Review of ESD 2019-43 by Mohammad Shamsudduha and Richard G. Taylor

The authors use the results of three different GRACE-based TWS methods and 4 Land surface models to generate an ensemble of groundwater storage anomalies. These are subsequently analyzed by a non-parametric statistical method to separate seasonal signals from non-linear trends and residuals.

The main message of the paper is that trends in GWS anomalies (ΔGWS), if existing, are non-linear in the vast majority of main aquifer systems and that rainfall anomalies play an important role in explaining these non-linear trends.

I enjoyed reading the paper. I find that it is a well-written with an important message that deserves publication. However, I have a few comments.

Moderate comments

1. I find the lack of reference to estimates based on global hydrological models (GHMS) remarkable. The first spatially distributed global assessment of depletion rates where based on such models and, albeit indirect, should be used in the discussion. They are the basis for the "narratives on global groundwater depletion" that are mentioned in the discussion and the abstract (See https://iopscience.iop.org/article/10.1088/1748-9326/ab1a5f/meta for an overview of these studies). This is the more remarkable, given that the authors do use Land Surface Models (LSMs) to estimate ΔGWS from GRACE ΔTWS.

2. Regarding the estimation of ΔGWS from GRACE ΔTWS (Equation 1): I am quite doubtful that the surface water storage from integrating LSM runoff on a monthly basis is sufficiently accurate. Even a small basin as the Rhine has a discharge peak routing time of a week, while that of the Amazon amounts to 3 months. Apart from the lack of river routing, GLDAS LSMs do not include the storage and delayed discharge from reservoirs, lakes and inundated floodplains (GHMs do a better job in that respect; see https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL081836). This fact may lead to underestimation of ΔSWS and subsequently an overestimation of ΔGWS and its noisiness. Granted, comparison with piezometric data in the Limpopo and the Ganges-Brahmaputra is favourable, but this can be scaled easily by changing specific yield.

3. The discussion related to the "narrative of global groundwater depletion" needs elaboration:
    - Not only piezometric studies show that groundwater depletion can be very local; this is also true for model-based estimates of groundwater depletion. See for instance results from Wada et al. 2012 (Figure S5) https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2012GL051230 and De Graaf et al 2017 (Figure 11): https://www.sciencedirect.com/science/article/pii/S030917081630656X#fig0011 . This means that at the aquifer scale anomalous rainfall may cause an overall increase in groundwater storage, while groundwater depletion may locally still

persist. Thus, the “narrative of global groundwater depletion” pertains to “groundwater depletion as a global phenomenon”.

- The current consensus seems to be that global ΔTWS has been increasing between 1950-1995 by dam building, decreasing from 1995-2005 by groundwater depletion and has been increasing again since then by increased land water storage due to a climate-change induced increase in precipitation: see the review by Wada et al: https://link.springer.com/chapter/10.1007%2F978-3-319-56490-6_7 Yet, at the same time groundwater depletion at the current hotspots has persisted. How do your findings relate to these insights?

4. Line 146: I don’t understand the 20 realisations. I would think: 3 GRACE products, 4 LSM estimates of ΔSMS and ΔSWS and one LSM with ΔSNS amount to 3x4x1 = 12 realisations? Or did you combine e.g. ΔSWS from one LSM with the ΔSMS from another? If you did this, this seems to be inconsistent as it would not preserve mass and overestimate the errors due to the LSM corrections.

Small remarks:

- The first sentence of the introduction: Doell et al (2012) is only one model-based study providing these numbers. I would advise using less significant numbers based on an overview of estimates by Hanasaki et al (2018): https://www.hydrol-earth-syst-sci.net/22/789/2018/
- Lines 212-215: Trying out different smoothing parameters. I feel that the results of this exercise should be shown, at least in the Supplementary Information (SI). Perhaps report the statistics of the residuals for a number of settings of the smoothing parameters to justify the values chosen.
- On a related note: Looking at some of the plots in the SI I see that residuals are often far from white. In the time series literature this would be seen as a serious model insufficiency. Some discussion on how this would affect results is warranted as well.