For data description manuscript [essd-2026-180]

1st round author's response to the reviewer #2

General comment. This study addresses a critical challenge in high-altitude climatology by constructing TPHH, a long-term (1901–2023), high-resolution (1/30°) near-surface humidity, temperature, and surface pressure dataset for the Tibetan Plateau. The work is highly significant as it bridges the pre-satellite era data gap, providing much-needed spatial detail for a region characterized by extreme topographic complexity and elevation-dependent warming. The methodological innovation lies in the application of the FourCastNet architecture, which synergistically integrates Vision Transformers and Adaptive Fourier Neural Operators to map coarse-resolution signals from the CRU dataset onto a high-resolution grid governed by terrain constraints. The validation effort is commendable, employing not only a network of 135 independent meteorological stations but also an innovative biological cross-verification using 41 tree-ring proxy groups to ensure temporal fidelity in the early 20th century. This dataset could provide a foundational material for studying long-term hydrological cycles, glacier dynamics, and ecosystem responses to climate change on the Tibetan Plateau. To meet the stringent standards of ESSD, it is essential that the authors thoroughly address the concerns raised below, specifically regarding the physical grounding of the FourCastNet architecture and the reliability of reconstructions in the data-sparse western Tibetan Plateau.

Authors' response: We sincerely thank the reviewer for the positive assessment of the scientific value of TPHH and its potential applications. We also appreciate the reviewer's emphasis on the interpretation of the FourCastNet framework and on the reliability of the observationally sparse western and early-period reconstruction.

We have revised the response and manuscript to describe FourCastNet as a data-driven statistical downscaling model. The model does not impose thermodynamic or dynamical equations through its architecture or loss function; instead, it learns statistical relationships between coarse CRU_TS predictors and the terrain-related spatial patterns represented in the TPMFD target fields.

We have removed language that could imply hard physical enforcement and now distinguish learned terrain-related spatial relationships from explicit physical constraints.

We have also limited the reliability assessment to the evidence actually provided: target-referenced reconstruction-performance diagnostics, Pettitt breakpoint analysis, monthly trend analysis, DEM sensitivity, station-based consistency evaluation, and tree-ring proxy-window diagnostics.

The 135-station CMA network comprises 125 stations with observations extending to the pre-1979 period and 10 additional stations with observations only from 1979 onward; all station comparisons are described as station-based consistency evaluations because indirect information overlap with TPMFD cannot be excluded. Tree-ring analyses are presented as proxy-based support for temporal plausibility rather than validation of local-scale early-century accuracy.

The newly added analyses provide reconstruction-performance, breakpoint, trend, DEM-sensitivity, and proxy-window diagnostics. These results support temporal consistency and large-scale plausibility, but do not establish absolute accuracy or a comprehensive uncertainty estimate for the observationally sparse pre-

1950 period. Revised locations: Sections 2–6; Supplementary Table S1 and Figures S8–S47.

Major comment 1. Quantification of uncertainty in the pre-1950s observational void. The authors utilize CRU TS 4.08 as the primary predictor. It is well-documented that before the 1950s, ground-based observations within the Tibetan Plateau were virtually non-existent, and the CRU signal in this period relies heavily on spatial interpolation from distant stations outside the plateau. While the deep learning model successfully learns fine-scale "terrain textures" from the TPMFD training period (1979–2023), there is a risk that the reconstructed high-resolution features in the early 20th century are merely reflections of static topographic constraints rather than actual climatic fluctuations. The authors must provide a deeper discussion or a sensitivity analysis comparing the model's error characteristics before and after the densification of the observation network (e.g., around 1950).

Authors' response: We sincerely thank the reviewer for emphasizing the pre-1950 observational void. We agree that this is the least observationally constrained part of TPHH and that fine-scale early-century patterns cannot be interpreted on the same basis as post-1979 fields.

We have added a dedicated discussion explaining that FourCastNet can refine the spatial expression of signals present in CRU_TS but cannot recover mesoscale early-period variability that is absent from the coarse predictors. Some fine-scale features may therefore reflect stable statistical spatial modes learned from the modern TPMFD period.

A direct comparison of reconstruction errors before and after 1950 was not possible because suitable pre-1950 observations are unavailable. We therefore added indirect stability and plausibility diagnostics—Pettitt tests and monthly trends for five 30-year windows, plus target-referenced performance maps for two modern windows—and expanded the Discussion to explain this limitation. These analyses do not quantify pre-1950 error. These are stability and plausibility diagnostics rather than direct pre-1950 error estimates.

For the 1961–1990 window containing the 1978/1979 transition, 4.0% of valid temperature grid cells, 2.1% of specific-humidity grid cells, and 21.9% of pressure grid cells have Pettitt $p < 0.05$. Of all valid grid cells, significant Pettitt-selected breakpoints in 1978 or 1979 occurred in 1.43% for temperature, 0.03% for specific humidity, and 8.56% for pressure.

The revised manuscript states that these results support regional-scale temporal consistency, especially for temperature and humidity around the transition, but cannot establish absolute accuracy for the pre-1950 period or validate local fine-scale variability in the western and central high-elevation Plateau. Revised locations: Section 4.3.2; Discussion Section 6.3; Supplementary Figures S30–S47.

Major comment 2. Theoretical justification of physical consistency. The manuscript repeatedly emphasizes that the hybrid-structure deep learning approach is physically-consistent. However, FourCastNet is essentially a data-driven super-resolution model. While it effectively captures "terrain governance," it is unclear whether the model architecture or loss functions explicitly enforce thermodynamic laws (e.g., the Clausius-Clapeyron equation or the relationship between air pressure and hypsometry). The authors should clarify if this consistency is merely a spatial correlation with topography or if the model satisfies fundamental meteorological equations during the downscaling process.

Authors' response: We thank the reviewer for this important conceptual clarification. We agree that FourCastNet is a data-driven model and that its architecture and loss function do not explicitly impose the Clausius–Clapeyron relationship, the hypsometric equation, surface-energy balance, water-vapor conservation, or other meteorological governing equations.

We have revised the terminology throughout the Abstract, Introduction, Methods, Results, Discussion, and Conclusions to describe the method as data-driven statistical downscaling that reproduces terrain-related spatial relationships learned from TPMFD.

The model learns a mapping between coarse-resolution CRU_TS predictors and high-resolution TPMFD target fields during 1979–2023. Consequently, the fine-scale structures in TPHH are interpreted as learned statistical patterns associated with elevation gradients, mountain–valley contrasts, and regional hydroclimatic organization, rather than as fields produced by explicit equation-based constraints.

The DEM ablation experiment provides an additional diagnostic: explicitly adding DEM changes R^2 and IOA by less than 0.001 and yields only very small changes in RMSE and MAE. This result indicates that the explicit DEM channel supplied little additional predictive information beyond the learned CRU_TS–TPMFD mapping, and it does not demonstrate enforcement of physical laws. Revised locations: Abstract and Introduction; Methods Section 3.1; Discussion; Supplementary Figures S24–S29.

Major comment 3. Spatial Bias in Validation and Representative Uncertainty。 As shown in Figure 1 and Figure 7, the 135 validation stations are predominantly clustered in the eastern and southern edges or river valleys of the plateau. The western interior and the high-altitude “No Man's Land” (Qiangtang) are almost entirely unrepresented in the ground-truth network. Given that the model's RMSE is higher in complex terrain, the authors should provide a spatial uncertainty map or an analysis of residuals across elevation gradients to quantify the reliability of the data in these unmonitored high-altitude regions.

Authors’ response: We sincerely thank the reviewer for emphasizing the spatial bias of the station network. We agree that the eastern and southern concentration of stations limits direct evaluation over the western interior and Qiangtang.

We have added spatial maps of R^2, RMSE, MAE, and IOA for temperature, specific humidity, and pressure during 1979–2008 and 1994–2023, together with DEM/no-DEM comparisons. These maps provide spatially complete target-referenced reconstruction-performance and sensitivity diagnostics during the TPMFD overlap period.

We have also revised Figure 1 and the station terminology. Figure 1 now shows the 125 stations with observations extending to the pre-1979 period as circles and the 10 additional stations observed only from 1979 onward as triangles; together they form the 135-station network available after 1979.

The revised Discussion identifies the western and central high-elevation Plateau as weakly constrained by direct observations. The metric maps are not presented as direct uncertainty estimates for unmonitored areas, and users are advised to interpret local early-period variability cautiously. Revised locations: Sections 4.1–4.2 and Figures 1, 6, and 7; Discussion Section 6.3; Supplementary Figures S24–S29.

Major comment 4. Phenological logic of tree-ring variable selection (equation 5). In the ternary regression model for tree-ring validation, the authors selected April-May-June (AMJ) humidity as a key factor. The choice of this specific seasonal window requires stronger phenological justification. In many parts of the Tibetan Plateau, tree growth may be more sensitive to JJA (Summer) humidity or have lagged responses to winter moisture. The authors should explain why the AMJ window was selected as the primary moisture constraint across all 41 proxy groups.

Authors’ response: We thank the reviewer for requesting a clearer phenological basis for the AMJ humidity predictor. We agree that a single seasonal window should not be assumed to be optimal at every chronology.

We have expanded the Methods to explain that AMJ represents early-growing-season moisture conditions relevant to cambial reactivation, leaf expansion, root water uptake, and the initiation of xylem formation in many cold and seasonally dry tree-line environments.

We compared AMJ, MJJ, and JJA windows across the proxy groups shown in Supplementary Figures S9–S23. Group 3 is presented separately in Figure 12 for the selected window. The full group-by-group results are provided in Supplementary Figures S9–S13 (AMJ), S14–S18 (MJJ), and S19–S23 (JJA).

The revised text describes AMJ as a representative early-growing-season moisture indicator rather than a universally optimal window. The tree-ring analysis is presented as proxy-based evidence of temporal plausibility, not as proof of local climate accuracy. For the representative Group 3 example in Figure 12, the revised panel reports Corr. = 0.844 and $p < 0.001$. Revised locations: Methods Section 3.5 and Results Section 4.4; Figure 12; Supplementary Figures S8–S23.

Minor comment 1. Line 22 & Supplement (Fig S1, S2 titles): There is an inconsistency in the dataset's acronym. The supplement uses "THPP" while the main text uses "TPHH". Please standardize to TPHH throughout.

Authors' response: We have corrected "THPP" to "TPHH" in Supplementary Figures S1–S3 and checked the acronym throughout the manuscript and supplementary captions. Location: Supplementary Figures S1–S3 and captions.

Minor comment 2. Line 34: The text mentions "87 independent ground-based meteorological stations", but Line 56 and Figure 1 refer to "135 stations". Please verify the exact number used for historical validation (1901–1978).

Authors' response: We have removed the earlier 87-station category. The revised manuscript reports a total of 135 CMA stations, of which 125 have observations extending to the pre-1979 period and the remaining 10 have observations only from 1979 onward. Because indirect overlap with TPMFD cannot be excluded, neither network is labeled as an independent validation set. Revised locations: Section 2.3 and Figures 1 and 7.

Minor comment 3. Line 225 (Equation 1): The formatting for the R2 formula is slightly garbled in the PDF, with overlapping characters in the denominator. Please re-typeset using standard LaTeX.

Authors' response: We have re-typeset Equation 1 using standard mathematical notation and checked the denominator, summation limits, and superscripts in the revised PDF.

Minor comment 4. Line 290: The text states, "As demonstrated in Table 1...", but the performance metrics are actually presented in Table 3. This is a citation error.

Authors' response: We have corrected the citation from "Table 1" to "Table 3".

Minor comment 5. Lines 505, 520, 530 of References: Several citations are listed for 2025 or 2026 (e.g., Jiang et al., 2025; O'Neill et al., 2025). Please indicate if these are preprints or currently in press.

Authors' response: We have checked the cited 2025/2026 references and identified each as a published article, preprint, or in-press item as appropriate, with DOI and bibliographic details added where available.

Minor comment 6. Figure 6 (RMSE Heatmap): The color scale for specific humidity is quite narrow (max 0.40 g/kg). It is difficult to discern spatial variance in the plateau interior. Consider using a more contrast-enhancing color ramp for the humidity panel.

Authors' response: We have revised the specific-humidity RMSE color scale in Figure 6 to improve contrast over the Plateau interior while preserving the full data range.

Minor comment 7. Figure 12: The Pearson correlation coefficient (0.82) is unexpected high. Please include the significance level (p-value) to strengthen the statistical argument.

Authors' response: We have revised Figure 12 and its caption to report Corr. = 0.844 and $p < 0.001$ for the representative Group 3 experiment.

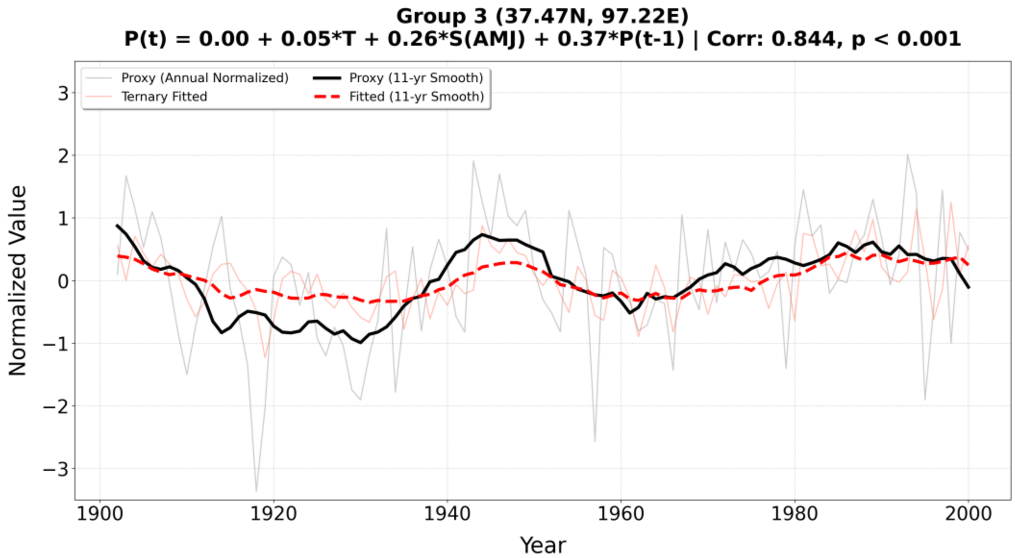

Figure 12. Temporal consistency between the actual tree-ring proxy and the TPHH-based ternary fitted series (1901–2000). The fitted series is generated via a standardized ternary regression model integrating TPHH JJA temperature (T), AMJ specific humidity (S), and lag -1 biological memory of the tree-ring proxy (P). Bold lines denote the 11-year moving average, highlighting the low-frequency correlation. See Supplementary **Figures S9–S23** for the complete validation across proxy groups.