Author Comment (AC2)

Reply to Referee comment #2 (RC2)

Manuscript: essd-2026-109

Title: A Comprehensive Global Aquatic N_2O Emission Database (GANED): Unravelling N_2O Emission Patterns from Different Water Bodies

Response to: Anonymous Referee #2 (RC2)

Original comment date: 11th June 2026

Dear Referee:

We sincerely thanks reviewer for your thorough and constructive comments on our manuscript. The reviewer has independently verified aspects of our archived data and identified reproducibility issues, data quality problems, and interpretive overstatements that require substantive correction. This Author Comment (AC) is posted by the contact author on behalf of all co-authors, in which we have carefully considered each comment and provide detailed responses below, including specific actions we will take in the revised manuscript and database. A complete point-by-point description of the changes, together with the revised manuscript will be provided upon formal resubmission following the ESSD final submission guidelines.

Reviewer comments are reproduced in *italics*, and our responses follow in regular text.

RC2 Major comments:

Comment 1

The distinction between concentration-derived (52%) and chamber-based (48%) fluxes is critical for interpretation and reuse. Users need documentation of the k and C_eq assumptions to interpret correlations involving temperature, wind, and flow. Please add fields such as flux_origin, k_value, k_method, C_{eq}_method, solubility_formula, and temperature_for_conversion, and document C_{eq} calculation assumptions.

Response: We fully agree. This is essential for users to interpret and filter flux data appropriately. In the revision, the Fluxes table will include explicit metadata describing the origin and calculation

of each flux estimate. For records derived using concentration-gradient approaches, we will revisit the original studies and add the following fields: *'Flux_Origin'* (chamber_static, chamber_floating, conc_gradient_*k*600, conc_gradient_direct_k, both_methods, or NR), *'k_value'* (reported gas transfer velocity in md^{-1}); *'k_method'* the parameterization applied (e.g., Wanninkhof 2014, Raymond 2012); *'Ceq_method' (*Method used to calculate equilibrium N_2O concentration (e.g., Weiss-1980; Weitz, 2019*)*; *'Solubility_formula'; Temp_used_conversion' (°C), 'Wind_used_for_k'* (in situ, regional station, ERA5, or not reported), and *'Atm_N_2O_used'* (the atmospheric N_2O concentration assumed in the source study). When the original publication does not provide the required information, the field will be coded as *'NR/NA'*. A new Section 2.3.3, entitled '*Flux Provenance and Methodological Comparability'*, will describe the calculation assumptions, measurement principles, and limitations associated with each flux origin. Regarding the suggested comparative analysis of chamber-based vs. concentration-gradient fluxes: we will explore this using the subset of studies that report both methods at the same sites. We acknowledge such studies may be scarce, but even a qualitative comparison would be informative. Method-stratified analyses will therefore be provided as a robustness assessment, and users will be encouraged to filter records by flux origin and calculation method for downstream applications.

Comment 2:

The correlation analysis supporting Table 2, Fig. 10, and the main driver conclusions must be fully reproducible. The unusually high correlation values (R = 0.943 for NH_4^+; R = −0.977 for DOC) require explicit and reproducible methodology; the archived script appears to contain pre-specified correlations. The current R-code should be revised so that these results can be reproduced directly from the deposited database. the table caption broadly mentions ~190 filtered rows but lacks specificity. The manuscript also contains internal inconsistencies: DOC is described in the text as a positive relationship (L453 and L615), whereas Table 2 reports a very strong negative correlation.

Response:

This is a critical reproducibility issue, and we thank the reviewer for independently checking our results. We agree that the correlation analysis must be fully reproducible and that the unusually large coefficients require careful verification. The correlation analysis in Table 2 was based on a filtered subset of records where both N_2O flux and the predictor variable had non-NA values. We will repeat the entire correlation analysis using a more robust approach: (a) we will compute

correlations for each predictor using all available pairwise complete observations; (b) we will clearly report the sample size (n) for each predictor; (c) we will flag correlations based on <30 observations as unreliable and exclude them from interpretive conclusions; (d) we will compute both Pearson (R) and Spearman rank correlations to account for skewness. We will also revise the archived R scripts so that all correlation values are computed directly from the archived data files, with fully documented filtering criteria. We will ensure that running the script on the deposited data reproduces every value in Table 2 and Fig. 10. In this way we will also resolve the DOC inconsistency in (L453 and L615), and correct all text passages where the correlation direction contradicts the table. We will explicitly document the input file, filtering criteria, handling of NA/zero/negative values, any transformations applied, and the aggregation level used. We will recheck every statement concerning the direction, magnitude, and significance of the correlations across the Abstract, main text, Table 2, and Figure 10.

Comment 3:

The overall Random Forest model has an R^2 of 0.13. This means that the model explains only about 13% of the variance in N_2O flux, leaving about 87% unexplained. This level of performance is not sufficient to support strong statements such as substrate availability being the primary global control on aquatic N_2O production. RF should be presented as exploratory. Repeated measures at the same site should be handled explicitly (e.g., 'Min River Estuary MP1 D1-D13'). Random Forest performance may improve with spatial/geomorphic predictors.

Response:

We fully agree. An R^2 of 0.13 means the model explains only 13% of the variance, and feature importance rankings from such a model cannot be interpreted as evidence for dominant global controls. The statement that the RF demonstrated "robust predictive capability" (L481) was overstated and will be removed. The RF analysis will be reframed as an exploratory demonstration of how GANED can be used, not as a definitive mechanistic analysis. We will also report feature importance with appropriate caveats: "Feature importance from this low-skill model indicates which variables contribute most to within-model prediction, but should not be interpreted as evidence for global mechanistic controls." We will add spatial predictors (coordinates, depth, surface area, elevation) as supplementary variables in a separate model version, as the reviewer suggested, and report whether predictive performance improves. This will help illustrate the

potential contribution of geomorphic and system-level factors. We will implement study-blocked cross-validation (CV) to avoid overestimating model performance due to spatial autocorrelation among records from the same site or study. We will clarify how repeated sites, multiple depths, and closely related sampling points were handled in the RF model

Comment 4:

Line 314: The '41.04% rivers' figure should be clarified as record share, not emission share. Suggested wording revisions include stating rivers accounted for the largest share of observational sites or were among the most frequently represented water-body types, rather than implying their emission contributions.

Response:

The reviewer is correct. The 41.04% figure represents the proportion of observations (flux records) in GANED that come from rivers, not rivers' actual contribution to global N_2O emissions. Presenting observational shares as emission contributions is misleading. We apologize for this error. We will revise all instances where observational proportions are presented as emission contributions. Specifically: Fig. 3b will be relabeled as "Proportion of observational records by aquatic type" rather than "proportion of water bodies contributions to net N_2O emission". All text references to percentage shares will clearly state these reflect observational coverage, not emission contributions.

Comment 5:

Several data-quality issues are visible: (i) TP range 0.04 to 4941 suggests unit mixing; DO clustering at 85–95 suggests % saturation rather than mg L^{-1}; (ii) negative NH_4^+ and NO_2^- values; (iii) Site_ID 2061, "Brisbane River Estuary 11," has an impossible latitude of −277.5509. Given the location and longitude, this is likely a misplaced digit and should be approximately −27.5509.; (iv) The authors should clarify the scope of the >3 SD filtering rule. Did this rule apply only to N_2O concentration and flux, or also to ancillary water-quality variables?

Response: We are extremely grateful to the reviewer for these careful observations, each has been verified and will be corrected in the revised manuscript. (i) For TP, value ~0.04-4941 seems more plausible as µg L^{-1} than mg L^{-1}. Our current checking identified records reported in both mg L^{-1} and

µg L^{-1}. The revised data sheet will therefore retain a source-unit field, *'TP_unit'*, and provide a harmonized concentration field, *'TP_ugL^{-1}'*, after verified conversion. Also, 36 DO records were reported as % saturation rather than mg L^{-1}. A *'DO_unit'* field will be added and a harmonized *'DO_mgL'* column derived using the Weiss (1970) formulation with in-situ temperature and salinity; (ii) for negative NH_4^+ and NO_2^- values: We will identify all negative values in the ancillary variables. Where these represent below-detection-limit (BDL) conventions (commonly reported as negative values after blank correction), we will flag them with a "Data_Flag" field. Where they appear to be errors, we will set them to NA. ; (iii) For '*Site_ID'* 2061 latitude error (-277.5509): We confirm this is a data entry error. The correct latitude is approximately *-27.5509 (Brisbane River Estuary)*. We will correct this entry and conduct a systematic validation of all latitude and longitude values against the range -90 to +90 and -180 to +180, flagging any remaining errors.; and (iv) Regarding the scope of >3 SD filtering rule: In the original version, the >3 SD filtering was applied only to N_2O concentration and flux values, not to ancillary water-quality variables. We will clarify this in the revised text and will also evaluate whether additional QC is needed for ancillary variables (as the reviewer's findings clearly indicate).

Comment 6:

Merging of nearby sites and multi-depth records (e.g., Min River Estuary MP1 D1–D13): please clarify whether these were averaged, retained separately, or treated as separate sites.

Response: We will clarify the handling of such records. Records like "Min River Estuary MP1 (D1-D13)" represent depth-specific measurements at a single station. In the current database, these are retained as separate observations with distinct Flux-Names. They were *not* averaged. The numbering (D1-D13) indicates individual depth-specific measurements. We will add a *'Depth_m'* field to the concentration and flux datasets, and we will clearly report that records with suffixes like D1, D2, etc. represent different depths at the same station. Where averaging was applied (specific cross-sectional sampling cases), we will add a *'Merged_Sites'* field indicating the number of points averaged, and the *"Aggregated-Space"* field will be marked *"Yes."*

Comment 7:

The manuscript states that The database spans '8 continents', which is misleading; Central America is a sub-region of the Americas. Terms 'Oceania' and 'Australia' should be used consistently.

Response: We agree. We will change “8 continents” to “7 continents and selected geographic regions" or simply “geographic regions.” Central America will be treated as a sub-region of North America (or the Americas) rather than as a separate continent. We will ensure consistent use of “Oceania” (rather than switching between Oceania and Australia) across the manuscript, appendix, and all data files. We have verified that the Sites dataset also contains “East Africa” as a continent entry, which will be corrected to “Africa” with “East Africa” recorded in the Site-Name field.

Comment 8:

Mean and median values should both be reported for highly skewed variables such as N_2O concentrations and fluxes. The manuscript should therefore report medians and interquartile ranges alongside means and standard deviations. This is especially important when comparing water-body types.

Response: We agree with the reviewer for this suggestion. N_2O concentrations and fluxes are strongly right-skewed, and means are strongly influenced by extreme observations. We will report medians and interquartile ranges alongside means and standard deviations in all summary tables (including Tables S2, S3, S4, and Table 2). Where we report mean values in text, we will also provide medians in parentheses.

We thank the Reviewer once again for the constructive and thoughtful comments. Following this final-response phase, we will finalize the interactive discussion via the ESSD portal and prepare a fully revised manuscript for peer-review completion, addressing all points outlined above. A complete point-by-point description of the changes will accompany the revised manuscript upon resubmission.

On behalf of all co-authors,