Introduction

First of all, I want to thank both referees for their constructive comments. In the following document I have answered all comments and listed the changes in our manuscript based on these comments. The new manuscript is added together with the final response. This final response is structured as follows:

- Reaction on comments from referee 1 (C5050)
- Reaction on comments from referee 2 (C5257)
- References

Reaction on comments from referee 1 (C5050):

Comment 1

The overall strategy for this study is not clearly explained. Some people use GRACE data to inform hydrological models in a calibration and/or data assimilation approach (e.g. (Milzow et al., 2011)*. Others compare GRACE data to hydrological model output to check consistency and identify weaknesses in both the GRACE data and the models. What exactly is the purpose here? This should be explained upfront. If the purpose is to inform the model, then it is essential to document model improvement in an independent validation period.*

Reaction

The hypothesis and strategy were indeed not given explicitly in our manuscript and improved the overall strategy and motivation in the new version. The main goal of this study is to show that the water mass changes in a hydrological model and surface water in the region can explain a large part of the mass variations, as observed by GRACE. This is in contrast with other papers like (Voss et al., 2013), which claim that a large part of the water mass changes are caused by anthropogenic groundwater depletion. Therefore, we want to show that it is also possible to explain the water mass variation largely by depletion of lakes and reservoirs and by natural groundwater variations.

The main differences between our study and other studies like (Longuevergne et al., 2013; Voss et al., 2013) are:

- The inclusion of the local geology in our current model, which gives a better insight in the groundwater flows and storages.
- The use of streamflow measurements for at least part of the catchment, which can be used to calibrate our model.
- Exclusion of smaller lakes by earlier researches, which still can have an influence on the total water balance due to large water level variations.

The reason that we included GRACE measurements in the model calibration was because we wanted to show that a hydrologic model can mimic water mass variations from GRACE. However, for an independent comparison between the model and GRACE data, this would indeed need a validation period, which is not available in our case because a similar drought as the one between 2007 and 2009 did not occur during the lifespan of the GRACE satellite mission.

Therefore, in this version of the manuscript we have chosen to use only streamflow measurements as a calibration for our hydrologic model in first instance, but at the same time we also compare the results with the case where GRACE data is included in the calibration process. This results in almost identical results of both calibration methods for the snow reservoir, unsaturated reservoir and fast runoff reservoir and a difference of about 8 mm EWH for the groundwater reservoir. The total decline from the calibration with only streamflow data was 39 mm EWH and the total decline including GRACE as a calibration parameter is 48 mm EWH. This shows that model results are quite consistent for both methods and lead to the same conclusions, although there is a small difference in water mass decline.

Changes made to manuscript

- Hypothesis/strategy and motivation for research edited in (1 Introduction line 80-90)
- Model approach updated (1 Introduction line 108-119)
- Calibration method explanation added (3.4 Model Calibration line 382-398)
- Method for model selection updated with new calibration method (3.4 Model Calibration line 410-440)
- Comparison two different calibration methods added (4.4 Natural groundwater variations line 524-541)
- Comparison two different calibration methods added (4.6 GRACE and modelled mass variations)
- Model results and figures updated based on new calibration method

Comment 2

GRACE processing: There are basically two ways of processing GRACE data in the literature: Either the level 1 range rate data is inverted for spherical harmonic coefficients or for mass changes on a grid. The second approach is usually termed the MASCON approach (Rowlands, 2005). *The method you present here seems to be a hybrid between the two, in the sense that mascon parameters are fitted to level 2 spherical harmonic coefficients instead of range rate data. I know that GRACE processing is not the focus of this paper, but it would still be nice to explain the pros and cons of doing it this way. For instance, I do not understand if you can retrieve the 6 MASCONS shown in Figure 2 independently of the rest of the planet or if you always have to invert for a global set of mascons? Also, why do the MASCONS have circular shape, why not adapt them to the geometry of the basins of interest as for instance in*(Krogh et al., 2010) *?*

Reaction

To explain the GRACE processing more clearly, the following text is added to the manuscript on line 163-170:

"This method is based of GRACE level-2 data from the Center of Space Research (CSR) and includes modifications for the gravitational flattening term C20 from satellite laser ranging. Furthermore the method also considers degree 1 terms associated with geo-center motions as a result of geophysical loading phenomena. The used GRACE method is not a spatial averaging kernel method, instead to obtain equivalent water levels over a region one has to add up the signal from the individual mascons. The signal at these mascons is obtained via a global inversion method."

This means that we use a global set of mascons and that the signal at these mascons was obtained by means of a global inversion method as described in (Schrama et al, 2014). Our original manuscript lacked a description at this point

A comparison between different GRACE methods for mass calculation can be found in (Shepherd et al., 2012). As long as the starting conditions of these methods are the same the differences turn out to be small. The method in (Schrama et al 2014) is an extention of one of the GRACE methods contained in (Shepherd et al 2012), and it shows nearly identical results.

The advantage of the approach in (Schrama et al 2014) is that resulting EWH values are part of a standard GRACE product and will be consistent with other regional solution worldwide. The disadvantage is that we have to adapt our study area to the given mascons. We are thus not able to use approaches like (Krogh et al., 2010), which adapt the mascons to the geometry of the studied basin. However, the changes in EWH due to adaption to the exact study area are small. We tested this using a comparison between the GRACE results for the current dataset and two extended datasets, see Figure 1 in this rebuttal. This shows that the GRACE signal remains largely the same, with only small differences in yearly fluctuation and water depletion over the whole period. If these relatively large changes in study area induce only small changes in the total GRACE signal it is likely that the differences due to adaptation to geometry are negligible.

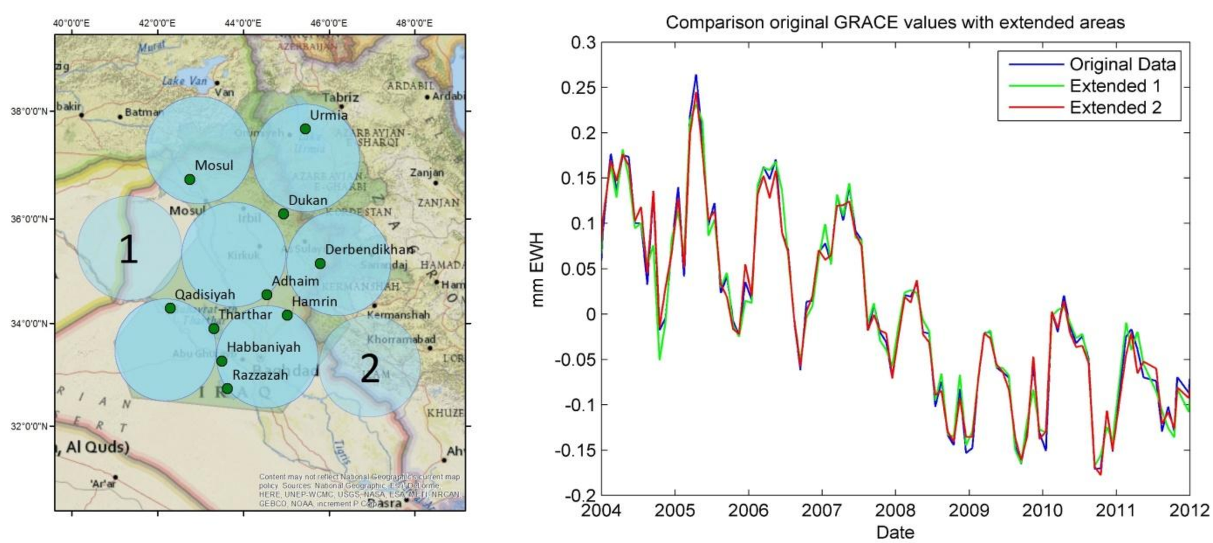

Figure 1 - Comparison of EWH values for current and extended areas.

In the old version a method was given to refine our results (page 11538 line 15-22), but it proved to have negligible influence and was therefore omitted in the new version.

Changes to manuscript

- Description of GRACE processing updated (3.1 GRACE mass calculations line 160-174)
- Comparison between our and other GRACE processing methods added (3.1 GRACE mass calculations line 175-186)

Comment 3

The purpose and basis of the various corrections described on page 11539 are not really clear to me. Why does the contribution of the lakes have to be scaled by ½ and 1/3? How did these numbers come about? Why is the GRACE region extended into the southwestern desert, if this is really not part of the study area of interest? All this needs to be motivated and explained much better.

Reaction

The two corrections made to the GRACE results are those based on the extension of our study area to the south west and the correction for lakes at the border of the study area, which are both related to water mass leakage

The water mass depletion of Lake Tharthar is more than 50 % of the total water mass depletion from lakes in our area (figure Figure 2). Therefore we extended the study area to the south-west to be sure that the water mass depletion from this lake was fully captured in our GRACE mass calculations.

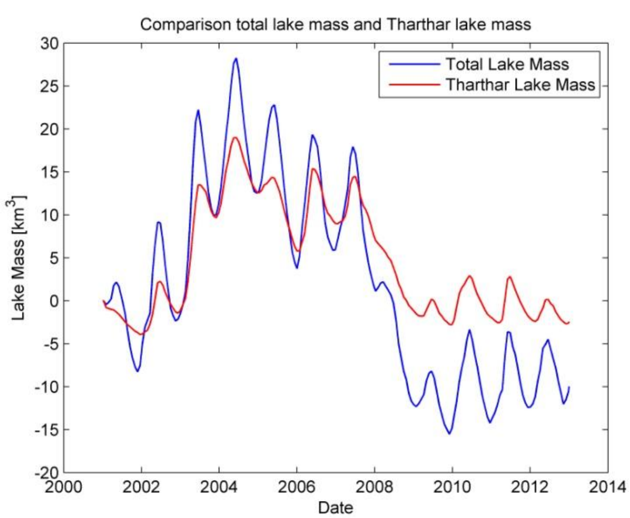

Figure 2 - Mass variation from Lake Tharthar compared with total mass variation from lakes in study area

Because Lake Urmia and lake Razzazah have much less impact on the total lake mass, we have chosen to correct their mass leakage out of our study area based on the method described in (Longuevergne et al., 2013). In this paper it is demonstrated how point masses within a circular mascon contribute to this mascon, based on the size of the mascon and the distance from the mascon center. For this correction we geometry of our whole study area and calculated the correction factors based on the distance of the lakes from the centre of our region. During the revision of this paper we studied the locations of the mascons and lake centres in more detail, which resulted in a slightly larger contribution of lake Urmia of 0.6 and a contribution of Lake Razzazah of 0.65. Corrections for other lakes in the region were also calculated, but the mass changes due to the correction factors of these lakes are negligible.

Changes to manuscript

- More detailed description and motivation of study area extension given (3.1 GRACE mass calculations line 185-240)
- Explanation and motivation of lake mass correction added (3.1 GRACE mass calculations line 198-204 and Derivation of lake mass line 291-297)

Comment 4

It is stated that the aquifers are karstified, and highly transmissive. Water level and storage variations in such aquifers are generally suppressed because of the high transmissivity and water is effectively drained over short time scales. Lines 4-9 on page 11548 seem to suggest the opposite. It would be nice to discuss if the simulated groundwater storage variations are reasonable given the available hydrogeological knowledge and observations from the region.

Reaction

This point indeed deserves an extended elaboration. There is quite some research on this because the water supply in Northern Iraq is dependent on springs that emerge from these aquifers (Ali and Stevanovic, 2010; Ali et al., 2009b). These studies show that the discharge from these karstic aquifers can be split up in two components. The first component is a rapid discharge within a month after rain events, which is related to channels and large fractures in the limestone aquifer. The second component is a much slower discharge, which is generally stable till the new wet season. While the first component is covered by the fast reservoirs in our hydrologic model, the second component is modelled as the slow groundwater reservoir. The recession value of these groundwater reservoirs is based on the recession values of three large springs in the area, the Bestansur, Zalim and Sachinar (Ali and Stevanovic, 2010; Ali et al., 2009a, 2009b), which respective recession coefficients are 0.0049, 0.0032 and 0.0038. The chosen recession bandwidth is chosen between 0.0035 and 0.0045.

Changes in manuscript:

- More elaborate description about groundwater flow added (3.3 Rainfall Runoff model 359-377)

Comment 5

Throughout the manuscript, language and grammar should be checked and clarity of the wording should be improved, see also details listed below. Please always call the same things by the same names. For example, the manuscript sometimes talks about the "Lesser Zab catchment" and sometimes about the "smaller Dukan area", although, I believe, those two names refer to exactly the same thing.

Reaction

Detailed comments are corrected and paper is checked for grammar and inconsistent wording. Also, tributary names and lake names are added to figure 2 and figure 5 of the manuscript. The Dukan area referred to in the text is a part of the Lesser Zab catchment upstream from Lake Dukan, which is now explained in the text and caption of figure 5 of the manuscript.

Changes in manuscript

- Lake names added to figure 2
- Tributary names added to figure 5

A lot of place names are used in the text, but cannot be found on any of the maps. I think the paper would benefit from a detailed base map that shows and names all places, rivers and lakes referred to in the text.

Reaction

tributary names are added in figure 2 and lake names are added in figure 5

Changes in manuscript

- Lake names added to figure 2
- Tributary names added to figure 5

Comment 7

Please explain how the uncertainty bands for the surface water storage and snow storage in fig 8 were derived.

Reaction

The errors due water level measurements are derived based on the estimated errors in lake levels given by (Crétaux et al., 2011; USDA/FAS, 2013). These time series of water level errors are then multiplied with the total area of these lakes and added up:

$$\sigma^t = \sqrt{\frac{\sum_{i=1}^{n}\left(\sigma_i^t A_i^t\right)^2}{n}}$$

Where n is the number of lakes, σ_i^t is the estimated error in water level for one of the lakes at time t and A_i^t and the calculated lake area of a particular lake at time t.

This uncertainty does not include the errors in lake area calculation, but these are difficult to quantify and are likely small due to the use of stage-area curves. For example, the difference in lake size from Lake Mosul and the derived values from (Issa et al., 2013) are about 10 km^2 but it only causes significant differences in the stage-volume curve for high water levels. Additionally, the relative error for larger lakes is much smaller because the ratio between lake shore length and lake area becomes smaller with increasing lake size. In figure Figure 3 this is illustrated using the resulting stage-area curve from lake Tharthar and Lake Qadasiyah, which have an large contribution to the total water mass depletion. Through the use of a linear/cubic fit of 192 measurements, most of the errors due to lake area and lake level calculations are filtered out.

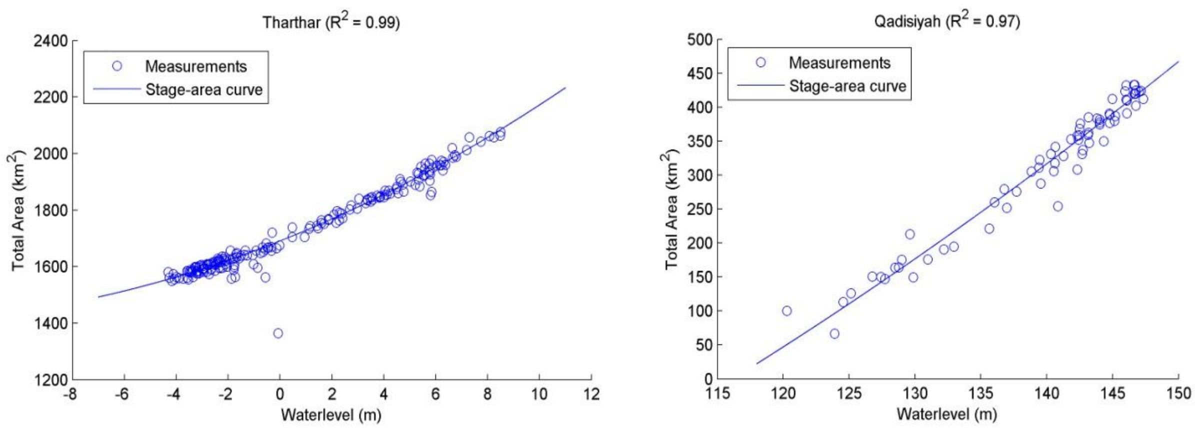

Figure 3 - Stage-area curve Lake Tharthar and Lake Qadisiyah

We replaced the lake mass graph with a new one based on a 95% confidence interval instead of 68%, to ensure consistency with the GRACE graphs. This error does not include the leakage of water mass of lakes outside the study area due to GRACE mass calculations, because these values cannot be clearly quantified.

The uncertainty of the snow storage is based uncertainties in the precipitation from the GLDAS model. These uncertainties were derived from the comparison between monthly rainfall from the GLDAS model and four rainfall stations in the region. In the new version, snow mass is added to the rainfall-runoff model, although differences are small compared to the original method.

Changes in manuscript

- Derivation of lake mass uncertainty added (3.2 Derivation of lake mass line 274-281)

- Description of possible additional uncertainties in lake mass derivation added (3.2 Derivation of lake mass line 282-297)

Comment 8

Why use snow from GLDAS? Why not run a simple snow accumulation and melt routine on top of the hydrological model? At least that would ensure consistent precipitation input. How exactly is the TRMM product corrected for snow from GLDAS?

Reaction

Snow mass from the GLDAS model was used because snow mass in the study area does not have much effect on the total mass decline and it is not likely that a snow accumulation and melt routine would have much effect on the total modelled snow mass. Therefore, the simplest approach was used. Together with the addition of snow mass, the snowfall from GLDAS was subtracted from TRMM precipitation and the snowmelt was added to TRMM precipitation.

Addition of a snow accumulation and melt routine would add to the consistency of the model and was therefore added in the new version. Figure Figure 4 shows a comparison between the snow mass derived from GLDAS with our rainfall-runoff model, which confirms that the results are similar. The added variables are a threshold temperature and a melting coefficient (mm/C/day), which were kept the same for all three geologic zones. Temperature values were derived from GLDAS on a daily basis (Rodell et al., 2004). Because temperature differences are large due to height differences in the basin, it was not possible to model this in a lumped approach. Therefore a semi-distributed model setup is used, with a distributed approach for the snow and unsaturated reservoir.

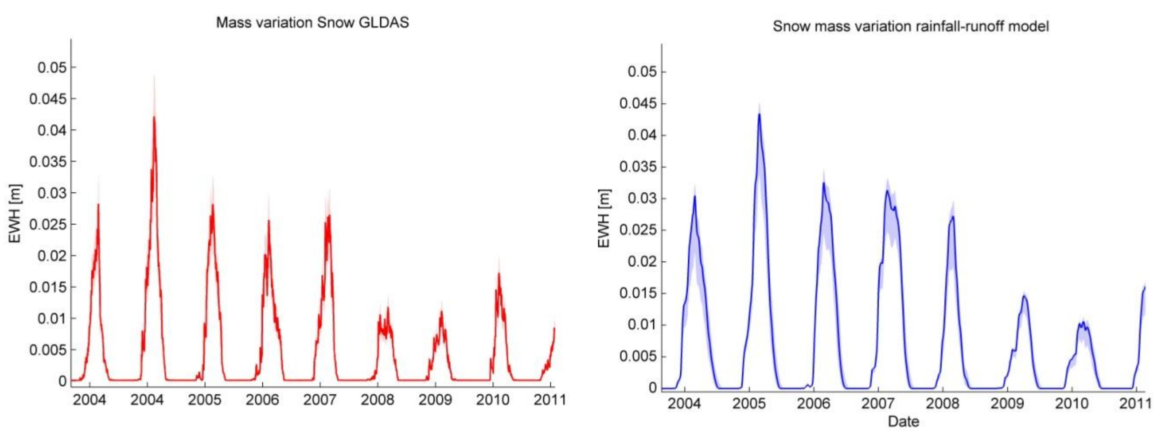

Figure 4 - (left) Snow mass from GLDAS. (right) Snow mass based on a snow accumulation and melting routine

Changes in manuscript:

- Description about snow mass derivation from GLDAS removed

- Description of snow mass reservoir and distributed model approach added (3.3 Rainfall-runoff model line 339-348)

Comment 9

Very little is said about how the uncertainties of the simulated storages have been determined (fig 9). To me, these uncertainty bounds look very narrow. Can they be justified? For instance, were the errors due to parameter transfer from gauged to ungauged parts of the catchment taken into account? Can these uncertainty estimates be hold up against real observations? It would be nice to see the comparison of simulated and observed hydrographs at least for the single available station…

Reaction

The uncertainty band are formed by the ensemble of all models on the pareto front (like in Werth et al., 2009), which is based on the Nash-Sutcliffe and log Nash-Sutcliffe performance. We used the difference between the maximum and minimum values of the ensemble as the uncertainty band. A discharge graph for 2005-2008 is given in figure Figure 5 and shows an ensemble of all model runs from the pareto front. This graph shows that the differences between modelled and measured values can still be substantial, but the baseflow values during the dry season are mainly within the minimum and maximum discharge bands. The pareto ensemble is a measure of uncertainty of the parameter values of the model, but does not represent other uncertainties based on the forcing, up scaling or model structure.

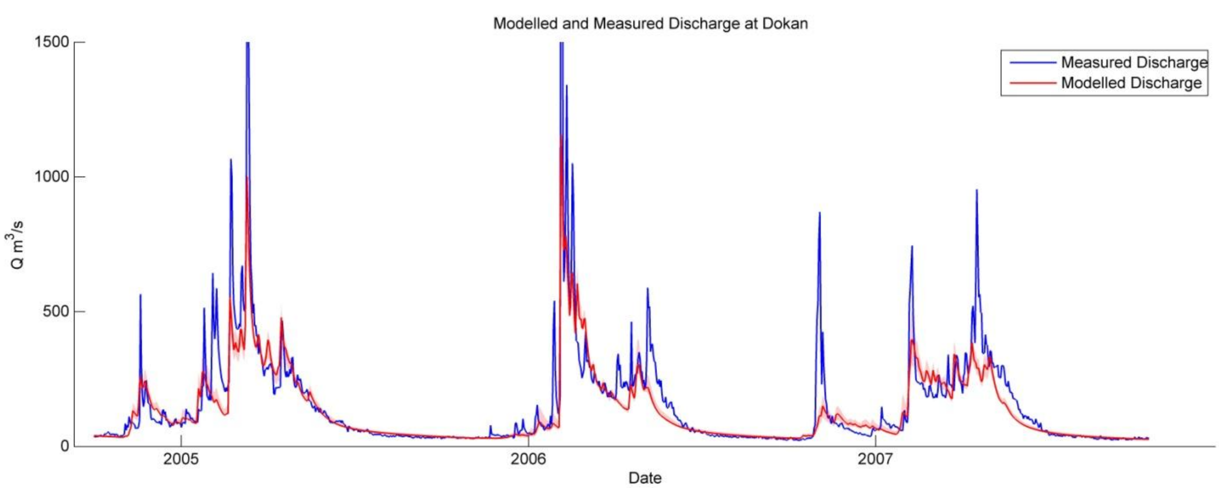

Figure 5 - Modelled and measured discharge at inflow of Lake Dukan

For instance, the effect of errors in rainfall and temperature measurements is not included in this procedure. Also, the spreading of rainfall varies over each TRMM pixel, which could cause lower peak flows in our model. Another influencing factor is the water use upstream from our gauging station, which likely causes a reduction in streamflow, but was not considered in our model. This can result in an underestimation of the groundwater variation, because baseflow values are likely higher than those used in our model calibration.

Because the mixture of uncertainties between GRACE, lake mass and the model output is questionable the residual graph is removed.

Changes in manuscript:

- Section on model discharge is added (4.3 Modelled discharge Dukan catchment line 500-522)
- Uncertainty bands of rainfall runoff model explained (3.4 Model Calibration line 435-438)
- Added figure 8 about modelled and measured discharge.
- Removed figure about residual water mass
- Discussion about upscaling of catchment added (4.4 Natural groundwater variations line 563-571)

Details

1. Page 11534, line 11: "Corrected for" should probably be "estimated". Total mass variation should include lakes. **> Sentence changed**
2. Page 11534, line 19: "Depletion of geology"? Please re-word. I guess you refer to natural depletion of groundwater. **> Sentence changed**
3. Page 11535, line 6: Is this predicted decrease due to anthropogenic climate change? **> changed to climate models**
4. Page 11535, line 20: "riparian" should be "upstream". **> Changed**
5. Page 11537 line 9: Please give the version of the TRMM 3B42 product used here. Versions 6 and 7 are quite different in other parts of the world, here too? **> V7 added**
6. Page 11537 line 22: Please reword. The lakes influence the GRACE signals**. > Changed**
7. Page 11538 line 9: "dises" should be circles?? **> This part is removed**.
8. Page 11542, line 5: "bias-corrected" may be better than "calibrated" **> Changed**
9. It is not unambiguously stated in the text how GRACE and the model are compared (e.g. page 11546, line 23). It is stated that GRACE and the model are compared, while in fact the comparison is between GRACE minus surface water storage and the model. **> Changed**
10. Figure 1: Both maps should have coordinate systems/scale bars and colorbar legends. **> Added**
11. Figure 2: Needs coordiates/scales **> Added**

Reaction on comments from referee 2 (C5257):

Comment 1

The method should generally be explained in more detail, especially the assumptions/constraints, etc. The way of GRACE processing using the mascon approach is only vaguely given. Moreover, GRACE is used for calibration and data reduction. Does this not imply internal correlations?

Reaction

The assumptions/constrains for several model steps were indeed only shortly discussed. In the new version of our manuscript modelling steps are explained in more detail and assumptions are given explicitly.

To explain the GRACE processing more clearly, the following text is added to the manuscript on line 163-170:

> *"This method is based of GRACE level-2 data from the Center of Space Research (CSR) and includes modifications for the gravitational flattening term C20 from satellite laser ranging. Furthermore the method also considers degree 1 terms associated with geo-center motions as a result of geophysical loading phenomena. The used GRACE method is not a spatial averaging kernel method, instead to obtain equivalent water levels over a region one has to add up the signal from the individual mascons. The signal at these mascons is obtained via a global inversion method."*

The reason that we included GRACE measurements in the model calibration was because we wanted to show that a hydrologic model can mimic water mass variations from GRACE. However, for an independent comparison between the model and GRACE data, this would indeed need a validation period, which is not available in our case because a similar drought as the one between 2007 and 2009 did not occur during the lifespan of the GRACE satellite mission.

Therefore, in this version of the manuscript we have chosen to use only streamflow measurements as a calibration for our hydrologic model in first instance, but at the same time we also compare the results with the case where GRACE data is included in the calibration process. This results in almost identical results of both calibration methods for the snow reservoir, unsaturated reservoir and fast runoff reservoir and a difference of about 8 mm EWH for the groundwater reservoir. The total decline from the calibration with only streamflow data was 39 mm EWH and the total decline including GRACE as a calibration parameter is 48 mm EWH. This shows that model results are quite consistent for both methods and lead to the same conclusions, although there is a small difference in water mass decline.

Changes in manuscript

- Description of GRACE processing updated (3.1 GRACE mass calculations line 163-170)

- Description of GRACE processing updated (3.1 GRACE mass calculations line 160-174)
- Hypothesis/strategy and motivation for research edited in (1 Introduction line 80-90)
- Model approach updated (1 Introduction line 108-119)
- Calibration method explanation added (3.4 Model Calibration line 382-398)
- Method for model selection updated with new calibration method (3.4 Model Calibration line 410-440)
- Comparison two different calibration methods added (4.4 Natural groundwater variations line 524-541)
- Model results and figures updated based on new calibration method

Comment 2

The limitations and error contributions of the various model reductions and assumptions should be given explicitly. Only then the usefulness and quality of the proposed "rainfall-runoff" model can be evaluated.
Did the authors compare their results with independent GRACE-based estimates of mass changes in that region (see, e.g., Sneeuw et al. 2014 for lake Urmia).

Reaction

Uncertainty of the rainfall-runoff model is given explicitly in the new version of the manuscript. Also, limitations and possible errors in the model are added.

The uncertainty band of the rainfall-runoff model are formed by the ensemble of all models on the pareto front (like in Werth et al., 2009), which is based on the Nash-Sutcliffe and log Nash-Sutcliffe performance indexes. We used the difference between the maximum and minimum values of the ensemble as the uncertainty band. A discharge graph for 2005-2008 is given in figure Figure 6 and shows an ensemble of all model runs from the pareto front. This graph shows that the differences between modelled and measured values can still be substantial, but the baseflow values during the dry season are mainly within the minimum and maximum discharge bands. The pareto ensemble is a measure of uncertainty of the parameter values of the model, but does not represent other uncertainties based on the forcing, up scaling or model structure.

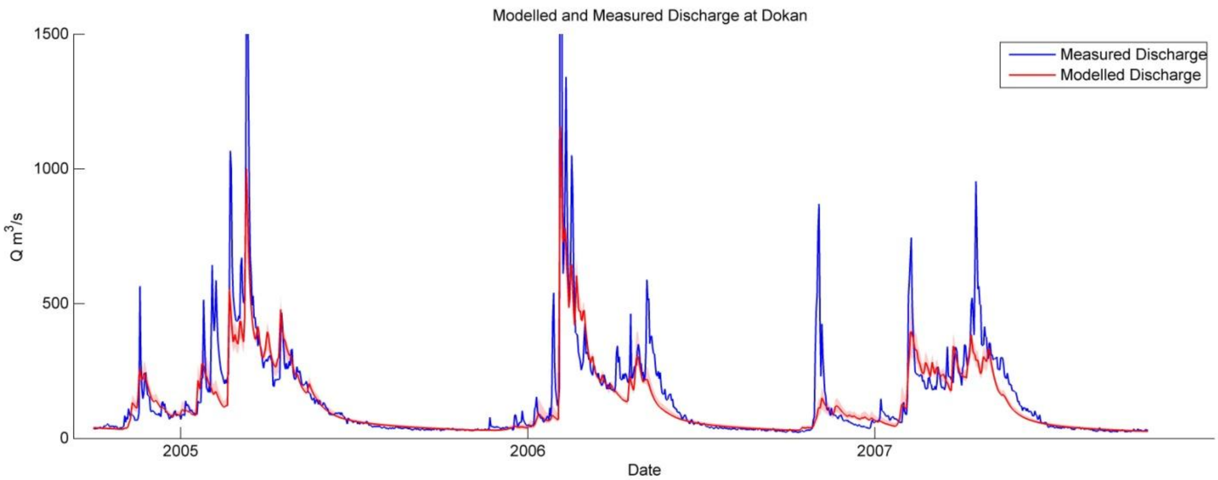

Figure 6 - Modelled and measured discharge at inflow of Lake Dukan

For instance, the effect of errors in rainfall and temperature measurements is not included in this procedure. Also, the spreading of rainfall varies over each TRMM pixel, which could cause lower peak flows in our model. Another influencing factor is the water use upstream from our gauging station, which likely causes a reduction in streamflow, but was not considered in our model. This can result in an underestimation of the groundwater variation, because baseflow values are likely higher than those used in our model calibration. Because the mixture of uncertainties between GRACE, lake mass and the model output is questionable the residual graph is removed.

We did compare our GRACE result with different other GRACE-based estimates from (Crétaux et al., 2011; Longuevergne et al., 2013; Voss et al., 2013). These studies obtain similar results, but the extent of our study area is not exactly the same. Also, comparison between our method and other common methods by (Shepherd et al., 2012) showed that differences are generally small. Comparison with results from (Tourian et al., 2015) ,which represents a part of our study area were not made. Figure Figure 2 gives a comparison between the resulting lake mass variation from our study and from (Tourian et al., 2015). This shows that although yearly variations between those models differ (especially 2006-2008), the trend in mass decline is almost the same. Differences between the results from different study are possibly related with the used satellite data, which was mainly from Envisat in (Tourian et al., 2015) and Jason 1&2 in our study.

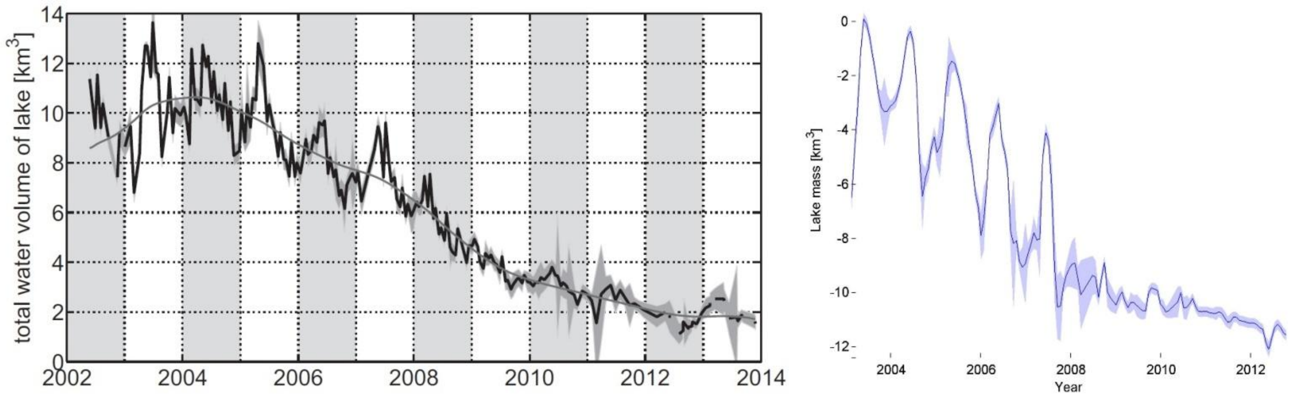

Figure 7 - Modelled water mass from Sneeuw and this study

Changes in manuscript

- Section on modelled discharge is added. In this section the derivation of the uncertainty band of the rainfall runoff model is given. Also other possible sources of uncertainty are discussed. (4.3 Modelled discharge Dukan catchment line 500 - 522)
- Uncertainty bands of rainfall runoff model explained (3.4 Model Calibration line 435-438)
- Discussion about upscaling of catchment added (4.4 Natural groundwater variations line 563-571)
- Added figure 8 about modelled and measured discharge.
- Removed figure about residual water mass

Comment 3

Throughout the paper starting with the abstract, the authors use different (partly incomplete) units for representing mass variations. For example, mass loss is represented sometimes in mm and sometimes in km3, for the first probably mm in EWH is meant, where the second is a volume change. It should be used consistently.

Reaction

Mass variations are always given as mm EWH in the new manuscript. Masses in km^3 are also given in some cases, but always together with a value in mm EWH.

Comment 4

On page 11539 (lines 8-9), the reason for selecting the weight 1/2 for the lake mass of Urmia and 1/3 for lake Razazzah is not obvious. Do these coefficients come from some empirical model?

Reaction

The correction for the lake masses of Lake Urmia and Lake Razzazah are based on work from (Longuevergne et al., 2013). In this paper is demonstrated how point masses within a circular mascon contribute to this mascon, based on the size of the mascon and the distance from the mascon center. For this correction we geometry of our whole study area and calculated the correction factors based on the distance of the lakes from the centre of our region. During the revision of this paper we studied the locations of the mascons and lake centres in more detail, which resulted in a slightly larger contribution of lake Urmia of 0.6 and a contribution of Lake Razzazah of 0.65. Corrections for other lakes in the region were also calculated, but the mass changes due to the correction factors of these lakes are negligible.

Changes in manuscript

- Explanation and motivation of lake mass correction added (3.1 GRACE mass calculations line 195-205 and 3.2 Derivation of lake mass line 291-297)

Comment 5 [i]

Are the estimated lake mass variations reliable and accurate?

Reaction

The errors due water level measurements are derived based on the estimated errors in lake levels given by (Crétaux et al., 2011; USDA/FAS, 2013). These time series of water level errors are then multiplied with the total area of these lakes and added up:

$$\sigma^t = \sqrt{\frac{\sum_{i=1}^{n}\left(\sigma_i^t A_i^t\right)^2}{n}}$$

Where n is the number of lakes, σ_i^t is the estimated error in water level for one of the lakes at time t and A_i^t and the calculated lake area of a particular lake at time t.

The estimated errors are based on the footprint from the different satellites and comparison between in-situ measurements and satellite altimetry . Because the error estimates came with the data from (Crétaux et al., 2011; USDA/FAS, 2013), it is quite difficult for us to give an idea of the accuracy of the estimated errors. For example, the graphs in figure Figure 2 show that the water levels difference of two separate satellite products can be much larger than the estimated errors in those products.

The presented uncertainty does not include the errors in lake area calculation, because these are difficult to quantify and are assumed to be relatively small due to the use of stage-area curves. For example, the difference in lake size from Lake Mosul and the derived values from (Issa et al., 2013) are about 10 km^2 but it only causes significant differences in the stage-volume curve for high water levels. Additionally, the relative error for larger lakes is much smaller because the ratio between lake shore length and lake area becomes smaller with increasing lake size. In figure Figure 3 this is illustrated using the resulting stage-area curve from lake Tharthar and Lake Qadasiyah, which have an large contribution to the total water mass depletion. Through the use of a linear/cubic fit of 192 measurements, most of the errors due to lake area and lake level calculations are filtered out.

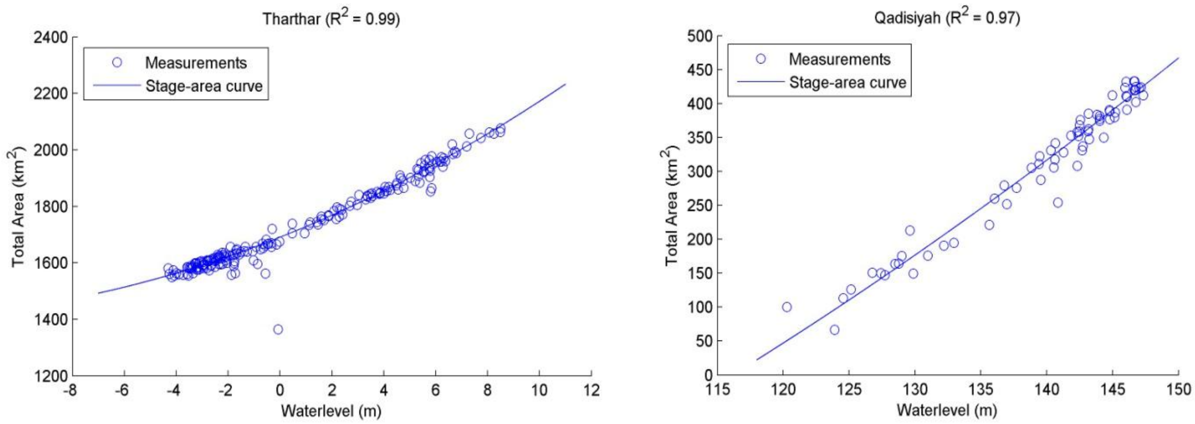

Figure 8 - Stage-area curve Lake Tharthar and Lake Qadisiyah

We replaced the lake mass graph with a new one based on a 95% confidence interval instead of 68%, to ensure consistency with the GRACE graphs. This error does not include the leakage of water mass of lakes outside the study area due to GRACE mass calculations, because these values cannot be clearly quantified.

Changes in manuscript

- Derivation of lake mass uncertainty added (3.2 Derivation of lake mass line 274-281)
- Description of possible additional uncertainties in lake mass derivation added (3.2 Derivation of lake mass line 282-297)

Comment 6

For the snowfall and snowmelt calculations, the authors used the GLDAS model. How reliable is that model for such calculations?

Reaction

Snow mass from the GLDAS model was used because snow mass in the study area does not have much effect on the total mass decline and it is not likely that a snow accumulation and melt routine would have much effect on the total modelled snow mass. Therefore, the simplest approach was used.

Addition of a snow accumulation and melt routine would add to the consistency of the model and was therefore added in the new version. Figure Figure 4 shows a comparison between the snow mass derived from GLDAS with our rainfall-runoff model, which confirms that the results are similar. The added variables are a threshold temperature and a melting coefficient (mm/C/day), which were kept the same for all three geologic zones. Temperature values were derived from GLDAS on a daily basis (Rodell et al., 2004).

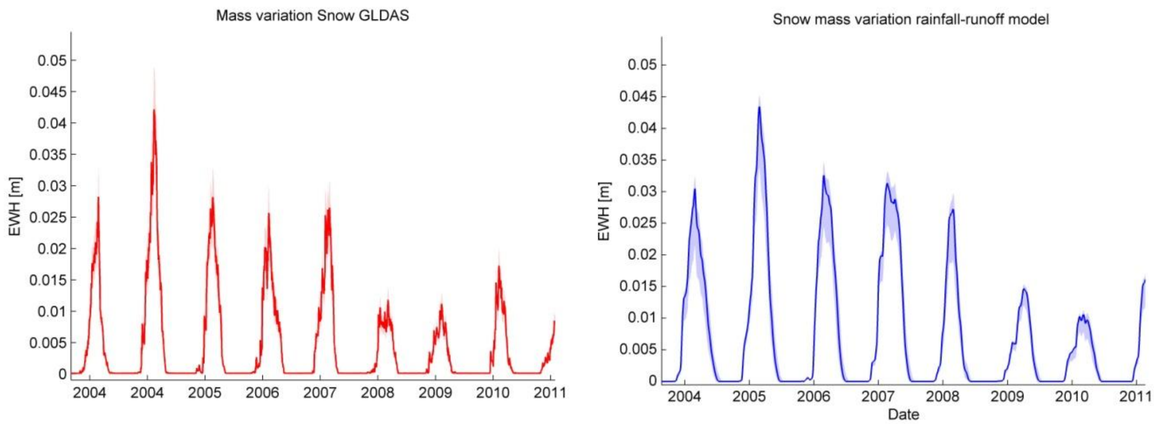

Figure 9 - (left) Snow mass from GLDAS. (right) Snow mass based on a snow accumulation and melting routine

Changes in manuscript

- Description about snow mass derivation from GLDAS removed
- Description of snow mass reservoir and distributed model approach added (3.3 Rainfall-runoff model line 339-348)

Comment 7

For groundwater level estimation, how many stations are used and how reliable are the data?

Reaction

Groundwater levels are based on the groundwater reservoirs in our rainfall-runoff model, which was stated on line 20-25 of page 11545 and in the caption of figure 9.

Uncertainty of these groundwater levels are therefore based on the bandwith of the pareto ensemble.

Details

1. All abbreviations (e.g. GRACE) should be explained at the first time of appearance. For example, GRACE is explained more often, see page 11535, line 28. But other abbreviations were never defined, e.g., WGHM, GGP and SD, etc. **> Checked and changed**
2. All data used, incl. background models should be summarized in a table. **> Added**
3. On page 11537 (line 12), there is one more “and” that should be removed. **> Changed**
4. The word “River” is sometimes written in capital and sometimes in small letters. **> Checked and changed**
5. In the section 3.1, the title “GRACE mass calculations” should be changed to “GRACE mass variation calculation” or to something similar. In addition, it should be said which GSM model from which analysis center has been used in the GRACE calculations. **> Changed and added**

6. On page 11543 in the formula section, punctuation should be used at the end of the formulas. **> This comment is not clear for me. Should I add a dot to end the formula?**

Figures

1. In Fig. 1, the legend for the colours should be included to specify the range of rainfall and topography variations. **> Changed**
2. For Fig. 6, it should be explained how to read and how to understand what is shown there. What can be learnt from such a representation?

   The pareto front is a measure for the uncertainty in model parameters, based on different performance indicators. Description of the pareto front/ensemble is added in section: 4.3 Modelled discharge Dukan catchment.
3. The x-axis of Fig. 8 is not labelled and has to be corrected for the starting year (year 2004 is used two times). **> Changed**
4. The words "Left" and "Right" in the caption of Fig. 9 are differently used as in Fig. 5. **> Changed**
5. The residuals in terms of EWH that are represented in Fig. 10 are rather big. Any explanation for this? Did the authors consider soil moisture at all levels down to the depth of 2 m?

   Differences in soil moisture could indeed be the reason of these residuals as is stated in section GRACE and modelled values. Other causes could be errors in rainfall data, model structure, anthropogenic activities and errors the mass leakage from lakes in the region.

References:

Ali, S. S. and Stevanovic, Z.: Time Series Analysis of Saraw Springs - SE of Sulaimaniya, Iraqi Kurdistan Region, in Advances in Research in Karst Media, edited by B. Andreo, F. Carrasco, J. J. Durán, and J. W. LaMoreaux, pp. 89–94, Springer, Leipzig., 2010.

Ali, S. S., Stevanovic, Z. and Al-Jabbari, M. H.: Auto and Cross Correlation Analysis of Bestansur spring- Sharazoor Plain Iraqi Kurdistan Region, in Proceeding of 3rd scientific conference of the College of Science, University of Baghdad, pp. 1941–1948., 2009a.

Ali, S. S., Stevanovic, Z. and Jemcov, I.: The Mechanism and Influence on Karstic Spring Flow - Sachinar Spring Case Example, Sulaimaniya, NE Iraq, Iraqi Bull. Geol. Min., 5(2), 87–100, 2009b.

Crétaux, J.-F., Jelinski, W., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero, M.-C., Nino, F., Abarca Del Rio, R., Cazenave, A. and Maisongrande, P.: SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data, Adv. Sp. Res., 47(9), 1497–1507, doi:10.1016/j.asr.2011.01.004, 2011.

Issa, I. E., Al-Ansari, N. and Knutsson, S.: Sedimentation and new operational curves for Mosul Dam, Iraq, Hydrol. Sci. J., 58(7), 1456–1466, doi:10.1080/02626667.2013.789138, 2013.

Krogh, P. E., Andersen, O. B., Michailovsky, C. I. B., Bauer-Gottwein, P., Rowlands, D. D., Luthcke, S. B. and Chinn, D. S.: Evaluating terrestrial water storage variations from regionally constrained GRACE mascon data and hydrological models over Southern Africa – preliminary results, Int. J. Remote Sens., 31(14), 3899–3912, doi:10.1080/01431161.2010.483483, 2010.

Longuevergne, L., Wilson, C. R., Scanlon, B. R. and Crétaux, J. F.: GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage, Hydrol. Earth Syst. Sci., 17(12), 4817–4830, doi:10.5194/hess-17-4817-2013, 2013.

Milzow, C., Krogh, P. E. and Bauer-Gottwein, P.: Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment, Hydrol. Earth Syst. Sci., 15(6), 1729–1743, doi:10.5194/hess-15-1729-2011, 2011.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin*, J. K., Walker, J. P., Lohmann, D. and Toll, D.: The Global Land Data Assimilation System, Bull. Am. Meteorol. Soc., 85(3), 381–394, doi:10.1175/BAMS-85-3-381, 2004.

Rowlands, D. D.: Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements, Geophys. Res. Lett., 32(4), L04310, doi:10.1029/2004GL021908, 2005.

Shepherd, A., Ivins, E. R., a, G., Barletta, V. R., Bentley, M. J., Bettadpur, S., Briggs, K. H., Bromwich, D. H., Forsberg, R., Galin, N., Horwath, M., Jacobs, S., Joughin, I., King, M. a, Lenaerts, J. T. M., Li, J., Ligtenberg, S. R. M., Luckman, A., Luthcke, S. B., McMillan, M., Meister, R., Milne, G., Mouginot, J., Muir, A., Nicolas, J. P., Paden, J., Payne, A. J., Pritchard, H., Rignot, E., Rott, H., Sørensen, L. S., Scambos, T. a, Scheuchl, B., Schrama, E. J. O., Smith, B., Sundal, A. V, van Angelen, J. H., van de Berg, W. J., van den Broeke, M. R., Vaughan, D. G., Velicogna, I., Wahr, J., Whitehouse, P. L., Wingham, D. J., Yi, D., Young, D. and Zwally, H. J.: A reconciled estimate of ice-sheet mass balance., Science, 338(6111), 1183–9, doi:10.1126/science.1228102, 2012.

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 Sens. Environ., 156, 349–360, doi:10.1016/j.rse.2014.10.006, 2015.

USDA/FAS: Global Reservoirs and Lake Monitor, [online] Available from: http://www.pecad.fas.usda.gov/cropexplorer/global_reservoir/ (Accessed 25 November 2014), 2013.

Voss, K. a, Famiglietti, J. S., Lo, M., Linage, C., Rodell, M. and Swenson, S. C.: Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region., Water Resour. Res., 49(2), 904–914, doi:10.1002/wrcr.20078, 2013.

Werth, S., Güntner, A., Petrovic, S. and Schmidt, R.: Integration of GRACE mass variations into a global hydrological model, Earth Planet. Sci. Lett., 277(1-2), 166–173, doi:10.1016/j.epsl.2008.10.021, 2009.