucial for a correct understanding of the freshwater system.

Based on the good fit of WGHM variant RS_Q_GW_NA to four types of observational data, we ~~are confident~~found that human water use reduced lake inflow that would have occurred without human water use during 2003-2013 by about ~~41~~39-45%. About 52-57% of the total water storage loss in Lake Urmia basin and only ~~40~~39-43% of lake water loss during this time period was due to human water use, and the 43-48% and ~~60~~57-61%, respectively, to climate variations. 87-90% of groundwater storage loss is estimated to be caused by human water use but this value may be somewhat overestimated by WGHM because climate-driven loss under naturalized conditions may be underestimated due to the simplified representation of groundwater-surface water exchanges in the model.

GRACE TWSA data indicate an increasing trend in water storage in the basin during 2014-2017 due to both less water use due to water management (ULRP, 2015b) and the wet years 2015/2016. This trend is about half as strong as the decreasing trend during 2003-2013. Further strengthening of efforts for decreasing human water use in the basin should be undertaken, while at the same time, global-scale mitigation of climate change by reducing greenhouse gas emissions to prevent strong decreases of precipitation and runoff. Our study has shown that the management of the Lake Urmia basin should be based on a comprehensive assessment of all water storages and flows in the basin, including human water uses of groundwater and surface water. We recommend refining the estimated net abstractions from surface water and groundwater by a basin-wide spatially explicit quantification not only of water abstractions but also return flows to groundwater and surface water.

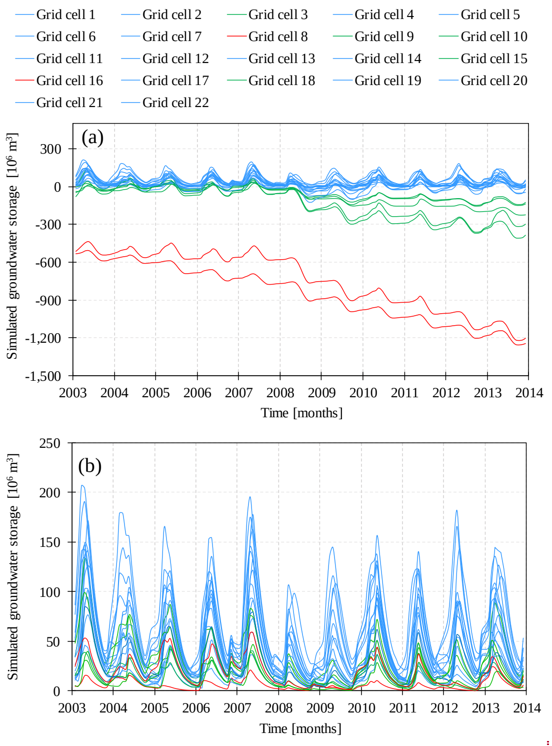
Data availability. In-situ data from "Iran Water Resources Management Company" including groundwater levels, precipitation and temperature are available upon request from the corresponding author. All other data are available in supplementary. Also, GRACE data is available through http://www2.csr.utexas.edu/grace/RL05_mascons.html (last accessed: 17 Jul. 2018). Lake water surface extents and water levels are available at http://hydrosat.gis.uni-stuttgart.de/php/index.php (last accessed: 17 Jul. 2018). All simulation results are available from the corresponding author.

Author contributions. S.M.H.M performed the modeling and writing. S. A., M. J. T., and K. E. contributed to editing the manuscript. P. D. contributed to analysis of the results, the discussion, and editing the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. The authors are grateful to Urmia Lake Restoration Program and Iran Water Resources Management Company for providing in-situ data. They also acknowledge the Center for Space Research (CSR) and Potsdam Institute for Climate Impact Research that provide GRACE data and global climate data, respectively. The authors graciously acknowledge Felix T. Portmann and Hannes Müller Schmied from Goethe University Frankfurt for their valuable contributions throughout the WaterGAP calibration, and thank the editor and reviewers for their comments and suggestions that helped to improve the paper.

~~Appendix A: Simulated groundwater storage in individual grid cells~~



~~**Figure A1: Simulated groundwater storage in each of the 22 0.5° grid cells in Lake Urmia basin under anthropogenically altered (Fig. A1a) and naturalized conditions (Fig. A1b).**~~

References

Abbaspour, M., and Nazaridoust, A.: Determination of environmental water requirements of Lake Urmia, Iran: an ecological approach, International Journal of Environmental Studies, 64, 161-169, doi:10.1080/00207230701238416, 2007.

AghaKouchak, A., Norouzi, H., Madani, K., Mirchi, A., Azarderakhsh, M., Nazemi, A., Nasrollahi, N., Farahmand, A., Mehran, A., and Hasanzadeh, E.: Aral Sea syndrome desiccates Lake Urmia: Call for action, Journal of Great Lakes Research, 41, 307-311, doi:10.1016/j.jglr.2014.12.007, 2015.

Ahmadzadeh, H., Morid, S., Delavar, M., and Srinivasan, R.: Using the SWAT model to assess the impacts of changing irrigation from surface to pressurized systems on water productivity and water saving in the Zarrineh Rud catchment, Agricultural Water Management, 175, 15-28, doi:10.1016/j.agwat.2015.10.026, 2016.

Alborzi, A., Mirchi, A., Moftakhari, H., Mallakpour, I., Alian, S., Nazemi, A., Hassanzadeh, E., Mazdiyasni, O., Ashraf, S., Madani, K., Norouzi, H., Azarderakhsh, M., Mehran, A., Sadegh, M., Castelletti, A., and AghaKouchak, A.: Climate-informed environmental inflows to revive a drying lake facing meteorological and anthropogenic droughts, Environmental Research Letters, 13(8), 084010, doi:10.1088/1748-9326/aad246, 2018.

Alizadeh-Choobari, O., Ahmadi-Givi, F., Mirzaei, N., and Owlad, E.: Climate change and anthropogenic impacts on the rapid shrinkage of Lake Urmia, International Journal of Climatology, 36, 4276-4286, doi:10.1002/joc.4630, 2016.

~~Azarnivand, A., Hashemi-Madani, F. S., and Banihabib, M. E.: Extended fuzzy analytic hierarchy process approach in water~~Amiri, V., Nakhaei, M., Lak, R., and Kholghi, M.: Geophysical, isotopic, and ~~environmental management (case study: Lake~~hydrogeochemical tools to identify potential impacts on coastal groundwater resources from Urmia ~~Basin,~~hypersaline Lake, NW Iran~~), .~~. Environmental ~~Earth Sciences, 73, 13-26,~~ Science and Pollution Research, 23(16), 16738-16760, doi: 10.1007/~~s12665-014-3391-6, 2015~~s11356-016-6859-y, 2016.

Azarnivand, A., and Banihabib, M. E.: A Multi-level Strategic Group Decision Making for Understanding and Analysis of Sustainable Watershed Planning in Response to Environmental Perplexities, Group Decision and Negotiation, 26, 629–648, doi: 10.1007/s10726-016-9484-8, 2017.

Azarnivand, A., Camporese, M., Alaghmand, S., and Daly, E.: Simulated response of an intermittent stream to rainfall frequency patterns. Hydrological Process, doi: 10.1002/hyp.13610, 2019.

Banihabib, M. E., Azarnivand, A., and Peralta, R. C.: A new framework for strategic planning to stabilize a shrinking lake, Lake and Reservoir Management, 31, 31-43, doi: 10.1080/10402381.2014.987409, 2015.

Bavil, S. S., Zeinalzadeh, K., and Hessari, B.: The changes in the frequency of daily precipitation in Urmia Lake basin, Iran, Theoretical and Applied Climatology, 133(1-2), 205-214, doi:10.1007/s00704-017-2177-7, 2018.

Beven, K., and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of hydrology, 249, 11-29, doi:10.1016/S0022-1694(01)00421-8, 2001.

Chaudhari, S., Felfelani, F., Shin, S., and Pokhrel, Y.: Climate and anthropogenic contributions to the desiccation of the second largest saline lake in the twentieth century, Journal of Hydrology, 560, 342-353, doi:10.1016/j.jhydrol.2018.03.034, 2018.

Danesh-Yazdi, M., and Ataie-Ashtiani, B.: Lake Urmia Crisis and Restoration Plan: Planning without Appropriate Data and Model Is Gambling. Journal of Hydrology, 576, 639-651, doi: 10.1016/j.jhydrol.2019.06.068, 2019.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A. M. T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation, 6(2), 182-197, doi:1089-778X(02)04101-2 , 2002.

Delju, A. H., Ceylan, A., Piguet, E., and Rebetez, M.: Observed climate variability and change in Urmia Lake Basin, Iran, Theoretical and Applied Climatology, 111, 285-296, doi:10.1007/s00704-012-0651-9, 2012.

Döll, P., ~~and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrology and Earth System Sciences Discussions, 12, 863-885, doi:10.5194/hess-12-863-2008, 2008.~~

~~Döll, P.,~~ Douville, H., Güntner, A., Müller Schmied, H., and Wada, Y.: Modelling freshwater resources at the global scale: Challenges and prospects. Surveys in Geophysics, 37, 195-221, doi:10.1007/s10712-015-9343-1, 2016.

Döll, P., ~~Fritsche, M., Eicker, A., Müller Schmied, H.: Seasonal water storage variations as impacted by water abstractions: Comparing the output of a global hydrological model with GRACE and GPS observations, Surveys in Geophysics, 35(6), 1311-1331, doi:10.1007/s10712-014-9282-2, 2014b.~~

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

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

~~Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: 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, 5698-5720, doi:10.1002/2014WR015595, 2014a.~~

~~Eicker, A., Schumacher, M., Kusche, J., Döll, P., and Schmied, H. M.: Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an ensemble Kalman filter: first results, Surveys in Geophysics, 35, 1285-1309, doi:10.1007/s10712-014-9309-8, 2014.~~

Eimanifar, A., and Mohebbi, F.: Urmia Lake (northwest Iran): a brief review, Saline Systems, 3, 5, doi:10.1186/1746-1448-3-5, 2007.

Farajzadeh, J., Fakheri Fard, A., and Lotfi, S.: Modeling of monthly rainfall and runoff of Urmia lake basin using "feed-forward neural network" and "time series analysis" model, Water Resources and Industry, 7-8, 38-48, doi:10.1016/j.wri.2014.10.003, 2014.

Fathian, F., Morid, S., and Kahya, E.: Identification of trends in hydrological and climatic variables in Urmia Lake basin, Iran, Theoretical and Applied Climatology, 119, 443-464, doi:10.1007/s00704-014-1120-4, 2014.

Forootan, E., Rietbroek, R., Kusche, J., Sharifi, M. A., Awange, J. L., Schmidt, M., Omondi, P., and Famiglietti, J.: Separation of large scale water storage patterns over Iran using GRACE, altimetry and hydrological data, Remote Sensing of Environment, 140, 580-595, doi:10.1016/j.rse.2013.09.025, 2014.

Ghale, Y. A. G., Altunkaynak, A., and Unal, A.: Investigation Anthropogenic Impacts and Climate Factors on Drying up of Urmia Lake using Water Budget and Drought Analysis, Water Resources Management, 32, 325-337, doi:10.1007/s11269-017-1812-5, 2018.

~~Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., and Koirala, S.: Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, Journal of Hydrometeorology, 13, 604-620., doi:10.1175/JHM-D-11-083.1, 2012.~~

~~Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377(1-2), 80-91, doi: , 2009.~~

~~Hamzekhani, H., and Aghaie, M. M.: Miandoab plain Groundwater modelling with the use of remote sensing technology to assess the impact of changes in aquifer conditions on surface flow, Report of Urmia Lake Restoration Program (ULRP) (In Persian), 2015.~~

Hasemi, M.: A socio-technical assessment framework for integrated water resources management (IWRM) in Lake Urmia Basin, Iran, 2011.

Hassanzadeh, E., Zarghami, M., and Hassanzadeh, Y.: Determining the main factors in declining the urmia lake level by using system dynamics modeling, Water Resources Management, 26, 129-145, doi:10.1007/s11269-011-9909-8, 2012.

Hesami, A., and Amini, A.: Changes in irrigated land and agricultural water use in the Lake Urmia basin, Lake and reservoir management, 32, 288-296, doi:10.1080/10402381.2016.1211202, 2016.

Hosseini-Moghari, S.M., Araghinejad, S. and Ebrahimi, K.: Monthly Precipitation Assessment: a misleading tool for understanding the effects of climate change, 8th Global FRIEND-Water Conference, Beijing, China, 6-9 November 2018.

Jalili, S., Kirchner, I., Livingstone, D. M., and Morid, S.: The influence of large-scale atmospheric circulation weather types on variations in the water level of Lake Urmia, Iran, International Journal of Climatology, 32, 1990-1996, doi:10.1002/joc.2422, 2012.

~~Kamali, M., and Youneszadeh Jalili, S.: Investigation of landuse changes in Lake Urmia Basin using remotely sensed images, Report of Urmia Lake Restoration Program (ULRP) (In Persian), 2015.~~

Khazaei, B., Khatami, S., Alemohammad, S. H., Rashidi, L., Wu, C., Madani, K., and Aghakouchak, A.: Climatic or regionally induced by humans? Tracing hydro-climatic and land-use changes to better understand the Lake Urmia tragedy, Journal of Hydrology, 569, 203-217, doi:10.1016/j.jhydrol.2018.12.004, 2019.

Landerer, F. W., and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water resources research, 48, doi:10.1029/2011WR011453, 2012.

~~Lange, S.: EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI), GFZ Data Services, doi:10.5880/pik.2016.004, 2016.~~

~~Lorenz, C., Kunstmann, H., Devaraju, B., Tourian, M. J., Sneeuw, N., and Riegger, J.: Large-scale runoff from landmasses: a global assessment of the closure of the hydrological and atmospheric water balances, Journal of Hydrometeorology, 15, 2111-2139, doi:10.1175/JHM-D-13-0157.1, 2014.~~

Madani, K., AghaKouchak, A., and Mirchi, A.: Iran's socio-economic drought: challenges of a water-bankrupt nation, Iranian Studies, 49, 997-1016, doi:10.1080/00210862.2016.1259286, 2016.

~~Mahab Ghodss consulting engineering: Report on the balance of water resources and consumption (40-year-average) in Lake Urmia Basin, Ministry of energy (In Persian), 2013.~~

Mirjalili, S.: Genetic algorithm. Evolutionary Algorithms and Neural Networks, Springer, 43-55, 2019.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: 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, 3511-3538, doi:10.5194/hess-18-3511-2014, 2014.

Pengra, B.: The drying of Iran's Lake Urmia and its environmental consequences, UNEP-GRID, Sioux Falls, UNEP Global Environmental Alert Service (GEAS), 2012.

~~Prudhomme, C., Parry, S., Hannaford, J., Clark, D. B., Hagemann, S., and Voss, F.: How Well Do Large-Scale Models Reproduce Regional Hydrological Extremes in Europe?, Journal of Hydrometeorology, 12, 1181-1204, doi:10.1175/2011JHM1387.1, 2011.~~

Saemian, P., Faridzad, M., Youneszadeh, S., Emadzadeh, M., Jalilvand, E., Farahmand, H., and Taheri, M.: Estimation of Evapotranspiration, water saving and agricultural water consumption at Urmia Lake Basin, 2010, Report of Urmia Lake Restoration Program (ULRP)(In Persian), 2015.

Scrucca, L.: GA: a package for genetic algorithms in R, Journal of Statistical Software, 53(4), 1-37, doi:10.18637/jss.v053.i04~~Save, H., Bettadpur, S., and Tapley, B. D.: High-resolution CSR GRACE RL05 mascons, Journal of Geophysical Research: Solid Earth, 121, 7547-7569, doi:10.1002/2016JB013007, 2016.~~
, 2013.

Shadkam, S., Ludwig, F., van Oel, P., Kirmit, Ç., and Kabat, P.: Impacts of climate change and water resources development on the declining inflow into Iran's Urmia Lake, Journal of Great Lakes Research, 42, 942-952, doi:10.1016/j.jglr.2016.07.033, ~~2016a~~2016.

Siebert, S., Burke, J., Faures, J.-M., Frenken, K., Hoogeveen, J., Döll, P., and Portmann, F. T.: Groundwater use for irrigation– a global inventory, Hydrology and Earth System Sciences, 14, 1863-1880, doi:10.5194/hess-14-1863-2010, 2010.

Sima, S., and Tajrishy, M.: Using satellite data to extract volume–area–elevation relationships for Urmia Lake, Iran, Journal of Great Lakes Research, 39, 90-99, doi:10.1016/j.jglr.2012.12.013, 2013.

~~Strassberg, G., Scanlon, B. R., and Chambers, D.: Evaluation of groundwater storage monitoring with the GRACE satellite: Case study of the High Plains aquifer, central United States, Water Resources Research, 45, doi: 10.1029/2008WR006892, 2009.~~

Tabari, H., and Talaee, P. H.: Analysis of trends in temperature data in arid and semi-arid regions of Iran, Global and Planetary Change, 79, doi:1-10, 10.1016/j.gloplacha.2011.07.008, 2011a.

Tabari, H., and Talaee, P. H.: Temporal variability of precipitation over Iran: 1966–2005, Journal of Hydrology, 396, 313-320, doi:10.1016/j.jhydrol.2010.11.034, 2011b.

~~Tallaksen, L. M. and Stahl, K.: Spatial and temporal patterns of large-scale droughts in Europe: Model dispersion and performance, Geophysical Research Letters, 41, 429-434, doi:10.1002/2013GL058573, 2014.~~

Tourian, M. J., Elmi, O., Chen, Q., Devaraju, B., Roohi, S., and Sneeuw, N.: A spaceborne multisensor approach to monitor the desiccation of Lake Urmia in Iran, Remote Sensing of Environment, 156, 349-360, doi:10.1016/j.rse.2014.10.006, 2015.

Tsou, C. S.: nsga2R: Elitist Non-dominated Sorting Genetic Algorithm based on R, R package version 1.0, Available at: https://CRAN.R-project.org/package=nsga2R, 2013.

ULRP: Evaluation and deduction of Urmia Lake Restoration Program need at least 3 years, Available at: http://ulrp.sharif.ir/en/news/news-1435-evaluation-and-deduction-urmia-lake-restoration-program-need-least-3-years, 2015b.

~~ULRP: Status of water resources and water consumption in the basin, Available at: http://ulrp.sharif.ir/fa (In Persian), 2015c.~~

ULRP: The trend of Lake Urmia changes, Available at: http://ulrp.sharif.ir/fa (In Persian), 2015a.

~~Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., and Landerer, F. W.: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons, Journal of Geophysical Research: Solid Earth, 120, 2648-2671, doi:10.1002/2014JB011547, 2015.~~

ULRP: Urmia Lake - Causes of shrinkage and potential threats. Available at: http://ulrp.sharif.ir/fa (In Persian), 2015c.

Vaheddoost, B., and Aksoy, H.: Interaction of groundwater with Lake Urmia in Iran. Hydrological processes, 32(21), 3283-3295, doi:10.1002/hyp.13263, 2018.

Werth, S., and Güntner, A.: Calibration analysis for water storage variability of the global hydrological model WGHM, Hydrology and Earth System Sciences, 14, 59, doi:10.5194/hess-14-59-2010, 2010.

~~Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, International Journal of Climatology, 32(13), 2088-2094, doi:10.1002/joc.2419, 2012.~~

~~Zarghami, M.: Effective watershed management; Case study of Urmia Lake, Iran, Lake and Reservoir Management, 27, 87-94, doi:10.1080/07438141.2010.541327, 2011.~~

Zeinoddini, M., Tofighi, M. A., and Vafaee, F.: Evaluation of dike-type causeway impacts on the flow and salinity regimes in Urmia Lake, Iran, Journal of Great Lakes Research, 35, 13-22, doi:10.1016/j.jglr.2008.08.001, 2009.