Responses to the comments from the 1st Reviewer

This manuscript presents an exceptionally comprehensive soil geochemical dataset that addresses a critical gap in global biogeochemical databases by systematically characterizing 1,300+ samples across 30 mountain regions spanning five climatic zones in China. The authors' methodological rigor is evident in their stratified sampling design across three pedogenic horizons (A, B, and C), standardized analytical protocols for 24 macro- and microelements, and integration with ancillary environmental variables including climatic indices, vegetation parameters, and human activity factor. The dataset's particular strength lies in its unprecedented spatial coverage of montane ecosystems, combined with vertical resolution that captures pedogenic gradients crucial for understanding soil formation processes and biogeochemical cycling.

Overall, the authors' efforts in assembling this high-resolution, multi-horizon, and climatically contextualized soil dataset are timely and scientifically significant for researchers in soil science, biogeochemistry, ecology, and Earth system modeling. Moreover, the manuscript is generally well organized, and it is suitable for publication in the journal after some minor revisions. Please find my comments below.

Responses: We appreciate your agreement with the significance of our study and the helpful suggestions to improve the manuscript. We carefully revised the manuscript based on your comments and suggestions. Please see our point-to-point response to your comments below.

Specific comments:

I recommend the authors should store the valuable data in the Zendo website.

Responses: We sincerely thank you for the suggestion regarding data archiving. Our dataset has been deposited in the National Tibetan Plateau/Third Pole Environment Data Center, and a corresponding DOI has been generated and included in the manuscript (https://doi.org/10.11888/Terre.tpdc.302620 or https://cstr.cn/18406.11.Terre.tpdc.302620). This data repository has been used by many publications in Earth System Science Data (e.g., Li et al., 2024; Jin et al., 2023; Ma et al., 2024). Importantly, this platform provides DOI-linked data access, similar to Zenodo, and allows free and direct data download without registration. Accordingly, we do not repeat the submitting our dataset on this platform.

References:

Jin, Y., Wang, H., Xia, J., Ni, J., Li, K., Hou, Y., Hu, J., Wei, L., Wu, K., Xia, H., and Zhou, B.: TiP-Leaf: a dataset of leaf traits across vegetation types on the Tibetan Plateau, Earth Syst. Sci. Data, 15, 25-39, https://doi.org/10.5194/essd-15-25-2023, 2023.

Li, X., Jin, H., Feng, Q., Wu, Q., Wang, H., He, R., Luo, D., Chang, X., Şerban, R.-D., and Zhan, T.: An integrated dataset of ground hydrothermal regimes and soil nutrients monitored in some previously burned areas in hemiboreal forests in Northeast China during 2016–2022, Earth Syst. Sci. Data, 16, 5009–5026, https://doi.org/10.5194/essd-16-5009-2024, 2024.

Ma, Y., Xie, Z., Chen, Y., Liu, S., Che, T., Xu, Z., Shang, L., He, X., Meng, X., Ma, W., Xu, B., Zhao, H., Wang, J., Wu, G., and Li, X.: Dataset of spatially extensive long-term quality-assured land–atmosphere interactions over the Tibetan Plateau, Earth Syst. Sci. Data, 16, 3017-3043, https://doi.org/10.5194/essd-16-3017-2024, 2024.

Line 123: Replace “was” with “were”. Please check other grammar issues in the manuscript.

Responses: We are very grateful for your detailed suggestions. We have corrected the sentence on line 123, replacing "was" with "were" as suggested. Following your advice, we have also performed a thorough proofread of the entire manuscript to identify and correct other grammatical issues and typos. We appreciate your help in improving the quality of our paper.

Line 132: Please specify the extraction method for pH measurement (e.g., water, KCl, or $CaCl_2$). This is essential for comparability with other pH datasets and can influence interpretation of cation exchange and element mobility.

Responses: Thank you for your valuable comment. We have revised the manuscript to specify the pH measurement method. The pH was determined using a water extraction method (with a 1:2.5 soil-to-water ratio). This clarification has been added to the Methods section (Line 132). The revised description is as follows:

Soil pH was measured using a pH meter (Mettler-Toledo FE28, Switzerland) after shaking the soil samples with deionized water at a 1:2.5 soil-to-water ratio.

Lines 154-158: The calculation of the Chemical Index of Alteration (CIA) should be more explicitly explained. Please clarify how CaO* was estimated, and whether the method has followed that of Nesbitt & Young (1982) directly or been corrected.

Responses: Thank you for pointing out a clearer explanation of the CIA calculation. In the revised manuscript, we provided a more detailed and explicit description of the method. The Chemical Index of Alteration (CIA) was calculated using the widely accepted formula proposed by Nesbitt & Young (1982):

$$CIA = \frac{Al_2O_3}{(Al_2O_3 + Na_2O + K_2O + CaO^*)} \times 100$$

where all oxides are expressed in molar proportions. As you rightly noted, CaO* should reflect only the amount of Ca derived from silicate minerals, excluding contributions from carbonates, phosphates, or exchangeable forms. To address this, we adopted a correction method following McLennan (1993), which has been applied in numerous geochemical studies to improve the reliability of CIA values. Specifically, CaO* was estimated as follows: when the measured CaO content is less than or equal to that of Na_2O, CaO* is assumed to be equal to the measured CaO; when the measured CaO content is greater than that of Na_2O, CaO* is assumed to be equal to Na_2O. We have updated the Methods section accordingly to reflect this correction.

References:

Nesbitt, H. W., and Young, G. M.: Early Proterozoic climates and plate motions inferred from major element chemistry of lutites, Nature, 299, 715-717, https://doi.org/10.1038/299715a0, 1982.

McLennan, S. M.: Weathering and Global Denudation, J. Geol., 101, 295-303, https://doi.org/10.1086/648222, 1993.

Lines 164-165: The strict coordination has been carried out, but it was not clearly defined. Does this refer to harmonization of sampling protocols across sites, or post-hoc statistical adjustments (e.g., normalization, transformation, unit standardization) to ensure cross-site comparability?

Response: Thank you for raising this important point. The "rigorous harmonization procedures" mentioned in the manuscript refers specifically to the harmonization of sampling protocols and laboratory measurement procedures across all sites. All soil samples were collected following a unified field sampling protocol and analyzed using consistent laboratory methods and instrumentation to ensure comparability of the physical and chemical data across different mountain regions. We would like to clarify that no post-hoc statistical adjustments (such as normalization, transformation, or unit conversion) were applied to the raw data. The consistency in methodology at both the field and laboratory stages eliminate the need for such adjustments and ensures that the observed variations reflect actual environmental differences rather than methodological artifacts. We have revised the manuscript to clarify this point and avoid misunderstanding:

The dataset integrates information from extensive field surveys, laboratory analyses, high-resolution satellite-derived vegetation indices, and ancillary environmental data compiled from national and global databases. To ensure data consistency and comparability across sites, all soil samples were collected following standardized sampling protocols and analyzed using uniform laboratory procedures and instrumentation.

Line 103: The manuscript would benefit from a concise description of the statistical or visualization methods used to generate Figures 2-6. This addition will help readers better interpret the trends and distributions presented.

Response: Thank you for your helpful suggestion. In the revised manuscript, we have added a new subsection titled "2.5 Statistical analysis" in the Materials and Methods section to provide a clear and concise description of the statistical and visualization methods used throughout the study. This addition aims to improve transparency and help readers better understand the analytical approaches and interpretation of the trends and patterns presented in the results. The added content is as follows:

2.5 Statistical analysis

All statistical analyses were conducted using R (version 4.3.1). To test differences in element concentrations among soil horizons, we employed linear mixed-effect models using the "lmer" function from the "lme4" package, where soil horizon was treated as a fixed factor and sampling site as a random factor. Regression analyses were conducted to examine the spatial distribution characteristics of each element. To explore the compositional differences in elemental assemblages across soil horizons and to assess the influence of environmental variables on soil element variation, redundancy analysis (RDA) was conducted using the “rda” function in the “vegan” package. Correlation analyses were conducted separately for each soil horizon to identify horizon-specific relationships between elemental concentrations and environmental drivers. Furthermore, simple linear regression was employed to quantify the individual explanatory power (R^2) of each environmental variable for each element. The cumulative explanatory power of all environmental factors was also calculated to evaluate their combined influence on element variation.

Line 260: The authors provided horizon-level sampling and vertical stratification but did not elucidate the implications for soil development modeling. Given the presence of C-horizon data and CIA indices, this dataset could serve as a valuable benchmark for soil formation modeling (e.g., using SoilGen or CLORPT frameworks). A short paragraph in Section 4 may highlight this point.

Response: Thank you for this insightful comment. Recognizing the potential value of horizon-specific data and weathering indices (e.g., CIA), we have added a short paragraph in the subsection of “Potential applications of the dataset” (Section 4) to highlight how this dataset could be used to support soil formation modeling efforts. The added content is as follows:

In addition, the inclusion of horizon-specific data (O, A, and C horizons), weathering indices, and lithological information provides valuable input for soil formation and rock weathering models. Process-based models like SoilGen or conceptual frameworks such as CLORPT (climate, organisms, relief, parent material, and time) can benefit from the dataset’s vertical resolution and environmental coverage to simulate pedogenesis, profile evolution, and mineral nutrient release across climate gradients. Accordingly, the dataset can serve as a regional benchmark for calibrating and validating long-term soil development models, particularly in mountainous regions where such data are scarce yet critically needed.

Line 316: Add a sentence summarizing the dataset structure (e.g., file formats, variable descriptions, metadata schema) to assist users in quickly understanding how to work with the data.

Response: Thank you for this thoughtful suggestion. In response, we have revised the corresponding section of the manuscript to include a brief summary of the dataset structure.

5 Data availability

The database is freely accessible via the National Tibetan Plateau/Third Pole Environment Data Center at https://doi.org/10.11888/Terre.tpdc.302620 or https://cstr.cn/18406.11.Terre.tpdc.302620 (Wu et al., 2025b). The dataset provides comprehensive information for each sample, including mountain affiliation, geographical coordinates, climatic characteristics, vegetation type, soil type, parent rock type, normalized difference vegetation index, atmospheric nitrogen deposition rates, soil physicochemical properties, chemical weathering indices, and concentrations of 24 soil elements. The data are stored in Excel spreadsheet format, accompanied by a separate data documentation file that describes variable names, units, and definitions.

Line 249: The value "Fe (>200%)" as explanatory power in redundancy analysis seems inconsistent (R^2 cannot exceed 100%). Please double-check this statement or clarify if it refers to cumulative variance.

Response: Thank you for the valuable feedback. The y-axis of Figure 6 represents the cumulative explanatory power of all environmental variables for each individual element. Therefore, the total explanatory value may exceed 100%. To avoid confusion, we have clarified this point in the Statistical Analysis subsection of the Methods section and revised the figure legend accordingly.

Fig. 6 Explanation of elemental variation by environmental factors based on regression modelling. Columns with different colors represent different environmental variables. Total height of each bar indicates the cumulative explanatory power. MAP, mean annual precipitation; MAT, mean annual temperature; AI, aridity index; DIN, dissolved inorganic nitrogen; NDVI, normalized difference vegetation index; CIA, chemical index of alteration

Tables 1 and 2: Several abbreviations used in these tables (e.g., MAT, MAP) are not defined within the table notes. As tables should be interpretable independently of the main text, please add a legend or footnotes explaining all abbreviations.

Response: Thank you for your feedback. To improve the clarity of Tables 1 and 2, we improved the legends and defined all abbreviations used in the tables.

Figures 2 and 3: Both figures lack x-axis labels, which impairs interpretability. Ensure all figures include complete and clear axis annotations, including units.

Response: Thank you for raising this important question. We have revised Figures 2 and 3 to include complete and clear x-axis labels, along with appropriate units where applicable. These additions improve the readability and interpretability of the figures. The revised figure is shown below.

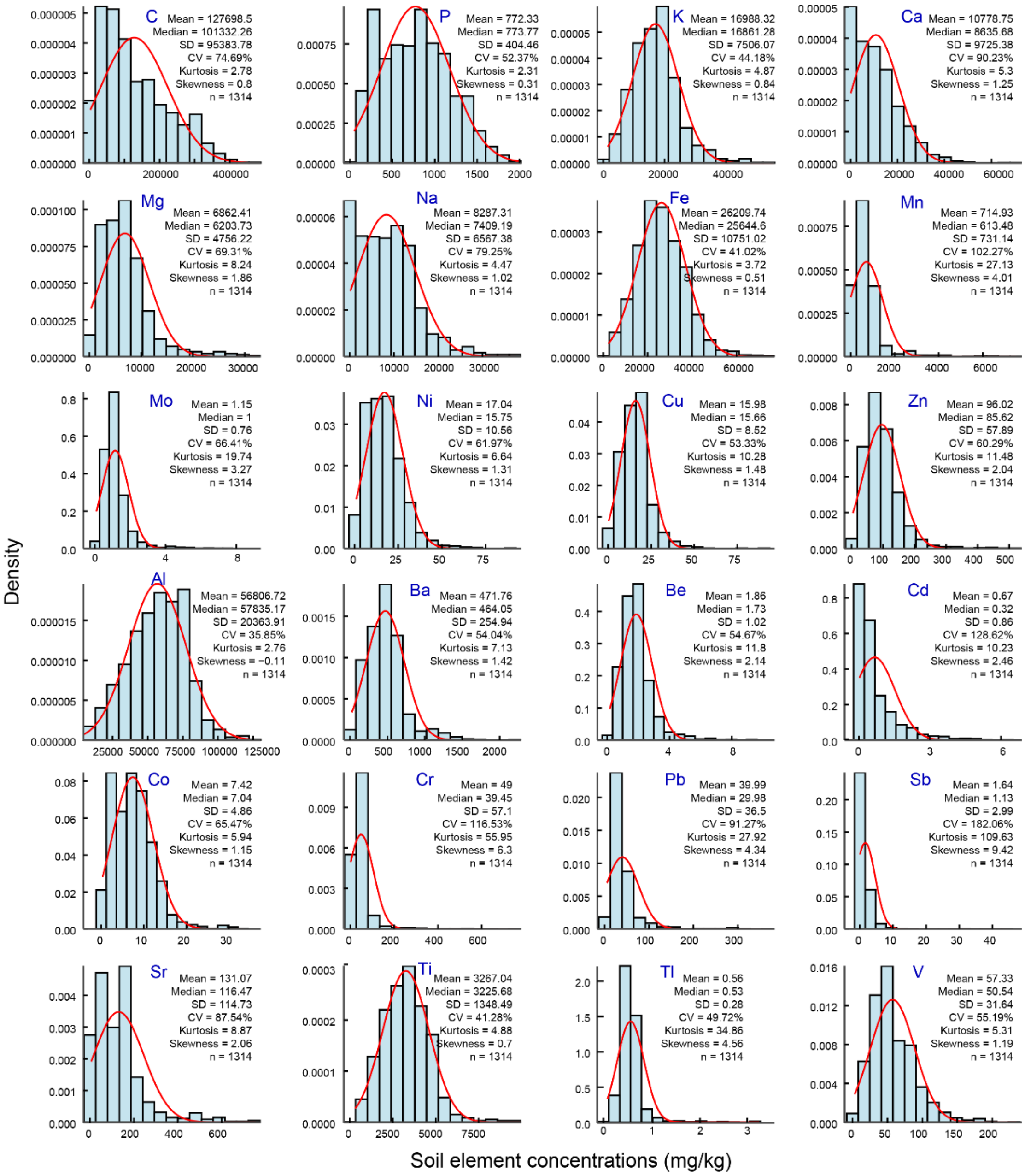

Fig. 2 Frequency distribution of soil elements across the China's mountains. Red curve on each histogram represents the fitted normal distribution. The statistical parameters

of the corresponding element are annotated in the upper left of each sub-figure.

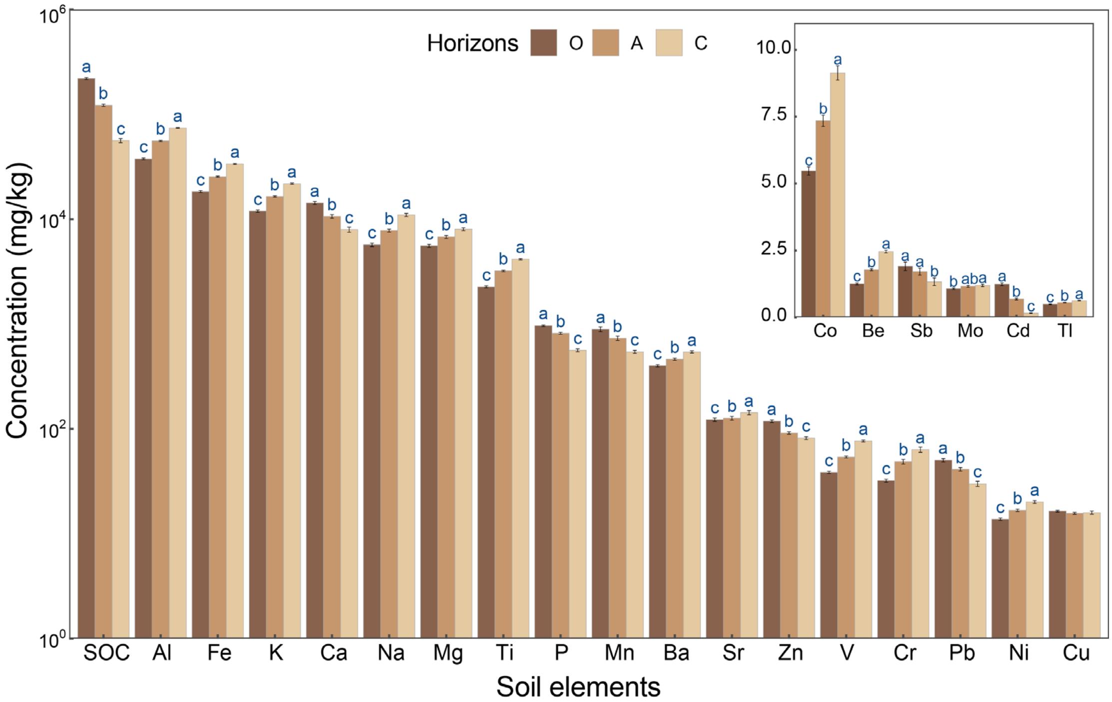

Fig. 3 Mean concentrations of 24 elements across different soil horizons. Lowercase letters indicate significant differences in each element among soil horizons ($p < 0.05$), and error bars represent the standard error.