Response to reviewers

To Reviewer 1

This study used machine learning models combining satellite data and geographical variables to reconstruct historical variations in Chemical Oxygen Demand (COD, a proxy for ROM) along the Yellow River's Headstream (YRHS), and to predict future changes under typical Shared Socioeconomic Pathways.

Response: We sincerely thank the reviewer for reading and commenting on this manuscript. We have carefully considered all comments and addressed them point by point in the revised manuscript.

The specific suggestions are as follows:

Lines 13-17: In the abstract background, it is mentioned that riverine organic matter is important and sensitive to climate change, but the specific gap in current scientific understanding is not clearly identified (e.g., lack of long-term observations? lack of high-resolution spatial distribution? lack of predictive capability for future changes?). The "However" sentence reads more like a general issue of data scarcity rather than an unsolved problem directly targeting the "response mechanisms to climate change." It is recommended that the authors address this.

Response: We thank the reviewer for this helpful comment. We agree that the original abstract background did not clearly identify the specific knowledge gap. We have therefore revised the abstract to explicitly frame the unresolved issue as the response of riverine organic matter (ROM) to climate change in the headwaters of the Qinghai–Tibetan Plateau (QTP). The added sentence reads:

“However, how ROM responds to climate change in the headwaters of the Qinghai–Tibetan Plateau (QTP) remains poorly understood, largely owing to a lack of long-term, high-spatial-resolution in situ measurements.”

This revision directly targets the response-mechanism gap rather than presenting data scarcity as the gap itself; the lack of long-term, high-resolution observations is now described as a key constraint on understanding that mechanism. We believe this clarifies the scientific gap raised by the reviewer.

Lines 13-17: It is recommended that the authors briefly include the research implications at the end of the abstract.

Response: We thank the reviewer for this valuable suggestion. We have added a brief statement of the research implications at the end of the abstract, as follows:

“Although our study provides a comprehensive and scalable understanding of how climate change is reshaping riverine organic matter dynamics, substantial uncertainties remain regarding future changes in riverine organic matter (ROM), owing to the complex interactions among precipitation, warming, and their combined effects on carbon cycling in alpine catchments.”

Lines 35-52: The first paragraph of the introduction uses a lengthy description to present the research background. However, the core of the paper should focus on organic matter in high-altitude headwater rivers. It is recommended that the authors condense the background information.

Response: We thank the reviewer for this constructive comment. We agree that the first paragraph

of the Introduction contained overly lengthy background information that was not sufficiently focused on the central topic of the paper. We have therefore removed several redundant or less relevant sentences and condensed the background discussion. The revised paragraph now more concisely introduces the research context and places greater emphasis on organic matter dynamics in high-altitude headwater rivers, thereby better aligning the Introduction with the core focus of the study.

Line 76: The introduction states that the source region of the Yellow River is located in the eastern Qinghai-Tibet Plateau. Why was the Yellow River source selected as the study area, and what makes this region unique compared with other basins on the Qinghai-Tibet Plateau (e.g., the source region of the Yangtze River or the Lancang River)? It is recommended that the authors add one sentence outlining the environmental uniqueness of the Yellow River source region.

Response: We thank the reviewer for this valuable comment. We agree that the rationale for selecting the Yellow River source region and its environmental uniqueness relative to other basins on the Qinghai–Tibetan Plateau should be clarified. Accordingly, we have added the following sentence to the revised manuscript:

“The YRHS lies in a transition zone between permafrost and seasonally frozen ground, where climate warming has driven substantial permafrost degradation and triggered a cascade of environmental and ecological issues (Yang et al., 2023).”

This addition emphasizes the region’s distinctive cryospheric and hydrological setting, which makes it particularly sensitive to climate change and helps distinguish it from other basins on the Qinghai–Tibetan Plateau.

Line 88: In the last paragraph of the introduction, the authors state that "according to previous studies, ROM is commonly estimated using Chemical Oxygen Demand (COD) as a proxy," citing two references (Liu et al., 2023; Wang et al., 2026). Do these two studies originate from the Qinghai-Tibet Plateau or similar high-altitude environments? The authors should briefly clarify whether this proxy is applicable in plateau regions.

Response: We thank the reviewer for this valuable comment. We have updated the references to studies that focus on the relationship between COD and dissolved carbon in fresh waters, including rivers. These studies provide direct evidence that COD is closely related to dissolved organic carbon (DOC) in riverine systems. Although these studies are not exclusively from the Qinghai–Tibetan Plateau, the COD–DOC relationship is generally applicable to freshwater environments. Therefore, COD can be used as a proxy for riverine organic matter in the alpine study area. We have added this clarification to the revised manuscript.

Figure 1: Is the division of the upper, middle, and lower reaches of the Yellow River headstream based on any specific criteria? Are there specific place names associated with these divisions? If so, it is recommended that the authors provide this information.

Response: We thank the reviewer for this helpful comment. The division of the upper, middle, and lower reaches of the Yellow River headstream was defined based on three criteria: (1) landscape and geomorphological characteristics, (2) hydrological conditions, and (3) the distribution of national monitoring sites in the study area. In the revised manuscript, we have added the specific place names associated with each segment to Figure 1 and its caption, as follows: the upper reach

extends from source to Deang; the middle reach from Deang to Longwu; and the lower reach from Longwu to Xunhua. This information is now clearly presented in the revised figure, allowing readers to better understand the spatial framework of our study.

Lines 156-158: In Section 2.2.3, it is mentioned that historical and future discharge data were obtained from multiple published studies (Long et al., 2024; L. Liu et al., 2025; M. Ma et al., 2023; Wang, 2024). Are the discharge data from these studies consistent? Do they include discharge projections under different scenarios? The authors should provide additional clarification.

Response: We thank the reviewer for this helpful comment. We have clarified this point in Section 2.2.3. The historical discharge data are all based on in-situ measurements and are therefore consistent in terms of the underlying data source. The reason we compiled historical discharge data from multiple published studies is that these sources provide complementary data. For the future discharge projections, and in response to the reviewer's comment, we have now used data from a single study. These projections include the SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios.

Section 4.1.1: It is stated that "the midstream region has higher soil organic carbon, vegetation cover, fine particle content, and precipitation, and therefore higher COD." This describes a spatial correlation but directly implies a causal relationship. In fact, higher COD may also be influenced by human activities (e.g., urban areas or grazing activities). This should be addressed in the discussion.

Response: We thank the reviewer for this insightful comment. We agree that the original wording described a spatial correlation and could be misread as implying a direct causal relationship. We

have therefore revised the relevant sentence to avoid overstating causality and to acknowledge that COD is influenced by multiple factors, including potential human activities.

At the same time, we would like to clarify that the study area is a high-altitude alpine headwater region of the Qinghai–Tibetan Plateau, characterized by sparse population, limited urbanization, and minimal industrial activity. Therefore, anthropogenic impacts are relatively minor compared with natural controls such as soil organic carbon, vegetation cover, fine particle content, and precipitation. Nevertheless, local human disturbances such as grazing may still exert some influence on organic matter inputs, and we have now acknowledged this possibility in the discussion.

Line 39, Line 48: The authors should arrange references in chronological order when citing them within the text.

Response: Thanks for this comment. We have arranged references in chronological order.

To Reviewer 2

This study used machine learning models combining water quality measurements, satellite reflectance data, and geographical variables to reconstruct historical variations in COD in the source region of the Yellow River and to predict its future changes and exports under three Shared Socioeconomic Pathway (SSP) scenarios. The approach is a valuable response to the scarcity of monitoring data in this region; however, several issues concerning the study design should be addressed before the main conclusions can be considered robust.

Response: We sincerely thank the reviewer for reading and commenting on this manuscript. We have carefully considered all comments and addressed them point by point in the revised manuscript.

The specific comments are as follows:

Major comments:

1. Clarification or correction of the study-area delineation. The source region of the Yellow River is conventionally delineated at Tangnaihai hydrological station (~122,000 km2), whereas this study analyses ~195,000 km2 (Line 102), extending through the Longyangxia and reservoir reach to site M4, and states that this area contributes "nearly 49%" of the Yellow River's total discharge (Line 111, uncited). The authors should explain the basis for this broader delineation, verify the area and runoff-contribution figures against published values, and either restrict "source region" statistics to the conventional delineation or rename the study area; otherwise, the reported means, trends, and exports are not comparable with previous source-region carbon budgets (e.g., L. Li et al., 2025).

Response: We thank the reviewer for this valuable comment. We agree that the previous study-area delineation required clarification. The larger area reported in the original manuscript represented the headwater catchment within the Qinghai–Tibetan Plateau, which extends beyond the conventional source region delineated at Tangnaihai (M2). In the revision, we have addressed this issue as follows.

(1) Study-area definition. We now define the headwater region of the Yellow River as the catchment upstream of Xunhua (M4), located at the eastern edge of the Qinghai–Tibetan Plateau. This definition was adopted to cover the portion of the Yellow River basin within the plateau, including the Longyangxia reservoir reach. The catchment area has been corrected to approximately 133,000 km^2, which is controlled by M4.

(2) Discharge contribution. We have removed the uncited statement that the study area contributes "nearly 49%" of the Yellow River's total discharge.

(3) Discharge data. We have included the mean discharge for Tangnaihai (M2) and clarified that discharge at M2 is used to represent the natural hydrology of our study area as downstream M2 is affected by damming. Additionally, M2 and M4 have comparable discharge over the historical period. For example, the 1987–2016 mean discharge is 184.90×10^8 m^3 at M2 and 180.64×10^8 m^3 at M4. Furthermore, M2 has readily available long-term discharge observations, which is critical for COD_{Mn} calculation and projection.

2. Conversion from runoff to river discharge in the flux calculation. The compiled hydrological products do not directly provide discharge at the sites used in Eq. 4: Long et al. (2024), for example, provide gridded runoff projections at 0.25-degree resolution from an LSTM model

trained on cryosphere-hydrology model output, and gridded runoff must be accumulated over the upstream area to obtain discharge at a river cross-section. The manuscript presents only the flux equation using river discharge and concentration (Eq. 4) after citing these runoff data, without explaining how the published runoff was converted or assigned to sites M1–M2 and C1–C3 (e.g., grid accumulation, routing, or drainage-area scaling); the terms "runoff" (Fig. 9c) and "river discharge" (Line 239) are also used interchangeably. The authors need to provide the specific conversion procedure and discuss the associated uncertainties to ensure the reproducibility of the flux results.

Response: We thank the reviewer for this thorough and constructive comment. In the revision, we have addressed these issues as follows:

(1) Direct use of discharge data. Rather than deriving discharge from gridded runoff products, we now use discharge data obtained directly from published sources for the relevant stations and reaches. Specifically, we collected discharge data from studies that provide gauged or modeled discharge at the sites of interest (M1–M4 and C1), and we have clarified the exact sources in the revised Methods section. This avoids the need for grid accumulation, routing, or drainage-area scaling and ensures that the discharge values used in Eq. 4 are consistent with the spatial locations of the sampling sites.

(2) Recalculated COD fluxes. Using these direct discharge data, we recalculated the COD fluxes throughout the study. Figure 9 has been updated accordingly to present the new flux results. The flux equation (Eq. 4) now explicitly uses river discharge (not runoff) for all calculations.

(3) Terminology. We have replaced “runoff” with “river discharge” throughout the manuscript, including in Figure 9 and the relevant text, to ensure consistent and accurate terminology.

3. Reliability of projections built on different data sources. Because the SSP126 runoff is taken from Ma et al. (2023) while the SSP245 and SSP585 runoff is taken from Long et al. (2024), the differences among scenarios in Fig. 9d (a decline under SSP126 versus increases under SSP245 and SSP585) may reflect differences between the source studies rather than climate, as the caption itself acknowledges (Lines 369–370). The authors should either adopt one internally consistent runoff source for all three scenarios or, at a minimum, quantify and discuss the uncertainty introduced by mixing data sources before interpreting the differences among scenarios.

Response: We thank the reviewer for this important comment. We agree that mixing projected discharge data from different sources could confound the interpretation of inter-scenario differences in Fig. 9d. In the revision, we have therefore adopted a single, internally consistent data source for all future scenarios. Specifically, projected discharge data for SSP126, SSP245, and SSP585 were all obtained from Ma et al. (2023), and the projections from Long et al. (2024) are no longer used. We have updated the Methods, Figure 9d, and its caption accordingly, and we have removed the previous caveat regarding mixed data sources. This ensures that the differences among scenarios reflect scenario-driven climate and hydrological differences rather than inconsistencies between source studies.

Minor comments:

1. Lines 53–54: Ran et al. (2021) quantifies CO_2 evasion from Chinese inland waters, so attributing "61% of total riverine ROM efflux" to that study appears inaccurate. Please verify this citation.

Response: We thank the reviewer for identifying this citation issue. We have corrected the citation attribution and updated the associated references.

2. Lines 122–126 and Figure 3 (Lines 262–271): Sites M1–M4 are included among the 107 training sites, and the model reliability check uses only 2023 data at M2 and M3. Why were matchups from M1 and M4 not included? Were the M2 and M3 matchups excluded from the training set, and if so, on what basis? If they were not excluded, the evaluation is not independent of the training data.

Response: We thank the reviewer for this important comment. We clarify that sites M1–M4 are included among the 107 training sites, and the model performance and reliability for COD prediction on the training set is shown in Fig. 2a. The comparison of in-situ measured and predicted COD values is intended to illustrate the reliability of the trained model, not to serve as an independent validation of this model. The 2023 measured and predicted COD data at M2 and M3 were selected because these sites have more matched observations, allowing a clearer site-level illustration of model performance. M1 and M4 were not included in Fig. 3 because they had too few concurrent 2023 matchups for a meaningful comparison.

3. Lines 128–131 and Eq. 4 (Lines 237–239): The analytical method described yields the permanganate index (COD_Mn), which should be labeled consistently instead of 'COD,' because COD measured by the dichromate method (COD_Cr) oxidizes organic matter much more completely and yields systematically higher values. In addition, the exported quantities are gigagrams of COD_Mn (oxygen equivalents) without conversion to organic matter or carbon, so

direct numerical comparison with carbon budgets requires a justified conversion.

Response: We thank the reviewer for this important clarification. We agree that the analytical method measures the permanganate index (COD_{Mn}), not dichromate COD (COD_{Cr}), and that the two are not interchangeable because COD_{Cr} oxidizes organic matter more completely and yields systematically higher values. We have therefore revised Section 2.2.1 and the rest of the manuscript to label the parameter consistently as COD_{Mn} and to avoid using “COD” as an abbreviation for COD_{Mn}. We also now state explicitly that COD_{Mn} is an oxygen-equivalent index, rather than a direct measure of organic matter or carbon. To permit comparison with carbon budgets, we converted COD_{Mn} fluxes to DOC equivalents using equation developed by previous studies (Fig. S4), and we discuss the associated uncertainty.

4. Lines 144–145, 159–164, and the Fig. 7 caption: Historical climate is taken from an observation-based dataset (Peng, 2025a, b), while future climate comes from a single GCM (BCC-CSM2-MR) with no bias correction described, and the caption's "based on CMIP6 data" is misleading because only one CMIP6 model is used. Since CMIP6 models show substantial temperature and precipitation biases over the Tibetan Plateau, it is recommended that the projections be bias-corrected against the observational baseline and that a multi-model ensemble be used, or at least that the uncertainty introduced by relying on a single model be discussed; the caption should be corrected accordingly.

Response: We thank the reviewer for this important comment. Historical climate is taken from the observation-based dataset of Peng (2025a, b), while future monthly temperature and precipitation data was collected from Hu et al. (2025). Bias correction for future monthly temperature and

precipitation is performed using the Delta Change Method described in Section 2.4 and Text S1 of the Supplementary Material. In this approach, the projected change from each CMIP6 model relative to its own baseline is applied to the observation-based historical baseline, thereby reducing systematic biases in temperature and precipitation over the Tibetan Plateau.

5. Lines 183–184 and 251–252: Since longitude and latitude are the two most important predictors, could these two variables dominate the model and mask the contribution of the spectral information? What would the model performance be if the coordinates were removed from the inputs?

Response: We thank the reviewer for this insightful comment. We agree that longitude and latitude are among the most important predictors and could potentially dominate the model and obscure the contribution of spectral information. To test this directly, we performed an ablation experiment in which longitude and latitude were removed from the input variables while all other predictors, including the spectral bands/indices, were retained.

The results show that model performance decreased from [R^2 = 0.925, RMSE = 0.326, MAE = 0.421] in the full model to [R^2 = 0.425, RMSE =0.911, MAE =0.620] after removing longitude and latitude. This confirms that geographic location carries substantial site-specific information related to catchment characteristics, climate, and anthropogenic influences, which is consistent with the site-specific nature of riverine COD dynamics.

6. Section 2.3 (Lines 186, 222, 243, 274): "Anaconda3 environment/platform" is a package manager, not a method; the actual software versions and configuration (e.g., Python) should be

reported instead. The reflectance extraction protocol (e.g., buffer size and minimum water-pixel counts around each site) should also be described.

Response: We thank the reviewer for this correction. We agree that "Anaconda3 environment/platform" is a package manager rather than a methodological description, and we have removed this phrasing from the revised manuscript. In its place, we now report the actual software versions and configuration used for model training: Models for COD_{Mn} prediction were trained on Landsat-derived spectral features using AutoGluon-Tabular (v1.3.1) via its Python API in Python 3.8.3.

We also agree that the reflectance extraction protocol required a clearer description. We have added the following details to Section 2.2.2: Landsat surface reflectance was extracted within a 90 m buffer around each site, and only pixels classified as water were retained. A minimum of 3×3 water pixels was required for a site–date matchup to be included; otherwise, the observation was excluded.

7. Lines 218 and 300–301: Historical COD changes are computed as differences between two-year endpoint averages (2014–2015 vs. 2023–2024) without uncertainty estimates. Given the site-level RMSE of 0.58 mg L^{-1} at M2 (Line 266), it is unclear whether the +0.441 mg L^{-1} midstream change (Line 296) is distinguishable from retrieval error. Trend estimation from the full annual series (e.g., the Mann–Kendall test with Sen's slope) with confidence intervals is recommended.

Response: We thank the reviewer for this important comment. We agree that comparing two endpoint averages does not provide a formal estimate of temporal trend uncertainty and may make relatively small changes difficult to distinguish from the uncertainty associated with COD retrieval.

We have therefore revised the analysis of historical COD changes to use the full annual time series rather than relying solely on differences between the 2014–2015 and 2023–2024 endpoint averages.

Specifically, we now calculate annual COD values for each site and estimate monotonic temporal trends using the Mann–Kendall test and Sen's slope estimator. Confidence intervals for the Sen's slope are also reported to quantify the uncertainty of the estimated temporal changes. This approach uses information from the entire observation period and provides a more statistically robust assessment of whether COD has undergone a systematic increase or decrease.

We also acknowledge that the RMSE of the first-stage Landsat-based COD model represents an important source of uncertainty. The revised trend analysis therefore does not interpret a small difference between two endpoint averages as evidence of a significant temporal change. Instead, historical changes are evaluated based on the estimated trend magnitude and its confidence interval, together with the statistical significance of the Mann–Kendall test. We have accordingly revised the text describing the +0.441 mg L^{-1} midstream change and have avoided interpreting this endpoint difference as a significant historical increase unless supported by the revised trend analysis.

We have also added a clarification in the Discussion that the temporal trends are derived from model-retrieved COD and therefore remain subject to retrieval uncertainty. Although the Mann–Kendall/Sen's slope framework provides uncertainty estimates for the temporal trend, it does not eliminate systematic errors associated with the COD retrieval model. We therefore interpret the historical trends conservatively and acknowledge this limitation explicitly.

8. Section 2.4 (Lines 220–229): The second model is trained on outputs of the first model rather than on observations, so errors propagate without independent ground truth. Moreover, applying relationships learned from spatial gradients to temporal climate change is a space-for-time substitution that should be stated and justified.

Response: We thank the reviewer for identifying this important methodological issue. We agree that our two-stage modeling framework does not provide independent ground truth for the second-stage model. In the first stage, models for COD prediction were trained using Landsat-derived spectral features and observed COD measurements. The resulting model was then used to generate satellite-based COD estimates. In the second stage, these satellite-based COD estimates, rather than directly observed COD measurements, were modeled as a function of geographical and environmental variables to enable spatially continuous prediction and future projection at the 160 sites. Therefore, prediction uncertainty from the first-stage remote-sensing model can propagate into the second-stage model, and the second-stage predictions cannot be regarded as independently validated against observations.

We have revised the Methods and Discussion to make this two-stage structure explicit and to acknowledge the associated uncertainty. In particular, we now clarify that the second-stage model should be interpreted as learning the spatial relationships between geographical/environmental characteristics and Landsat-derived COD estimates, rather than directly learning an observed COD response. Consequently, the uncertainty of the first-stage COD retrieval is inherited by the second-stage predictions.

We also agree that the application of these spatial relationships to future conditions constitutes a space-for-time substitution. The second-stage model derives relationships from spatial differences

among the 160 sites under present-day conditions and applies these relationships to projected future environmental conditions. This approach assumes that the spatial relationships observed across the current environmental gradient provide a reasonable approximation of the temporal response of COD to future environmental changes. We now explicitly state this assumption in the revised manuscript.

Because this assumption may not hold under nonstationary future conditions, we have tempered our interpretation of the future COD projections. We now describe them as model-based scenario projections conditional on the spatial stationarity assumption, rather than as independently validated forecasts. We have also added this limitation to the Discussion and highlighted the need for long-term, independently observed COD records and temporally explicit process-based or hybrid models to evaluate the robustness of these projections in future studies.

9. Line 529: 'The data will be available upon request' does not meet the Copernicus data policy. The reconstructed COD data and related datasets, together with the model code, should be deposited in a public repository with a DOI.

Response: We thank the reviewer. To comply with the Copernicus data policy, the reconstructed COD dataset and all related datasets used in this study will be deposited in Zenodo, a public persistent repository, which will assign a DOI. The DOI and repository links will be included in the revised Data Availability statement and cited in the manuscript.

10. Language: Please correct "rives" (Line 49), "ongoing climate changing" (Line 75), and "Yellowing River" (Line 87). Also, "the China's Yellow River" (title) does not seem to be

commonly used, please consider alternative phrasing.

Response: We thank the reviewer for catching these language errors. We have corrected them in the revised manuscript. We have also checked the manuscript for similar typographical and grammatical issues. We appreciate the reviewer's careful reading.