March 4, 2025

Dear Editor,

Thank you for the opportunity to submit a minor revision. In response to the remaining reviewer comments on drought return period on S_r, we added a new discussion paragraph (lines 276 - 285) highlighting why requiring perfectly matched return periods is not necessary, given that these S_r estimates are typically presented and interpretated as measures of *storage capacity* rather than time-dependent drought indicators. The added discussion clarifies the comparability between different S_r products and strengthens our key conclusions. A detailed response addressing the reviewer's remaining concerns is included in this letter. In the following pages, reviewer's comments are reproduced in their entirety in black, and our responses are noted in blue.

Additionally, we have updated the Code availability and Data availability sections (lines 377 - 380) to include DOI links to our $S_r^{GRACE/FO}$ dataset and python code.

We hope you find our revised manuscript suitable for publication.

Best regards,

Meng Zhao

On behalf of all co-authors

Reviewer 2: In the revised version of this work, the authors have nicely addressed most of my questions. I am glad they incorporated the GLEAM ET estimates as validation as this adds robustness to the analysis.
However, I have one lingering question: in my previous comments I said that the methods used were not readily comparable. The authors argue that they are comparable because of their shared definition of the physical processes involved. I agree that there is a shared definition of the processes, but I don't think the estimates are truly comparable because:
1) Sraccum calculates Sr based on "cumulative water deficit extremes occurring with a return period of 80 years".
2) The return period for SrRDxWHC is not clear.
3) SrGRACEFO uses the maximum multiyear drydown on record.

I think that the timescales used to analyze how extreme the drydown period is will affect the storage. For example, if we assume that the return period for SrGRACEFO is a 200-years. It is reasonable to assume that Sraccum calculated for a 200-yer return period drought would be significantly larger. The estimates of Sr must be done with "comparable" droughts.

I am not 100% sure how you can directly address this issue. Perhaps if we knew the return period of the droughts you analyze in your work and then used the Sraccum estimates for that return period?

Response: We are glad that our effort to make the analysis more robust and improve the overall quality of this manuscript is appreciated.

Thank you for your comment regarding the comparability of S_r estimates due to differences in drought return periods. While we agree that the return period over which S_r is calculated will influence the values estimated, we view this return period as part of the methodology of each S_r estimate. That is, because most literature interpreting prior estimations of S_r does not consider the time period over which it is calculated (interpreting it instead as an absolute maximum storage capacity), our key finding - that the GRACE-derived S_r shows that storage capacity is larger than previously estimated – is not sensitive to this assumption. Nevertheless, we agree that a comparison that is as similar in time period as possible would be ideal.

Additionally, the S_r^{accum} used in our comparison is statistically inflated from the raw S_r^{accum}, which is directly derived from Earth observations from 2003-2018 before being scaled to an 80-year return period under the assumption of a Gumbel's distribution. However, it is uncertain whether real-world drought return times follow this distribution. A more appropriate comparison would be between the unscaled S_r^{accum} and $S_r^{GRACE/FO}$, as the latter was derived from 2002-2022 data, a similar period as that of the unscaled S_r^{accum}. This comparison would ensure that both estimates reflect droughts that occurred during a similar period, rather than relying on an imposed statistical return period.

Unfortunately, the unscaled S_r^{accum} is not publicly available. Even so, it is evident that the unscaled S_r^{accum} would be smaller than the scaled S_r^{accum} used in this study. Given that the scaled S_r^{accum} is already lower than $S_r^{GRACE/FO}$, the unscaled version would be even smaller. This further supports our conclusion that S_r^{accum} underestimates root-zone storage capacity compared to $S_r^{GRACE/FO}$.

We now explicitly discuss the return period assumption in S_r^{accum} and its impact on the S_r^{accum} - $S_r^{GRACE/FO}$ difference in lines 276 – 285, which we have reproduced below for the reviewer's convenience:

> *The discrepancy between $S_r^{GRACE/FO}$ and S_r^{accum} is further influenced by differences in how drought return periods are defined. S_r^{accum} values reported in Stocker et al. (2023) are statistically scaled to represent an 80-year drought return period, rather than being directly derived from observed drought events. This extrapolation assumes a fixed probability distribution of drought occurrence, which may not fully capture real-world hydrological variability. In contrast, $S_r^{GRACE/FO}$ is based on observed multiyear TWS drawdowns from 2002-2022, directly reflecting droughts that occurred over the past two decades. A more comparable approach would require using the unscaled S_r^{accum}, which was derived using Earth observations of precipitation and ET from 2003-2018, without statistical adjustments. However, the unscaled S_r^{accum} is not publicly available. Given that the 80-year return period scaling inflates S_r^{accum} values, and $S_r^{GRACE/FO}$ is still substantially larger, it follows that the unscaled S_r^{accum} would be even lower, further supporting the conclusion that the S_r^{accum} approach underestimates root-zone storage capacity compared to $S_r^{GRACE/FO}$.*