**Response to comments from Anonymous Referee #1**

Dear reviewer:

Thank you very much for your valuable comments which help to improve the manuscript ESSD-2025-497 greatly. We have replied the comments one by one and revised the manuscript as well as the supplements accordingly. Our response consists of two parts: (1) response to overall remarks; (2) response to specific comments. The details are listed below.

With best regards,

Zongxia Wang, Suxia Liu, Xingguo Mo

1. **Response to Overall Remarks**
2. The manuscript sometimes reads like the authors are just using and comparing lots of data without a clear objective of why all the data are required. They use lots of datasets, then compare the datasets without any reasonable justifications on why such comparisons are needed.

**Reply**: Thanks for the suggestion. Detailed justifications on why all the datasets are required and the objectives of the comparisons between different datasets have been supplemented in the revised manuscript. To be specific, a brief section has been added at the beginning of Chapter 2 to summarize the datasets utilized for generating igGWSA in this study. Besides, justifications for comparing several kinds of non-improved GWSA estimations with igGWSA have been added to Section 3.3.1. Moreover, justifications for comparing SMSimproved to fixed-depth SMS (i.e., SMS200 and SMS289) have been added to Section 4.2.1. For details, please refer to the "Response to Specific Comments" below (Comment 3, 12, and 17).

1. It also contains a large amount of information, which sometimes make it challenging to follow how the study was done and identify the key findings. The authors should restructure the different sections ensuring they only keep key information in the main document while moving any supporting text to the supplements.

**Reply**: Thanks for the suggestion. To improve the readability of the main manuscript, we have moved some supporting information to the supplements. To be specific, the detailed principles of PSM simulation in CLSM have been moved to Text S4 in the revised supplements. The discussion on the influence factors of PSM modelling has been moved to Text S9 in the revised supplements.

1. **Response to Specific Comments**
2. LN22: “… would lead to misinterpretations of groundwater storage variations in glacier-covered regions” – misinterpretation of GW storage? You detail some of the misinterpretations in the last sections but you could also briefly state some here.

**Reply**: Thanks for the comment. A brief introduction regarding the misinterpretations of groundwater storage variations caused by the incomprehensive considerations of non-groundwater components has been added to the abstract in the revised supplements.

Line 21–27 in revised manuscript with changes marked:

*Compared to igGWSA, simplified estimation would lead to misinterpretations of groundwater storage variations. For example, neglecting glaciers and permafrost would erroneously interpret glacier melting and permafrost degradation as part of changes in groundwater storage, respectively, thereby exaggerating groundwater mass loss in global glaciated areas and concealing groundwater mass gain in global permafrost zone. Failure to account for lakes and reservoirs would also lead to an overestimation of groundwater decline globally.*

1. LN26: “https://doi.org/10.5281/zenodo.16871689 (Wang et al., 2025)” – The data is provided as a .mat file. Matlab is not freely available, making the data un-accessible to many. Can the authors provide the dataset in an open format; e.g. in .nc or a multi-band .tiff?

**Reply**: Thanks for the suggestion. We have provided the igGWSA dataset in NetCDF format on Zenodo. The new version of igGWSA dataset is publicly available through https://doi.org/10.5281/zenodo.18216946.

1. Chapter 2: As mentioned above, this chapter presents a lot of datasets without necessarily justifying why such data are needed. Add a brief section at the beginning of the chapter summarizing what data are presented/detailed in the later stages.

**Reply**: Thanks for the comment. A brief section has been added at the beginning of Chapter 2 in the revised manuscript to summarize the datasets utilized for generating igGWSA in this study.

Line 114–121 in revised manuscript with changes marked:

*Four categories of data were utilized in this study to generate igGWSA: (1) terrestrial water storage data, which was the fundamental data source for estimating igGWSA; (2) non-groundwater components data, including glaciers, permafrost, snow, lakes, reservoirs, surface runoff, plant canopy water, and profile soil moisture, which would be deducted from terrestrial water storage to isolate igGWSA; (3) predictor variables for profile soil moisture, which served to reduce the uncertainty in single-source profile soil moisture simulation; and (4) in situ-observed and model-simulated groundwater data, which acted as independent benchmarks for validating igGWSA. Detailed descriptions of the aforementioned datasets were presented below.*

1. Chapter 2: All the datasets used in this study have varying resolutions (mostly varied spatial resolutions). The authors briefly state that they resample data using the bilinear interpolation method. They should add a bit more detail acknowledging the scale issues.

**Reply**: Thanks for the comment. A more detailed description of the spatial resolution issues has been added to the final paragraph of Chapter 2 in the revised manuscript.

Line 219–227 in revised manuscript with changes marked:

*... The spatial resolution of these gridded datasets varied significantly, ranging from 0.083º🞩0.083º to 1.25º🞩0.9375º. As a compromise, a uniform spatial resolution of 0.5º🞩0.5º was adopted for this study. For datasets with resolution finer than 0.5º🞩0.5º, e.g., 0.1º🞩0.1º, the raw data can be upscaled by averaging the values within every non-overlapping 5🞩5 pixel block. For datasets with resolution coarser than 0.5º🞩0.5º, e.g., 1º🞩1º, scale harmonization was achieved by applying bilinear interpolation. Bilinear interpolation incorporates the values of the four surrounding pixels, producing a smoother output than nearest neighbour interpolation (Hu et al., 2013; Ruan et al., 2025). Therefore, it is more appropriate for continuous data, such as soil moisture data used in this study.*

1. LN133: ‘Snow water equivalent simulations from seven reanalysis products’ - which metric is used for the SWE ensemble-wighted average used in this study? average, median …?

**Reply**: Thanks for the comment. The weighted average of SWE was obtained by using BTCH method. Specifically, the weight assigned to each dataset was determined by its error variance. Descriptions about weighted average were not placed in the Datasets section here, but rather in the Methods section and the supplement.

1. LN161-178: ‘2.2.7. Profile soil moisture (PSM) …’ – since you do not present the other datasets/variables with this much detail, this block of text could be well suited for the supplements (only keep a short summary here).

**Reply**: Thanks for the suggestion. Details regarding the principles of PSM simulation have been moved to Text S4 in the supplements, with only a brief summary retained in the revised manuscript.

Line 172–186 in revised manuscript with changes marked:

*In this study, PSM simulated by GLDAS CLSM was utilized. CLSM relies on the concepts of TOPMODEL to generate soil moisture patterns, and the principles were detailed in Text S4 and Gascoin et al. (2009).*

1. LN180: ‘meteorological variables, and vegetation index’ - You mention the meteorological variables and indices used later in the text (i.e. precipitation and ndvi) but you should also list them here.

**Reply**: Thanks for the comment. The specific meteorological variables and vegetation index used as predictor variables for PSM have been listed in the first sentence of this paragraph.

Line191–193 in revised manuscript with changes marked:

*Root zone soil moisture (RZSM), meteorological variables (air temperature and precipitation), and normalized difference vegetation index (NDVI) were selected as predictor variables for PSM.*

1. LN181: ‘…selected as predictor variables’ - Why was evapotranspiration (which is a main component of the water cycle, thus a likely driver of water storage anomalies) not used as one of the predictors/covariates? there are many global ET products that are freely available.

**Reply**: Thanks for the comment. We selected RZSM, air temperature, precipitation, and NDVI as predictors of PSM. ET was not considered for the following reasons. (1) According to Seneviratne et al. (2010, https://doi.org/10.1016/j.earscirev.2010.02.004), soil moisture–climate interactions primarily manifest as soil moisture–temperature and soil moisture–precipitation feedbacks. Therefore, we consider air temperature and precipitation as representative meteorological factors influencing soil moisture. (2) Generally, soil moisture acts as one of the key controlling factors of ET (Seneviratne et al., 2010). For example, Budyko (1956) defined ET regimes as a function of soil moisture. Therefore, it would be less appropriate to consider ET as a predictor of PSM instead. (3) ET is closely related to air temperature and vegetation conditions (NDVI). Therefore, including ET as one of the predictors is highly likely to introduce multicollinearity.

1. LN208: Why not use the much simpler nearest neighbour. Unlike bilinear, it does not create new data (i.e. keeps the original data as-is)?

**Reply**: Thanks for the comment. Nearest neighbour interpolation assigns the value of the nearest pixel without creating new values and is suitable for discrete data, such as land use types. In contrast, bilinear interpolation incorporates the values of the four surrounding pixels, producing a smoother output. Therefore, bilinear interpolation is more appropriate for continuous data, such as soil moisture and snow water equivalent data used in this study. The advantages of bilinear interpolation have been added to the final paragraph of Chapter 2 in the revised manuscript.

Line 223–227 in revised manuscript with changes marked:

*... For datasets with resolution coarser than 0.5º🞩0.5º, e.g., 1º🞩1º, scale harmonization was achieved by applying bilinear interpolation. Bilinear interpolation incorporates the values of the four surrounding pixels, producing a smoother output than nearest neighbour interpolation (Hu et al., 2013; Ruan et al., 2025). Therefore, it is more appropriate for continuous data, such as soil moisture data used in this study.*

1. LN219: ‘generate ensemble simulations’ - do you mean to ‘generate the weighted average from the ensemble simulations’ ?

**Reply**: Thanks for the comment. The phrasing here may be confusing. We actually mean to ‘obtain the weighted average of multi-source simulations from different reanalysis products’. The phrasing has been corrected in the revised manuscript.

Line 238–240 in revised manuscript with changes marked:

*In this study, BTCH was employed to obtain the weighted average of multi-source simulations of snow water equivalent, plant canopy water and RZSM from different reanalysis products (Fig. 1, Text S6).*

1. LN220: ‘… (Fig. 1, Text S5’ - In Text S5 in the supplementary document you write ‘𝜎𝑖 is the error variance of the 𝑖th dataset’ – 𝜎𝑖 is the standard deviation; change 𝜎𝑖 to 𝜎𝑖2, which is the variance.

**Reply**: Thanks for the correction. We have changed 𝜎𝑖 to 𝜎𝑖2 in Text S5 in the supplementary document.

Line 65 in revised supplement with changes marked:

*...* 𝜎𝑖2 *is the error variance of the ith dataset.*

1. LN280: ‘To quantify the impacts of incomprehensive considerations of non-groundwater components, five kinds of non-improved GWSA were further estimated as listed below.’ - Seems arbitrary; any justifications for selecting these 5 and not any other combinations?

**Reply**: Thanks for the comment. Detailed justifications for constructing these non-improved GWSA estimations have been supplemented in the revised manuscript.

Line 298–326 in revised manuscript with changes marked:

*To quantify the potential impacts of incomprehensive considerations of non-groundwater components (e.g., neglect of glaciers, permafrost, deep-layer soil moisture, etc.), six kinds of non-improved GWSA were further estimated and compared to igGWSA. The estimation methods and their respective purposes are listed below.*

*(1) GWSAsimplified: Similar to most previous studies (Liu et al., 2023; Peng et al., 2021; Zhao et al., 2023), GWSAsimplified was separated from TWSA by subtracting only soil moisture storage in the 0–200 cm soil layer, snow water equivalent, and plant canopy water simulated by GLDAS Noah. The potential impacts of this most simplified approach adopted by most previous studies on the accuracy of GWSA estimation can be quantified by comparing GWSAsimplified and igGWSA.*

*(2) GWSACLSM: CLSM-simulated PSM instead of improved PSM was used to estimate GWSA, with all other non-groundwater components identical to those for igGWSA. By comparing GWSACLSM and igGWSA, we can determine whether machine learning improves the accuracy of GWSA estimation by reducing the uncertainty of CLSM-simulated PSM.*

*(3) GWSA289: Soil moisture storage in the 0–289 cm soil layer derived from ERA5-Land instead of improved PSM was used to estimate GWSA, with all other non-groundwater components identical to those for igGWSA.*

*(4) GWSA200: Soil moisture storage in the 0–200 cm soil layer derived from FLDAS Noah instead of improved PSM was used to estimate GWSA, with all other non-groundwater components identical to those for igGWSA. The potential impacts of neglecting deep-layer soil moisture on the accuracy of GWSA estimation can be quantified by comparing GWSA200, GWSA289 and igGWSA.*

*(5) GWSAnoGLA: GlacierWSA was ignored when estimating GWSA, with all other non-groundwater components identical to those for igGWSA. The potential impacts of neglecting glaciers on the accuracy of GWSA estimation can be quantified by comparing GWSAnoGLA and igGWSA.*

*(6) GWSAnoPER: PWSA was ignored when estimating GWSA, with all other non-groundwater components identical to those for igGWSA. The potential impacts of neglecting permafrost on the accuracy of GWSA estimation can be quantified by comparing GWSAnoPER and igGWSA.*

1. LN298: ‘five globally recognized hotspots of groundwater depletion, …’ – reference/citation needed.

**Reply**: Thanks for the comment. The relevant references have been supplemented in the revised manuscript.

Line 329–331 in revised manuscript with changes marked:

*..., validation of igGWSA was performed in five globally recognized hotspots of groundwater depletion, i.e., North China Plain, California Central Valley, High Plains of USA, Ganges-Brahmaputra River Basin, and the Middle East (Chen et al., 2016; Jasechko et al., 2024).*

1. LN304-306: ‘Accordingly, point-scale data were first converted into pixel-scale by averaging observations of wells located in the specific grid cell. Then in situ GWL and GWSA estimation at a 0.5° X0.5° resolution were upscaled to obtain basin-averaged time series.’ - not enough information for the reader to determine how this was done.

**Reply**: Thanks for the comment. The detailed procedure for scale transformation has been supplemented in the revised manuscript.

Line 338–344 in revised manuscript with changes marked:

*Scale transformation involved three steps. (i) Converting the in situ data from point scale to pixel scale. For example, for a given 0.5º🞩0.5º grid cell that contains 10 wells, the value of this grid cell was represented by the average of the observations of all these wells. (ii) Converting the in situ data obtained in step i from pixel scale to regional scale. For example, for a given region that contains 20 grid cells, the value of this region was represented by the average of the observations of all these cells. (iii) Converting the igGWSA data from pixel scale to regional scale as in step ii. Subsequently, validation was conducted based on the regionally-averaged time series of in situ data and igGWSA.*

1. LN342: ‘loess and chernozem zones worldwide’ – reference needed.

**Reply**: Thanks for the comment. The relevant references have been supplemented in the revised manuscript.

Line 386–387 in revised manuscript with changes marked:

*... encompassed the major loess and chernozem zones worldwide (Fu et al., 2017; Li et al., 2025).*

1. LN372: ‘Given this, evaluation of interannual trends in PSM was carried out additionally’ – grammar: rephrase or remove ‘additionally’.

**Reply**: Thanks for the correction. ‘additionally’ has been removed from this sentence to improve readability.

Line 417–418 in revised manuscript with changes marked:

*Given this, evaluation of interannual trends in PSM was carried out.*

1. LN380-387: section 4.2.1 - Why do you compare the era5-land-SMS289 (and other SMS) estimates to SMSimproved? Some justification needed here.

**Reply**: Thanks for the comment. The justification for comparing SMSimproved to fixed-depth SMS (i.e., SMS200 and SMS289) has been added to the very beginning of Section 4.2.1 in the revised manuscript.

Line 426–429 in revised manuscript with changes marked:

*To quantify the difference between soil moisture storage (SMS) within fixed-depth layers and that across the entire profile, as well as the potential significance of deep-layer SMS, our improved estimation of PSM (SMSprofile) was compared to SMS in 0–200 cm layer (SMS200) and 0–289 cm layer (SMS289) from the perspectives of magnitude and spatial pattern (Fig. 5).*

1. LN417: ‘Validation of igGWSA against GWDin situ and GWSAWGHM’ - Since you also compare your igGWSA with the GWSAWGHM, can you also provide other metrics for a more exhaustive comparison, e.g. bias …

**Reply**: Thanks for the comment. Comparison between igGWSA and GWSAWGHM based on relative error has been supplemented in the revised manuscript. The values of relative error have been added to the new Figure 7.

Line 352–355 in revised manuscript with changes marked:

*Moreover, to quantify the difference in their interannual variations, the relative error (RE) of the trend in igGWSA was computed, taking GWSAWGHM as the benchmark. For example, positive RE indicates that igGWSA overestimates groundwater decline compared to GWSAWGHM.*

Line 461–469 in revised manuscript with changes marked:

*In terms of relative error, igGWSA exhibited a more pronounced decline than GWSAWGHM in Ganges-Brahmaputra and the Middle East, while showing a weaker decline in North China Plain, Central Valley, and High Plains (Fig. 7). The differences in the declining rates can be attributed, to some extent, to the uncertainties inherent to groundwater simulation in WGHM. Döll et al. (2014) found that the performance of WGHM in simulating groundwater depletion was directly influenced by model assumptions. For example, groundwater depletion was simulated best assuming irrigation at 70% of optimal water requirement. Moreover, Döll et al. (2012) suggested that WGHM possibly overestimated groundwater withdrawals in the High Plains aquifer, which was consistent with the negative RE as illustrated by Fig. 7c.*

1. LN419-…: ‘section 4.3.2: Uncertainty analysis’ - recall the metric used to quantify the uncertainty here. Is it GTCH, similar to how uncertainties in PSM are quantified?

**Reply**: Thanks for the comment. The metric used to quantify the uncertainty here is derived from GTCH, which is similar to that of PSM. The necessary description of this metric has been supplemented in the revised manuscript.

Line 476 in revised manuscript with changes marked:

*The global-averaged relative uncertainty based on GTCH ranked as: ...*

1. LN440: ‘This pattern was found in Region 5, 9, and 12’ - igGWSA in region 12 does not appear to show a decreasing trend.

**Reply**: Thanks for the comment. igGWSA in Region 12 exhibited a very slight and non-significant decreasing trend (−0.04 cm/yr, as shown in Table S2 in the supplements).

1. LN447: ‘Therefore, absence of glaciers would inevitably’ - not clear …rephrase.

**Reply**: Thanks for the comment. This sentence has been rewritten in the revised manuscript.

Line 502–504 in revised manuscript with changes marked:

*Therefore, failure to account for glacier water storage would inevitably result in inaccurate estimation of GWSA, and thereby lead to misinterpretation of GWS evolution in glacier-covered regions.*

1. LN456-458: Interesting observations. Can you expound a bit on this?

**Reply**: Thanks for the comment. These findings have been elaborated in the revised manuscript.

Line 529–536 in revised manuscript with changes marked:

*Lake water storage was observed to decline remarkably in Aral Sea (−12.60 cm/yr, p<0.01), Caspian Sea (−5.58 cm/yr, p<0.01), Malawi (−3.92 cm/yr, p<0.01), and Baikal (−1.38 cm/yr, p<0.01). When estimating GWSA in these lakes without considering changes in lake water storage, these strong signals of lake shrinkage would then be attributed to groundwater changes, thereby reversing the actual increasing trend in igGWSA (ranging from 0.68 cm/yr to 11.46 cm/yr) into the decreasing trend in GWSAsimplified (ranging from −4.82 cm/yr to −0.70 cm/yr) and indicating unrealistic risks of groundwater depletion.*

1. Chapter 4: What about permafrost? The authors seem to have left out presenting results on effects of ignoring permafrost when estimating the GWSA.

**Reply**: Thanks for the comment. The impacts of ignoring permafrost on GWSA estimation have been supplemented in the revised manuscript. To be specific, the estimation method of GWSA ignoring permafrost (GWSAnoPER) has been added to Section 3.3.1. The permafrost zonation scheme has been added to Section 3.3.3. The quantitative analysis results have been added to Section 4.4.2.

Line 324–326 in revised manuscript with changes marked:

*...* *GWSAnoPER: PWSA was ignored when estimating GWSA, with all other non-groundwater components identical to those for igGWSA. The potential impacts of neglecting permafrost on the accuracy of GWSA estimation can be quantified by comparing GWSAnoPER and igGWSA.*

Line 375–377 in revised manuscript with changes marked:

*For permafrost zones, four high-latitude and high-altitude regions were involved: (1) North America; (2) Europe; (3) North Asia; (4) Tibetan Plateau. Intercomparison between igGWSA and GWSAnoPER was conducted in permafrost zones.*

Line 508–521 in revised manuscript with changes marked:

*4.4.2. Permafrost zones*

*Significant decrease in permafrost water storage was found in all the four permafrost zones, resulting in great discrepancies between GWSAnoPER and igGWSA in these regions (Fig. 10, Table S3). In Europe and the Tibetan Plateau, both GWSAnoPER and igGWSA exhibited an upward trend. However, the increase in GWSAnoPER was weaker as it was confounded by the permafrost change signal. In North Asia, permafrost mass loss even reversed the potential increasing trend in igGWSA into a decreasing trend in GWSAnoPER. In North America, both GWSAnoPER and igGWSA showed a downward trend, but GWSAnoPER declined more markedly given the additional contribution of changes in permafrost.*

*Regarding to global permafrost zone, igGWSA increased significantly (0.08 cm/yr, p<0.05) while GWSAnoPER decreased significantly (−0.36 cm/yr, p<0.01), showing a pattern similar to that in North Asia. GWSAnoPER interpreted global permafrost degradation (−0.44 cm/yr, p<0.01) as part of changes in GWS, which concealed the actual GWS mass gain across the global permafrost zone.*

1. LN522: ‘Mann-Whitney U test…’ - Reference needed or provide a bit more details in the supplements.

**Reply**: Thanks for the comment. The relevant reference has been supplemented in the revised manuscript.

Line 600–602 in revised manuscript with changes marked:

*Mann-Whitney U test (Mann and Whitney, 1947) was adopted to further identify the differences between igGWSA and GWSACLSM.*