1. Regarding the intensification of climate and its link to surface and groundwater resources that the Authors mention (e.g., "Intensifying climate change and human activities are substantially altering frozen ground conditions, disrupting both surface water regimes and groundwater connectivity."), please discuss how strong is the intensification and which water source is the strongest; for example, based on the study by Koutosyiannis (2020; doi:10.5194/hess-24-3899-2020), the impact of the over-exploitation of groundwater resources is identified as one of the most severe anthropogenic actions with direct effect in the sea-level rise.

Response: We thank the reviewer for highlighting the need to evaluate the magnitude of these impacts and for recommending the highly relevant study by Koutsoyiannis (2020). We agree that our manuscript should explicitly address "how strong the intensification is" and "which water source is the strongest impacted." We will incorporated this evaluation into the Discussion section (Section 4.1).

To directly answer your questions, we first expanded our discussion at the broader hydrological scale. Drawing on Koutsoyiannis (2020; doi:10.1038/s41467-020-16757-w), alongside Taylor et al. (2013; doi:10.1038/nclimate1744) and Rodell et al. (2018; doi:10.1038/s41586-018-0123-1), we explicitly state that groundwater is the water source most severely impacted by anthropogenic intensification. Human extraction often far exceeds climate-driven natural recharge variations, resulting in an over-exploitation severe enough to shift the global water mass balance and contribute to sea-level rise. We use this global baseline to contextualize the severe groundwater depletion observed in our study area. Conversely, our statistical analysis shows that seasonal surface water (SSW) remains overwhelmingly driven by natural climate variability, responding highly to extreme meteorological events (e.g., floods).

Unlike SSW, the dynamics of permanent surface water (PSW) are governed by a complex mix of both human and natural forces. This complexity is exactly why our quantitative attribution model (Geographical-XGBoost) was specifically applied to PSW. Based these attribution results, we further clarify which specific PSW are most strongly impacted by natural or human factors. Our findings show that human activities control the dynamics of permanent river systems (71.0%) and reservoirs (56.0%), whereas natural climate factors remain the primary driver for permanent lakes (93.6%).

By structuring the discussion this way in Section 4.1, we comprehensively clarify the distinct magnitudes of climate and anthropogenic intensifications across groundwater, seasonal surface water, and specific types of permanent surface water.

2. The Authors mention that "Across distinct water types, natural factors control 93.6% of lake dynamics, whereas human activities dominate river systems (71.0%) and reservoirs (56.0%)."; please further explain how did the Authors estimate the 93.6% of natural factors impact and the 6.6% for the human activities by also considering the uncertainty error in the maps themselves (please calculate this error and include it in the estimations with surplus values).

Response: We appreciate the reviewer for this insightful comment. We agree that reporting point estimates (e.g., 93.6%) without uncertainty bounds overlooks the inherent classification errors. To address your request, we detail the mathematical estimation of these percentages and explain how we incorporated map uncertainty into the final results.

(1) How the percentages were estimated

These values were calculated by aggregating the local SHAP (SHapley Additive exPlanations) values generated by the Geographical-XGBoost model. Specifically, for a given water type, we calculated the global importance (I_j) of each driving variable j by averaging its absolute SHAP values across all N spatial units:

$$I_j = \frac{1}{N}\sum_{i=1}^{N}\left|SHAP_{i,j}\right|$$

Then, the percentage contribution for natural factors (P_{NF}) and human activity factors (P_{HAF}) were calculated as the ratio of their summed absolute SHAP values to the total absolute SHAP values of all input variables:

$$P_{NF} = \frac{\sum_{j\in NF} I_j}{\sum_{j\in All} I_j} \times 100\%$$

$$P_{HAF} = \frac{\sum_{j\in HAF} I_j}{\sum_{j\in All} I_j} \times 100\%$$

We will add these exact formulations to Section 2.3.5 of the revised manuscript to ensure methodological transparency.

(2) Considering the uncertainty error in the maps

Regarding map uncertainty, the Geographical-XGBoost framework (Grekousis, 2025; 10.1007/s10109-025-00465-4) is a deterministic model and lacks a native error propagation module. However, we agree that classification errors inevitably propagate into the final attribution results. To quantify this without altering the algorithm's core architecture, we applied a Monte Carlo-based sensitivity analysis.

Rather than relying on Overall Accuracy, which is dominated by the non-water background, we used the standard error derived from the area estimation adjustment method (Olofsson et al., 2014; 10.1016/j.rse.2014.02.015). Based on our validation, the relative standard error (RSE) for the estimated surface water area is approximately 20.3%. Strictly speaking, this 20.3% error applies to the cross-sectional area estimation. Applying it directly to the time-series trend slope of Permanent Surface Water (PSW) represents a highly conservative (worst-case) assumption, as time-series trend fitting inherently smooths out random year-to-year classification noise. Nevertheless, we deliberately used this margin as the perturbation scale to rigorously test our model under extreme noise.

Strictly speaking, this 20.3% error applies to the cross-sectional area estimation. Applying it directly to the time-series trend slope of Permanent Surface Water (PSW) represents a highly conservative (worst-case) assumption. Time-series trend fitting (e.g., Theil-Sen slope) inherently smooths out random year-to-year classification noise, and permanent water bodies typically exhibit much higher classification accuracy than the overall water extent. Nevertheless, we deliberately chose this 20.3% error margin as the perturbation scale to thoroughly test the stability of our model under extreme noise.

We introduced random normal perturbations (with a standard deviation equal to 20.3% of the observed trend value) to our target variable across all spatial grids. Regarding the implementation, Geographical-XGBoost is exceptionally computationally intensive because it constructs an independent sub-model for every single spatial unit (over 11,000 units in our study). Given the substantial computational overhead and time costs, we will performed 20 independent Monte Carlo iterations to balance statistical reliability with computational feasibility. However, should the reviewer or editor deem it necessary, we are fully prepared to increase the number of iterations in the final revision.

By averaging the outputs of these iterations, we will derived the final attribution percentages with explicit uncertainty bounds. For example, the natural factors' control over lake dynamics is now reported as 93.6%±[X.X]%, and human activities dominate river systems at 71.0%±[X.X]%. They will directly answer your request to include the uncertainty error in our estimations.

3. Please show Figures 4 and 5 along with rainfall and runoff satellite maps, taken for example from products such as IMERG (NASA, 2020; https://catalogue.ceda.ac.uk/uuid/47c32530265d4d6e8fdb6c08b2330371) and ERA5 (Hersbach et al., 2020; doi.org/10.1002/qj.3803), to justify the changes in surface water area and availability. Also, additional data analysis may be required to justify the very large peaks in these Figures, what is this sudden break in Figures 4's timseries, please further explain what are the shaded areas in their curves and how are they calculated.

Response: We thank the reviewer for this constructive suggestion. While we had provided a time-series plot of mean annual precipitation in the original Supplementary Information (Figure S3), we agree that providing precipitation and runoff data is essential to justify the observed temporal changes and extreme peaks in surface water area and availability.

Regarding the data selection, rather than introducing the IMERG dataset (which only begins in mid-2000 and misses the historical 1998 flood), we chose to maintain internal consistency by using the high-resolution (1-km) precipitation dataset already employed in our attribution model (Table 1). For runoff, we extracted Surface Runoff (SR) from the ERA5-Land dataset, as SR is the direct hydrological driver for rapid surface water expansion.

To address your specific inquiries regarding the temporal features in Figures 4 and 5, we provide the following clarifications:

(1) The "very large peaks": These reflect the regional hydrological response to extreme events, most notably the well-documented 1998 Songhua and Nen River severe flood. To visually justify this, we will add spatial anomaly maps of precipitation and SR for these extreme years. These anomaly maps will directly provide clear evidence for the spikes in surface water area. We will also add a brief historical context for this peak in Section 3.2.

(2) The "sudden break": The shift around 2012 represents a structural regime transition in the regional hydrology, which is closely tied to changes in the spatial distribution of precipitation. As detailed in our response to Reviewer #1, we have validated this breakpoint using rigorous statistical tests (both a permutation-based CUSUM test and the Pettitt test, $p < 0.01$), confirming it is a statistically valid transition from shrinkage to expansion.

(3) The shaded areas: The shaded areas around the curves represent the 95% confidence intervals of the linear regression lines. We will explicitly state this calculation in the figure captions to avoid any confusion.

4. A length of at least 30 years is required to capture the long-term climatic variability of the key hydrological-cycle dynamics (Dimitriadis et al., 2021; doi:10.3390/hydrology8020059); the Authors may use this info to justify the adequate length of the 37 years of data-maps they apply (i.e., 1988 to 2024) at the SHRZ. Additionally, please enhance the information on the data analysis by including Tables regarding the spatiotemporal timeseries used in the analysis with primary information (e.g., resolution, exact time-range, missing maps, etc.), spatiotemporal primary statistics (e.g., total mean, standard-deviation, skewness, kurtosis, etc.), seasonal statistics (e.g., primary statistics for each season), autocorrelation structures, etc; this could help the Authors identify several similarities among the data statistics and perform a pooled-analysis that can reveal additional interesting conclusions.

Response: We thank the reviewer for pointing us to Dimitriadis et al. (2021), which provides excellent theoretical backing for our study design. We will cite this work in Section 2.3 to explicitly justify that our 37-year observation window (1988-2024) is adequate to capture long-term hydroclimatic variability.

We also agree with the suggestion to provide more descriptive statistics. In the revised Supplementary Information, we will include a new table summarizing the time series used in our analysis. Specifically, this table will report:

(1) Primary information: Spatial resolution, exact time range.

(2) Primary statistics: Mean, standard deviation, skewness, and kurtosis for surface water area (TSW, PSW, SSW) and main hydroclimatic drivers (e.g., precipitation, runoff).

(3) Autocorrelation: The lag-1 autocorrelation coefficient to assess interannual memory effects.

All hydroclimatic and surface water time series are temporally continuous across their respective periods. Regarding seasonal statistics, we should note that the SHRZ is located at the margin of the Eurasian permafrost, meaning surface water is completely frozen and snow-covered during winter. To avoid spectral confusion, our remote sensing extraction is restricted to the ice-free window (April to October). Consequently, the calculated statistics will reflect this specific observing period rather than the four standard meteorological seasons.

Adding these descriptive statistics will make the data properties more transparent and help us identify similarities across the time series, as the reviewer suggested.

5. Please consider simplifying each conclusion since they seem very difficult to understand, and specify what is their practical meaning, where they can be used and under what conditions, etc.:

Response: We agree with the suggestion. In the revised manuscript, we will rewrite the Conclusion section in plain, direct language, explicitly emphasizing the practical implications of our findings for water resource management.

(a) "The IOWDM-ENC demonstrated more superior applicability than existing methods in the SHRZ, achieving an overall accuracy of 99.03%."; what is a practical application for this.

(a) **Response:** We will clarify that the practical application of IOWDM-ENC is to provide local water management agencies with a reliable tool for continuous, near-real-time monitoring of water bodies in heavily agricultural regions, without relying on delayed global datasets.

(b) "The Geographical-XGBoost-SHAP framework effectively resolved the stationarity limitations of global-scale models. Rather than assuming a uniform influence, this spatially explicit GeoAI framework quantitatively identifying specific locations where factors function as promoters or inhibitors. This capability provides a reliable technical basis for revealing localized hydrological controls that vary significantly across this complex permafrost margin."; please note that data cannot be stationary or non-stationary but rather a model can be either stationary or non-stationary, and one can select both models and then decide which one best describes the observed series (see for example, discussion in Serinaldi and Kilsby, 2015; doi:10.1016/j.advwatres.2014.12.013).

(b) **Response:** We sincerely thank the referee for pointing out this critical epistemological distinction regarding "stationarity." The referee is correct that data itself is not stationary or non-stationary; rather, these are properties of the mathematical models we choose to describe the data. In the revised text, we will remove the phrase "spatial non-stationarity of hydrological drivers." Instead, we will state that the Geographical-XGBoost model captures "spatially varying relationships" (spatial heterogeneity). We will also cite Serinaldi and Kilsby (2015) in the methodology to acknowledge the distinction between data dynamics and model assumptions.

(c) "Surface water dynamics are governed by distinct regulation mechanisms. At the regional scale, a divergence exists between the antagonistic pattern in the permafrost region (where permanent water expansion is constrained by environmental barriers) and the synergistic pattern in the seasonal frozen ground region (where climate, terrain, and human factors amplify expansion). At the typological scale, permanent lake dynamics are primarily controlled by natural factors (55.7% CEF and 37.9% TSF). In contrast, anthropogenic factors dominate both RCF (71.0%) and RP (56.0%), while exerting opposing influences by constraining RCF expansion but promoting RP expansion."; this is quite confusing (please simplify), what do the Authors mean by "distinct regulation mechanisms (are they not natural ones, and why they are distinct).

(c) **Response:** We will simplify this text. The practical meaning is that surface water in the permafrost region is primarily expanding because humans are building reservoirs, while the natural terrain actually limits expansion. Conversely, in the seasonal frozen region, water is expanding naturally because flat terrain and increased precipitation act together. We will rewrite this conclusion to avoid confusing expression.

(d) "The hydrological trade-off emerged between surface water expansion and groundwater storage depletion. The observed increase in surface water area coincides with a decline in GWSA, suggesting that surface water expansion has occurred at the potential expense of subsurface storage. This trade-off is particularly significant in the seasonal frozen ground region, highlighting the urgent need for surface-groundwater management strategies tailored to the distinct regulation mechanisms identified in this study."; this can be quite confusing since I would expect when the groundwater resources increase, and after a threshold point, the surface ones would also start to increase (please further explain the practical meaning of this).

(d) **Response:** The referee raises a very logical point: under natural conditions, surface and groundwater should increase or decrease together. The practical meaning of our "trade-off" conclusion is exactly that this natural linkage has been broken by human intervention. Surface water is expanding because humans are retaining water in reservoirs and continuously pumping groundwater for agricultural irrigation, preventing natural recharge. Therefore, the "trade-off" reflects an artificial, unsustainable anthropogenic diversion of water from the subsurface to the surface. We will rewrite this conclusion to state this physical reality clearly.

6. Please further explain how Equations 1-4 are related to the water surface area and availability estimations extracted from the maps; please show the equations for the whole methodology followed in this study so the Readers can better comprehend it.

Response: We will clarify this in Section 2.3.1. Equations 1-4 are used to calculate the pixel-wise spectral index values. These values are then fed into our logical threshold conditions (e.g., MNDWI > EVI) to generate binary water maps (1 for water, 0 for non-water). The final surface water area is calculated by multiplying the total count of identified water pixels by the area of a single Landsat pixel. We will add the explicit mathematical formulation for this area aggregation step to ensure the readers fully comprehend the entire workflow.

7. Please perform a strong polishing of the text by following the SI standards units and regulations (for example, please do not use italics in the Equations 1-4 when more that 1 letter is used for a symbol).

Response: We thank the referee for catching this formatting error. According to SI and ISO formatting standards, multiletter variables (such as MNDWI, NDVI, EVI) should indeed be set in roman (upright) type, not italics. We will perform a strict review of all equations and text in the revised manuscript to ensure complete compliance with SI standards.