

The article entitled "Water storage trends derived from the GRACE/-FO global gravity-based groundwater product (G3P)" discusses prospects and limitations of the newly available satellite-based global groundwater change assessment. I find the paper interesting and well within the scope of the journal. However, I also noted a few weaknesses that should be thoroughly addressed through a revision.

We thank you for the constructive assessment of our manuscript. Please find our responses to your comments below.

I think that it would be good to find indicative names for the clusters depicted in Fig. 2. Using the colors is not really helpful, since those can only be used with the Figure. There are common features in most of the clusters, and I strongly urge the authors to come up with a narrative to distinguish those groups.

We see the point in referring to clusters mainly by color or number might be insufficient. We will revise Fig. 2 and Sect. 3.1.1 by adding descriptive cluster names and clearer information on the dominant time-series patterns. We also want to clarify that the clusters are not intended as hydrological attribution classes (c.f. comments to referee #1). They summarize time-series characteristics, such as seasonality, interannual variability, and trend-like behavior. Thus, regions with different physical drivers may fall into the same cluster if their GRACE/-FO time series are statistically similar. Possible labels will be kept descriptive, for example: For TWS: background seasonal/interannual variability; weak drying with seasonality; tropical seasonal variability; high-amplitude tropical variability; mixed depletion / cryospheric-loss regime; strong cryospheric mass-loss regime. For GWS: background GWS variability; high-latitude seasonal/interannual variability; moderate GWS drying; positive GWS / recharge-like regime; strong GWS decline / possible leakage-affected regime; Amazon high-amplitude seasonal regime. We will shorten the clustering discussion, better connect it to the subsequent trend-significance analysis, and consider moving detailed cluster material to the Supplement. This revision also addresses the related comment by Referee #1 concerning the added value and interpretation of the clustering analysis.

There are quite a few inconsistencies in the nomenclature (e.g., identical units are abbreviated in different ways) and various rather sloppy formulations. Some examples are given below. A careful revision of the full manuscript in terms of language and style seems urgently necessary.

We will perform a full consistency check of units, notation, storage-component names, figure captions, table headings, and terminology. Volume-change rates will be harmonized throughout the manuscript, to $\text{km}^3 \text{ yr}^{-1}$. We will also revise ambiguous or awkward formulations throughout the text, not only the examples listed by the referee.

The paper is heavily relying on abbreviations that are introduced at some point but later hardly used again. This applies in particular to abbreviations of certain mathematical methods, which have hardly any value apart from expert readers who might recognize the abbreviations from earlier encounters outside this manuscript. I think that this should be changed to make the article better readable.

We agree that the manuscript currently introduces too many abbreviations, especially in the methods section. We will reduce abbreviations that are not used repeatedly or are not essential for reproducibility. Essential abbreviations such as TWS, GWS, RZSM,

SWE, SWS, GRACE/-FO, and possibly GLS/OLS will be retained where needed, but better explained. For statistical terms used only briefly, we will either write them out or move technical detail to the Supplement.
Additionally, we will address the following issues:

line 29: "close to persistent" is not very precise wording

line 58 and 69: different ways for volume change units are given without an obvious reason. Please harmonize throughout the article.

line 84: There is an extensive body of (recent) literature from Donald Argus and colleagues on the use of GRACE data in California which should be cited here.

line 109: "long-period-transient-type apparent-secular trend" is a very much German construct. Please revise.

line 148: I am still not sure what an "uncertainty aware" assessment should be. Please elaborate and explain better.

line 152: Two times "limited" in a single line

line 155: demonstrated scalability has not been achieved, I believe. Please re-consider your statement.

line 165: I think that this is rather a sequential than a hierarchical approach.

line 199: It should rather read w.r.t. the temporal mean over the reference period. Same applies to line 204 (and elsewhere).

line 236: pre-analysis sounds wrong here

line 238: I think that overtones of the annual signal are meant here. Harmonics basically can have any frequency.

line 266: "gaps as-is" is a non-information. Please explain or remove.

line 277 and following: Various abbreviations listed here could be omitted since those are never used again: GLS, RMLE, MLE, OLS, PLWN.

line 301: space are missing in the equation.

line 311: "surface is calculated geodetically" sounds quite blatant for a trivial latitude weighting.

line 313: what is a "direct physical integration"?

line 625: why do you use bold font in the subsection on Asia only?