RC2 - Anonymous Referee #2 – May 10, 2022

Review of the manuscript egusphere-2022-22

This is a review for the article: "Influence of intensive agriculture and geological heterogeneity on the recharge of an arid aquifer system (Saq-Ram, Arabian Peninsula)" by Seraphin and colleagues. Seraphin and colleagues present a method that combines GRACE satellite products with Global land data assimilation model outputs into a simple regional water balance model for the estimation of regional groundwater recharge rates. Furthermore, the importance of artificial recharge from irrigation return flows is evaluated and compared to the estimated natural groundwater recharge rate, as is the importance of recharge over the comparably limited geographic extents of volcanic deposits (with both artificial recharge and recharge over volcanic deposits being hugely important for the aquifer system). The study is very well researched, presented and written, and can provide guidance to similar estimations for other data scare regions. I have only few minor concerns that I list below. Once these points have been addressed, I recommend moving forward and accepting the article for full publication in HESS.

Minor Comments:

General:

- The abstract is generally well written, clearly describes the goals, the data sources as well as the results. However, no details on the applied method are provided. I suggest reducing mentioning the different datasets and GRACE solutions in such detail, as writing out the different names plus providing the abbreviations consumes way too much space. The freed up space I suggest using for a sentence on the applied methodology, for example, linking to the first sentence of section 2.4: "The data was used to build a regional-scale water mass-balance and estimate recharge from variations in groundwater storage"

Since the different versions of the products are listed in Table 1, we shortened lines 242-247 and 266-269 by removing the version numbers and writing the products names and citations in line. However, even if this may seem fastidious, the mention of the different datasets and their associated versions and citations are important to us since both GRACE and GLDAS are often updated with new measurements, but also with new treatments that can cause differences between results obtained for the same study area by using different versions of the datasets. In fact, this is one of the differences (among others) that may explain discrepancies with the study by Fallatah et al. (2019) mentioned in the introduction.
The suggested sentence has been added at the beginning of section 2.4.

- I appreciate that uncertainties are provided for every number, a sign of a thorough analysis and something that is too often absent missing in similar analyses. Nevertheless, the uncertainties seem very small for satellite based assessments of recharge over such a large domain. Based on the numbers provided (i.e., 4.4 +/- 2.6%), Seraphin et al suggested that recharge rate of the entire Saq-Ram Aquifer lies between ~2-7% of the average annual rainfall. Isn't this range a little narrow considering that the datasources are satellite based data and global land data assimilation model outputs? In the cited study by MacDonald et al 2021, who synthesised recharge rates in arid africa based on more local and therefore

generally more accurate methods for the studied regions, assume recharge rates of 3.3 +/- 5.5% of the annual average rainfall and thereby provide a more conservative range of ~0-10%. What I want to say is that I believe that the uncertainty estimates are too small and neglect some intrinsic uncertainty in the source products used for these calculations. This being said, I believe that the order of magnitude of the estimated recharge is very reasonable. To conclude, I expect that the authors add a thorough discussion of how reasonable these uncertainty estimates are given the uncertainty in the source products and the applied method.

We agree that the range that we obtain for the natural recharge rate can seem narrow considering a satellite-based approach. However, this approach is integrative and probably best suited to the large studied domain. By contrast, the synthesis reported by MacDonald et al. (2021) gathers various local methods known to be less integrative (mostly chloride mass balance and environmental tracers), and we averaged the results from multiple areas (with sometimes several studies for a same site) including seven of them yielding a zero recharge (out of 23 with rainfall < 150 mm yr^{-1}). Thus, the resulting average uncertainty is greater (i.e. 3.3 +/- 5.5%), and cannot be compared with the one from a single, and more or less homogeneous, study site using a different approach (i.e. 4.4 +/- 2.6% using the gravity-based water budget).
Moreover, GRACE products do not provide associated uncertainties for the raw data used here (they do so for the JPL's scaling factors for example, but we did not use those since it would be incoherent with the study of such an arid domain with TWS variations mainly driven by groundwater mass variations as stated Lines 248-252). And as you said, even if they were providing uncertainties, it would still be lower than averaging the products as we did (i.e. JPL and CSR).
However, to illustrate this, we can add a comment about uncertainties at the beginning of the discussion section, and more specifically about Blazquez et al. (2018) exploring the uncertainty in GRACE estimates:
"This study provides an estimate of the 2002-2019 domain-average natural recharge with associated uncertainty accounting for variations in natural discharge, groundwater pumping and irrigation return flow. The uncertainties associated with the calculation of the ΔGWS long-term trends with the GRACE and GLDAS products have also been considered. However, the error associated with the GRACE Mascons could not be propagated to the results since it is not provided. Blazquez et al. (2018) investigated the uncertainty of GRACE data by solving a global water budget using trends in ocean mass, ice loss from Antarctica, Greenland, arctic islands and trends in water storage over land and glaciers. The authors estimated a 0.27 mm yr-1 uncertainty for the GRACE data, a figure significantly lower than the uncertainties of the ΔGWS trends used in this study (Table 1)."

A Blazquez, B Meyssignac, JM Lemoine, E Berthier, A Ribes, A Cazenave, Exploring the uncertainty in GRACE estimates of the mass redistributions at the Earth surface: implications for the global water and sea level budgets, Geophysical Journal International, Volume 215, Issue 1, October 2018, Pages 415–430, https://doi.org/10.1093/gji/ggy293

- Lines 191-193: Related to the above comment, rather than just assuming that unquantified outlets constitute minor outflows and then neglecting them, I suggest considering the impact of such outflows along the boundaries of the system quantiatively by adding a term to the water balance and extending the uncertainty analysis. This can be done very quickly and would put a number on that assumption, rather than neglecting such terms. The beauty of a

simple water balance analysis is that such terms can easily be considered qualitatively. This could provide yet more realistic uncertainty estimates and help in resolving the aforementioned issue.

We believe that adding an unknown random natural discharge to the water budget would artificially raise the uncertainty on the recharge and would be less rigorous than ignoring it. Indeed, this outflow (towards the southeastern Khuff aquifer) could easily be null (or even negative, meaning an inflow). This small limit is indeed located next to the Al Qasim area which is known to be the one presenting the most intensive groundwater withdrawals. This creates a large drawdown cone almost reaching the eastern aquifer limit. Moreover, the total natural discharge fluxes being two orders of magnitude lower than groundwater pumping (Table 1), accounting for this hypothetical outflow would have an insignificant impact on the resulting estimate of natural recharge.

However, to answer this legitimate question, we suggest to add the following sentence at the end of section 2.2.2 (Line 195):
"Finally, with regard to historical piezometric maps of the Saq aquifer (Sharaf and Hussein, 1996; Lloyd and Pim, 1990), it can be assumed that the southeastern limit with the Khuff aquifer is likely inactive given the large drawdown cone created by the intensive pumping of the Al Qasim area."

- Sharaf, M. A. and Hussein, M. T.: Groundwater quality in the Saq aquifer, Saudi Arabia, Hydrol. Sci. J., 41, 683–696, https://doi.org/10.1080/02626669609491539, 1996.
- Lloyd, J. W. and Pim, R. H.: The hydrogeology and groundwater resources development of the Cambro-Ordovician sandstone aquifer in Saudi Arabia and Jordan, J. Hydrol., 121, 1–20, https://doi.org/10.1016/0022-1694(90)90221-I, 1990.

- I miss a discussion of the importance of ephemeral and intermittent streams in arid regions. These are often the main sources of recharge in arid regions as they collect and distribute the rainfall rapidly throughout the system, making infiltration available also to regions where it didn't rain locally. As stated in the beginning of section 2.4, permanent surface water bodies are almost completely absent from the Saq-Ram aquifer system, but surely intermittent systems are not. In other words, what happens to all the rainfall that doesn't form recharge prior to it being evaporated or transpired (i.e., ~95% of the AAR, according to the calculations in this study)? Before that water evaporates or is consumed and transpired by vegetation, it certainly forms intermittent stream networks. I suggest adding a short paragraph on their importance for (eco)hydrology and especially groundwater recharge in arid regions, supported by the at least the three references listed below, and drawing a link to intermittent streams on the arabian peninsula and the Saq-Ram Aquifer system. This would nicely round off the discussion of the importance of artificial recharge from agriculture and of the geology on the regional recharge. Suggested references:
Bourke et al., 2020, doi: 10.1002/wat2.1504
Dogramaci et al., 2015 , doi: 10.1016/j.jhydrol.2014.12.017
Schilling et al., 2021, doi: 10.1029/2020WR028429

Thank you for the useful references. We can add this comment at the beginning of the discussion section:
"Even if the Saq-Ram domain is devoid any permanent surface water bodies, ephemeral streams are known to be important for (eco)hydrology and local groundwater recharge in arid

regions (Shanafield et al., 2021; Dogramaci et al., 2015; Schilling et al., 2021). However, runoff coefficients were estimated at about 1% in the region (Al-Hasan and Mattar, 2013) while more than 90% of this runoff is lost by evaporation in the low lands. Thus, accounting for recharge redistribution through ephemeral streams in the water budget of the large Saq-Ram aquifer system would be quantitatively insignificant."

Al-Hasan, Abdul Aziz Saleh and Yousry El-Sayed Mattar. "Mean runoff coefficient estimation for ungauged streams in the Kingdom of Saudi Arabia." Arabian Journal of Geosciences 7 (2013): 2019-2029.

- Lines 270-273: This is not well explained and it's completely unclear to me which polygons were used to spatially average the GRACE and GLDAS products over the studied domain. In Figure 3, which is referenced here, no polygons or maps are provided. Explain.

How anomalies were computed (both with GRACE and GLDAS datasets) was also unclear for Reviewer #1 so the first sentence of this paragraph can be edited as:
"The SWS anomalies were computed in the same way as the GRACE TWS anomalies, i.e. by subtracting the January 2004 to December 2009 average from each monthly value of the time series."
Regarding the spatial average, we weighted the monthly spatial average of each signal by the proportion of each GRACE/GLDAS polygon within the domain (cut through by the aquifer system limit). This is stated in the next sentence:
"Both the GRACE products and GLDAS solutions were spatially averaged using the surface weight of each polygon within the Saq-Ram Aquifer domain (Figure 3)."
So, a polygon fully within the domain has a weight of 1, while a polygon 30% within the domain has a weight of 0.3 in the computation of the spatial domain averages.
We tried to provide the GLDAS/GRACE polygons on figure 3, but the fact that each product has a different size and shape of grid cell (listed section 2.3) makes the later unreadable, and adding multiple maps for this purpose seemed unnecessary (especially since this is an easily accessible information: https://grace.jpl.nasa.gov/data-analysis-tool, and https://ccar.colorado.edu/grace/gsfc.html).

Figures:

- Figure 1: provide coordinate reference system

The caption of the Figure 1 can be modified as such:
"Context map of the Saq-Ram Aquifer System (coordinates shown by straight dotted line every 5 degree; Shorelines and country borders extracted from Wessel and Smith, 1996; Administrative regions extracted from www.gadm.org)"

- Figure 2: from the legend and the caption it is not clear what the different lines refer to. Extend the caption to provide more information than just references without any additional info. It should be clear from the caption alone what is meant, the reader should not have to go and dig through the manuscript for the explanation of those references first.

The colored lines correspond to different sources of agricultural withdrawal data. The figure caption can be modified as:
"Annual average rainfall (Climatic Research Unit; mm yr^{-1}) of the Saq-Ram Aquifer System and agricultural withdrawal (from different sources represented by colored lines; 10^6 m^3 yr^{-1}) of its Saudi part (except for Othman et al.'s (2018) data corresponding to Al-Qassim, Ha'il and Al-Jouf regions of KSA)."

- Figure 3: Change axis titles to all lower case titles or use the already introduced abbreviations right away.

Done.
And "Monthly" can be added to the figure caption as such:
"Monthly times series of (a) the GRACE-JPL, -CSR, -GSFC terrestrial water storage anomalies (TWS; mm), and (b) the GLDAS-VIC, -CLSM, -NOAH soil water storage anomalies (SWS; mm) of the Saq-Ram domain."

- Figure 7: A very nice analysis and presentation

Thank you.

Abbreviations:

- The Kingdom of Saudi Arabia (KSA) is sometimes called 'Saudi Arabia', sometimes with the abbreviation KSA or sometimes simply referred to as 'Saudi'. Double check and make the naming consistent throughout the manuscript.

We will only use the terms 'Saudi Arabia' and 'Saudi' (referring to what "belongs" to Saudi Arabia).

- The GRACE abbreviation is introduced at least three times: once in the abstract, once in the intro, once in the methods. Introduce it once in the intro, that is sufficient. Remove the introduction of the abbreviation in the abstract to save space for more relevant info.

Done.

- Other abbreviations such as terrestrial water storage (TWS), GWS and SWS are also introduced multiple times (three or four times at leas). Moreover, the full names are written with capital first letters in the figures' axis titles (e.g, Figure 3), rather than without capital letters or with the abbreviations. Avoid introducing abbreviations so many times and make the naming consistent throughout the manuscript.

Done.
But, if it is possible, we would like to keep the definition of TWS and SWS both in the data section (2.3) and the method section (2.4) so the reader have a clear understanding of the methodology without having to refer to the prior data section (2.3).

References:

- The intro is a little light on recent references, particularly on available methods for groundwater recharge quantification (in arid regions). Since all methods are subject to different sources of considerable uncertainty, it would be good to provide more references and to direct the reader to this information. I would suggest adding the following references to lines 62-65:
Shanafield and Cook, 2014, doi: 10.1016/j.jhydrol.2014.01.068
Banks et al., 2020, doi: 10.1016/j.jhydrol.2020.125753

Done.
Thank you for providing such useful references.

- For the discussion of the importance (and dominance) of intermittent streams on groundwater recharge in arid regions, see comment above

Acknowledged.

- Lines 249-252: provide a reference for this statement

Done. One statement can also be added to justify that the use of scale factors is not suitable for this study:
"In fact, these downscaling factors are based on the mass distribution calculated by land surface model (LSM) accounting for surface and subsurface water transfers (Landerer and Swenson, 2012) while TWS variations in such arid regions are expected to be chiefly controlled by groundwater mass variations. Moreover, as stated by the authors, the use of such gain factors is not suitable to quantify trends."

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