Item-by-item responses to all review comments

<u>NOTE</u>: To facilitate the evaluation of our responses, original comments are listed first in their originals (in black), followed by our itemized responses (in red). An annotated version of revised manuscripts is attached.

Referee's Comments

The manuscript entitled "Role of Forest Stand Structure in Groundwater Storage Decline in the Three-North Shelterbelt Forest Region, China" by Luo et al. investigates the impacts of vegetation restoration and forest stand structure on groundwater storage variation in the TNSFR. The manuscript includes extensive datasets and large-scale analyses. The topic is highly relevant for understanding the ecological consequences of large-scale revegetation programs.

I appreciate the effort invested in data collection, processing, and visualization. It seems that there are gaps between figures and the accompanying text. Although the figures contain a considerable amount of valuable information, the Results section does not fully describe or synthesize the key patterns shown in the figures, which prevents the reader from fully appreciating the contribution of the study.

Response: First of all, we would like to thank you for your constructive comments to improve our manuscript. We really appreciate your precious time and great efforts on reviewing the manuscript. We have carefully addressed all the comments, which have greatly improved the quality of this manuscript. Please find below our detailed responses to the comments point by point.

Specific comments:

1. This manuscript stated three objectives:

   (1)*To examine ground water storage (GWS) changes driven by vegetation restoration and climatic factors across the TNSFR from 2003 to 2023;*

   (2)To assess the impact of vegetation restoration on regional GWS while accounting for stand age and dominant species;

   (3)To quantify the direct and indirect impacts of climate change, vegetation restoration, and other anthropogenic activities on GWSA (GWS anomalies) using SEM.

   However, the current ms only partially addresses these objectives.

   a) Objective 1: Sections 3.1 and 3.2 mainly describe the temporal variations in GWS, climate variables, and vegetation indices. However, the analyses primarily remain descriptive. Groundwater, climate, and vegetation changes are largely presented independently, making it difficult to evaluate their respective contributions to GWS variations in the TNSFR.

   Response: We thank the reviewer for this constructive comment. We have reorganized the Results section and strengthened the quantitative attribution analysis. The revised Results now first present the spatiotemporal patterns of GWSA, then examine the effects of climate variability and vegetation restoration, followed by a structural equation modelling analysis to quantify the direct and indirect pathways through which hydroclimatic, vegetation, and anthropogenic factors affect GWSA. Additionally, we added a forest stand-structure analysis based on the 9th NFI data and used relative importance analysis to quantify the roles of dominant species, forest type, stand age, and their interactions in explaining spatial variations in GWSA trends.

   These revisions provide a more integrated framework for assessing the relative roles of climate variability, vegetation restoration, human activities, and forest stand structure in groundwater storage changes, rather than presenting the different variables as separate descriptive patterns. Please see the revised Sections 3.1-3.4.

   b) Objective 2: Although several interesting findings are reported in section 3.3, it remains unclear how stand age, forest type, and species jointly regulate groundwater dynamics, or which of these factors plays the dominant role. Additionally, the stated objective is to assess their impacts on regional GWS; however, the implications of these stand-level differences for the regional groundwater decline across the TNSFR are not clearly established.

Consequently, the connection between the stand-scale analyses and the regional conclusions remains relatively weak.

Response: Based on the comments, we have conducted a *Lindeman–Merenda–Gold*-based relative importance analysis to quantify the contributions of stand age, forest type, and dominant species to GWSA trends. The results show that dominant species was the primary factor explaining GWSA variations, contributing 33.22% of the explained variance, while forest type and stand age had relatively minor contributions. Crucially, the analysis revealed a significant interaction between stand age and dominant species (explaining an additional 16.22% of the variance), indicating that stand age regulates groundwater dynamics in a species-specific manner rather than following a uniform pattern across all forests.

Furthermore, to strengthen the connection between these stand-scale findings and regional implications, we added a dedicated discussion paragraph in Section Discussion. Specifically, we highlight that the dominant tree species extracted from the 9th NFI sample plots (poplar, pine, spruce, larch, and cypress) coincide with the most common and representative afforested tree types across the TNSFR. Thus, the species-specific mechanisms identified at the plot level provide key mechanistic insights for regional groundwater management, effectively connecting local stand attributes to regional-scale groundwater responses. Please see Lines 223-241 for the methodology, Lines 392-402 for the updated results, and Lines 505-514 for the discussion.

Lines 223-241: "*To quantify the relative importance of forest stand-structure variables in explaining GWSA trends, we conducted a Lindeman – Merenda – Gold-based relative importance analysis within a multiple linear regression framework. GWSA trend was used as the response variable, while stand age, forest type, and dominant species were used as explanatory variables. The relative importance of each explanatory variable was assessed using the Lindeman–Merenda–Gold method, which decomposes the model* R^2 *by averaging the sequential sums of squares over all possible orderings of regressors (Fu et al., 2025; Jiao et al., 2021). This approach reduces the dependence of variance attribution on the order in which predictors enter the model and is therefore suitable for comparing the explanatory contributions of correlated predictors. The analysis was performed using the relaimpo package in R (Groemping, 2007). The dominant explanatory factor was identified as the variable contributing the largest proportion of the variance in GWSA trends.*"

Lines 392-402: "*To further quantify the relative importance of stand age, forest type, and dominant species, we conducted a Lindeman–Merenda–Gold-based relative importance analysis. The main-effect model explained 37.02 % of the variance in GWSA trends. Among the three predictors, dominant species contributed the largest share of variance (33.22 %), far exceeding the contributions of forest type and stand* *age. Forest type and stand age explained only 2.73 % and 1.07 % of the variance. These results indicate that species identity was the dominant forest stand-structure variable associated with GWSA trends, whereas broad forest type and stand age alone had much weaker explanatory contributions. When age-related interaction terms were included, the model explanatory power increased to 43.05 %, corresponding to an additional 6.03 % of explained variance. In this interaction model, dominant species remained the largest contributor, explaining 22.22 % of the variance, while the interaction between stand age and dominant species accounted for 16.22 %. These results indicate that the relationship between stand age and GWSA trends differed substantially among dominant species, rather than following a consistent pattern across all forest stands.*"

Lines 505-514: "*Furthermore, the dominant tree species extracted from the 9th NFI sample plots used in this study—specifically poplar (Populus spp.), pine (Pinus spp.), spruce (Picea spp.), larch (Larix spp.), and cypress (Cupressus spp. / Platycladus spp.)—coincide with the common and widely distributed afforested tree species across the TNSFR (Zhai et al., 2023; Zhu et al., 2017). Because these species accurately represent the primary silvicultural tree types in the region, the stand-scale empirical relationships and species-specific mechanisms unraveled here provide valuable mechanistic insights for understanding regional forest-water dynamics. Consequently, these findings establish a meaningful link between local stand attributes and regional groundwater responses, offering targeted guidance for sustainable groundwater management and species selection across the TNSFR. As more fine-grained plot-level data become available in the future, further field observations and higher-resolution modeling will help deepen our understanding of these stand-scale forest-water interactions across diverse regional micro-environments.* "

c) Objective 3: Although total effects are reported, the indirect pathways among these drivers are only briefly mentioned, limiting the mechanistic insight that SEM is intended to provide. More importantly, the forest stand characteristics, extensively analyzed in Section 3.3 (e.g., forest type, stand age, and species composition), are not incorporated in SEM (please correct me if I am wrong). As a result, the analyses of forest structure remain largely disconnected from the driver attribution analysis, and the relationships among the three stated objectives are not fully integrated.

Response: Based on the comments, we added explanations regarding the indirect pathways among these drivers. Please see Lines 332-350: "*To quantify the impacts of climatic, vegetation, and anthropogenic drivers on GWSA dynamics, a Structural Equation*

Model (SEM) was constructed (Fig. 10a). The SEM fit statistics indicated high model performance and reliability, with CFI = 0.973, TLI = 0.926, RMSEA = 0.065, and SRMR = 0.025. Climatic factors exhibited the largest total standardized effect on GWSA in the TNSFR. Among them, Temperature exhibited the strongest total negative effect (−57.6%), which was primarily attributable to its direct effect (−56.4%). In addition, temperature indirectly affected GWS through its influence on irrigation and vegetation dynamics. Precipitation showed a positive direct effect on GWS but also exerted a negative indirect effect by promoting vegetation growth and associated water consumption, resulting in only a weak total effect. These results indicate that climate influences groundwater storage not only through direct hydroclimatic forcing but also through indirect pathways involving vegetation and agricultural water use.

Among anthropogenic factors, NDVI emerged as the primary driver of GWSA change, with a total effect of −20.9%. NDVI affected GWSA both directly and indirectly through ET. Irrigation also exerted a strong negative effect on GWSA, with comparable direct (−8.3%) and indirect (−7.0%) effects. It directly reduced groundwater storage through groundwater abstraction and indirectly affected GWSA through its positive effects on NDVI and ET. Population had no direct pathway to GWSA but influenced groundwater storage indirectly through irrigation. GDP affected GWSA through both direct and indirect pathways, with its indirect effects operating through NDVI and coal mining. Coal mining, in contrast, primarily exerted a direct negative effect on GWSA.

Overall, the SEM results demonstrate that climate warming exerts the primary control on GWSA decline in the TNSFR, while vegetation greening and agricultural irrigation serve as the main non-climatic contributors through both direct and indirect pathways."

Moreover, the SEM is specifically designed to analyze time-series dynamic causal pathways of hydroclimatic variables, vegetation growth, and human activities on interannual GWSA variations. In contrast, forest stand characteristics (e.g., dominant species, forest type, and stand age from the 9th NFI) represent static spatial attributes in our dataset and do not vary continuously over the annual time steps used in the SEM.

To integrate these stand-scale factors into the broader attribution framework, we addressed them in Section 3.4 using Lindeman–Merenda–Gold-based relative importance analysis and multiple linear regression to explain the spatial variations in GWSA trends. Together, the SEM and stand-scale analyses provide a complementary, two-step framework: the SEM captures the temporal/dynamic attribution, while the stand-structure analysis explains the spatial heterogeneity in groundwater trends. Please see Lines 392-402 for the updated results:

"*To further quantify the relative importance of stand age, forest type, and dominant species, we conducted a Lindeman–Merenda–Gold-based relative importance analysis. The main-effect model explained 37.02 % of the variance in GWSA trends. Among the three predictors, dominant species contributed the largest share of variance (33.22 %), far exceeding the contributions of forest type and stand age. Forest type and stand age explained only 2.73 % and 1.07 % of the variance. These results indicate that species identity was the dominant forest stand-structure variable associated with GWSA trends, whereas broad forest type and stand age alone had much weaker explanatory contributions. When age-related interaction terms were included, the model explanatory power increased to 43.05 %, corresponding to an additional 6.03 % of explained variance. In this interaction model, dominant species remained the largest contributor, explaining 22.22 % of the variance, while the interaction between stand age and dominant species accounted for 16.22 %. These results indicate that the relationship between stand age and GWSA trends differed substantially among dominant species, rather than following a consistent pattern across all forest stands.*"

2. A considerable portion of Section 3.1 focuses on validating GWSA and SMSA estimates rather than describing groundwater dynamics themselves. These analyses resemble model or data validation and may be more appropriately presented in the Methods section or supplementary materials.

   Response: Based on the comments, we revised Section 3.1 to improve the organization of the Results. In the revised manuscript, Section 3.1 mainly focuses on the spatiotemporal dynamics of GWSA in the TNSFR. Meanwhile, the detailed results of the SMSA product comparison and validation have been moved to the Supplementary Information. Please see Supplementary Lines 67-73: "*The results indicated that although both ERA5 and GLDAS could capture the spatiotemporal dynamics of soil water, GLDAS showed higher consistency with ESA CCI. At the temporal scale, the correlation coefficient between GLDAS and ESA CCI was 0.834, which was 0.13 higher than that of ERA5, accompanied by a lower RMSE (0.012 m^3 m^{-3}). At the spatial scale (Supplementary Fig. S4), the correlation coefficient between ERA5 and ESA CCI was 0.530, whereas that between GLDAS and ESA CCI reached 0.554. Spatial comparisons further demonstrated that GLDAS achieved higher consistency and smaller errors across most of the study area.*"

3. Several interpretations, explanations, and quotations are included within the Results section. In my opinion, the Results section should primarily report observations, statistical relationships, and quantitative findings, whereas mechanistic explanations and broader implications should be reserved for Discussion.

   Response: Based on the comments, we have thoroughly revised the Results section

(Sections 3.1–3.4) by removing mechanistic explanations, broader implications, and literature citations, and relocating them to the Discussion section. . These revisions make the Results section more concise and focused on the main quantitative findings. Please see the revised Sections 3.1–3.4 and Section 4 (Discussion).

4. In several sections, conclusions or interpretations are introduced before the corresponding observations and analyses are presented. As a result, the Results section sometimes reads as a sequence of predetermined conclusions followed by supporting evidence, rather than a progression from observations to interpretation. For example, Line 274: "The results revealed that pure forests exhibited more pronounced groundwater depletion than mixed forests. Specifically, conifer pure stands (CP) showed the most negative GWSA trends, reaching a peak of -1.06 mm month^{-1} at the mature stage, which was significantly higher than that of conifer mixed stands(CM, -0.31 mm month^{-1}) and broadleaf pure stands (BP, -0.67 mm month^{-1}) at the same age class." A clear progression from evidence to interpretation would improve the scientific logic and readability of the manuscript.

   Response: Based on the comments, we revised the Results section to follow a clearer evidence-to-interpretation structure. We also removed or weakened several interpretive statements from the Results section to avoid presenting predetermined conclusions before the supporting evidence. These revisions improve the logical flow and readability of the Results section. Please see Lines 357-361: "*To further investigate how forest structure type and stand age affect GWS, we analyzed GWSA trends across different forest structure categories using the Ninth NFI data (Fig. 11). At the mature stage, conifer pure stands showed the most negative GWSA trends, reaching -1.06 mm month-1, compared with -0.31 mm month-1 for conifer mixed stands and -0.67 mm month-1 for broadleaf pure stands. Across comparable age classes, pure forests generally showed more negative GWSA trends than mixed forests.*"

5. Many figures that appear essential for supporting the main conclusions are placed in the Supplementary Materials despite being repeatedly cited in the main text. If these figures are critical to the interpretation, they should be moved to the main manuscript. Additionally, figures are not positioned close to the corresponding text, which reduces the readability.

   Response: Based on the comments, we moved the figures that are critical for supporting the main conclusions (e.g., Fig. S3 and S5) from the Supplementary Materials to the main text and renamed them as Fig. 4 and Fig. 5. Additionally, we have re-arranged the layout of the revised manuscript to ensure that all figures are placed in close proximity to their corresponding text descriptions, thereby improving the overall readability.

6. The manuscript contains numerous abbreviations, long paragraphs, and limited

visual separation between different analyses (paragraphs).

Response: Based on the comments, we have made the following revisions:

We thoroughly checked the use of abbreviations throughout the manuscript, ensuring that all key terms are defined at their first occurrence and minimizing redundant acronyms.

We restructured and shortened several overly long paragraphs, ensuring that each paragraph focuses strictly on a single topic or analytical step.

We thoroughly reorganized the content and updated the Results section to improve the narrative flow and logical separation among different analyses.

Please see the revised Sections 3 and 4.

7. I am curious about Figure 3b which showed a totally reversed temporal trend between observed and estimated GWSA, could you explain a little bit about this phenomenon?

   Response: Thank you for this important comment. After rechecking our processing workflow, we identified that this artifact was caused by an improper sequence in data normalization and spatial averaging. Specifically, we mistakenly normalized the station-level time series before spatial aggregation, rather than calculating the regional average first and then applying normalization. Because different monitoring stations possessed varying baselines and variances (especially with missing values), normalizing individual stations prior to averaging introduced significant distortion, which artifactually inverted the temporal trend in Figure 3b.

   We have corrected this data processing sequence by first calculating the regional mean of observed GWSA and then normalizing the aggregated series. Furthermore, minor residual discrepancies between observed and estimated GWSA can be partly attributed to scale mismatches, as well-scale point measurements inherently differ from satellite-derived regional grid estimates. Nevertheless, the corrected results now show highly consistent long-term trends between the two datasets.