

Response to reviewer

REVIEWER 1

We thank the reviewer for the detailed comments and suggestions. We have revised the manuscript accordingly, with particular attention to the description of the database construction, the statistical analysis, the spatial coverage of the data, and the terminology used throughout the manuscript. We have also revised several interpretations to better reflect the limitations of the dataset and strengthened the description of the quality control and uncertainty. We would also like to highlight that this is primarily a database paper, not a hypothesis-based paper. The purpose of the statistical analysis is to illustrate the potential use of the compiled endmembers and to guide users in assessing their suitability for applications such as End-Member Mixing Analysis (EMMA) and isotope hydrograph separation.

We would also like to emphasize that the main novelty and scientific contribution of this manuscript lie in the compilation and harmonization of stable water isotope data from glacierized environments into a single, accessible database. By providing observed ranges of $\delta^{18}\text{O}$ and $\delta^2\text{H}$ across different hydrological endmembers, the database offers a practical reference for researchers and field hydrologists working in glacierized catchments. These ranges can help users establish an initial understanding of the isotopic variability that may be expected in different water sources, place new field measurements in a broader context, and assess whether locally observed isotope signatures are consistent with previously reported values. The database can therefore support the planning of field investigations, the preliminary identification of potential endmembers, and the interpretation of isotope measurements in data-scarce regions. Importantly, these ranges are intended as reference values rather than universally transferable endmember signatures, since local hydroclimatic conditions, elevation, seasonality and hydrological processes can substantially influence isotope compositions. The principal contribution of the manuscript is thus not to establish universal isotopic distinctions among water sources, but to provide a consolidated empirical reference that can guide future isotope-hydrology investigations and support applications such as End-Member Mixing Analysis (EMMA) and hydrograph separation.

We have also corrected the mistakes in the database and submitted it as a new version in Zenodo: <https://doi.org/10.5281/zenodo.23184536>

Comment 1 The database construction and literature-search methodology needs substantially more detail. The manuscript states that the database was constructed from 63 publications using searches primarily in Web of Science and Scopus and provides several search terms. However, the procedure is not sufficiently reproducible. The authors should provide: exact search strings; dates on which the searches were conducted; databases searched; inclusion and exclusion criteria; screening procedure; number of records/publications identified initially; number excluded at each stage and reasons for exclusion; procedure for identifying duplicate datasets; criteria for deciding whether a dataset from a publication was sufficiently documented for inclusion; procedure for handling supplementary datasets; procedure for incorporating unpublished data generated by the authors. A PRISMA-like flow diagram, even if adapted for a database-construction study rather than a systematic review, would substantially improve transparency. This is particularly important because

the paper describes the database as a global compilation, while the resulting coverage is highly uneven.

We agree. The original manuscript did not provide sufficient information to reproduce the literature-search and screening procedure. We have therefore expanded Section 2.2 to describe the databases searched, search dates, eligibility criteria, screening procedure, treatment of duplicates and supplementary datasets, requirements for sample-level data, and the incorporation of unpublished datasets contributed by the authors. The revised text now reports the number of publications identified, screened, excluded and retained, together with the principal reasons for exclusion. We also clarify that publications reporting only summary statistics, without recoverable individual isotope observations, were not incorporated as individual records. We thank the reviewer for pointing out the additional literature from the Himalaya and Hindu Kush, including the studies by Jeelani et al. and Lone et al. (2021, 2022). We reviewed these studies during the revision. However, they report isotope results as aggregated values and do not provide the individual sample-level $\delta^{18}\text{O}$ and $\delta^2\text{H}$ measurements required for inclusion in the present database. As the database is restricted to individual observations and does not include means or other aggregated values, these studies could not be incorporated as database records. We have clarified this inclusion criterion in Section 2.2.

Comment 2 – There are important inconsistencies concerning the temporal and geographic coverage. Several statements concerning the temporal and spatial extent of the database appear inconsistent. For example, the Abstract states that the database contains data from 19 countries, whereas Section 2.2 states that it covers 20 countries. The Abstract also states that isotope measurements span 1978-2023, whereas Section 2.2 states that the temporal coverage extends from 1981 to the most recent publications up to 2025. These discrepancies need to be resolved throughout the manuscript, including: number of countries; number of catchments; earliest isotope observation; latest isotope observation; publication-year range; sampling-year range. The authors should clearly distinguish publication year from sampling year, as these are not equivalent. Besides lot of work on glaciers/cryosphere using stable water isotopes in the Indus River Basin has been carried out which is completely missing when it comes to Himalaya or Hindu-kush (e.g. Jeelani et al., Lone et al., 2021; 2022; etc.).

We agree and thank the reviewer for identifying this ambiguity. The manuscript previously mixed two different temporal descriptors: the years in which samples were collected and the publication years of the source studies. These are now reported separately throughout the manuscript. The country count has also been harmonized throughout the manuscript. Specifically, the database currently covers **18 countries (Line 50; 275)**. We now explicitly distinguish the **sampling-year range** from the **publication-year range (line 51; 240; 908)**, rather than describing both as “temporal coverage.”

Comment 3 – The manuscript needs a much clearer distinction between observations, samples and records. The manuscript repeatedly uses terms such as individual entries, measurements, samples, records, and catchments. It is not sufficiently clear whether one database row represents: one physical water sample; one isotope measurement; one sampling event; one observation extracted from a publication; or one sample aggregated from a publication table. The statement that the database contains 12,348 “individual entries” and excludes averages needs clarification. This distinction is particularly important because the statistical analysis appears to treat the individual

isotope values as observations. I recommend adding a precise definition of record, sample, sampling event, study, catchment, and publication.

We agree. We have added explicit operational definitions for the terminology used in the manuscript and database. In particular, we now distinguish a database record from a physical sample, sampling event, study, publication and catchment. This distinction is important because the number of database records should not be interpreted as the number of statistically independent sampling units. See paragraph 2.2.2.

Comment 4 – Statistical independence and pseudo replication are major concerns. This is, in my view, one of the most important issues in the manuscript. The database combines thousands of measurements from a relatively small number of publications and catchments. Consequently, the individual observations cannot automatically be considered statistically independent. For example, hundreds of measurements from the same catchment, publication, monitoring station or laboratory may contribute disproportionately to a particular continental or endmember distribution. Nevertheless, Section 4.6 reports pairwise statistical comparisons and Dunn's tests, concluding that most endmembers are statistically distinguishable. The authors need to explain: What constitutes the statistical unit? Are individual isotope measurements treated as independent? Were observations weighted equally? Were measurements grouped by catchment? Were multiple observations from the same publication treated as repeated observations? Were differences among studies/catchments/laboratories considered? Was hierarchical structure accounted for? Were sample sizes dominated by a few highly sampled catchments? A conventional Dunn test applied directly to thousands of pooled measurements may produce very small p-values simply because of the very large sample size, even when the practical separation among endmembers is modest. The authors should therefore reconsider whether the reported significance testing is statistically appropriate. A hierarchical or study/catchment-aware approach may be more appropriate, or the authors should explicitly justify why individual observations can be treated as independent.

We thank the reviewer for raising the issue of statistical independence and potential pseudo-replication. We agree that individual isotope records originating from the same study, catchment, sampling campaign, or monitoring location should not automatically be regarded as statistically independent observations. This is now stated explicitly in the revised Methods (2.2.2. Terminology), where a database record is defined as an individual isotope observation associated with its available metadata, and where we clarify that multiple records may originate from the same study or catchment. We would also like to clarify the objective of the manuscript. This is primarily a descriptive database paper presenting a harmonized compilation of stable water isotope measurements from glacierized catchments worldwide. The database intentionally preserves the observed natural variability of isotope signatures across different environments, seasons, elevations, hydrological components and climatic settings. Its structure also reflects the practical availability of published data and the accessibility of glacierized field sites, which are inherently uneven across regions. The manuscript explicitly discusses this uneven spatial coverage and notes that research infrastructure, site accessibility and historical sampling priorities contribute to differences in data density among regions. The statistical analyses are therefore intended as supporting and exploratory tools to guide users in understanding broad differences and overlap among hydrological and cryosphere endmembers, rather than as the principal scientific outcome of the paper or as formal tests of hypothesis. The manuscript also reports distributional overlap and effect sizes so that p-values are not interpreted alone as

evidence of hydrological separability. We therefore do not consider a full hierarchical or mixed-effects analysis necessary for the objectives of this database paper. This paper is a database style paper, and it includes only descriptive statistics for the database overview purpose according to the journal scope. Such approaches would be appropriate for a dedicated hypothesis based study seeking to quantify study-, catchment-, laboratory- or site-level effects, but this is beyond the scope of the present database paper. The purpose here is to provide a transparent, harmonized reference dataset that captures the available range and variability of stable isotope observations and can subsequently be used by researchers for more specific statistical or process-based analyses (line 345).

Comment 5 – The statistical methodology is inadequately described. Section 4.6 reports: “The post-hoc Dunn’s test results indicate that most cryosphere and hydrological endmembers are statistically distinguishable...” However, the Methods section does not adequately describe the statistical workflow used to obtain these results. The manuscript should explicitly state: which omnibus test was performed before Dunn's test; whether Kruskal–Wallis tests were used; assumptions/checks; treatment of missing values; multiple-comparison correction; correction method used; significance threshold; software/package used; exact statistical model; whether analyses were conducted globally and separately by continent; sample size used in each comparison; whether outliers were retained. Without these details, the analysis cannot be independently reproduced.

We agree. A dedicated statistical-methods subsection has now been added. (2.4)

Comment 6 – Statistical significance is sometimes interpreted as hydrological separability. The manuscript states that statistically significant differences demonstrate “robust separation among major water sources.” This interpretation needs to be substantially moderated. A statistically significant difference in isotope distributions does not necessarily mean that the two endmembers are sufficiently separated to be useful in an isotope mixing model. For isotope hydrograph separation or EMMA, the relevant question is not simply: Are the two distributions statistically different? but rather: Are they sufficiently distinct relative to their within-source variability and analytical uncertainty to provide useful discriminatory power? The authors themselves later acknowledge that endmembers should be defined carefully and that uncertainty needs to be considered in mixing analyses. This is an important point, but it should be integrated more explicitly into the statistical interpretation. I recommend reporting effect sizes and distributional overlap in addition to p-values.

We have moderated the relevant statements throughout the revised manuscript to avoid equating statistical significance with hydrological separability. In particular, we no longer interpret significant pairwise differences as demonstrating “robust separation among major water sources.” Instead, the statistical comparisons are presented as exploratory and descriptive indicators of differences among endmember distributions. We would also like to reiterate that this is primarily a database paper, not a hypothesis-based paper. The purpose of the statistical analysis is to illustrate the potential use of the compiled endmembers and to guide users in assessing their suitability for applications such as End-Member Mixing Analysis (EMMA) and isotope hydrograph separation. We agree that statistically significant differences alone are not sufficient to establish that two endmembers are suitable for mixing-model applications. Their practical discriminatory value also depends on the magnitude of separation relative to within-endmember variability, distributional overlap, analytical uncertainty, and the specific hydrological context. Accordingly, the revised manuscript now emphasizes that

endmembers should be evaluated using their full distributions and associated uncertainty rather than p-values alone. Effect sizes and distributional overlap are also reported and discussed alongside the statistical tests. The statistical results are therefore intended to demonstrate the potential utility of the database for identifying candidate endmembers, while the final selection of endmembers for EMMA or hydrograph separation must remain application- and catchment-specific. Statistical analysis is only described in a short paragraph in point 4.6 and results are presented in annexes as guidance for future users, highlighting that this study emphasizes on the description of the database.

Comment 7 – The treatment of spatial sampling bias requires stronger quantitative assessment. The authors correctly acknowledge that isotope observations are unevenly distributed geographically, with substantial clustering in Europe, Asia and North America and very limited representation in Africa and Oceania. This is an important limitation, but the implications are not fully incorporated into the subsequent global analyses. For example, the manuscript compares continental distributions despite extremely different numbers of observations: Europe: 5,899; Asia: 3,132; North America: 3,022; South America: 249; Africa: 32; Oceania: 12. With only 12 observations in Oceania and 32 in Africa, continental-scale conclusions are necessarily very limited. The authors should distinguish clearly between: “distribution of available observations” and “representative distribution of isotope compositions across the continent.” The latter cannot be inferred from the current sampling density.

We agree that the continental distributions presented in the manuscript represent the distribution of the available observations in the database and should not be interpreted as statistically representative distributions of isotope compositions across entire continents. We have revised the manuscript accordingly and moderated the wording throughout the continental-scale analysis. The objective of this database paper is to compile and describe the currently available stable isotope observations from glacierized catchments rather than to construct a geographically balanced sampling design. The unequal number of observations among continents reflects differences in published data availability, monitoring infrastructure, and accessibility of glacierized environments. These limitations are inherent to the available observational record and are now stated more explicitly. We have therefore revised the text to distinguish clearly between the “distribution of available observations” and continental representativeness. In particular, conclusions for Africa and Oceania (line 673, 719) have been further restricted because of their very limited sample sizes, and continental comparisons are now described as exploratory patterns within the compiled database rather than representative estimates for entire continents. We have also emphasized that regions with sparse spatial coverage should not be used to infer continent-wide isotope characteristics without additional observations. We agree that a quantitative assessment of geographic representation is important. The manuscript already reports the number of records and catchments by continent and compares the distribution of sampled catchments with the global distribution of glacier area and ice volume. We consider these metrics appropriate for the scope of this descriptive database paper, while a formal assessment of spatial representativeness would require a dedicated spatial sampling framework beyond the purpose of the present compilation. We do not consider that additional figures or a complex spatial weighting analysis are necessary to address this comment. The manuscript already provides a quantitative description of the geographic coverage through the number of records and catchments represented in each region, and Table 2 explicitly compares the distribution of sampled catchments with global glacier area and ice volume. The revision therefore focuses on clarifying what the available observations can and

cannot represent, and on systematically moderating continental-scale statements so that they are interpreted as patterns within the compiled database rather than as fully representative isotope distributions for entire continents.

Comment 8 – Table 2 requires clarification concerning what constitutes a “catchment”. Table 2 compares the number of catchments in the database with global glacier area and ice volume. However, the manuscript needs to define precisely how a catchment was identified. For example: Does a publication containing samples from three sites within the same drainage basin count as one catchment? Are sub-catchments counted separately? Are multiple studies in the same catchment counted once or multiple times? Are catchment boundaries derived from the source publications or standardized using a GIS dataset? This is particularly important if the table is intended to quantify geographical representativeness.

We agree that the definition of “catchment” should be made more explicit, particularly because Table 2 uses the number of catchments as an indicator of the spatial coverage of the database. We have therefore clarified the terminology and counting procedure in the revised Methods. In this database, a catchment refers to the drainage unit associated with the sampling locations as identified from the source publication and available geographic information. Multiple sampling sites located within the same identified drainage basin are assigned to the same catchment and are therefore counted once in Table 2. Similarly, multiple publications or studies referring to the same catchment do not constitute additional catchments. Where the source information distinguishes hydrologically separate drainage units or sub-catchments, these may be retained as separate catchments when they represent distinct sampling systems. Catchment boundaries were not systematically re-delineated using a common global GIS watershed dataset; the database preserves the catchment attribution available from the original studies and harmonized geographic metadata. A definition of catchment was added in section 2.2.2 Terminology. We have also clarified that Table 2 is intended to describe the geographic coverage of the available database rather than to demonstrate statistical representativeness of glacierized catchments globally.

Comment 9 – The claim that the database provides “characteristic isotopic signatures” should be moderated. The manuscript repeatedly refers to characteristic isotope values of different cryosphere endmembers. For example, the Abstract describes glacier ice and glacial meltwater as typically having $\delta^{18}\text{O}$ around -14‰ and $\delta^2\text{H}$ around -100‰ . However, the manuscript itself demonstrates substantial overlap among several endmembers. This raises an important conceptual issue: there may not be a globally universal isotopic fingerprint for a particular endmember such as glacier meltwater or snowmelt. The observed isotope composition depends on: elevation; latitude; precipitation season; moisture source; continentality; accumulation history; snowpack evolution; sublimation; refreezing; melt timing; glacier age; geographic setting. Therefore, I suggest replacing language implying universal “fingerprints” with terminology such as observed isotopic distributions, regional ranges, or empirical isotope signatures within the compiled dataset, where appropriate.

We agree that the isotope composition of a given cryosphere or hydrological endmember cannot be represented by a single globally universal value or “fingerprint”. Stable isotope compositions vary with latitude, elevation, climatic setting, moisture source, seasonality, accumulation history, melt and refreezing processes, and other local hydrological controls. The substantial variability and overlap observed in the compiled dataset itself supports this interpretation. We have revised the Abstract to

avoid implying that the reported central values for glacier ice or glacial meltwater represent universal values. The purpose of the database is not to prescribe fixed isotope values for individual components, but to document the range and variability of stable isotope compositions reported for cryosphere and hydrological components across different environmental settings. These empirical distributions can provide a reference for evaluating potential endmembers in applications such as End-Member Mixing Analysis or isotope hydrograph separation, but endmember suitability must ultimately be assessed within the specific geographic and hydrological context of each application. Expressions implying characteristic or universal values have been replaced with wording referring to observed isotopic distributions, ranges and signatures within the compiled dataset or specific regional settings. We have also clarified that the reported isotope values represent empirical distributions influenced by geographic, climatic and hydrological controls, rather than fixed endmember values applicable across all glacierized catchments.

Comment 10 – The interpretation of d-excess is too categorical. The manuscript interprets d-excess largely in terms of moisture-source relative humidity, sea-surface temperature and kinetic fractionation. This is broadly reasonable, but the interpretation of d-excess in snow, glacier ice, groundwater and meltwater is more complicated because secondary processes can substantially modify d-excess. The authors themselves recognize the roles of: sublimation; snow metamorphism; evaporation; refreezing; atmospheric exchange; groundwater mixing. Therefore, statements such as elevated d-excess indicating particular moisture-source conditions should be framed more cautiously, especially for secondary hydrological endmembers. For example, the statement that elevated Asian d-excess “reflects a strong influence of moisture derived from relatively dry, cool oceanic sources” requires stronger evidence than the isotope distribution alone provides.

We agree that d-excess cannot be interpreted solely as an indicator of moisture-source relative humidity or sea-surface temperature, particularly for secondary hydrological and cryosphere endmembers. While d-excess in precipitation can provide information on conditions at the moisture source, its value in snowpack, glacier ice, meltwater, groundwater and other secondary endmembers may also be modified by post-depositional and hydrological processes, including sublimation, snow metamorphism, evaporation, refreezing, atmospheric exchange and mixing. We have therefore moderated the interpretation of d-excess throughout the revised manuscript in paragraph 4.2. Statements that previously attributed elevated or reduced d-excess directly to specific moisture-source conditions have been reformulated to indicate that such values are consistent with, or may reflect, those conditions rather than being uniquely diagnostic of them. We now distinguish more clearly between primary atmospheric controls on d-excess and subsequent modification during storage, melting, refreezing, evaporation and mixing. In particular, the interpretation of elevated d-excess in Asia has been revised. We no longer state that the observed values demonstrate a strong influence of relatively dry, cool oceanic moisture sources. Instead, we interpret the pattern as potentially reflecting a combination of moisture-source conditions, atmospheric transport, precipitation seasonality and post-depositional processes. We would also like to emphasize that this is a database paper. The purpose of the interpretative discussion is to guide users in understanding the possible processes that may influence the observed isotope distributions and to support informed use of the database. We do not intend to draw definitive process-based conclusions from d-excess alone. The revised text therefore presents these interpretations as possible explanations and points of guidance for subsequent site-specific studies, rather than as conclusions established by the database itself.

Comment 11 – The interpretation of Figure 2 and the regression requires revision. The manuscript reports a regression slope of approximately 8.23 and intercept of approximately 15‰, and interprets this as indicating particular fractionation and moisture-source processes. However, several issues need clarification: Was ordinary least squares regression used? Were all 12,348 measurements included? Were observations weighted? Were different endmembers pooled? Were repeated measurements from the same catchment treated independently? Was uncertainty in both $\delta^{18}\text{O}$ and $\delta^2\text{H}$ considered? Were outliers included? What are the confidence intervals and R^2 ? Is the reported slope statistically different from 8? Pooling different water types from different environments and then fitting a single global regression can produce a line that does not necessarily represent a physically meaningful meteoric water line. I recommend reporting the regression methodology and uncertainty explicitly and avoiding a strong process interpretation unless supported by additional analysis.

We agree that fitting a single regression to pooled isotope observations from different hydrological and cryosphere endmembers, geographic regions and sampling contexts should not be interpreted as defining a physically meaningful global meteoric water line. The primary purpose of Figure 2 is descriptive: to show the distribution of the compiled $\delta^{18}\text{O}$ and $\delta^2\text{H}$ observations in dual-isotope space and their general position relative to the established Global Meteoric Water Line (GMWL). We have therefore revised this section and removed the strong process-based interpretation of the pooled regression slope and intercept. As this is a database paper, our objective is to characterize the range and distribution of the available observations and to guide users in their interpretation, rather than to derive a new global meteoric water line or make inferential conclusions regarding fractionation processes from a pooled regression. We also recognize that the database contains observations from multiple endmembers, studies and catchments, with multiple records potentially originating from the same study, catchment, sampling location or monitoring period. We have therefore revised the associated discussion of Figure 2 and to emphasize comparison with the established GMWL as a reference rather than interpretation of the pooled regression as a new meteoric water line.

Comment 12 – Figure 3 requires greater transparency regarding sample size. The figure caption states that sample sizes are shown for each category. However, because the sample sizes differ substantially among endmembers, the figure should be accompanied by either: an explicit table of n values; median and interquartile range; preferably confidence intervals; and, where relevant, number of independent catchments/studies contributing to each category. A category with thousands of observations from a few studies should not be visually or statistically interpreted in the same way as a category with tens of observations from several independent catchments.

We agree that the number of individual isotope observations differs substantially among hydrological and cryosphere components and that the number of records should not be interpreted as equivalent to the number of independent studies or catchments. Figure 3 is intended as a descriptive visualization of the distributions represented in the compiled database. The figure already reports the number of individual observations (n) for each component and the boxplots themselves show the median, interquartile range, overall spread and outliers. We have clarified this explicitly in the revised caption and text. We also agree that a component represented by many measurements from a limited number of studies or catchments should not be interpreted in the same way as a component represented across a larger number of independent spatial units. This issue is now stated more clearly in the manuscript. The database contains multiple records that may originate from the same study,

catchment, sampling location or monitoring period, and these records are therefore not equivalent to statistically independent studies. Because the purpose of Figure 3 is to describe the empirical distribution of the observations available in the database rather than to provide population-level estimates for each component, we have not added confidence intervals to the figure. Confidence intervals could imply a degree of statistical independence and representativeness that is not warranted by the heterogeneous observational structure of the database. Instead, we retain the median, interquartile range, data spread and sample size as descriptive measures and have strengthened the accompanying text to caution users regarding differences in sampling intensity and study/catchment representation. Where statistical comparisons are performed elsewhere in the manuscript, the analytical unit and limitations related to non-independence are described separately.

Comment 13 – The manuscript should distinguish measurement uncertainty from environmental variability. The uncertainty section is conceptually useful, but the manuscript sometimes combines several fundamentally different uncertainty sources. Analytical uncertainty, sampling variability, spatial heterogeneity and between-study methodological differences are not interchangeable. For example: analytical precision is measurement uncertainty; storm-to-storm isotope variability is environmental variability; different sampling seasons represent sampling design; differences among laboratories represent inter laboratory uncertainty; differences among catchments represent spatial variability. The manuscript should provide a clearer uncertainty framework, perhaps separating: Measurement uncertainty-sampling uncertainty-environmental variability-representativeness uncertainty- database harmonization uncertainty. This would make the database substantially more useful to downstream users.

We agree that the different sources of uncertainty should be distinguished more explicitly because they have different origins and implications for downstream use of the database. The manuscript already treats these sources separately in Sections 3.1–3.5, including analytical precision, environmental variability, temporal sampling variability, spatial heterogeneity, and metadata/data-harmonization limitations. However, we agree that the terminology used to describe them was not sufficiently systematic and could give the impression that these fundamentally different sources of variability were being combined. We have therefore revised the uncertainty framework to distinguish more clearly between: (1) measurement uncertainty associated with analytical precision and inter-laboratory differences; (2) sampling uncertainty related to temporal sampling design and differences in sampling strategy; (3) environmental variability associated with natural spatial and temporal isotope variability; (4) representativeness uncertainty arising from uneven spatial coverage and differences in the number of observations among catchments and regions; and (5) database harmonization uncertainty associated with differences in reporting formats, metadata completeness, terminology and integration of heterogeneous source datasets. We also clarified that these categories should not be considered interchangeable or combined into a single numerical uncertainty estimate. Rather, they describe distinct sources of uncertainty or variability that users should consider depending on the intended downstream application of the database.

Comment 14 – Quality-control procedures need to be documented more rigorously. The authors state that outliers were flagged rather than removed and provide thresholds for $\delta^{18}\text{O}$, $\delta^2\text{H}$ and d-excess. However, the rationale for these thresholds is not sufficiently explained. For example: Why was $\delta^{18}\text{O}$ outside -40 to +20‰ selected? Why was $\delta^2\text{H}$ outside -320 to +120‰ selected? Why is d-excess outside 0-25‰ flagged? Are these physical thresholds, statistical thresholds or screening

thresholds? Are flagged observations included in Figures 2-6? Are they included in the statistical analyses? How many records were flagged? How many were subsequently verified or retained? This should be reported quantitatively.

We agree that the purpose and application of the quality-control thresholds should be described more explicitly. The thresholds of -40 to $+20\text{‰}$ for $\delta^{18}\text{O}$, -320 to $+120\text{‰}$ for $\delta^2\text{H}$, and 0 to $+25\text{‰}$ for d-excess were used as screening thresholds to identify values requiring verification; they were not treated as physical limits or statistical outlier criteria and were not used for automatic data exclusion. The objective was to detect potentially unusual values, transcription errors or reporting inconsistencies while avoiding removal of environmentally plausible isotope compositions. We have clarified this distinction in the revised Methods. Values outside the screening ranges were checked against the original source information where possible.

Comment 15 – Harmonization to VSMOW/VSMOW2 needs more explanation. The manuscript states that isotope values were reported relative to VSMOW/VSMOW2 and that data were reported to VSMOW-SLAP using two-point normalization. However, the database spans several decades and multiple laboratories. The authors should explain how measurements reported using older standards or calibration schemes were harmonized. In particular: Were original isotope values converted? Were values simply retained as reported? Were laboratory-specific corrections available? Were data measured before adoption of current reference materials treated differently? This is important for a global dataset where differences of less than 1‰ may be scientifically relevant.

We agree that the description of isotope-scale harmonization was insufficiently precise and could imply that all historical measurements were retrospectively recalibrated by the authors. This was not the case. During database compilation, $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values were retained as reported in the original publications or contributed datasets, provided that they were reported in δ -notation relative to the VSMOW/VSMOW-SLAP framework. We did not apply retrospective numerical conversions between VSMOW and VSMOW2, nor did we apply laboratory-specific correction factors, because the information required for such corrections was not consistently available across the source studies. Measurements generated before the adoption of the current VSMOW2/SLAP2 reference materials were therefore not recalculated or treated differently when the original study reported the values on the VSMOW scale. We have revised the manuscript to distinguish between harmonization of reporting conventions and analytical recalibration. In this database, harmonization refers to standardization of units, δ -notation, parameter names and metadata, rather than reprocessing of the original laboratory measurements. Where analytical method and reported measurement uncertainty were available, these were retained as metadata to allow users to evaluate comparability among datasets. We also now acknowledge that the database spans several decades and multiple laboratories and that differences in reference materials, calibration procedures, instrumentation and laboratory practice may introduce inter-study uncertainty. This is particularly relevant for applications involving differences approaching analytical precision. Users undertaking high-precision comparisons should therefore consult the original source information and consider the reported analytical uncertainty and methodological context.

Comment 16 – The use of “EMMA” terminology should be reconsidered. The manuscript states that endmember labels were standardized to a controlled vocabulary used in endmember mixing

analysis (EMMA). However, the database contains categories that are not necessarily conventional endmembers in a strict mixing-model sense. For example: groundwater; stream water; lakes; talus water; permafrost thaw; supraglacial water. Some of these can themselves represent mixtures rather than chemically or isotopically discrete sources. I suggest distinguishing between: source/endmember categories and observed hydrological compartments. This distinction would make the dataset much more useful for mixing-model applications.

We agree that the term “endmember” should not be used as a generic synonym for all hydrological categories represented in the database. In the revised manuscript, we now distinguish explicitly between “hydrological and cryosphere components,” which describe the observed water and ice compartments represented in the database, and “endmembers,” which refers to components considered as potential source waters in a specific isotope-mixing or hydrograph-separation application. We have clarified in Section 2.1 that components such as streams, lakes, groundwater, supraglacial water, permafrost thaw water and talus-slope water may themselves integrate or mix several water sources and therefore should not automatically be considered discrete endmembers. Their suitability as endmembers depends on the specific catchment, hydrological context and degree of isotopic separation among potential sources. EMMA is therefore retained as one potential application of the database, but the database categories themselves are not assumed a priori to represent valid endmembers. We have revised the terminology throughout the manuscript accordingly.

Comment 17 – Some interpretations are stronger than the data support. Several conclusions appear to infer hydrological processes from isotope distributions alone. For example, the manuscript states that stream waters show intermediate values indicating mixing between cryosphere and meteoric sources and that these values imply dominant glacier/snowmelt contributions in some regions. An intermediate isotope composition does not by itself quantify the contribution of each source. Without discharge-weighted source sampling, temporal sampling, or mixing-model analysis, isotope values alone should not be interpreted as demonstrating a dominant contribution. The manuscript should carefully distinguish: “consistent with mixing” from “demonstrates/indicates a dominant contribution.”

We agree that isotope distributions alone cannot be used to quantify the relative contributions of different water sources or demonstrate dominance of a particular source without appropriate temporal sampling, discharge information, or formal mixing-model analysis. We have therefore revised the relevant interpretations throughout the manuscript to distinguish more clearly between isotope patterns that are “consistent with” particular hydrological processes and evidence that would be required to quantify source contributions. In particular, statements implying that intermediate stream-water isotope values demonstrate a dominant glacier or snowmelt contribution have been removed or rephrased. The revised text now states that intermediate stream-water isotope compositions are consistent with mixing between cryosphere-derived and meteoric waters, but that the isotope distributions compiled in this database do not by themselves allow the relative proportions of these sources to be determined. More generally, process interpretations are now presented as possible or plausible explanations of the observed isotope distributions rather than as quantitative attribution of hydrological sources. This distinction is consistent with the scope of the study, which is to characterize and compare the isotope distributions represented in the database and to provide a reference for

subsequent catchment-specific mixing analyses, rather than to quantify source contributions at individual sites.

Comment 18 – The continental analysis should be substantially revised. The continental analysis is potentially interesting, but several interpretations are based on very different sample sizes and climatic/geographic mixtures. For example, Europe contains 5,899 observations while Africa contains only 32. Consequently, statements about continental differences should explicitly include: n observations; n studies; n catchments; geographical coverage; elevation range; climatic zones represented. The authors should avoid implying that the observed distributions are representative of entire continents.

We agree that continental comparisons must be interpreted in the context of the strongly uneven sampling intensity and geographical coverage of the database. We have substantially revised the continental analysis to make these limitations explicit. Section 2.3 now reports the number of observations by continent and clarifies that the database represents the distribution of currently available isotope observations rather than a statistically or geographically representative sampling of entire continents. We also explicitly note that sampling density reflects differences in research effort, monitoring infrastructure, accessibility and data availability. To further improve transparency, we added Supplementary Table S2, which reports the number of observations, studies and catchments for each continent × hydrological/cryosphere component combination. This allows the reader to distinguish, for example, between categories represented by many observations from relatively few catchments or studies and those supported by broader spatial or study-level coverage. We have also revised Section 4.3 to describe the continental boxplots as distributions of the observations available in the compiled database rather than representative continental isotope distributions. Statements implying continent-wide representativeness have been moderated accordingly, with particular caution for regions with sparse coverage such as Africa and Oceania. Elevation and climatic setting are discussed as important controls on isotope variability and are retained in the database metadata where available. However, because elevation information is incomplete for some records and a standardized climatic-zone classification was not consistently available across all source studies, we did not introduce a new retrospective climatic classification that would require additional assumptions. Instead, geographical and climatic limitations are now explicitly discussed when interpreting the continental patterns.

Comment 19 – The 60° latitude division for North America needs stronger justification. The manuscript separates North America into >60° and ≤60° latitudes to distinguish Arctic/sub-Arctic from temperate/alpine settings. This is a useful exploratory division, but the choice of 60° is somewhat arbitrary and may combine highly different mountain and climatic systems within each group. The authors should either: justify the threshold more rigorously; conduct sensitivity analyses using alternative boundaries; or describe this explicitly as an exploratory stratification.

The division was not selected solely on the basis of the isotope data. The 60th parallel is commonly used as a geographic delimiter of northern North America; for example, Environment and Climate Change Canada defines northern Canada as the region north of 60°N, and 60°N also forms part of the North American boundary used by the Arctic Monitoring and Assessment Programme. From a hydroclimatic perspective, this division also broadly separates the predominantly cold, sub-Arctic/Arctic environments represented in the northern part of our database from the temperate–

alpine glacierized environments farther south. These regions differ in temperature regime, snow and ice seasonality, permafrost occurrence, atmospheric transport and melt-season characteristics, all of which influence stable water isotope distributions. We have clarified that 60°N is used as a practical analytical grouping and is not intended to imply an abrupt or universal climatic boundary. The Arctic region is generally defined by the Arctic Circle, which sits at approximately 66° 33' N (or 66.5°) latitude.

Comment 20 – Figure 4 and the continental comparisons should include independent sample counts. The figures show boxplots of continental isotope distributions, but the manuscript would be much stronger if the figure or supplementary information reported: n observations / n studies / n catchments for every continent × endmember combination. This would allow readers to immediately distinguish well-constrained distributions from poorly sampled categories.

We agree that the number of individual observations alone does not adequately describe the strength or independence of the continental distributions, because many observations may originate from a limited number of studies or catchments. We have therefore added Supplementary Table S2, which reports the number of observations, studies and catchments for each continent × hydrological/cryosphere component combination. This allows readers to distinguish categories supported by broad study and catchment coverage from those represented by many observations from only a few locations or studies. The caption of Figure 4 has also been revised to direct readers to Supplementary Table S2 for these independent sample counts. We have further clarified in the text that the continental boxplots describe the distributions of observations available in the compiled database and should not be interpreted as equally sampled or statistically representative continental populations.

Comment 21 – The manuscript should provide a comprehensive data dictionary. The manuscript states that the database contains 44 fields. A complete data dictionary should be provided in the Supplementary Information or Zenodo repository. This is particularly important for an ESSD data publication.

We agree that a comprehensive data dictionary is essential for the transparency and reuse of the database, particularly for an ESSD data publication. In the revised version, we have clarified that the current database contains 39 fields. A complete data dictionary is provided with the database and describes the variables included in the dataset, including their definitions, units where applicable, data type or format, and the meaning of categorical fields (See the sheet called “Metadata” in the database excel file available in Zenodo). We have also revised Section 2.2.4 to direct readers explicitly to the data dictionary accompanying the archived dataset. This ensures that users can interpret the database structure and metadata consistently and facilitates reuse of the dataset in accordance with FAIR data principles.

The conclusions should be more focused on the database rather than broad hydroclimatic conclusions. The strongest contribution of the manuscript is the database and its harmonization, rather than the broad global interpretation of isotope patterns. The conclusions would therefore benefit from emphasizing: database scope; spatial and temporal coverage; endmember categories; metadata; quality-control procedure; uncertainty; limitations; applications; future updates. Some broader process interpretations should be shortened or moved to a more clearly qualified discussion. Data and code availability should be made more explicit. The manuscript provides Zenodo DOIs for both data and code. This is a major strength. However, the data availability

statement currently contains a grammatical error: “Data described in this manuscript can be publicly available...” The statement should be corrected and should specify: exact dataset version; exact code version; whether the Zenodo record is versioned; whether all supplementary tables are included; whether raw source data can be traced to the original publications. The manuscript currently identifies the dataset and software repositories, which is good practice.

We agree that the principal contribution of the manuscript is the development, harmonization and documentation of the stable-water-isotope database, and that the Conclusions should emphasize this contribution more clearly. We have therefore revised the Conclusions to focus primarily on the scope and structure of the database, including its spatial and temporal coverage, the hydrological and cryosphere components represented, standardized metadata, quality-control procedures, uncertainty and limitations, potential applications, and the need for continued updates as new isotope data become available. We have also revised the Data and Code Availability statements.

REVIEWER 2

General comments

This manuscript presents a compilation of more than 12,000 $\delta^{18}\text{O}$ and $\delta^2\text{H}$ measurements from glacierized catchments around the world. Rock glaciers, ice-cored moraines, talus slopes and permafrost thaw are rarely sampled, and having them gathered in one place is very useful. I also think it was the right decision to keep individual measurements rather than site averages. A dataset like this has clear value for isotope hydrology and glaciology, and it is a good fit for ESSD. I enjoyed reading the paper.

I went through the manuscript and also looked at the Zenodo file. My first three comments concern the dataset itself and how it is documented (quality control, compilation and metadata, terminology), since users will rely on this paper to understand what they download. The next two concern the geographical coverage and the interpretation of pooled global distributions, which I think goes further than the data allow. The last one deals with the d-excess discussion and with places where the text does not match the figures.

None of this is out of reach, but together it amounts to a fair amount of work, so I recommend major revision.

We thank the reviewer for the careful assessment of both the manuscript and the archived dataset, and for recognizing the value of compiling individual stable-water-isotope observations from comparatively rarely represented cryosphere and hydrological components. We have undertaken a substantial revision of the manuscript and database documentation. In particular, we strengthened the description of database construction and quality control, clarified terminology and metadata, revised the uncertainty framework, moderated interpretations of pooled and continental isotope distributions, added information on study- and catchment-level representation, and revised the d-excess discussion. We have also refocused the Conclusions on the scope, limitations, reproducibility and future development of the database. We have also corrected the mistakes in the database and submitted it as a new version in Zenodo: <https://doi.org/10.5281/zenodo.23184536>

1) Quality control and data file

Section 2.2 states that outliers were flagged rather than removed (L179–189). This is good practice, but I could not find these flags in the file: there is no flag column, and some values far outside the stated bounds are not marked (see below). The "Parameters definition" sheet of the file is a very useful start. It would help to add the flag field and the units where they are missing, and to align it with the text, which mentions 44 fields (L166) while the file and the sheet contain 39. The "std", "error" and "uncertainty" fields also overlap in meaning and are unevenly filled ("std" is almost empty), so a clearer definition, or merging them, would help users. At L320–323, the authors also suggest that the spread of values within each category gives insight into both environmental variability and potential analytical uncertainty. I would nuance the second part, since that spread is dominated by differences between sites and climates, which makes analytical uncertainty hard to see.

The manuscript has been corrected to state that the current database contains 39 fields, and the parameter dictionary has been expanded to define the database fields, units where applicable, and the meaning of the analytical-information fields. We also agree that the previous wording concerning analytical uncertainty was too broad. The revised uncertainty framework now distinguishes measurement uncertainty from environmental variability, sampling uncertainty, spatial/representativeness uncertainty and database-harmonization uncertainty. We no longer imply that the overall dispersion within a hydrological component provides a direct measure of analytical uncertainty, because the observed spread is predominantly influenced by real spatial, temporal and environmental heterogeneity. Regarding quality control, we have clarified that the specified isotope ranges were used as screening criteria for verification, rather than as physical limits or statistical outlier criteria. Values outside these ranges were not automatically excluded and were checked against the source information where possible.

A few things that are not mentioned in the manuscript came up when I checked the file. Some of them may have a simple explanation.

The d-excess distribution stops just above 0 ‰, the lower flag bound, with no negative value anywhere, while high values are kept without any mark. This looks as if negative values were removed rather than flagged. It matters for the interpretation too. At L360–364 and L382–385, the lack of strong isotopic enrichment in lakes is explained by cold conditions, in contrast with arid settings, whereas Sect. 4.4 (L526–530) attributes their lower d-excess to evaporation and variable residence times. It would help to reconcile these two readings. Aridity is not limited to warm regions, and several sites in the database may be cold and arid. There, evaporated waters with low or negative d-excess would be expected, and where evaporation signals are absent, a short residence time may be the better explanation. I would suggest restoring the removed values, if any, and adding a nuance on this point.

We agree that negative d-excess values are physically possible and should not be excluded merely because they fall outside an operational screening range. We have therefore clarified that the 0–25‰ d-excess interval was used only as a quality-control screening window and not as an admissible physical range. The database and source information were rechecked to ensure that observations were retained unless an actual transcription or data-entry error could be established. We have also revised the discussion of lakes and d-excess. The previous text could imply that cold conditions alone explain

the absence of strong isotopic enrichment, while lower d-excess was separately attributed to evaporation. The revised interpretation is more cautious: lake isotope distributions may reflect a combination of temperature, aridity, evaporation, residence time, hydrological throughflow and source-water composition. In cold but arid environments, evaporative modification can still occur, whereas lakes with limited enrichment may instead reflect relatively short residence time or strong throughflow. The database itself does not provide sufficient information to distinguish these mechanisms uniquely.

A few entries look wrong, for example an unrealistic $\delta^{18}\text{O}$ value (record 7081) or publication years (and data-owner labels) increasing by one per row from 2015 to 2071 for Gordon et al. (2015). These are most likely cells shifted or auto-incremented during copy-paste in a spreadsheet, which is easy to fix, and a simple check of value ranges and coordinates should catch any remaining cases.

We have rechecked the affected entries against their source information and corrected confirmed transcription or spreadsheet-entry errors. We also repeated range and consistency checks for isotope values and associated metadata to identify similar issues.

About a dozen pairs of records share the same site, date and isotope values, mostly in Michelin et al. Some may be laboratory replicates, which would be worth indicating in the file; otherwise they should probably be removed.

We agree. Potential duplicate records were rechecked against the corresponding source datasets.

Could the authors explain how the groundwater samples from Mount Fuji fit the selection criteria? I may have missed the link with glacier or snow processes, and a sentence would help readers understand why they are included.

We have rechecked these records against the inclusion criteria and their source context. The revised manuscript now makes the inclusion criterion more explicit: observations must be associated with an identifiable hydrological or cryosphere component within the glacierized catchment context considered by the database. Where the connection of a dataset to the scope of the compilation could not be adequately established, the corresponding observations were reconsidered rather than retained solely because stable-isotope measurements were available.

2) Data compilation and metadata

Section 2.2 describes the compilation in a few lines: 63 publications, listed in Table S1, plus unpublished data, with individual values only, no averages (L156–166). The databases and keywords used for the search are given (L161–164). Since the paper will be the main reference for the dataset, I would like to see a little more on how it was put together: the date of the search, the number of studies found and screened, the criteria used to include or exclude them, how values were extracted (tables, supplements, digitized figures, contact with authors), how unpublished data were measured and checked, and how overlaps between papers or with GNIP/GNIR were handled.

We agree. This point is also consistent with the concerns raised by the other reviewer, and Section 2.2 has been substantially expanded. The revised manuscript now describes the literature-search procedure, databases and search terms, search period, screening and eligibility criteria, requirements for inclusion of sample-level paired $\delta^{18}\text{O}$ and $\delta^2\text{H}$ observations, reasons for exclusion, treatment of duplicate observations, and incorporation of directly contributed datasets. The database is restricted

to individual observations that can be associated with an identifiable hydrological or cryosphere component and sampling context; publications reporting only means or ranges without recoverable individual values are not entered as individual database records. We also distinguish a database record, physical sample, sampling event, study, publication and catchment so that the provenance and statistical structure of the compilation are clearer.

Also, only 63 publications seemed to me a little surprising given how much has probably been published, and a short note on why some studies were not included would be helpful. Section 3.4 states that the coordinates and elevation of sampling locations were retained (L294–297), and this is the case for most records. In some catchments, however, different water types share a single coordinate, sometimes rounded to the degree, which suggests a catchment-level location. Would it be possible to give the sampling location for these records as well, or at least a field stating whether coordinates refer to the sample or the catchment? This would make later reuse and reinterpretation much easier, for instance relating samples to elevation, distance from the glacier or local topography. It would also help to state where elevations come from and what counts as a catchment.

We agree that the number of source publications requires explanation. The current database contains 49 source publications/datasets rather than 63 publications, and the manuscript has been corrected accordingly. The relatively limited number primarily reflects the requirement for recoverable individual paired isotope observations with sufficient information to assign a hydrological/cryosphere component and sampling context. Studies reporting only summary statistics, figures without recoverable individual observations, or insufficient contextual information were not entered as individual database records. We have also added an operational definition of a catchment. A catchment is assigned primarily from the drainage unit reported in the original publication and available geographic metadata. Multiple sampling sites within the same identified drainage basin are treated as belonging to the same catchment, whereas hydrologically distinct drainage units reported separately may be retained as separate catchments. Catchment boundaries were not systematically redelineated using a single global GIS dataset. We agree that coordinate precision and spatial scale vary among sources. The database preserves coordinates and elevations as reported or recoverable from the source rather than generating artificial precision. This limitation is now discussed explicitly as part of spatial and metadata uncertainty. Where only a catchment-level or approximate location is available, users should therefore not interpret the coordinates as precise sample locations.

Regarding L303, it would be useful to indicate for each record whether the original study reported a calibration against VSMOW and SLAP (or VSMOW2/SLAP2), or whether this was assumed, as it is hard to verify for older data. The "number of samples" field mentioned at L169 does not appear in the file, and it would in any case be at odds with "single data" (L158). Finally, the number of countries differs between the abstract (19) and L160 (20), and the temporal coverage at L160–161 seems to refer to publication years rather than to the sampling period given in the abstract.

We agree. The manuscript has been revised to avoid implying that all historical measurements were retrospectively normalized by us. Stable-isotope values were retained as reported by the original sources in δ -notation relative to the VSMOW/VSMOW-SLAP framework. No retrospective conversion between VSMOW and VSMOW2 or laboratory-specific recalibration was performed where the information required for such corrections was unavailable. Differences in reference materials, normalization procedures and laboratory calibration are now explicitly identified as a source of inter-

study uncertainty. The terminology relating to individual records and sample numbers has also been revised. A database record is now defined explicitly as one row containing one paired $\delta^{18}\text{O}$ and $\delta^2\text{H}$ observation and its associated metadata. The geographic and temporal counts have been harmonized throughout the manuscript. The revised manuscript distinguishes sampling year from publication year and reports sampling-year availability separately. The current database contains 12,348 records distributed across six continents, and the country count is reported consistently in the revised text.

For precipitation, the text acknowledges that samples may be single events, daily or monthly composites (L257–259), and later states that they are mostly monthly (L625). The sampling interval and the representativeness of each sample are crucial here: an event sample and a seasonal composite do not carry the same information, and they cannot be meaningfully pooled or averaged without weighting by precipitation amount. The file has sample period fields, but they are filled for only a small part of the precipitation records. I would suggest completing them where possible, and adding the phase and amount of precipitation.

We agree that event, daily and composite precipitation samples are not equivalent and that their temporal representativeness should be considered when the database is used for process analysis or weighted precipitation-isotope estimates. Sampling-period information has been retained wherever it was reported in the original source, and the manuscript now treats incomplete temporal metadata as an explicit sampling limitation rather than assuming equivalence among precipitation observations.

However, precipitation amount and phase were not reported consistently enough across the source studies to introduce complete standardized fields retrospectively without substantial inference. We therefore did not estimate missing values. Instead, we preserve the available sampling-period metadata and explicitly caution that pooled precipitation observations should not be interpreted as precipitation-amount-weighted regional or continental isotope means.

3) Use of the term "endmember"

Throughout the paper, the 14 categories are called endmembers. I would suggest calling them "water types" and keeping "endmember" for the way the data may be used in mixing models. Section 2.1 itself defines endmembers as pure or unmixed sources, and Table 1 describes streams and groundwater as integrating several contributions. Streams are precisely the mixtures that EMMA aims to decompose, and lakes and groundwater are themselves mixtures, even though groundwater is often used as an endmember for streamflow: they integrate rain, snowmelt, ice melt or subsurface water in proportions that vary in space and time, so their composition is an outcome of mixing rather than an input to it. Even snowpack or glacier ice act as endmembers only with respect to a given mixture in a given catchment. A reader less familiar with mixing models could otherwise take lakes or groundwater for unmixed sources. Simple rules for assigning samples to categories (e.g. glacial meltwater vs. stream) would also help users. Note also that the file contains a fifteenth category, "snow" (two records), which is defined in the file but not in Table 1.

We agree and have revised the terminology throughout the manuscript. The database categories are now described primarily as hydrological and cryosphere components, referring descriptively to the sampled water or ice compartment. The term endmember is reserved for the situation in which a component is evaluated as a potential source in a catchment-specific isotope-mixing or hydrograph-separation application. Section 2.1 now explicitly states that streams, lakes, groundwater and other

components may themselves integrate multiple sources and therefore should not automatically be regarded as unmixed endmembers. Suitability as an endmember depends on the specific hydrological problem, spatial and temporal context, and isotopic separation among candidate sources. We have also harmonized the component terminology between the manuscript and database and reviewed the residual “snow” category so that the controlled vocabulary used in the archived dataset is consistent with the definitions presented in the manuscript. Assignment to components follows the description of the sampled material in the source publication; where the sampling context is insufficient to distinguish a component reliably, the observation is not assigned by inference.

4) Spatial coverage (Sects. 2.3 and 4.3)

Section 2.3 concludes that North America, with 18.8 % of the catchments, is "broadly consistent with its large global share of glacier area and volume" (L202–203). I am not sure this holds, since Table 2 gives North America about a third of the global glacier area and volume. Conversely, Europe provides nearly half of the records (L198), mostly from the Alps (L192), from only 13 catchments (Table 2). Counting catchments hides this imbalance: a single intensively sampled catchment weighs much more in the medians and tests than one with a handful of samples. The authors already acknowledge this kind of bias at L286–294, and note the lack of data from northern Europe at L484–485. I would suggest carrying this into Sect. 2.3, redoing the comparison with the number of records per region, and naming the glacierized regions that have no data. Users would then know where the dataset is representative, and where results should not be extrapolated.

We agree that the earlier wording was too strong. Section 2.3 has been revised to describe the geographical distribution of the available observations, rather than to imply that the continental coverage is proportional to global glacier area or volume. The manuscript now reports the large imbalance in record numbers explicitly and states that sampling density reflects data availability, research infrastructure, historical research effort and field accessibility rather than a balanced sampling design. Table 2 is now presented as a characterization of relative geographic coverage rather than as evidence of statistical representativeness. We have also added Supplementary Table S2, which reports the number of observations, studies and catchments for each continental group × component combination. This makes clear where large numbers of observations arise from relatively few studies or catchments. The continental discussion has accordingly been moderated, and poorly represented regions are explicitly identified as such. The revised text emphasizes that the compiled distributions should not be extrapolated to unsampled glacierized regions or interpreted as complete continental isotope distributions.

In Sect. 4.3, results are grouped by continent, while the discussion itself invokes different mechanisms: monsoon precipitation in Asia (L475), the influence of the tropical Pacific in the Andes (L448), westerly transport in Europe (L486). Each continent, however, covers several climate regimes. For Asia, for example, the monsoon and westerly domains described by Yao et al. (2013), already cited at L471, would be a natural basis for the grouping. It would also be useful to give the number of sites and sources behind each panel of Figs. 4 and 6. Several panels rest on a single site (Africa) or mainly on one study (North America >60°), and some boxplots are drawn from one value; these would be better shown as points and not interpreted.

We agree that continent is a broad geographical grouping that can encompass multiple hydroclimatic regimes. We have therefore revised Section 4.3 so that the continental figures are presented as

descriptive summaries of the observations available in the database rather than as homogeneous climatological populations. We have deliberately not introduced a new retrospective climate-zone classification because a consistent climatic-domain attribution was not available for all records and would require an additional classification framework beyond the original database metadata. To make the sampling structure transparent, Supplementary Table S2 now reports the number of observations, studies and catchments for every continental group × component combination, and Figure 4 directs the reader to this information. Categories supported by only one observation or one study/catchment are therefore identifiable and are not used to support broad continental conclusions. We also revised the text so that climatic mechanisms such as monsoonal influence, westerly transport or tropical-Pacific effects are discussed as possible contextual controls within the sampled regions rather than as properties of an entire continent.

5) Global comparisons between water types (Sects. 4.2 and 4.6)

Sections 4.2 and 4.6 explain differences between water types by processes (e.g. L379–385, L592–601), and recommend using in EMMA only the endmembers that show "a clear, significant separation" (L628–629). However, rare water types often come from one or a few sites (permafrost thaw from a single one), and precipitation was sampled over different intervals than the other components (L625–627). Pooled differences therefore mostly reflect where and when samples were taken, rather than processes. I would suggest presenting these global comparisons as descriptive, and basing any process interpretation on comparisons within the same catchment, for instance the difference between each water type and local precipitation or stream water, which removes most of the site effect.

We agree with the central point that pooled global differences should not be interpreted as direct evidence of specific hydrological processes. The revised manuscript now presents the global comparisons as descriptive empirical distributions of the observations represented in the database. We explicitly state that isotope composition depends strongly on latitude, elevation, moisture source, precipitation seasonality, accumulation conditions and melt/refreezing history, and that broad overlap among components prevents interpretation as universal isotopic fingerprints. We have also removed the recommendation that candidate endmembers should be selected simply because they show a statistically significant difference. We agree that within-catchment comparisons are preferable for process attribution. However, a systematic within-catchment source-normalization analysis would constitute a separate inferential analysis and is beyond the primary objective of this database paper. The purpose of the present analyses is therefore to characterize the compiled observations and guide users toward appropriate catchment-specific analyses using the archived metadata.

The Dunn's tests (L589, Figs. S1 and S2) should be described in the Methods, including any preceding global test and the correction for multiple comparisons. As samples are grouped by study and site and are very numerous, almost any difference becomes significant; effect sizes or overlap measures that account for site would be more informative. Similarly, the Fig. 2 regression mixes regional gradients, so its intercept should not be read as a moisture-source signal (L348–350). Local meteoric water lines per catchment would be a useful addition to the dataset.

We agree. A dedicated statistical-methods subsection has been added describing the analytical workflow, including the omnibus Kruskal–Wallis test, subsequent Dunn pairwise comparisons, multiple-comparison correction, significance threshold, treatment of missing values and the

global/continental analysis structure. Importantly, the revised statistical analysis does not treat the thousands of individual isotope records as independent replicates. The statistical comparisons are presented as exploratory support for a descriptive database paper rather than as population-level inference. We also agree regarding Figure 2. The strong interpretation of the pooled regression slope and intercept has been removed. Figure 2 is now used descriptively to show the observations relative to the established Global Meteoric Water Line, rather than to derive a new global meteoric water line or infer a unique moisture-source process from the pooled data. Local meteoric water lines would indeed be useful for specific catchments, but their robust derivation requires adequate local precipitation sampling and temporal coverage, which are not consistently available across the database. We therefore consider this an appropriate downstream application rather than a uniform addition to the present global compilation.

6) d-excess (Sect. 4.4) and agreement with the figures

Sect. 4.4 reads d-excess mostly as a moisture-source signal, while in the cryosphere post-depositional processes (sublimation, melt, refreezing) can play a major role, which the text acknowledges only briefly (L521–523). The text also states that continental recycling lowers d-excess (L556–562), whereas recycling tends to raise it, or at least not to lower it, and it is sub-cloud evaporation that lowers it. I think this section would benefit from a thorough revision, keeping in mind the cold-arid point raised above for lakes, for which lake type is not documented and the sample is small.

We agree and have substantially revised the d-excess discussion. The manuscript now distinguishes between the primary atmospheric controls on d-excess in precipitation and the subsequent modification that may occur in snow, glacier ice, meltwater, groundwater and surface waters through sublimation, snow metamorphism, evaporation, melting, refreezing, atmospheric exchange and hydrological mixing. We therefore no longer treat d-excess as a uniquely diagnostic moisture-source indicator for secondary hydrological or cryosphere components.

The interpretation of Asian d-excess has likewise been moderated and is now framed as potentially reflecting a combination of moisture-source conditions, atmospheric transport, seasonality and post-depositional modification rather than a unique dry/cool oceanic source.

We also agree with the reviewer regarding continental recycling and sub-cloud evaporation. The wording has been revised so that sub-cloud evaporation is not conflated with moisture recycling, and statements concerning recycling are presented more cautiously.

The lake interpretation has also been qualified. Given the limited number of lake observations and incomplete information on lake hydrological type and residence time, the observed distributions cannot uniquely distinguish the effects of evaporation, cold-arid conditions, throughflow or short residence time.

In several places the text and figures do not agree, which the authors will easily check. In Sect. 4.4, the ranking of snowpack, meltwater and streams (L517–526) and the description of North America >60° as elevated (L563) do not match Figs. 5 and 6. In Sect. 4.3, the medians given for Asia (L469–474), Europe (L485) and Africa (L458–459) differ from Fig. 4; for Asia, snowpack and streams appear inverted. L480 and L491 cite Fig. 5 instead of Fig. 4. Oceania is described as having a single endmember (L494), whereas the file contains three water types, and it is absent from Figs. 4 and 6. Finally, the

abstract highlights groundwater and firn as relatively enriched, whereas Fig. 3 shows snowpack melt as the most enriched type, and firn rests on samples from a single study. Other inconsistencies of this kind may exist elsewhere in the manuscript, and I would encourage the authors to cross-check the text against the figures and tables throughout.

We have cross-checked the descriptive text against the database, figures and supplementary tables and revised statements that were not supported by the plotted distributions. In particular, numerical descriptions of continental medians and the relative ranking of components have been checked against the corresponding figures; incorrect figure references have been corrected; and statements based on very small numbers of observations or a single study have been modified or removed. We have also revised the Abstract so that it no longer identifies single isotope values or individual components as representing globally characteristic enriched or depleted signatures. Instead, it emphasizes the observed distributions and substantial variability among regions and catchments. This is consistent with the revised scope of the manuscript as a harmonized database. For Oceania and other sparsely represented groups, we now avoid drawing regional conclusions from the limited observations.

Technical corrections

Please check the typography and make sure that all references cited in the text are listed in the reference section; about fifteen to twenty seem to be missing, about ten others differ in author name or year, and several listed references are never cited.

We have carried out a full cross-check of the in-text citations and reference list and corrected inconsistencies in author names, publication years and formatting. References cited in the text have been added where missing, and uncited references have been reviewed and removed where they are no longer required. The manuscript has also undergone a general typographical and consistency check, including terminology, figure numbering, isotope notation and tracked revisions.