Dear Editor and Reviewers,

We have revised the manuscript according to the insightful comments provided by the editor and reviewers as well as according to the short comments. All recommendations have been addressed in the revised manuscript. In addition to a point-by-point response which has been already uploaded in HESS website, all relevant changes made in the manuscript have been listed below. We would like to thank you for the thorough consideration and critical comments that helped us improve our manuscript.

Editor Comment

Editor Comment: You received high quality and detailed reviews to which you also provided detailed answers. Thanks to the referees and you for this discussion phase. To me the most important is the better highlight the added value of this work and embed it in the vast amount of literature on Lake Urmia. Also some more details on the model and model performance is required.

Response: *We have highlighted the innovation of our study in comparison with the vast amount of the literature on Lake Urmia. Also, we have added some details on the model and calculated two new model performance indicators.*

- *With respect to "better highlight the added value of the work": We completely revised the abstract, reformulated the objectives of the paper in the Introduction and included a section (section 4.2) where we discuss the shortcomings of past work on quantification of the relative impacts of humans and climate on the hydrology of Lake Urmia basin.*
- *With respect to "embed it in vast literature": We included and discussed (also in section 4.2) very recent publications.*
- *With respect to "more details on model and model performance": The WaterGAP description in section 2.1 was rewritten and extended, in particular regarding the simulation of lake dynamics and the performance of WaterGAP as compared to other global hydrological models. In addition, following the suggestion of a reviewer, the WaterGAP model performance for Lake Urmia basin was quantified by two additional performance indicators (Table 4).*

Point-by-point response to the Referee #1

Referee #1: This manuscript uses WaterGAP Global Hydrology model to quantify the effects of human water use on inflow to Lake Urmia, lake water volume, and groundwater. The model was manually calibrated 4 for time using different observation data sets (remote sensing of irrigated area, monthly total water storage anomaly, in-situ observations of stream flow, and groundwater levels from 284 wells). Strengths of the work include a focus on the pressing problem or Lake Urmia decline and identification of the effects on groundwater. With these strengths, there are also several issues that I feel need to be addressed to accept this manuscript for publication.

Response: *We would like to thank you for the thorough consideration and critical comments that helped us improving the manuscript. We have tried to do our best to consider all your recommendations in the revised version. Below, we have provided a point-by-point response to your comments.*

Referee #1: Is the finding that humans affected lake decline new? There have been several recent studies that report this finding (Alborzi et al. 2018; Chaudhari et al. 2018; Shadkam et al. 2016) and some of these studied used the same model inputs as this work and also report groundwater changes. What is new in this work?

Response: *We agree that different studies have been conducted on the Lake Urmia for quantifying the anthropogenic and climatic impacts on shrinking of the lake. But most of these studies focused on drought events, inflow into the lake or lake levels. None of these studies considered the effect of human water use on TWSA and groundwater storage (quantitatively), even distinguishing surface water from groundwater use. Also none of previous studies used such a diverse observation data that allow understanding of the different storage compartments of the basin. Furthermore, it should be noted the study does not only aim at quantifying the different impacts of climate and human water use but equally at understanding, for the example of the investigated basin, the value of different observational data types for calibrating a hydrological model and thus understanding dynamics and flows in a basin (see page 5, lines 30ff). Our study shows that for a comprehensive hydrological modeling that captures correctly the main dynamics in a basin, a multi-objective calibration is needed, and then previous studies missed some part of a comprehensive hydrological process modeling (e.g. the specific impact of groundwater pumping as compared to surface water use). Below please find the shortcomings of the studies you mentioned.*

Alborzi et al. (2018):

1) Did not use a hydrological model.

2) Did not have any discussion on groundwater, lake storage, total water storage under natural condition.

Chaudhari et al. (2018):

1) Used global hydrological model (HiGW-MAT) without calibration for Lake Urmia basin.

2) In hydrological modeling the focus is only on streamflow and there is no discussion on groundwater, lake storage, total water storage under natural condition.

3) Did not take into account impacts of domestic and industrial water use.

4) Did not use in-situ data for irrigation water requirement and used FAO Penman Monteith method for estimating irrigation water requirement.

5) Did not use in-situ climatic data for estimating irrigation water requirement (used 6 h atmospheric reanalysis data provided by Japanese Meteorological Agency (JMA) Climate Data Assimilation System (JCDAS) as meteorological data)

Shadkam et al. (2016):

1) Only considered a single objective calibration of VIC model based on the inflow into the lake.

2) In hydrological modeling the focus is only on streamflow and there is no discussion on groundwater, lake storage, total water storage under natural condition.

3) All irrigation requirements need to be fulfilled by available surface water.

4) Did not take into account the impact of domestic and industrial water use.

As mentioned above, none of previous study provide a comprehensive modeling such as our manuscript. We have described all modeling studies over Lake Urmia basin along with their limitations in Introduction of the revised version.

Referee #1: Is a global hydrologic model appropriate for a basin level analysis?

<u>Response</u>: *Even though global hydrological models have a coarse spatial resolution, they contain a lot of different data (climate, physiographic, water use etc.) for which local information may not be available or of better quality. The study has shown that for the relatively large Lake Urmia basin, the global WGHM model can be informative after multi-observation calibration as after calibration the fit to those different types of observations is rather good.*

Referee #1: The description of how the model simulates relevant processes is scant.

<u>Response</u>: *Due to the fact that the length of manuscript already is long and also there are some other literatures that have described the model simulates relevant processes, we explained the model in a summarized form. In the revised version we have added the following sentence to section 2.1:*

"WGHM simulates daily 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 (Werth and Güntner, 2010)."

Referee #1: Given resonance times, how appropriate is the temporal spacing (daily) relevant to the spatial grid size?

<u>Response</u>: *In our opinion, a daily time step fits well to simulation of water flows and storage at the 0.5° grid scale and is the usual time step in global hydrological modeling. Land surface models that also simulate energy flow require a smaller time step.*

Referee #1: Is it computationally efficient to run a global model for 15 or 20 grid cells of interest? There needs to be a much stronger justification for why the modeling and calibration methods are the correct approaches to use to answer the motivating questions.

<u>Response</u>: *It should be noted there is no need to run the model for whole globe. The model can be run for a specific basin, in the case of Lake Urmia for just 22 grid cells. Therefore, simulations are highly efficient. We have added this explanation to section 2.1. Using WaterGAP to answer the motivating questions is also efficient in that the model for the Lake Urmia basin was already set up at the beginning of the study as WaterGAP as a global model is ready for simulating any (large enough) basin around the world. We do not think that it is necessary to further explain in the text why the modeling and calibration approach is suitable, as e.g. section 2.1 describes that WaterGAP allows to consider a complex hydrological system (including surface water, groundwater, lake, human water use, etc).*

Referee #1: There is a lot of focus in the text on the multiple calibration variants run with different input data sets. What was learned from this activity? How do those result effect Lake Urmia management?

Response: *We described the lesson we learned from this activity in details in section "3.2 What we learn from the calibration?". We learned that there is no guarantee that a single objective calibration improves the model performance with respect to the simulation of other components of hydrological system. As a result, for defining the Lake restoration plans a comprehensive modeling framework like our study is needed. Specifically, quantification of return flow from irrigation is paramount for managing irrigation in the basin. We have added the following paragraph to the conclusion, as the last one:*

"Our study has shown that 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."

Also, to clarify of research objectives we have revised the objectives of the paper in the Introduction section as follows:

"The aim of our study was twofold. On the one hand, we wanted to quantify, by a holistic and reliable modelling approach, 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, 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)."

Referee #1: Also, what could one potentially learn from 4 model calibrations that use different calibration data and yield four different models?

Response: *It is very common to use a single or two objective calibrations for calibrating a hydrological model. In this study, we have tried to understand which level of data can reveal that our modelling is holistic. Based on the results, for a holistic modeling, at least remote sensing, discharge and groundwater levels data are required. In addition, we have investigated with adding each in-situ data in calibration process how the model performance improved in simulating different water resources components.*

Referee #1: What are the limitations of this study? The discussion of uncertainty in the results needs to go much deeper and be more specific. This uncertainty is real and likely plays a large role in the interpretation of the results.

Response: *In the revised manuscript, we have added, as section 4.3, a short discussion.*

"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 the basin as a whole and 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 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 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 uncertainties of the observation data. GRACE TWSA data are more reliable for larger (100,000 km2 (Landerer and Swenson, 2012)) areas than the basin area. 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 "observed" lake water volume decline, are be affected by a likely underestimation of the actual decline by the "observed" value derived 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."

Referee #1: I found the writing difficult to follow in numerous places, particularly the results section. There are lots of acronyms, run-on sentences, and text that digresses from the section headers or topic sentences of paragraphs. The writing here made it difficult for me to see the main results and findings of the work.

Response: *We have revised whole of manuscript and reformulate the results section for the revised version.*

Referee #1: pp. 2-4. The first three figures recount results from prior work. I would much prefer to see figures and tables focus on new insights gained from the work. For example, new figures that show uncertainties.

Response: *In our opinion these figures are needed to inform the reader about the story that happened at Lake Urmia basin. Also, only Figure 2 is taken from previous studies. Due to the length of manuscript, we prefer not to add new figures.*

Referee #1: p. 3, line 5. I think "somewhat recovered" is overstated. Hard to tell from Figure 3. Maybe stabilized.

Response: *We agree with you. So, we have revised the sentence as follow:*

"After 2015, lake extent and storage have stabilized".

Referee #1: p. 3, line 18. Is the value -11.2 mm/yr correct? It seems incredibly small. In The Hashemite Kingdom of Jordan, drawdowns are 1+ m/yr, in numerous wells. In the U.S., we talk about drawdowns of ft/year.

Response: *The value taken from the study of Forootan et al. (2014) refers to groundwater storage, not to a drawdown of the groundwater table. And it is the average loss over the whole basin and not a drawdown in an extraction well which of course can be much higher than an average drawdown over the whole basin. Change in groundwater storage can be calculated by multiplying change in groundwater level with the specific yield.*

Referee #1: p. 6, line 5. Only the anomalies? Or at all time periods? If the former, please explain what is meant by "anomaly", how determined, and why anomaly is the appropriate frame to discuss. I would want to calibrate a model across a range of conditions some of which might include anomalies.

Response: *The model has no information about the real bathymetry and initial value of lake storage; it can only simulate changes as compared to some initial condition. As a result, we can compare only lake storage change or lake storage anomalies. Lake storage anomaly at a given time is equal to the lake storage at that time minus the long-term average of lake storage. In this study, as mentioned in p. 20 line 24, anomalies were calculated with respect to the mean lake storage during 2004-2009 (baseline period used for the provided GRACE data).*

Referee #1: p. 6, lines 18-28. I'm not familiar with WaterGAP. How does this model actually work? Explain.

Response: *While the WaterGAP description in section 2.1 is brief, the reader is referred to publications that provide more information on the model. However, as written above, we plan to extend the WaterGAP description slightly, in particular with respect of lake modeling.*

Referee #1: p. 7. What is total water storage anomaly (TWSA)? This term seems rather central to the paper. Please explain.

Response: *Total water storage (TWS) is amount of water which is stored in different components of the continents, e.g. as follows (Scanlon et al., 2018):*

$$TWS = SnWS + CWS + SWS + SMS + GWS \quad (1)$$

where SnWS is snow water storage, CWS is canopy water storage, SWS is surface water storage, SMS is soil moisture storage, and GWS is groundwater storage. Neither hydrological models nor GRACE can compute the total amount of stored water. They can only compute varisations according to a temporal average. Therefore, TWS anomalies (TWSA) are evaluated, defined as TWS(t) – mean (TWS).

We have inserted the following sentence as the second sentence of the GRACE section (p. 7, line 9):

"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."

Referee #1: p. 8, lines 13-18. This method of applying (1- return flow multipliers) to the abstractions to estimate consumptive use assumes that water is used by only one water user. Is this a realistic assumption? If the return flow is used by another agricultural user and then again by a 3rd or 4th user, the basin-wide consumptive use fraction will be much different than the values reported. The large grid size magnifies this error. Table 1. How sensitive are study results to the values in this table?

Response: *In our opinion due to the fact that most return flow returns to groundwater, there is no concern in this regard. On the other hand, in arid area e.g. Urmia basin the return flow to surface water in each irrigation is not too much that can be used by another user. Anyway, we do not know exactly how the authors of the values determined them. However, the study results are not sensitive to the independent return flow estimates that were used only in one variant (RS_Q_GW_NA). We have written in section 4.1:*

"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 significantly and only leads to slight parameter adjustments. This indicates a reasonable simulation of per hectare water consumption for irrigation by the WaterGAP model."

Referee #1: p. 10. Lines 5-15. So the correction factors are needed because WaterGAP does not get the underlying physical hydrology correct? The correction is linear? Is the process causing the error also linear?

Response: *Irrigated area in the standard version of WaterGAP is constant during the period of investigation. The correction factor based on MODIS remote sensing time series adjusts the WaterGAP value in each year homogeneously across the basin. Standard WaterGAP irrigation area is multiplied by the correction factor.*

Referee #1: p. 11, line 1. Which parameters were varied to calibrate this model?

Response: *These parameters are presented in Table 3.*

Referee #1: p. 11, line 4. What is meant by optimal fit?

Response: *Optimal fit means the best possible match between observed and simulated time series. Performance of the fitted model is quantified by three different performance indicators in Table 4.*

Referee #1: p. 15, line 2 and Table 3. Shouldn't these parameter values be the same across all the model variants? What is physically changing in the system that these parameter values would change across the model variants?

Response: *These parameters should not be the same because model calibration means change in model parameters to achieve the best fit to observations. As different observations are used in each variant, the values of calibration parameters are different, too. There is no physical change in the basin but different parameter values indicate that model parameterization cannot be uniquely determined. Calibration just to streamflow observations, as is usually done in hydrological modeling, does not assure a correct simulation of water storage changes, for example. In our study, the parameters optimized by just using remote sensing information for model calibration (variant RS) lead to an unsatisfactory simulation of inflow into the lake (see Table 4).*

Referee #1: p. 16, lines 2-3. There could be a net groundwater abstraction but still areas where there is recharge. Is this an issue of coarse spatial resolution?

Response: *Yes, there could be areas of recharge due to irrigation with surface water. However, the study just analyzes the mean behavior over the whole basin, based on the results in 22 grid cells.*

Referee #1: p. 18, lines 1-10. The discussion of uncertainty here is missing a fundamental point. Calibration cannot help if the model structure is uncertain (or in error) or the temporal or spatial scaling of the model is mismatched to the modeled parameters of interest. This discussion also heads in a different direction than "what do we learn from the calibrations?" The text never explains what was learned. What was learned? Please discuss.

Response: *Regarding uncertainty of model structure, we plan to refer to multi-model ensembles in the new discussion section (see above). In this section we do explain what we learned from our calibration exercise or rather from adding more types of observations. It is outside the scope of the paper to discuss the impact of temporal or spatial scales on model results as we did not investigate this (or e.g. uncertainty due to the applied climate forcing). Thus, when we discuss different calibration variants we discuss how uncertainty can be reduced by additional observational data types.*

We believe that in the "What we learn from the calibration?" section, the key findings were reported. The most important of them reveals that a single objective/observation calibration cannot capture hydrological dynamics and there is no guarantee a well-simulated model based on a tuned variable can properly simulate other components of the model. As a result, for a general statement about water resources in a given region a multi-objective calibration is required.

Referee #1: p. 18, lines 11-20. These statements are better placed in the introduction to justify the use of global hydrologic models. Still, why is a global model the appropriate choice when the domain of study is limited to one hydrologic basin (Urmia)?

Response: *We prefer to leave this paragraph in the results section to clarify the context of the calibration exercise that is on the one hand done to efficiently analyze the specific situation in the Lake Urmia basin but also to evaluate the value of calibrating global hydrological models against multiple observation types. The reasons for using a global model have been given above.*

Referee #1: p. 19, line 22. What beta?

Response: *At the beginning of the sentence, there is the reference to Eq. 2. Beta adjusts the net abstraction from groundwater.*

Referee #1: p. 19, line 30. "much less overestimated" means what?

Response: *Its means that although the model variant still overestimates inflow to the lake, the degree of overestimation has decreased strongly.*

Referee #1: p. 19, line 31-32. This doesn't make sense to me. How come it is ok to change the parameter in one model variant but not others?

Response: *It is another thing we can learned from the calibration. As shown in Figure 6, in each variant the model is calibrated against different observations. For example, in the first variant the WGHM evaluated based on TWSA and in the second one WGHM evaluated based on TWSA and inflow into the lake. In the first variant we did not need to change the parameters which have no effect on simulated TWSA. But in the second variant we have to calibrate model against both TWSA and inflow into the lake. So, in this variant the parameters which have an effect on inflow into the lake also should be adjusted. It means that when we have a single objective calibration it is possible we reach to appropriate results via different combination of parameters values. When we add more and more objectives or observational data types in the calibration process, the number of parameters which should be changed.*

Referee #1: p. 20, lines 1-5. I would expect to see better calibration with more observational data (i.e., stream flows and lake levels).

Response: *We agree with you. Our results agree with your expectations, and this is what we express in the sentence starting in line 4: "Still, calibration to both observational data types leads to the best simulation of both annual lake inflow and lake volume anomalies."*

Referee #1: p. 20, line 22. I'm confused. The scenario "with reservoirs but without human water use" does not fit either of the two scenarios described in the prior sentence.

Response: *WHGM has the capability to assess the effect of dam building on water resources without considering human water use from reservoirs. Our results showed that the results with and without reservoirs has only 2% difference which indicated insignificant effect of reservoirs on water resources over the basin. So the run with or without reservoirs effect can be considered the same. As a result, we can consider the run without human water use as WGHM run under natural condition.*

Referee #1: p. 20, line 24. What is meant by anomalies? This term has still not been defined.

Response: *Anomalies is the difference from an average, or baseline for example in this study GRACE data are total water storage anomalies based average of its observation between 2004-2009. Added to the text when introducing GRACE (see response above).*

Referee #1: p. 21, line 17, "The lower lake water loss…" What are the loss terms besides evaporation? How are these other loss terms smaller when inflow is larger? Explain.

Response: *Evaporation is the only loss term of the lake. There was less decrease in lake storage due to more inflow into the lake. We have revised the sentence as:*

"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."

Referee #1: p. 21, lines 20-23. I don't follow this explanation. There are too many NAs in this sentence. What causes the difference between the naturalized and anthropogenic scenarios?

Response: *In naturalized runs, abstraction from the surface water or groundwater are assumed to be zero, while and water flows and storages vary only due to climate variations while in the anthropogenic scenario we simulate both the effect of water abstraction by human and of climate variations. We have deleted the sentence in lines 20-22.*

Referee #1: p. 21, lines 25-28. I don't follow. What is the connection between the first part of the sentence and the second part?

Response: *We wanted to state that inflow was less than the minimum environmental water requirements ($3{,}085 \cdot 10^6$ m^3) and that therefore a loss of lake water volume is expected. So, the first part of the sentence indicated under anthropogenic situation the inflow into the lake since 2008 never has reached to minimum environmental water requirements. The second part is related to naturalized condition which showed that under naturalized condition only in 2008 and 2009 the inflow into the lake were less than minimum environmental water requirements. We intend to reformulate the first sentence as follows:*

"Since 2008 the inflow into the lake has never reached 3,085·10⁶ m³/yr. This is the value estimated to be the minimum environmental water requirements that compensates the amount of annual evaporation from of the lake surface (Abbaspour and Nazaridoust, 2007). Therefore, a decrease of lake water storage can be expected for the best estimate of WaterGAP of 2,639·10⁶ m³/yr."

Referee #1: p. 21, lines 28-32. Is a run-on sentence.

Response: *We agree with you. We have revised it.*

Referee #1: p. 21, lines 32 – 21. Put these ratios in context. What is desirable? Undesirable? What has implications for lake health? What values are acceptable?

Response: *We think that it does not make sense to talk about acceptable values. From the perspective of lake health, a higher value may be more desirable, or rather that the values under anthropogenic conditions become closer to the values under naturalized conditions. However, we do not think it is necessary to state this explicitly in this scientific publication, it will be clear to the reader even without stating this explicitly.*

Referee #1: p. 23, line 15. How can water storage be negative?

Response: *Total groundwater storage is not computed in WaterGAP, only storage relative to a storage that occurs when the heads in surface water and groundwater are the same. Negative values of groundwater storage computed by WaterGAP indicate that net abstractions from groundwater are larger than natural groundwater recharge, while baseflow is zero. Groundwater levels can be assumed to have dropped below the surface water heads. In WaterGAP, groundwater recharge from rivers is not taken into account.*

Referee #1: p. 23, line 18. This is an interesting result. It needs to be much more strongly emphasized. These cells are the locations where groundwater declines and there could be problems.

Response: *As WaterGAP does not include reliable high-resolution information on irrigated areas and groundwater use infrastructure, the computed cell-specific net abstractions from groundwater are highly uncertain. This is why our study focuses on basin averages, and cell-specific results are less prominently shown.*

Referee #1: p. 23, lines 21-23. This qualification and limitation seems rather important. Why should the model results be trusted or used if the model does not get groundwater storage correct?

Response: *No model is perfect, and calibration as done here is a way to compensate for a lack of process accuracy. Due to calibration, we do get groundwater storage (more or less) right.*

Referee #1: p. 23, lines 23-25. Run-on sentence. What is meant by the clause with maximum?

Response: *As mentioned in p. 23 line 22, WaterGAP cannot simulate a possible drop of the groundwater table below the surface water level in the absence of groundwater abstractions, and groundwater storage might in reality have been lower before the start of groundwater abstraction than simulated in the naturalized run. Thus, contribution of human water use to groundwater storage decline might therefore be overestimated.*

Referee #1: p. 24, lines 18-19. Is this result surprising? More calibration data means a better fit model. How does this result improve understanding of the Lake Urmia system?

<u>Response</u>: *There are a few hydrological modelling studies on Lake Urmia basin but all of them (except Chaudhari et al. (2018) who has not implemented calibration at all) have only a singly observational data type/objective for calibration. So as first multi objective calibration study of Lake Urmia basin we show how a single objective calibration does not have the proper capability for a comprehensive modeling in the basin. As a result, this study provides a unique information for understanding the Lake Urmia basin system not only the Lake Urmia which does not reported in any previous studies. In addition, if the model structure or the input data are wrong, it may not be possible to improve the simulation of an increasing number of observational data type that are used for calibration. Some trade-off may occur; for example, total water storage anomaly simulation may decrease but streamflow performance may increase if streamflow is added as a second calibration data type, in addition to total water storage anomaly.*

Referee #1: p. 24, lines 25-28. Is this finding new? If so how? I feel the Urmia Lake Recovery Program has been working under the assumption that agricultural water use was a large contributor to lake decline and that they have been taking steps in recent years to address.

<u>Response</u>: *Regarding lake inflow, Shadkam et al. (2016) reported similar results as our study. However, results regarding TWSA, lake storage and groundwater storage are the new findings.*

Yes, Urmia Lake Restoration Program (ULPR) is working under the assumption that agricultural water use was a large contributor to lake decline. ULPR cannot manage climate variations or changes, so they certainly need to focus on management water use over the basin, if according to our study, human impact is significant.

Referee #1: p. 24, line 29. 90% of what?

<u>Response</u>: *90% of groundwater storage losses. We have reformulated the sentence as follows:*

"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."

Referee #1: p. 24, lines 19-24. I disagree. There are lots of other similar systems in the world – Great Salt Lake, Owens Lake, Dead Sea, Ural Sea, etc. each satisfy the first two criteria listed. What of these results is generalizable?

<u>Response</u>: *Thank you for pointing this out. We do not know what is generalizable as we have not done this calibration exercise in other basins. We intend to add one sentence at the end of the paragraph that provides a recommendation based on our study.*

"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, but such data may be crucial for a correct understanding of the freshwater system."

Referee #1: p. 24, lines 31-34. How do the model results inform the 2014-2017 trends? Also, how can climate change be constrained in this basin? Explain.

Response: *Unfortunately, almost all the input datasets are not available from 2014 onwards. Thus we did not have the model results for recent years. With "constraining climate change" we meant the global reduction of greenhouse gas emissions, nothing at the basin scale. We have revised this part as follows:*

"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."

Referee #1: Figure 9. What is being shown in panels A, B, and D? The y-axis labels were mentioned in the text but never explained.

Response: *All panels have been explained in following pages and lines:*

Figure 9a on p. 21 lines 1-5.

Figure 9b on p. 21 lines 15-16.

Figure 9d on p. 22 lines 1-14.

Referee #1: Figure A1a. The color scheme makes it difficult to differentiate grid cells. Use only three colors to differentiate the 3 types of storage. How can storage volume be negative?

Response: *We have changed the colors according to your suggestion. Total groundwater storage is not computed in WaterGAP, only storage relative to a storage that occurs when the heads in surface water and groundwater are the same. Without NAg, storage can therefore never become zero or less than zero. If NAg becomes larger than groundwater recharge, storage can obtain negative values.*

Referee #1: Data availability. I don't follow. If the authors do not have permission to share the data, then how can they share by author request? The HydroSat site underwritten by the University of Stuttgart is neat. What is the original source data for Urmia? Also, there is no water storage anomaly data for Urmia.

Response: *We have no permission to share data except by personal request for research purposes. We have indicated the source of all data in the manuscript. The data for Lake Urmia was obtained from various sources. Regarding HydroSat site, website you can download original data for lake level and extend with related reference which is Tourian et al. (2015) after registration. Also about water storage anomaly data for Urmia, for the lake water storage anomaly, it should be noted that anomalies can be calculated easily by subtracting the mean lake water storage in baseline (2004-2009) from the lake water storage time series.*

The section on data availability have been reformulated as follows:

"In-situ data from "Iran Water Resources Management Company" including groundwater levels, precipitation and temperature publicly are available upon request from the corresponding author. 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 in the supplement."

We have done our best to consider all your comments and suggestions in the revised version. Below you can see the changes have applied based on your comments.

Referee #1: The description of how the model simulates relevant processes is scant.

Response: *In the revised version we have described the model in more details as follows (section 2.1):*

"WGHM simulates daily 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 (Werth and Güntner, 2010)."

Referee #1: There is a lot of focus in the text on the multiple calibration variants run with different input data sets. What was learned from this activity? How do those result effect Lake Urmia management?

Response: *We described the lesson we learned from this activity in details in section "3.2 What we learn from the calibration?". We learned that there is no guarantee that a single objective calibration improves the model performance with respect to the simulation of other components of hydrological system. As a result, for defining the Lake restoration plans a comprehensive modeling framework like our study is needed. Specifically, quantification of return flow from irrigation is paramount for managing irrigation in the basin. We have added the following paragraph to the conclusion, as the last one:*

"Our study has shown that 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."

Also, to clarify of research objectives we have revised the objectives of the paper in the Introduction section as follows:

"The aim of our study was twofold. On the one hand, we wanted to quantify, by a holistic and reliable modelling approach, 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, 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)."

Referee #1: What are the limitations of this study? The discussion of uncertainty in the results needs to go much deeper and be more specific. This uncertainty is real and likely plays a large role in the interpretation of the results.

<u>Response</u>*: In the revised manuscript, we have added, as section 4.3, a short discussion.*

"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 the basin as a whole and 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 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 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 uncertainties of the observation data. GRACE TWSA data are more reliable for larger (100,000 km2 (Landerer and Swenson, 2012)) areas than the basin area. 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 "observed" lake water volume decline, are be affected by a likely underestimation of the actual decline by the "observed" value derived 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."

Referee #1: I found the writing difficult to follow in numerous places, particularly the results section. There are lots of acronyms, run-on sentences, and text that digresses from the section headers or topic sentences of paragraphs. The writing here made it difficult for me to see the main results and findings of the work.

<u>Response</u>: *We have revised whole of manuscript and reformulate the results section for the revised version.*

Referee #1: p. 3, line 5. I think "somewhat recovered" is overstated. Hard to tell from Figure 3. Maybe stabilized.

Response: *We agree with you. So, we have revised the sentence as follow:*

"After 2015, lake extent and storage have stabilized".

Referee #1: p. 7. What is total water storage anomaly (TWSA)? This term seems rather central to the paper. Please explain.

Response: *We have inserted the following sentence as the second sentence of the GRACE section (section 2.2.1):*

"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."

Referee #1: p. 8, lines 13-18. This method of applying (1- return flow multipliers) to the abstractions to estimate consumptive use assumes that water is used by only one water user. Is this a realistic assumption? If the return flow is used by another agricultural user and then again by a 3rd or 4th user, the basin-wide consumptive use fraction will be much different than the values reported. The large grid size magnifies this error. Table 1. How sensitive are study results to the values in this table?

Response: *In our opinion due to the fact that most return flow returns to groundwater, there is no concern in this regard. On the other hand, in arid area e.g. Urmia basin the return flow to surface water in each irrigation is not too much that can be used by another user. Anyway, we do not know exactly how the authors of the values determined them. However, the study results are not sensitive to the independent return flow estimates that were used only in one variant (RS_Q_GW_NA). We have written in section 4.1:*

"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 significantly and only leads to slight parameter adjustments. This indicates a reasonable simulation of per hectare water consumption for irrigation by the WaterGAP model."

Referee #1: p. 21, line 17, "The lower lake water loss…" What are the loss terms besides evaporation? How are these other loss terms smaller when inflow is larger? Explain.

Response: *Evaporation is the only loss term of the lake. There was less decrease in lake storage due to more inflow into the lake. We have revised the sentence as:*

"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."

Referee #1: p. 21, lines 25-28. I don't follow. What is the connection between the first part of the sentence and the second part?

Response: *We wanted to state that inflow was less than the minimum environmental water requirements (3,085·10^6 m^3) and that therefore a loss of lake water volume is expected. So, the first part of the sentence indicated under anthropogenic situation the inflow into the lake since 2008 never has reached to minimum environmental water requirements. The second part is related to naturalized condition which showed that under naturalized condition only in 2008 and 2009 the inflow into the lake were less than minimum environmental water requirements. We intend to reformulate the first sentence as follows:*

"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 that compensates the amount of annual evaporation from of the lake surface (Abbaspour

and Nazaridoust, 2007). Therefore, a decrease of lake water storage can be expected for the best estimate of WaterGAP of 2,639·10^6 m^3/yr."

Referee #1: p. 24, line 29. 90% of what?

Response: *90% of groundwater storage losses. We have reformulated the sentence as follows: "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."*

Referee #1: p. 24, lines 19-24. I disagree. There are lots of other similar systems in the world – Great Salt Lake, Owens Lake, Dead Sea, Ural Sea, etc. each satisfy the first two criteria listed. What of these results is generalizable?

Response: *Thank you for pointing this out. We do not know what is generalizable as we have not done this calibration exercise in other basins. We intend to add one sentence at the end of the paragraph that provides a recommendation based on our study.*

"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, but such data may be crucial for a correct understanding of the freshwater system."

Referee #1: p. 24, lines 31-34. How do the model results inform the 2014-2017 trends? Also, how can climate change be constrained in this basin? Explain.

Response: *Unfortunately, almost all the input datasets are not available from 2014 onwards. Thus we did not have the model results for recent years. With "constraining climate change" we meant the global reduction of greenhouse gas emissions, nothing at the basin scale. We have revised this part as follows:*

"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."

Referee #1: Data availability. I don't follow. If the authors do not have permission to share the data, then how can they share by author request? The HydroSat site underwritten by the University of Stuttgart is neat. What is the original source data for Urmia? Also, there is no water storage anomaly data for Urmia.

Response: *We have no permission to share data except by personal request for research purposes. We have indicated the source of all data in the manuscript. The data for Lake Urmia was obtained from various sources. Regarding HydroSat site, website you can download original data for lake level and extend with related reference which is Tourian et al. (2015) after registration. Also about water storage anomaly data for Urmia, for the lake water storage anomaly, it should be noted that anomalies can be calculated easily by subtracting the mean lake water storage in baseline (2004-2009) from the lake water storage time series.*

The section on data availability have been reformulated as follows:

"In-situ data from "Iran Water Resources Management Company" including groundwater levels, precipitation and temperature publicly are available upon request from the corresponding author. 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 in the supplement."

Point-by-point response to the Referee #2:

Referee #2: It can be considered as an interesting update in vast literature of Lake Urmia studies while authors tried to use a vast variety of data between 2003 to 2013 to evaluate the situation of Lake Urmia. I think a Major revision are needed prior to evaluating its technical quality. My comments are listed as bellow

Response: *We would like to thank you for the thorough consideration and critical comments that helped us improving the manuscript. We have tried to do our best to consider all your recommendations in the revised version. Below, we have provided a point-by-point response to your comments.*

General comments:

Referee #2: A technical proof reading is needed since some of the sentences are not understandable.

Response: *We have revised the whole manuscript and the manuscript will be checked by a native speaker.*

Referee #2: Your given figures and tables do not necessarily indicate to the discussion you have made.

Response: *We provided discussion on all figures and tables. We agree that some part of the discussion might not directly related to the figures and tables which is the lesson we learned from this study. However, we have re- written the result and discussion section with considering your comment.*

Referee #2: Relative error in your models is important since figures are dimensionless. Still, given figures seems to have unacceptable errors.

Response: *We have calculated Relative Absolute Error (RAE). But we think you were misled because all figures have dimension. We assume that you did not follow the different calibration variants. If you see some unacceptable errors in the figures, these errors are related to the variables which did not consider as an objective function in the calibration process (see different calibration variants).*

Referee #2: Methodology should be revised since it is not clear how you evaluated the figures in discussion

Response: *We do not understand the comment well. However, we have revised the methodology section.*

Specific comments:

Referee #2: You are suggesting that the Lake would have been vanished any way but there are a vast literature against your statement. What is your comments? You should also add sentences in the text about it.

Response: *We do not suggest at all that the lake would have vanished without human water use. For example, we stated the following:*

"Still, even in the WGHM-NAT, the average inflow into the lake from 2009-2013 of 3,670·10^6 m^3 would have been only enough to keep the lake from further loosing volume but would not have been enough for a recovery to conditions between 2003 and 2007 (Fig. 9b),"

Also in the revised version section 4.2, we have indicated why some studies were not in line with our statement and vice versa as follows:

"In order to define the lake restoration program, it is vital to know which factors contribute how much to 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 et al. (2018) concluded that human-induced changes accounted for 86% of the lake volume decline during 1995-2010, while we determined the value of 40% for 2002-2013. According to our study, human water use was the reason for 41% inflow reduction into the lake during 2003-2013 which is similar to the values of Shadkam et al. (2016) for the years 2003-2009 (comp. their Figs. 8). Discrepancies are likely due to different analysis methods but different analysis periods, as well as different 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 statistical analysis assumes that river inflow can be considered to reflect "anthropogenic impacts" while precipitation and evaporation changes reflect climatic variations while river inflow is in reality also affected by climate 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 explain the dramatic decline of the lake level. They did not use in-situ data (except lake water level data) for their analysis. Based on a analysis of Standardized Precipitation Index (SPI), a drought index, AghaKouchak et al.)2015) reported there was no significant trend in droughts over the basin during past three decades and concluded from this that human activities not climatic variations are the main reason 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 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 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.

In addition, 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. 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, 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 lake volume decline of 86% determined by Chaudhari et al. (2018) may arise from the poor performance of the uncalibrated model."

Referee #2: One of the main factors in your study is the calibration of satellite data and application of filters on data which are main issues. E.g. In GRACE data how did you manage to use 2degree precision into such a cristal clear results?

Response: *We agree with you. We used satellite data for three objectives including the irrigated area, lake level (and extent) and TWSA. Irrigated area and lake level were taken from previous studies which they needed filters on data before application. Hence, among all used satellite data, here we only discuss GRACE data. To deal with your mentioned issue about GRACE data, we have used CSR mascon solutions product which based on Save et al. (2016) do not need any filters. Also, as we mentioned in the manuscript, we know that it is recommended GRACE data products only for areas with at least 100,000 km^{2} (Watkins et al., 2015; Landerer and Swenson, 2012). But 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.*

Referee #2: Why did you use a time length between 2003 to 2013? Since the decreasing trend have already started from late 90's.

Response: *The main reason was shortage of data. The GRACE data and irrigated area which play important roles in this study were not available for late 90's. As a results we faced substantial missing data before 2000.*

Referee #2: There are a lot of missing data in historical time series records of the region. How did you manage to remove them? are they satisfactorily acceptable methods to be applied?

Response: *We disagree. There are no significant missing data for 2003-2013. The method used for filling the gaps are as follows:*

- *Irrigated area: There is no missing data in this data (except 2013).*
- *GRACE data: There are some missing data (8 months) in its observations which have filled using linear interpolation.*
- *Inflow into the Lake: There is no missing data in the dataset of annual inflow into the lake.*

- *Groundwater levels: We assessed data of 635 wells then we removed the wells with more than 12 months or six consecutive months missing values. After removing the wells with significant missing values, we have worked on 284 wells. If there are missing data in the dataset of these 284 wells, we used linear interpolation (if only one month is missing) or linear regression (with the nearest well in common period) for filling the gaps in data.*
- *Water withdrawals: There is no missing data in this data.*
- *Precipitation: Almost there is no missing data in the studied stations between 2003-2013. Few missing data filled in comparison with near stations.*
- *Temperature: There is no missing data in this data.*
- *Lake volume: There are some missing data for lake volume which have filled using linear interpolation.*

Referee #2: Page 9, Line 20, Calibration: You have to give the error evaluation if you have used "try and error" method.

Response: *We used trial and error to determine the most appropriate parameters of model in each calibration of variant based on the evaluating the model error with respect to the observations used, while trying to keep the number of adapted parameters at a minimum. We provide the final errors in Table 4. We do not think that it is interesting for the reader to provide the errors/model performances for all trials (there were many).*

Referee #2: Page 12, Performance criteria, Line 5: These criteria do not show relative error in your models since RMSE number is not necessarily satisfactory.

Response: *We have calculated Relative Absolute Error (RAE) in the revised version.*

Referee #2: Page 13, Figure 7: None of these figures indicate to an acceptable calibration.

Response: *This figure does not show any calibration results. This figure shows the inputs of model in different variants, not the model's output. Thus, there is no reason to fit the lines in this figure.*

Referee #2: Page 14, Figure 8d: The discrepancy and error is growing in your anomalies. How do you interpret?

Response: *It should be noted that Figure 8d shows change not anomalies. Anyway, we agree that the error in the second half of the time series is bigger than the first half. However, the difference in errors is minor. Also, discrepancies in groundwater storage (e.g. in peak of seasonality) can represent some minor discrepancy in groundwater storage changes. If you consider Figure 8c, which shows the normalized groundwater storage, there is no growing in discrepancy.*

Referee #2: Page 18-23: Your discussions are too long, yet none of them are visuable from given tables and figures. It is a very long article, yet given informations are narriative and reader should accept your sentences without having a chance to approve it.

Response: *We have re-written the results and discussions considering your comment.*

The changes have applied based on the comments of the Referee #2

We have done our best to apply all your comments and suggestions in the revised version. Below you can see the changes have applied based on your comments.

Referee #2: Relative error in your models is important since figures are dimensionless. Still, given figures seems to have unacceptable errors.

Response: *We have calculated Relative Absolute Error (RAE). But we think you were misled because all figures have dimension. We assume that you did not follow the different calibration variants. If you see some unacceptable errors in the figures, these errors are related to the variables which did not consider as an objective function in the calibration process (see different calibration variants).*

Referee #2: Methodology should be revised since it is not clear how you evaluated the figures in discussion

Response: *We have added the following sentences in section 2.4*

"Trends and overall behaviour of the time series were also analysed."

Referee #2: You are suggesting that the Lake would have been vanished any way but there are a vast literature against your statement. What is your comments? You should also add sentences in the text about it.

Response: *We do not suggest at all that the lake would have vanished without human water use. On page 21 line 28, for example, we state the following:*

"Still, even in the WGHM-NAT, the average inflow into the lake from 2009-2013 of 3,670·10^6 m^3 would have been only enough to keep the lake from further loosing volume but would not have been enough for a recovery to conditions between 2003 and 2007 (Fig. 9b),"

Also in the revised version section 4.2, we have indicated why some studies were not in line with our statement and vice versa as follows:

"In order to define the lake restoration program, it is vital to know which factors contribute how much to 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 et al. (2018) concluded that human-induced changes accounted for 86% of the lake volume decline during 1995-2010, while we determined the value of 40% for 2002-2013. According to our study, human water use was the reason for 41% inflow reduction into the lake during 2003-2013 which is similar to the values of Shadkam et al. (2016) for the years 2003-2009 (comp. their Figs. 8). Discrepancies are likely due to different analysis methods but different analysis periods, as well as different conceptualizations, make a direct comparison of the estimated relative contributions difficult.

"In order to define the lake restoration program, it is vital to know which factors contribute how much to 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 et al. (2018) concluded that human-induced changes accounted for 86% of the lake volume decline during 1995-2010, while we determined the value of 40% for 2002-2013. According to our study, human water use was the reason for 41% inflow reduction into the lake during 2003-2013 which is similar to the values of Shadkam et al. (2016) for the years 2003-2009 (comp. their Figs. 8). Discrepancies are likely due to different analysis methods but different analysis periods, as well as different 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 statistical analysis assumes that river inflow can be considered to reflect "anthropogenic impacts" while precipitation and evaporation changes reflect climatic variations while river inflow is in reality also affected by climate 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 explain the dramatic decline of the lake level. They did not use in-situ data (except lake water level data) for their analysis. Based on a analysis of Standardized Precipitation Index (SPI), a drought index, AghaKouchak et al.)2015) reported there was no significant trend in droughts over the basin during past three decades and concluded from this that human activities not climatic variations are the main reason 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 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 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.

In addition, 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. 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, annual inflow into the lake was estimated to be 3,700·106 m3 in 2003 (their Fig. 8) while observed inflow was much higher, 5,835·106 m3. In 2009, observed inflow, with 1,036·106 m3, was only half of the simulated one. Therefore, the very high human contribution to lake volume decline of 86% determined by Chaudhari et al. (2018) may arise from the poor performance of the uncalibrated model."

Beside reviewers' comments and suggestions, we also consider the short comments in the revised version. Below the changes have applied based on the short comments have been written.

The changes have applied based on the comments of S. Chaudhari as short comment

Comment: The authors should use contrasting colors in figures. It is difficult to distinguish the WGHM-ANT and WGHM-NAT lines in Fig 9 due to similar colors.

Response: We have changed the color of WGHM-NAT in the revised version.

The changes have applied based on the comments of S. Khatami as short comment

Comment: Other metrics such as Willmott's refined index of agreement [Willmott et al., 2012] and KGE [Gupta et al., 2009] shown to be better than NSE.

Response: *We have calculated the KGE in the revised version.*

Comment: Using the term validation is both semantically and theoretically wrong. As a matter of good practice, it's been recommended to use the term evaluation instead of validation.

Response: *"validation" was replaced with "evaluation".*

Comment: You have used CC and NSE (P 9 L 11-12) to cross compare precipitation and temperature records of difference sources. It is better to use (more) resistant alternatives such as Spearman ranked correlation (instead of Pearson correlation) and Willmott's refined index of agreement [Willmott et al., 2012]

Response: *We have calculated the Willmott's refined index of agreement in the revised version.*

Comment: this is an unsubstantiated claim. As far as I know there is no (reliable) evidence on the degree of awareness regarding this issue. Please remove it, or provide the evidence.

Response: *We have removed the sentence.*

Comment: P 25 L 9: It is better to explicitly acknowledge the organisations that provided you with GRACE and climate data, and the URL links if available.

Response: *We have explicitly acknowledged the organizations that provided the used data. The URL links are available in the text.*

Comment: P 28-29: The URLs in the ULRP references are not accurate. Please update them.

Response: *We have updated the URLs.*

****

References

Alborzi, et al.: 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.

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.

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.

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, 2014.

Farokhnia, A.: Impact of Land Use Change and Climate Variability on Urmia Basin Hydrology (Dissertation) Tarbiat Modares University, Agricultral department (in Persian), 2015.

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

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

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.

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.

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.

Scanlon, B. R. et al.: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data, Proceedings of the National Academy of Sciences, 115(6), 1080-1089, doi:10.1073/pnas.1704665115, 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.

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.

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.

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.

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 Leibnitz 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 ~~decreased precipitation,~~declining groundwater ~~levels~~tables and a very strong reduction in the volume ~~and more~~as well as recently ~~also~~ in the extent of Lake Urmia. ~~Human water use has exacerbated~~For the ~~desiccating impact of climatic variations. This~~ case of Lake Urmia basin, this study ~~quantifies~~explores the ~~contribution~~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 ~~to the reduction of inflow into Lake Urmia, to~~and climate variations. The WaterGAP Global Hydrology Model (WGHM) was for the ~~loss of~~first time calibrated against multiple in-situ and spaceborne data to analyse the decreasing lake water volume ~~and to the~~ , lake river inflow, loss of groundwater, and total water storage in the entire ~~Lake Urmia~~ basin during 2003-2013. ~~To this end, the WaterGAP Global Hydrology Model (WGHM) was manually calibrated specifically for the basin against multiple in-situ and spaceborne data, and~~Then the best-performing calibration variant was run with or without ~~taking into account~~considering water use to quantify the impact of human water use. ~~Observation data~~Observations encompass remote-sensing based time series of annual irrigated area 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~~ is~~, including the time series of lake volume not used for calibration, was achieved if the maximum amount of observations ~~is~~was used for calibration. Calibration against MODIS and GRACE ~~TWSA improves~~data alone improved simulated inflow into Lake Urmia but inflow and lake volume loss were still ~~overestimates it by 90%; it results in an overestimation of lake volume loss, underestimation of~~still overestimated, while groundwater loss was understimated and ~~a shifted~~ seasonality of groundwater storage~~.~~ was shifted as compared to observations. Lake and groundwater dynamics ~~can~~could only be simulated well if calibration against groundwater levels ~~leads~~led to ~~adjusting~~an adjustment 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, human water use was the reason for 50% of the total basin water loss of about 10 km^3 during 2003-2013, for 40% of the Lake Urmia water loss of about 8 km^3 and for up to 90% of the groundwater loss. Lake inflow was 40% less than it would have been without human water use. ~~We found~~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. These findings ~~may serve to~~can support water management in the basin and more specifically Lake Urmia restoration plans.

1 Introduction

Iran is a country with 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. ~~Furthermore, climate~~Climate change has resulted~~, in the majority of Iran's regions, not only~~ in increased temperatures ~~but also~~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 drastic decline of groundwater levels is a hidden disaster since there is little societal awareness of it in Iran.~~

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. ~~Lake water volume is now approximately 30·10^9 m^3 below its historical maximum (ULRP, 2015a).~~ Precipitation in the basin shows a decreasing trend over the period 1951-2013, with ~~particular~~particularly low values after 1995, ~~which together with increasing temperatures~~ and ~~thus~~ evaporation ~~is very likely to have contributed to the decrease in lake volume~~has increased (Alizadeh-Choobari et al., 2016). ~~Human activities include increasing surface and groundwater abstractions as well as numerous man-made reservoirs (Fig. 6 in Hassanzadeh et al., 2011).~~ Lake water volume is now approximately 30·10^9 m^3 below its historical maximum (ULRP, 2015a).

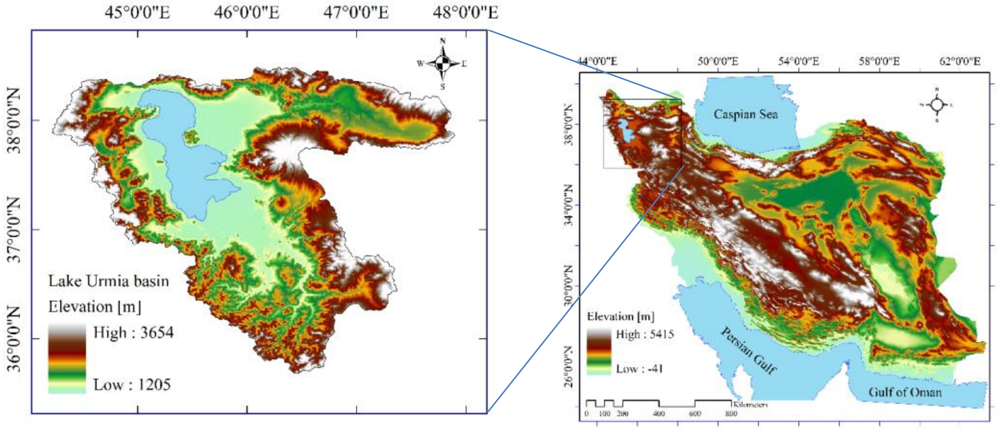


Figure 1: Location of Lake Urmia basin.

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., ~~2011~~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, 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). ~~From then on, in particular due to~~Afterwards, the water table dropped to 1,274 m in 2003 specially because of the severe drought in 1999-2001 ~~but~~ exacerbated by human water use~~, the water table dropped to 1,274 m in 2003~~ (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 ~~somewhat recovered~~stabilized (Fig. 3~~),~~) due to ~~both~~ the relatively high precipitation in 2015 and 2016, increased releases from reservoirs and management activities for decreasing water consumption (ULRP ~~(~~, 2015b).

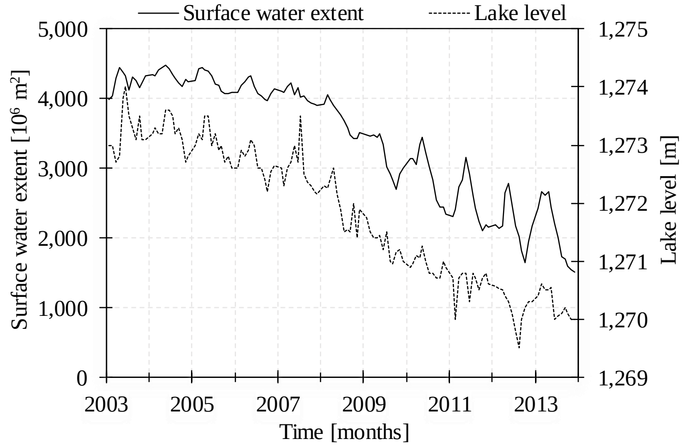


Figure 2: Time series of surface water extent and water table elevation of Lake Urmia (data from Tourian et al., 2015).

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; Alizadeh-Choobari et al., 2016; Ghale et al., 2018~~). Using~~; 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) ~~examined~~investigated the effect of irrigation system changes in the basin from the surface ~~systems~~ to pressurized systems~~ using the Soil and Water Assessment Tool (SWAT) model; eventually,~~; they found that such changes ~~in irrigation system~~would increase water productivity but would have no effect on~~ the~~ lake inflow and would reduce groundwater levels by 20%.

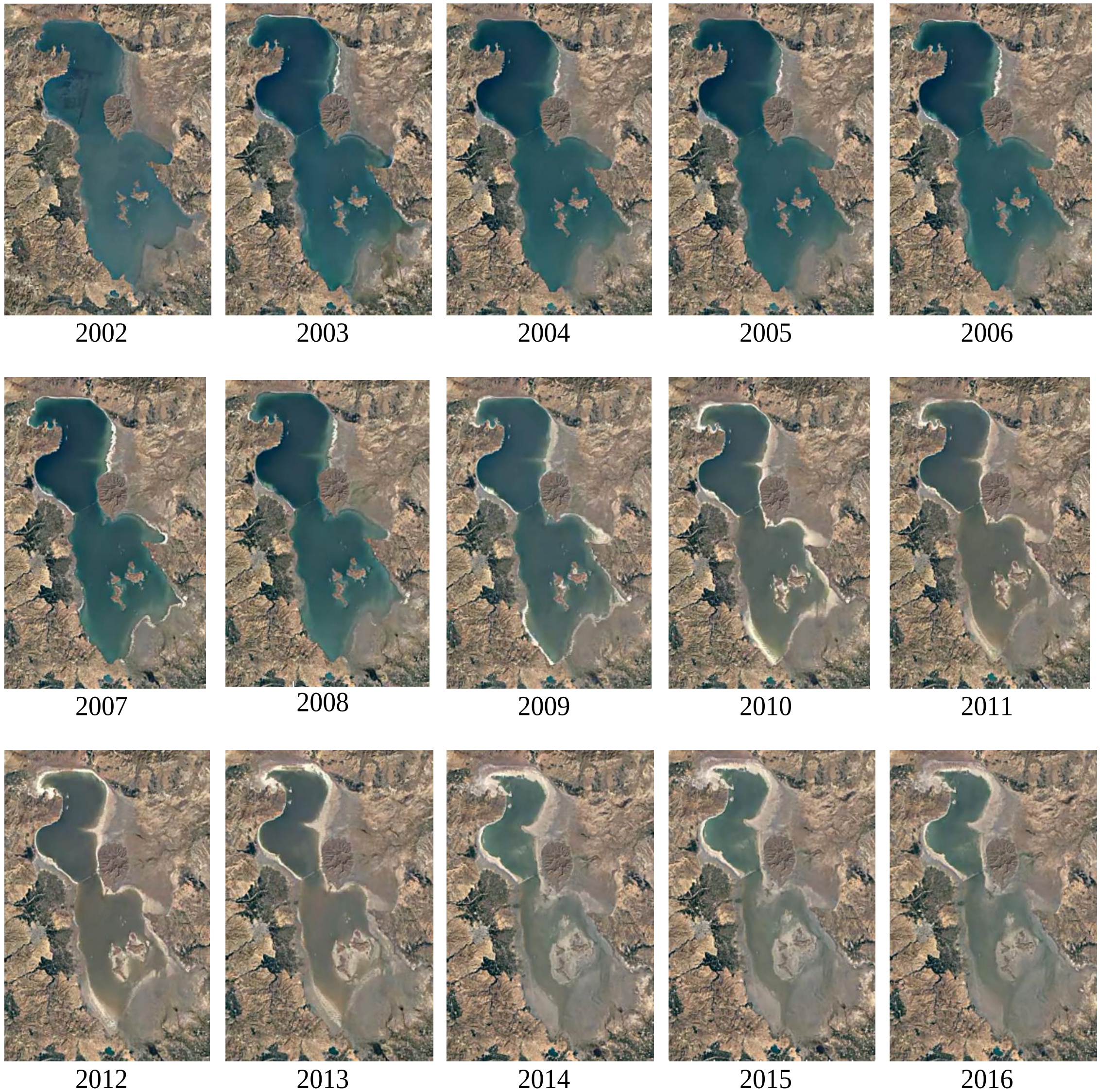


Figure 3: Lake Urmia during the time period 2002-2016 (Google Earth Timelapse, last accessed: 28 Apr. 2018).

~~Two~~Three hydrological modelling studies for Lake Urmia basin focused on quantifying the contributions of various factors on lake water volume (Hassanzadeh et al., ~~2011) or~~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. (~~2011~~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 ~~take into account~~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).~~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.

~~Thus, for quantifying the impacts of human water use and climate variations in the basin, a more holistic approach is required that takes into account all types of water use and distinguishes groundwater use from surface water use. In case of deep groundwater wells, groundwater can be pumped even in times of drought and thus low surface water availability. Groundwater abstractions may lead to a decline of groundwater storage and a decrease of groundwater discharge to surface water bodies, with negative impacts on freshwater ecosystems (Döll et al., 2014). With a spatial resolution of 0.5°×0.5°, the global water resources and use model WaterGAP (Water Global Assessment and Prognosis) 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). 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. 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). While WaterGAP provides information that is often not available at local to regional scales, a satisfactory simulation of individual basins with WaterGAP does require some adjustments based on regional information. In its standard version, WGHM~~The aim of our study was twofold. On the one hand, we wanted to quantify, by a holistic and reliable modelling approach, 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, 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~~ including annually varying correction factors for NAg and NAs. An important objective of this study was to~~. To understand the value of different observations or other regionally available data for understanding dynamics of water flows and storages in a basin~~. Therefore~~, 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~~, and~~ as well as regional data of basin-wide total withdrawals plus estimated return flow fractions (RS_Q_GW_NA variant). Model ~~validation~~evaluation was done by comparison of simulated lake water volume anomalies against observed anomalies. ~~Model~~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 ~~water~~

~~use and reservoirs, we quantified the differential~~ human impacts, we determined the contributions of human water use ~~(and reservoirs)~~ and climate variation on lake inflow ~~into Lake Urmia and lake water volume as well as on groundwater and total water storage~~and water storages in ~~Lake Urmia basin during~~ the period 2003-2013. In section 2, we describe the utilized data and the simulation setup. The results of the four calibration variants and the impacts of human water use are shown ~~and discussed~~ in section 3. Section 4 ~~presents the~~ discusses multi-observation calibration and the analysis of human impact ~~of human water use on the basin's water resources~~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 available for this period. In the following sections, WaterGAP, WaterGAP input data and observational data used for calibration as well as the calibration variants 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). ~~It computes~~With a spatial resolution of 0.5°×0.5°, it simulates water abstractions and consumptive water use (so-called net abstractions, i.e. ~~NA)~~the amount of water that evapotranspirates during use and ~~for~~does not flow to surface water bodies and groundwater afterwards) in five sectors ~~and consequently NAg~~(irrigation, livestock, domestic, manufacturing and cooling of thermal power plants); then net abstractions from either groundwater (NAg) or surface water bodies (NAs ~~by estimating return flows (~~) 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. ~~In WGHM, NAg and NAs are then~~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 and NAs are subtracted from either the groundwater or surface water bodies (lakes, reservoirs or rivers) (Müller Schmied et al., 2014).

WGHM simulates daily water storage as well as ~~fluxes~~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). ~~Representation of Lake Urmia basin by 22 0.5° grid cells agrees well with the actual basin shape and the drainage pattern within the basin.~~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 (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 specific basins only. In this study, it was run only for the 22 0.5° grid cells that represent the Lake Urmia basin in WGHM (Fig. 4).

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.

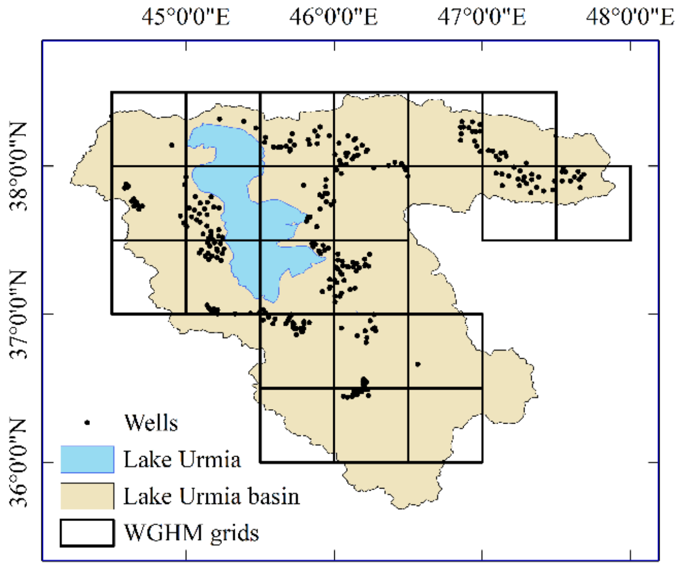


Figure 4: Grid cells in WGHM corresponds to Lake Urmia basin along with the location of groundwater wells across the basin.

2.2 Data

2.2.1 Remote sensing data

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. 45).

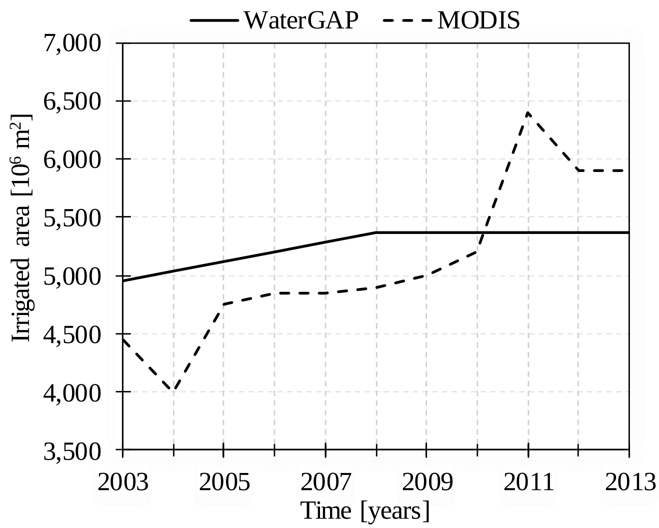


Figure 45: Irrigated area in Lake Urmia basin assumed in WaterGAP and derived from MODIS (data from Kamali and Youneszadeh Jalili, 2015).

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; http://www2.csr.utexas.edu/grace/RL05_mascons.html, 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 $8 \cdot 10^9$ 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 http://ulrp.sharif.ir/ (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.

2.2.3 Groundwater levels

For evaluating the groundwater status in Lake Urmia basin, we used groundwater head data of 284 wells during 2003-2013 (Fig. ~~5~~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 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. ~~Correlation~~The correlation coefficient (CC~~) and~~), Nash-Sutcliffe efficiency (NSE), and

Willmott's refined index of agreement (Willmott et al., 2012) were 0.985, 0.946, and 0.~~946~~897, respectively, for precipitation, and 0.996, 0.983, and 0.~~983,~~941 respectively, for temperature.

2.2.6 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 ~~validation~~evaluation of the model variants.

2.3 Calibration variants

Calibration was done by trial-and-error. It included the modification time series of irrigated area, 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. 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 ~~of~~ 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.

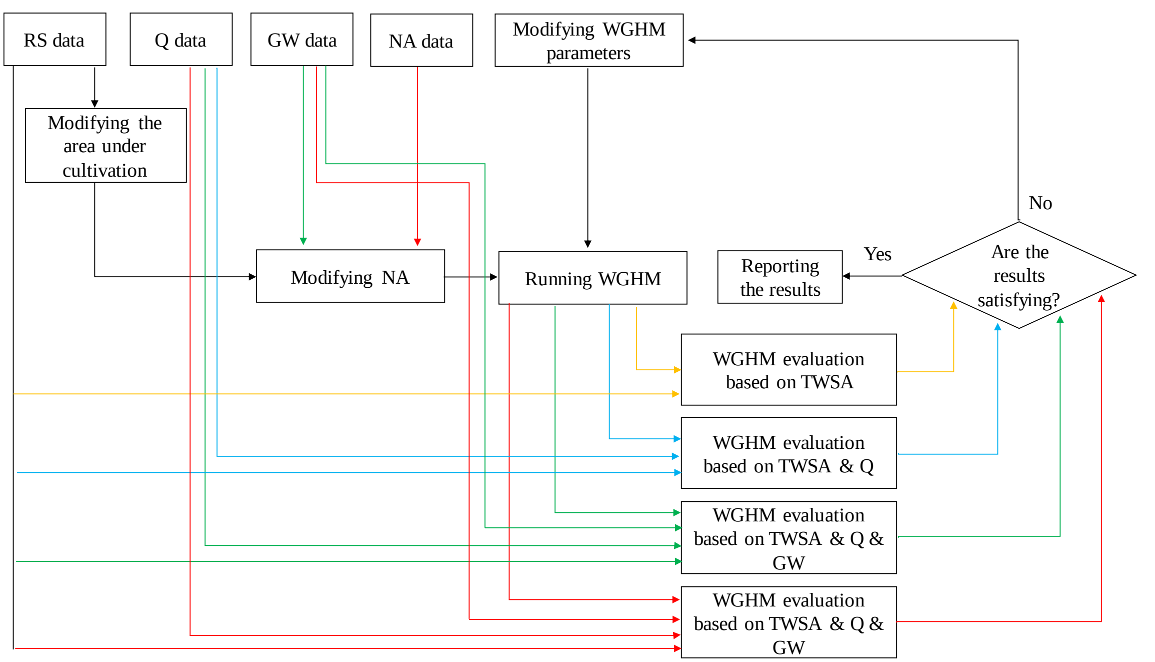


Figure 6: Flowchart for the four calibration variants. The black line is common in all variants, the mustard, blue, green and red lines represent calibration based on RS data (RS variant), RS data and inflow data (RS_Q variant), RS, inflow and groundwater level data (RS_Q_GW variant), and RS, inflow, 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 km^2, 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 http://www.fao.org/nr/water/aquastat, 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:

$$CF1(i) = \frac{Area_{irri}^{MODIS}(i)}{Area_{irri}^{WG}(i)} \tag{1}$$

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 ~~an optimal~~a good fit ~~to~~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:

$$NA(i) = \alpha(i) \times NAs(i) + \beta(i) \times NAg(i) \tag{2}$$

and new ~~optimal~~ values of temporally constant 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

$$CF2(i) = (\frac{Cu_{irri}^{WG}(i)}{Area_{irri}^{WG}(i)} - \frac{Cu_{irri}^{WG}(2009)}{Area_{irri}^{WG}(2009)}) \quad (3)$$

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, Eq. 4 was used for estimating water consumption time series over Urmia basin:

$$Cu_{irri}(i) = \left(\frac{Area_{irri}^{MODIS}(i)}{Area_{irri}^{MODIS}(2009)}\right) \times Cu_{irri}^{Obs}(2009) + CF2(i) \times Area_{irri}^{MODIS}(i) \quad (4)$$

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 ~~an optimal~~a good fit to the data also used in the RS_Q_GW variant (Fig. 6, red lines).

2.4 Performance indicators

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

$$CC = \frac{Cov\ (Obs.Sim)}{\sigma_{obs} \times \sigma_{Sim}} \quad (5)$$

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Sim_{(t)} - Obs_{(t)})^2}{\sum_{t=1}^{T}(Obs_{(t)} - \overline{Obs})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Obs_{(t)} - Sim_{(t)}\right)^2} \quad (7)$$

$$RAE = \frac{\sum_{t=1}^{T}|Obs_{(t)} - Sim_{(t)}|}{\sum_{t=1}^{T}|Obs_{(t)} - \overline{Obs}|} \quad (8)$$

$$KGE = 1 - \sqrt{(CC-1)^2 - \left(\frac{\sigma_{Sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\overline{Sim}}{\overline{Obs}} - 1\right)^2} \quad (9)$$

where *Cov* is covariance function, *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 ~~NSE~~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 calibration~~: results and discussion~~

~~3.1 Results~~

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 $4{,}185 \cdot 10^6$ m^3/yr in the standard version to $3{,}815 \cdot 10^6$ m^3/yr~~;~~, and increased from $4{,}233 \cdot 10^6$ m^3/yr to $4{,}781 \cdot 10^6$ m^3/yr in 2011-2013. ~~Increased~~However, increased net recharge of groundwater by return flows~~, however,~~ 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 $-2{,}294 \cdot 10^6$ m^3/yr in variants RS and RS_Q to $1{,}147 \cdot 10^6$ m^3/yr in variant RS_Q_GW (Fig. 7b). Keeping annual NA constant, NAs decreased accordingly from $6{,}373 \cdot 10^6$ m^3/yr to $2{,}931 \cdot 10^6$ m^3/yr. Total NA slightly decreased in variant RS_Q_GW_NA as compared to the other calibrations variants.

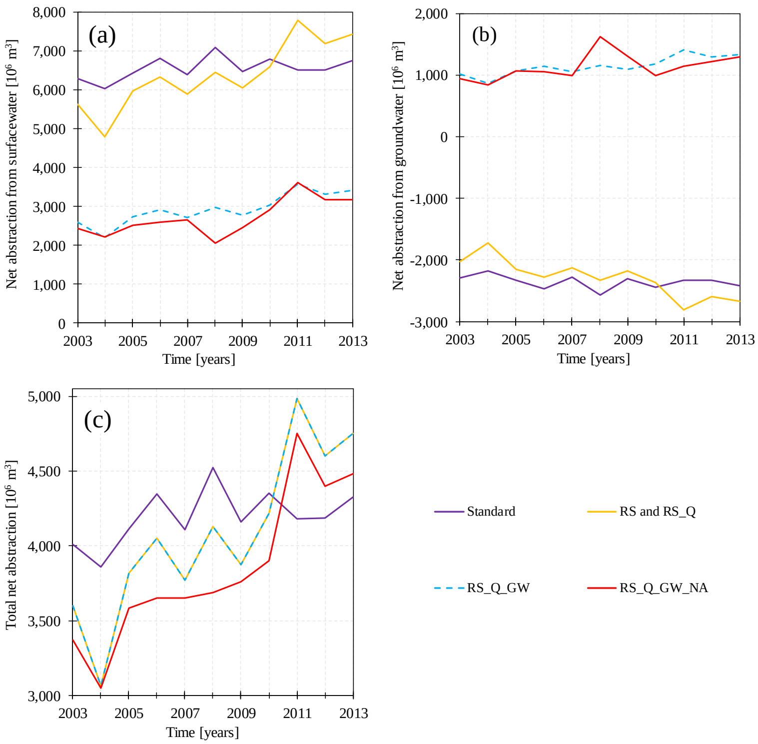


Figure 7: Time series of net abstractions from surface water (a) and groundwater (b), as well as total net abstractions (i.e. consumptive use) (c) in Lake Urmia basin in the standard version of WaterGAP as well as the various calibration variants.

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

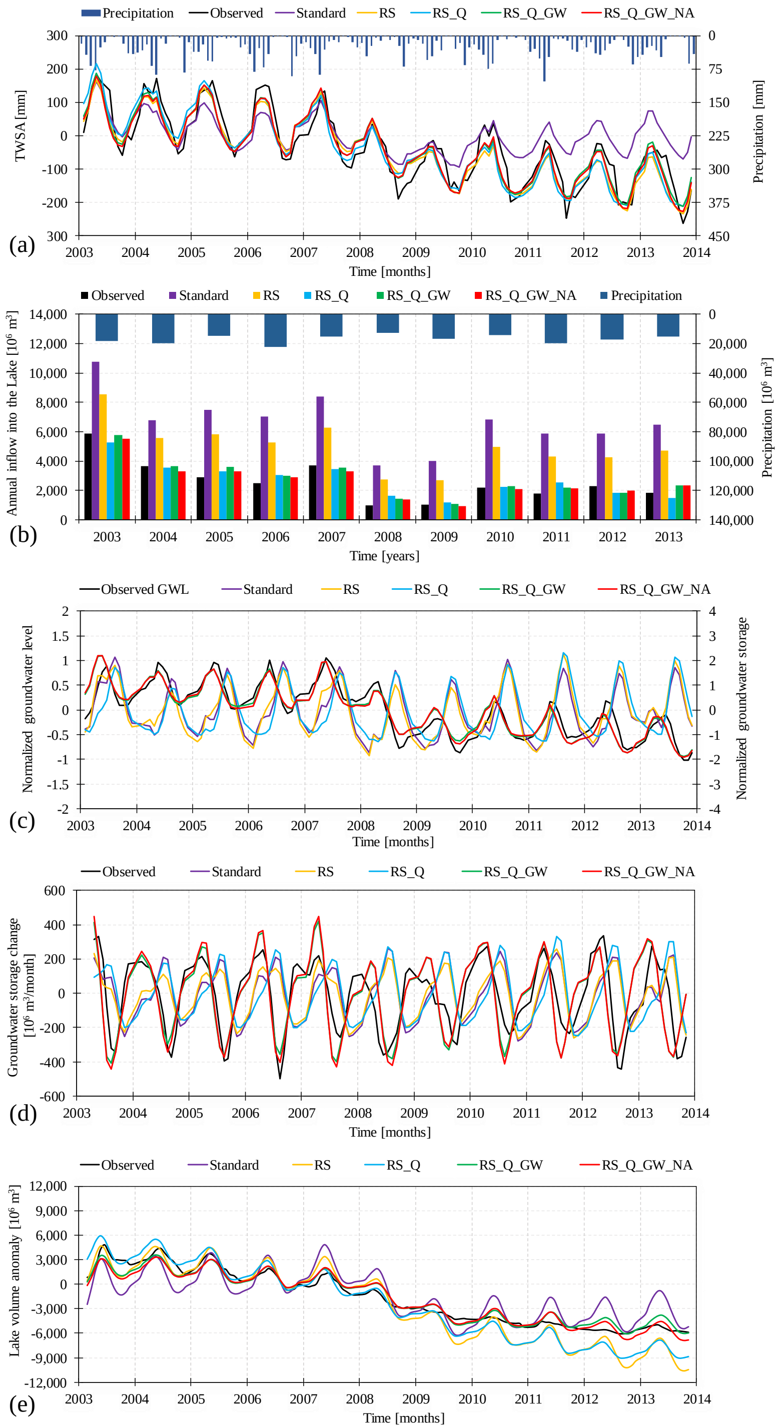


Figure 8: Time series of monthly TWSA of GRACE and WGHM (a), annual inflow into the lake Q from observations and WGHM (b) normalized observed groundwater level and normalized groundwater storage from WGHM (c), groundwater storage change GWSC from month to month from observations and WGHM (d) and the 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 ~~parameter~~parameters were re-set to the values listed in Table 3. ~~Please note~~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). ~~The~~When the more observational data types were considered in the calibration process, the ~~higher was the~~ number of parameters that needed to be adjusted increased whereas the required ~~changes in the~~ 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 WGHM version computes a much smaller loss. According to the data of Tourian et al. (2015), about 80 % of the total water loss in the basin was due to the loss of lake water. A stronger increase of human water abstractions over time (Fig. 7a), doubling of rooting depth and thus soil water capacity and a higher maximum canopy storage everywhere in the basin, as well as an increase of maximum active lake depth of Lake Urmia from 5 m to 9 m in ~~in~~ variant RS ~~result~~resulted in a good fit of WGHM TWSA to GRACE TWSA (Fig. 8a). 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. Still, simulated inflows into Lake Urmia computed in variant RS are still much higher than the observed values (Fig. 8b~~). Seasonality~~) and seasonality of groundwater levels 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

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 adjustment of the runoff coefficient and active lake depth (Table 3), while the fit to GRACE TWSA remains good (Fig. 8a). However, seasonality of groundwater table fluctuations is still not simulated properly. This 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). NAg in the standard, RS and RS_Q variants is negative, which means that there is 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 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 optimal parameter values (Table 3) and the fit to all observational data (Fig. 8).

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 estimatecompute an “observed” groundwater water storage anomaly, and thus an observed decline of groundwater storage between the year 2003 and 2013 of $1.8 \cdot 10^9$ 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_NAgNA 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, CCRMSE, RAE, and RMSEKGE with respect to monthly TWSA (Fig. 8a), annual Q (inflow to Lake Urmia, Fig. 8b) and monthly GWSC (Fig. 8d) are presented in Table 4 for the standard version and four calibrated variants. Regarding the fit to TWSA observations, NSE increased from 0.48 in the standard version to 0.84 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. This performance improvement is also reflected by CC, RMSE, RAE, and RMSE. PerformanceKGE. The performance with respect to observed inflow to the lake only improves marginally by calibration against TWSA, in variant RS. Only calibration against inflow observations strongly improves model performance, with NSE and KGE jumping from negative values for the standard and RS variantsvariant to values around 0.9. and RAE from 3.92 to 0.30. Integration of groundwater observations in again leadleads to a small performance improvement (see also RMSE). 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, all performance indicators show that consideration of remote sensing and streamflow observations only do not lead to an acceptable simulation of groundwater storage. Only the two variants 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 lake inflow, which may also be due to the uncertainty in estimating the basin-wide average 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).

For ~~validation~~model performance evaluation, we compared the lake volume simulated by WGHM with the observed lake volume of Tourian et al. (2015) (Fig.8e and Table 4). ~~All calibrated variants improve simulation of the lake volume as compared to the standard WGHM. While the~~The standard model underestimates the decline in both lake water and TWSA, all 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) due to assuming a net groundwater recharge due to surface water irrigation. Only variants RS_Q_GW and RS_Q_GW_NA simulate not only the groundwater dynamics but also the decline of lake water volume correctly. NSE for the monthly lake volume anomaly is 0.68 for the standard WGHM and improves to 0.77 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 of 0.94 or 0.95 (Table 4). We conclude that 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

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
	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
	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
Validation	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

3.2 What we learn from the 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 done 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, the well-known problem of equifinality (Beven and Freer, 2001) asks for additional types of observations to be added (Döll et al., 2016). Without additional data, more than one parameter combination can lead to a good fit to e.g. observed river discharge, but 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 input is uncertain. For example, climate input data are based on low-density climate observations and information on water use is often very scarce and outdated. It is generally not possible to obtain, for modelling at the global scale, the same detailed data for a specific region than in case of modelling just this region. Still, a global hydrological model includes all data for simulating water flows and storages in specific regions on interest everywhere on the globe, and model calibration against multiple (regional) observations is a means for improving 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 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. Please note, however, that 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 mean 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 storages. 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, 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 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, was not enough to adjust parameters of the hydrological model (in particular two groundwater recharge related model parameters, Table 3), but it was absolutely 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 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 of observations, TWSA, lake inflow and groundwater dynamics, while at the same time more and more parameters need to be adjusted (Tables 3 and 4). No trade-offs between the fits to the three observational data types occurred in the case of the Lake Urmia basin.

While introduction of annually varying corrections for NAg and NAs (Eq. 2, Table 2) for variants RS_Q_GW and RS_Q_GW_NA leads to an optimal 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 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 to all observations.

To assess the potential of using observed lake volume time series as calibration target and not only for validation, 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 did no change appreciably but not only simulated lake volume anomaly but also simulated 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 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 to 0.85. 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 Urmia, calibration against time series of lake volume anomalies could, in the absence of inflow data, help to improve simulation of inflow, while calibration against time series of inflow could, in the absence of lake volume observation, improve 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, we found that calibration aimed at optimizing the three criteria CC, NSE and RMSE with respect to monthly time series of observed total, groundwater and lake storages, with similar achieved performance values (Table 4), does not necessarily lead to similar estimates of total and compartmental water losses over the whole time period 2003 to 2013. For example, variants RS and RS_LV have the same values for all three performance criteria with respect to TWS (not shown) but TWS loss between 2003 and 2013 is simulated to be $11.15 \cdot 10^9$ m^3 and $7.86 \cdot 10^9$ m^3, respectively (Table 5). TWS loss according to variant RS_Q_GW_NA is, with $10.04 \cdot 10^9$ m^3, in between and quite different, even though NSE is only 0.04 better. We conclude that in case of relevant trends, the calibration criteria should include 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~~

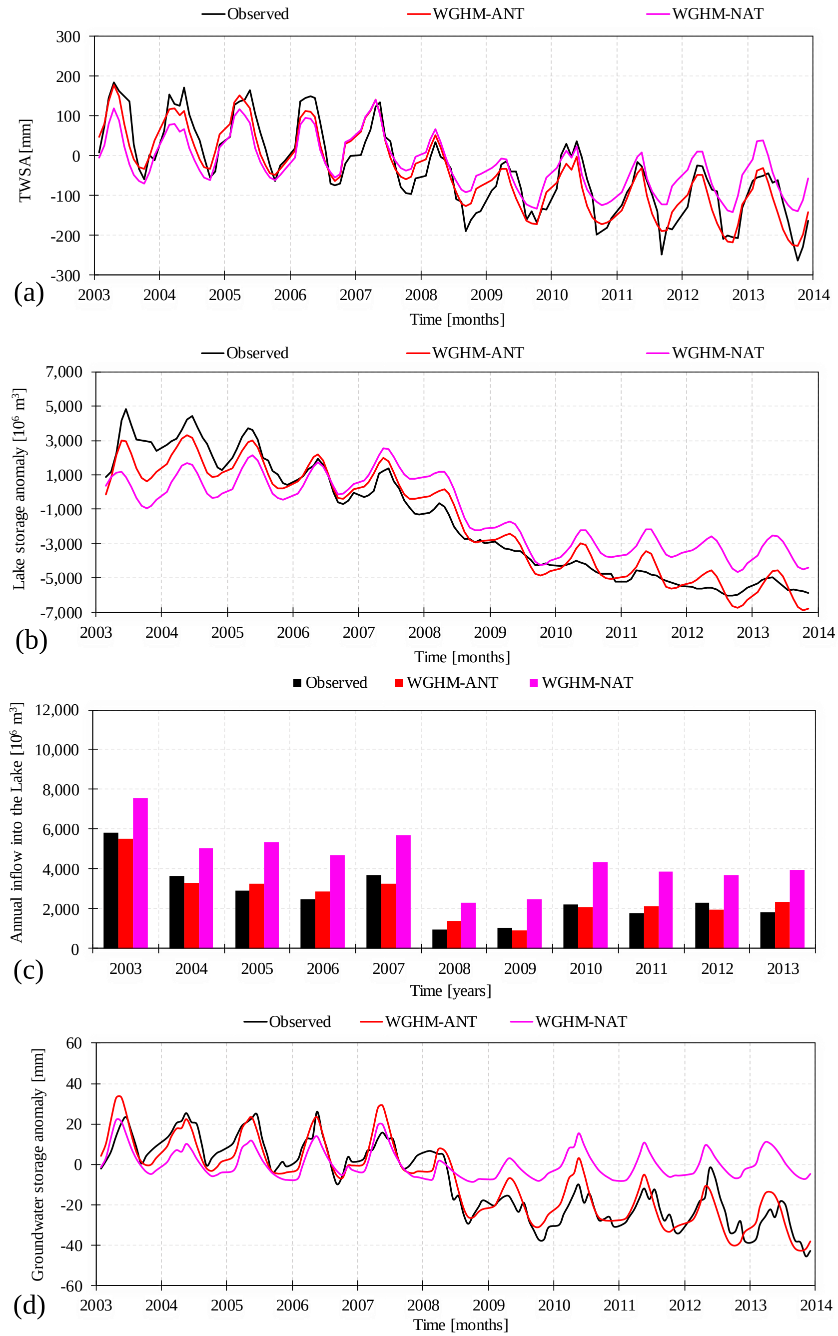
~~4~~3.2 Differential impacts of human water use and climate variation on Lake Urmia basin

The impact of human water use and man-made reservoirs on water flows and storages ~~can be~~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. ~~Note~~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~~)~~).

When comparing TWSA under anthropogenic and naturalized conditions in Fig. 9a, 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). 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 at a constant level after 2009, it decreases each year in WGHM-ANT due to consumptive increasing irrigation water use (Fig. 7c). The linear trend of WGHM-ANT and WGHM-NAT TWSA time series for the period 2003-2013 is -24.5 mm/yr (GRACE: -24.4 mm/yr) and -11.8 mm/yr, respectively. The TWSA trend for two sub-periods before and after 2008, 2003-2007 and 2009-2013 -14.2 and -16 mm/yr, respectively, for WGHM-ANT and only 0.7 and -3.85 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, $1{,}274 \cdot 10^6$ m^3 water/yr, while in the absence of human water use, it would have lost $614 \cdot 10^6$ m^3 water/yr, i.e. 52% less. Of this total water volume, $920 \cdot 10^6$ m^3/yr of lake water was lost, while only $548 \cdot 10^6$ m^3/yr would have been lost without human water use (Fig. 9b).

The ~~lower~~smaller decreasing trend for lake water ~~loss~~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 $4{,}454 \cdot 10^6$ m^3/yr under naturalized conditions, it decreases by 41% to $2{,}639 \cdot 10^6$ m^3/yr under anthropogenically altered conditions (Fig. 9c). The difference is only ~~half~~50% of NA as only a fraction of (potential) net abstractions from surface water NAs (required to allow optimal irrigation) could be ~~done~~made 1) due to a lack of water availability in the surface water bodies and 2) because ~~some part~~a fraction oft of NAg is provided a decrease in groundwater storage.~~ The computed percentage decline of lake inflow is similar to the values of Shadkam et al. (2016) for the years 2003-2009 (comp. their Figs. 8 and 9).~~

~~A loss of lake water volume is expected for these low inflows; since~~Since 2008~~,~~ the inflow into the lake has never reached $3,085\cdot10^6$ m^3~~,~~/yr. This is the value estimated to be the minimum environmental water requirements that compensates the amount of annual evaporation from of the lake surface (Abbaspour and Nazaridoust, 2007). Therefore, a decrease of lake water storage can be expected for the best estimate of WaterGAP of $2,639\cdot10^6$ m^3/yr. In WGHM-NAT, inflow was lower than $3,085\cdot10^6$ m^3 only in 2008 and 2009. Still~~, even in the WGHM-NAT~~, the average inflow into the lake from 2009-2013 of $3,670\cdot10^6$ m^3 would have been only enough to keep the lake from further loosing volume~~ but~~ (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 ~~still ~~relatively ~~dry climate~~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.



~~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.~~

While groundwater storage is estimated to decline by $251 \cdot 10^6$ m^3/yr during 2003-2013 in WGHM-ANT, the decline is only $27 \cdot 10^6$ m^3/yr in WGHM-NAT (Fig. 9d). Different from lake water storage, groundwater storage would have recovered after 2008/2009 if there had been no (increasing) net groundwater abstractions (Fig. 9d, compare Fig. 7b), even though mean groundwater recharge was on 2,579·10^6 m^3/yr during 2009-2013 as compared to 3,310·10^6 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-~~2007, and~~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. To summarize, human water use was the reason for 52% of the total water loss in the basin, for a maximum of 90% of the groundwater loss and for 40% of the Lake Urmia water loss during 2003-2013, and lake inflow was 41% less than it would have been without human water use.

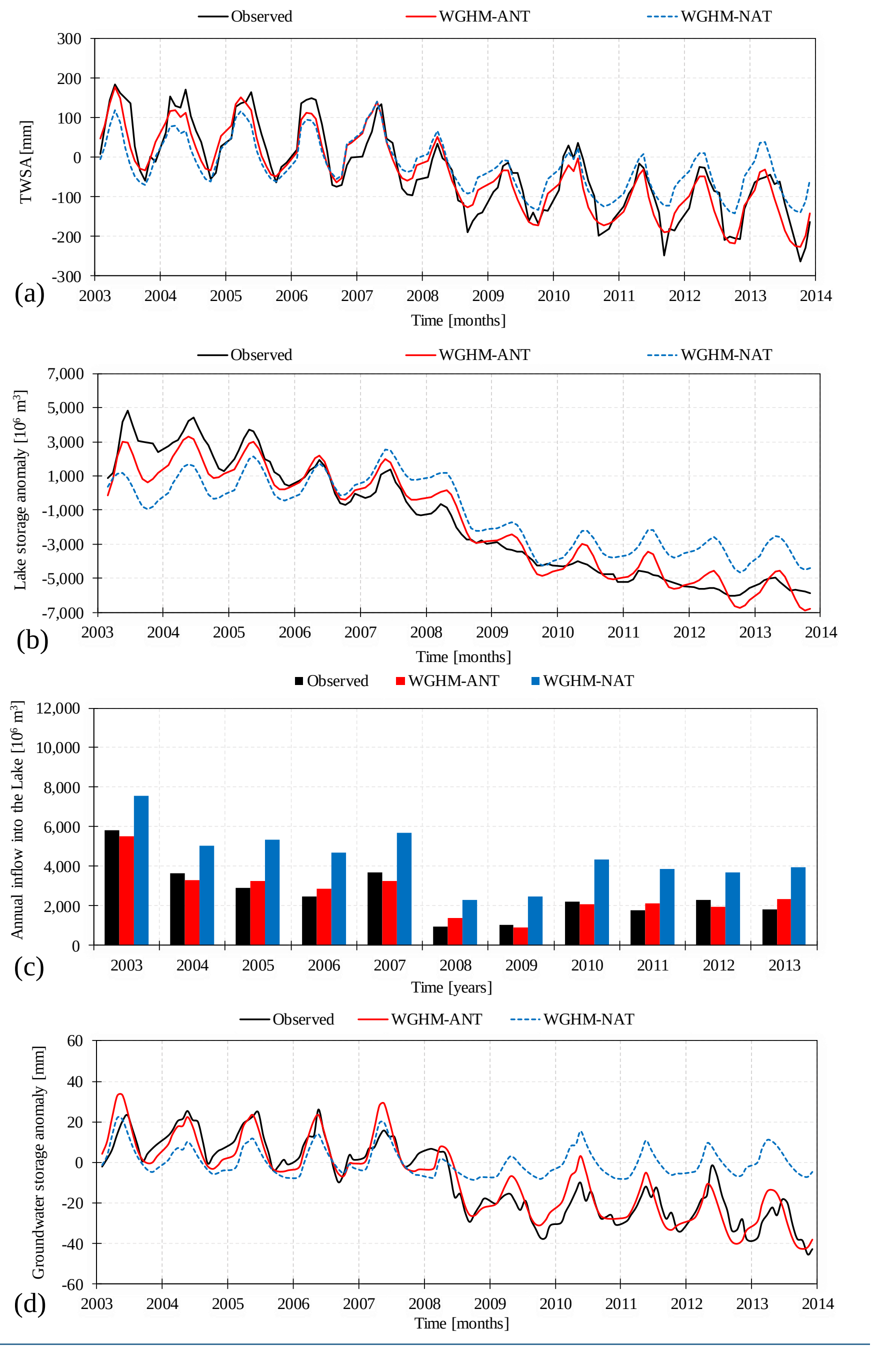


Figure 9: Time series of simulated (variant RS_Q_GW_NA) and observed monthly TWSA (a), lake water storage anomaly (b), annual inflow into the lake (c), and monthly groundwater storage anomaly (d), under anthropogenic (WGHM-ANT) and naturalized (WGHM-NAT) conditions.

4 Discussion

4.1 Multi-observation 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 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. 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, 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 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), 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 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 of observations, TWSA, lake inflow, and groundwater dynamics, while at the same time more and more parameters need to be adjusted (Tables 3 and 4). 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) for variants RS_Q_GW and RS_Q_GW_NA leads to the most suitable 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 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 did not change appreciably but not only simulated lake volume anomaly but also simulated 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 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 to 0.85. 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 Urmia, calibration against time series of lake volume anomalies could, in the absence of inflow data, help to improve simulation of inflow, while calibration against time series of inflow could, in the absence of lake volume observation, improve 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, we found that calibration aimed at optimizing the five criteria CC, NSE, RMSE, RAE and KGE with respect to monthly time series of observed total, groundwater and lake storages, with almost similar achieved performance values (Table 4), does not necessarily lead to similar estimates of total and compartmental water losses over the whole time period 2003 to 2013. For example, variants RS and RS_LV have the same values for all five performance criteria (expect KGE with 0.1 difference) with respect to TWS (not shown) but TWS loss between 2003 and 2013 is simulated to be $11.15 \cdot 10^9$ m^3 and $7.86 \cdot 10^9$ m^3, respectively (Table 5). TWS loss according to variant RS_Q_GW_NA is, with $10.04 \cdot 10^9$ m^3, in between and quite different, even though NSE and KGE are only 0.04 and 0.06 better, respectively. We conclude that in the case of relevant trends, the calibration criteria should include 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

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

In order to define the lake restoration program, it is vital to know which factors contribute how much to 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 et al. (2018) concluded that human-induced changes accounted for 86% of the lake volume decline during 1995-2010, while we determined the value of 40% for 2002-2013. According to our study, human water use was the reason for 41% inflow reduction into the lake during 2003-2013 which is similar to the values of Shadkam et al. (2016) for the years 2003-2009 (comp. their Figs. 8). Discrepancies are likely due to different analysis methods but different analysis periods, as well as different 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 statistical analysis assumes that river inflow can be considered to reflect “anthropogenic impacts” while precipitation and evaporation changes reflect climatic variations while river inflow is in reality also affected by climate 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 explain the dramatic decline of the lake level. They did not use in-situ data (except lake water level data) for their analysis. Based on a analysis of Standardized Precipitation Index (SPI), a drought index, AghaKouchak et al. (2015) reported there was no significant trend in droughts over the basin during past three decades and concluded from this that human activities not climatic variations are the main reason 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 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 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.

In addition, 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. 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, annual inflow into the lake was estimated to be $3{,}700 \cdot 10^6$ m^3 in 2003 (their Fig. 8) while observed inflow was much higher, $5{,}835 \cdot 10^6$ m^3. In 2009, observed inflow, with $1{,}036 \cdot 10^6$ m^3, was only half of the simulated one. Therefore, the very high human contribution to 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 the basin as a whole and 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 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 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 uncertainties of the observation data. GRACE TWSA data are more reliable for larger (100,000 km^2 (Landerer and Swenson, 2012)) areas than the basin area. 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 "observed" lake water volume decline, are be affected by a likely underestimation of the actual decline by the "observed" value derived 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.

5 Conclusions

This study investigated the differential impact of human water use and climate variations on water storage (total, groundwater, lake) in the Lake Urmia basin~~, including total water, groundwater and lake water storage,~~ as well as on ~~the~~ inflow into the lake ~~between~~during 2003~~ and ~~-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 takes into account the impact of human water use and man-made reservoirs on flows and storages. Using the best-performing model variant, the impact of human water use was determined by comparing the output of a naturalized run, where human water use was assumed to be zero, with the run with the historic water use. To understand the value of different observational data types for calibration, four calibration 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), and 4) statistical data on water withdrawals in the basin were added. A time series of observed lake volume was used for ~~validation~~evaluation.

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 model parameters could result in a good fit to observed TWSA. Consideration of these remote sensing data somewhat improved the dynamics of both inflow into Lake Urmia and lake water storage, inflow into the lake was still strongly overestimated by a factor of 0.92%, and groundwater dynamics should a 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 ~~the consideration of~~by using monthly time series of mean groundwater level variations ~~leads to the adjustment~~in the basins for calibration, we could adjust the fractions of human water use taken from groundwater and surface water ~~and thus to simulating the correct~~such that seasonality of groundwater storage~~.~~ was simulated correctly. Only ~~by calibrating against these observations was is~~then it was possible to simulate the observed groundwater loss, and~~ loss of~~ 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. ~~Based on this study, we~~We recommend ~~that~~to include, in case of relevant trends in observations, ~~one of the calibration criteria assess~~ the difference between observed and simulated trends~~, and~~ as one of the calibration criteria, not only differences between ~~e.g.~~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 ~~multiple~~specific observation data types for model calibration, including trade-offs among fit to multiple data types, can be derived from this study. ~~Conclusions are expected to be basin-specific, and~~ 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 that human water use reduced lake inflow that would have occurred without human water use during 2003-2013 by about 41%. About 52% of the total water storage loss in Lake Urmia basin and only 40% of lake water loss during this time period was due to human water use, and the 48% and 60%, respectively, to climate variations. ~~The human impact on~~ 90% of groundwater storage ~~losses of 90%~~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 ~~the wet years 2015/2016 with and~~ 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, ~~climate change must be constrained to prevent strong decreases of precipitation and runoff.~~ global-scale mitigation of climate change by reducing greenhouse gas emissions to prevent strong decreases of precipitation and runoff. Our study has shown that 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

~~The authors have no permission to share in~~In-situ data from "Iran Water Resources Management Company" including groundwater levels, precipitation and temperature publicly ~~but this data is~~are available upon request ~~to~~from the ~~correspondence~~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~~) and lake~~). Lake water surface~~ water~~ extents and water levels are available at http://hydrosat.gis.uni-stuttgart.de/php/index.php (last accessed: 17 Jul. ~~2018).~~2018). All simulation results are available in the supplement.

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 ~~organizations~~Center for Space Research (CSR) and Potsdam Institute for Climate Impact Research that provide GRACE data and global climate data~~. Last but not least the~~, 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 reviewers for their comments and suggestions that helped to improve the paper.

Appendix A: Simulated groundwater storage in individual grid cells

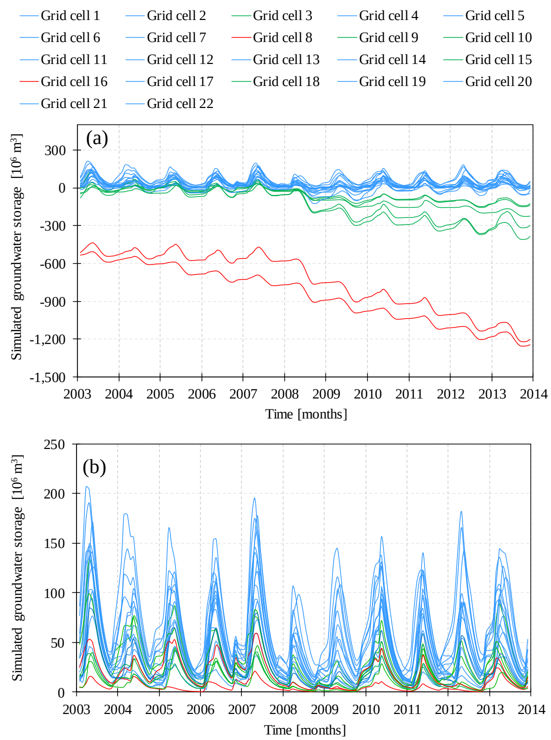


Figure A1: Simulated groundwater storage in each of the 22 0.5° grid cells in Lake Urmia basin under anthropogenically altered (Fig. A1a) and naturalized conditions (Fig. A1b).

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 and environmental management (case study: Lake Urmia Basin, Iran), Environmental Earth Sciences, 73, 13-26, doi: 10.1007/s12665-014-3391-6, 2015.

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.

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, ~~2014~~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: 10.1016/j.jhydrol.2009.08.003, 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, ~~2011~~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.

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.

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.

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.

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.

ULRP: Evaluation and deduction of Urmia Lake Restoration Program need at least 3 years, 2015b.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.irhttp://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.

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.