Referee #1

The paper describes a statistically based method to identify long and short term variability (trends and interannual variability) in TWS and GW. It also include analysis on the TWS components such as SWE, Soil moisture, surface water, and glacier. Th authors run their analysis on the G3P product. An already established framework that determines GWS from removing snow, soil moisture, etc. from GRACE/FO derived TWS estimates.

In general, I think the paper is a valuable contribution to the scientific community as it identifies trends and variabilities in GSW and TWS using the 21+ years of GRACE/FO record. The paper uses statistically robust methods to identify significance in the trends and compares their findings with previously defined trends.

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

I have a few recommendations, as noted next.

1. As a reader, I am lost to what the main novelty of the paper is as many previous references exist that determine long and short term variability using GRACE/FO data. I think it resides in determining trends using G3P, globally and for an extended time-period (21+years). I recommend the authors to simplify and reinforce what are the key novelties.

We agree that this should be outlined more clearly. The intended novelty is not only a general estimation of GRACE/-FO trends itself, but a first global assessment of statistically significant groundwater-storage trends from the G3P product over the extended GRACE/GRACE-FO record, and a joint interpretation with TWS and the subtracted storage components. The added value lies in combining the G3P groundwater product with a consistent stochastic-memory-aware trend-significance framework and in aggregating statistically significant trends to continental and global water-storage budgets.

We will sharpen this point in the abstract, introduction, results discussion, and conclusions. We will also more clearly separate the literature-synthesis part of the manuscript from the original contribution of this study; the literature review motivates the problem and places the results in context, whereas the new contribution is the G3P-based trend detection, significance assessment, and volumetric groundwater quantification. Such framing is already mostly coherent with the current manuscript, where G3P is introduced as a globally consistent satellite-gravimetry-derived GWS product and the trend framework is used to separate significant long-term trends from stochastic variability.

2. The reviewer (as well as other readers) might not as familiar with many of the statistical terms/methods that are described in 2.2. I recommend streamlining or explain in a more simple words these as much as possible and try to the best of their possibilities to make these statistics methods more easily digestible.

We agree that Sect. 2.2 can be made more accessible for the broad HESS readership. Although we included an illustrative example of the stochastic processes, we will further

streamline the section by adding short explanatory sentences before the formal definitions and equations. In particular, we will clarify the practical use of the method: for each grid-cell time series, we estimate a trend and seasonal terms, model the temporal correlation of the residuals, derive trend uncertainties from the calibrated stochastic model, and then test whether the trend differs significantly from zero.

3. While I think the cluster analysis (fig 2) makes sence from a statistical prespective, I am finding hard to connect or link it to a physically meaningful hydrological meaning. Can the authors be more explicit about why a specifc trend in one ares can be represented the same way as the trend in an other area? If we then show higher resolution trends (Fig 4), whats the needs to report the cluster analysis? Again, it is not clear what is the added value of the cluster analysis, please clarify or perhaps move to a supplementary text?

We agree that the role of the clustering analysis could benefit from a clarification. The clustering is not intended as a physical hydrological attribution or as a causal classification; it is intended as an explanatory tool to introduce important data characteristics, and to summarize dominant global time-series patterns, such as seasonality, interannual variability, and trend-like behavior. Therefore, physically different regions can belong to the same cluster if their GRACE/-FO time series have similar temporal structure. We will revise the clustering section accordingly, and we shall also introduce names for the clusters, as proposed by referee $2. This will help to make the discrimination of different phenomena more clear. We will make clear that the clustering provides an introductory global overview of the data before the formal trend-significance analysis, while the significance maps and regional trend interpretation remain the main quantitative results. We will shorten this section, and try to improve the connection to the later results, and consider moving detailed cluster material to the supplement.

4. In order to determine GWS, G3P removes single components of SWS, SWE, etc. While I think most of these are satellite derived, SWS is model based (according to 191-194) if so, how are human driven (diversions, dams, etc) fluxes/storages represented?

In G3P, SWS is taken from the OS LISFLOOD/GloFAS models.Therefore, human-managed surface-water storage, including reservoir operation, dams, diversions, regulated lakes, withdrawals, and return flows are considered; LISFLOOD is a high-resolution product, down to 0.05 degrees, driven by high-resolution ECWMF forcing inputs. Therefore, it can be considered generally as a reliable product. Since GWS is derived residually from TWS after subtracting SWS and the other components, unrepresented or imperfectly represented SWS trends can be aliased into the GWS estimate (as also mentioned for other products). We will add this discussion to Sect. 2.1 and revisit it in the discussion.

5. Along the same line, if only root zone soil moisture is used (182-185) how are irrigation driven changes of the SURFACE soil moisture being accounted for? Please add caveats about this.

The RZSM component used in G3P is derived from satellite surface-soil-moisture

observations with gap filling and propagation to the root zone. Thus, surface soil moisture is part of the input information, which shall also cover irrigation-driven surface wetting . Since GWS is computed as a residual, errors or omissions in irrigation-related RZSM variability can propagate into the inferred GWS trends. We will add this discussion to Sect. 2.1 and revisit it in the discussion.

6. Section 3.2 is the core of the paper in my opinion. I would emphasize it by adding a table to summarize also how the authors results agree or disagree with previously reported analysis. This will also help with framing the novelty point (above).

We agree that Sect. 3.2 is central to the manuscript. While Sect. 3.2.8 and Tables 1–2 already summarize main quantitative findings of this study, we acknowledge the point that the relationship to previous work can be made clearer. We consider adding a compact synthesis table focused on the main regional findings and their added value relative to earlier GRACE/-FO and groundwater studies. However, a numerical comparison is delicate, as published estimates differ in study period, processing strategy, spatial mask, component separation, leakage/GIA treatment, and significance criteria. We aim to indicate which patterns are confirmed by our analysis, which are newly resolved or strengthened by the G3P-based significance-enhanced framework, and which remain uncertain.

7. Finally, I found some typos here and there, please re-read accordingly to identify any grammar issues. Example: 111: “This [IS] a strong..” 715-716: some words are missing perhaps? 725: fix parathesis in the

We will carry out a language, grammar, nomenclature, unit-consistency, and formatting revision of the manuscript. We will correct the specific examples mentioned by the referee.