Responses to the community commenter's comments

1. CC1: The manuscript should clarify the novelty of the proposed framework relative to conventional hydrological regionalization, environmental similarity classification, and machine-learning transfer approaches. The authors should explain more explicitly how the DCD-C trajectory framework provides information that cannot be obtained from static permafrost or hydroclimatic classifications.

   AC1: We thank the commenter for noting that the novelty of the framework was not sufficiently clear. We have revised the Introduction to clarify that the contribution of this study is not a new clustering algorithm or machine-learning model, but a framework for identifying local vertical hydrothermal correlation structures and representing their temporal evolution across data-sparse permafrost regions.

   Conventional hydrological regionalization commonly transfers model parameters, hydrological signatures, or simulated responses. Environmental-similarity approaches classify regions using static climatic, topographic, or permafrost attributes, while machine-learning methods generally map environmental covariates to target variables. These approaches support regional prediction, but environmental similarity does not necessarily imply similarity in the vertical organization or temporal evolution of permafrost-related hydrological responses.

   The DCD-C framework addresses this gap by summarizing, within an information-rich reference basin, the depth distribution and integrated magnitude of the correlation between soil-ice loss and runoff. Successive DCD-C states are then organized as temporal trajectories, thereby retaining information on structural evolution that is absent from static permafrost or hydroclimatic classifications.

   The framework further addresses the information asymmetry between the reference basin and target regions. Soil-ice and runoff information is required only to define the reference trajectory archetypes. Publicly available climatic, topographic, and soil-thermal variables are then used to represent target units relative to these archetypes, rather than to reconstruct their DCD, C, or subsurface processes directly. OOD screening identifies where target environments fall outside the support of the reference domain. The novelty therefore lies in linking vertical structure identification, temporal trajectory representation, regional archetype transfer, and explicit applicability control within a single scale-transfer framework. Location of change: Lines 91-147 and 347-350.

2. CC2: The selection of a fixed three-month lag for calculating depth-dependent coupling requires stronger justification. Although a sensitivity analysis is mentioned, hydrological response times may vary spatially, seasonally, and among permafrost regimes. The authors should discuss whether using one lag for all grid cells could influence the estimated dominant control depth and coupling strength.

   AC2: We thank the commenter for noting that hydrological response times may vary among spatial units, seasonal conditions, and permafrost types. The original description of this setting as a “fixed three month lag” was inaccurate. The implemented analysis uses a common maximum lag horizon of τ_{max}=3 months. Specifically, correlations between the soil ice loss signal and precipitation adjusted runoff are calculated at lags of 0, 1, 2, and 3 months. The positive contributions across this range are then summarized to calculate DCD and C. This setting does not assume a 3-month response time for every spatial unit. It provides a common intraseasonal lag range for spatial comparison, while the correlation contributions may occur at different depths and lag positions within the 0-3 month range.

   We have clarified this procedure in the revised Methods and added a sensitivity analysis for τ_{max}=1-6 months (Fig. S1). The results show that the maximum lag horizon affects the absolute value of C and the class boundaries of some trajectory archetypes. Because C is the discrete sum of positive correlations across the included soil layers and lag values, an increase in C with a longer lag horizon is expected from its definition. The sensitivity analysis therefore focuses mainly on changes in DCD and the consistency of the trajectory classification. Changes in the absolute value of C are interpreted in light of this metric definition.

   For the settings neighboring the primary analysis, the median DCD values were 1.159, 1.210, and 1.179 m for τ_{max}=2, 3, and 4 months, respectively. Relative to τ_{max}=3, the matched label agreement was 71.3% and the ARI was 0.356 for τ_{max}=2. For τ_{max}=4, the matched label agreement was 83.4% and the ARI was 0.574. The complete results for 1-6 months are provided in the revised Fig. S1.

These results show that the overall DCD level remains similar under neighboring lag horizons, while the local assignment of trajectory archetypes shows some sensitivity to the selected maximum lag. We therefore retain τ_{max}=3 months as a common intraseasonal comparison range. It is no longer described as the uniquely optimal lag or as a fixed response time applicable to all spatial units. This common range supports comparison across spatial units but is not used to determine the optimal response time of each individual unit.

Location of change: Lines 255-299; Fig. S1.

3. CC3: The choice of four clusters appears to rely partly on physical interpretability. The manuscript would benefit from a clearer quantitative justification using cluster-validity statistics, stability analysis, or resampling. The authors should also indicate whether the four archetypes remain stable under reasonable changes in window length, time step, or clustering initialization.

AC3: We thank the commenter for the suggestions concerning the number of clusters, the temporal window configuration, and initialization stability. The revised manuscript now specifies the trajectory inputs, feature standardization, and clustering parameters. We evaluated the retained k=4 solution from four perspectives: candidate cluster numbers, random initialization, spatial subsampling, and sensitivity to window length and step size.

For each spatial unit, the DCD and C values from eight successive time windows were combined into a 16-dimensional trajectory vector. The 16 features were standardized separately across all reference basin units before K means clustering. Candidate values from k=2 to k=8 were compared using the silhouette coefficient, Calinski-Harabasz index, Davies-Bouldin index, within cluster sum of squares, minimum cluster proportion, and clustering stability. No single statistical metric uniquely supported k=4. Smaller values of k performed better for some separation metrics but merged structures with different trajectory characteristics. Larger values of k produced more fragmented classes. Under the retained k=4 solution, the four archetypes accounted for 18.1%, 19.0%, 34.5%, and 28.4% of the reference basin. No very small class was produced. By comparison, the minimum class proportion decreased to 12.0% for k=5, which mainly further divided the k=4 archetypes. The complete results are shown in Figs. S2-S4. We therefore retain k=4 as a compromise among statistical separation, class balance, stability, and trajectory interpretability. It is not treated as a uniquely optimal solution selected by a single metric.

The random initialization analysis used 100 random seeds with n_init=30 for each fit. Relative to the baseline solution, the median matched label agreement was 0.998, with a minimum of 0.961. The median ARI was 0.997, with a minimum of 0.910. In 100 independent analyses using 80% spatial subsamples, the median matched label agreement was 0.959 and the median ARI was 0.901. Their minimum values were 0.702 and 0.531, respectively. These results show that the k=4 solution is highly stable to random initialization and maintains high overall consistency under spatial sampling perturbations. Some boundary units can still change class. The complete distributions and statistics are provided in Fig. S6 and Table S1.

We also compared alternative window lengths and step sizes. Relative to the primary setting of a 20 year window and a 5 year step, the DCD spatial correlations were 0.918 and 0.861 under the 15 year/5 year and 25 year/5 year settings. The corresponding correlations for standardized C were 0.944 and 0.910. Matched label agreement was 0.656 and 0.817, and the ARI was 0.440 and 0.579, respectively. Under the 20 year/1 year and 20 year/7 year settings, the DCD correlations were 1.000 and 0.943, and the standardized C correlations were 1.000 and 0.943. Matched label agreement was 0.685 and 0.682, and the ARI was 0.525 and 0.481, respectively.

The 20 year/1 year setting uses the same 20 year data at the window centers shared with the primary setting. The continuous DCD and C fields are therefore identical at these shared centers. However, this setting includes more intermediate windows and produces a denser temporal trajectory. The independently clustered class boundaries can therefore differ. The complete results are provided in Fig. S5 and Table S2.

Overall, the continuous DCD-C structure remained highly consistent across the tested window lengths and step sizes. Discrete class membership was more sensitive to the temporal window configuration and to units near class boundaries. The revised manuscript therefore describes the four groups as trajectory archetypes with transitional boundaries. They are no longer presented as strictly separated natural classes that remain unchanged under all parameter settings.

Location of change: Lines 300-323; Figs. S2-S6; Tables S1-S2.

4. CC4: The authors are strongly recommended to cite recent work demonstrating the value of assimilating remotely sensed environmental information into integrated hydrological models. In particular, Zafarmomen, N., Alizadeh, H., Bayat, M., Ehtiat, M., and Moradkhani, H; is relevant to the manuscript's broader discussion of transferring and constraining hydrological process information using spatially distributed observation.

   AC4: We thank the commenter for recommending this relevant study. We have cited Zafarmomen et al. (2024) in the revised Introduction as a recent example of integrating remotely sensed information, data assimilation, and physically based hydrological modelling. The corresponding reference has also been added to the reference list.

   Location of change: Lines 85-90; References currently at lines 858-861.

5. CC5: The national transfer model achieves an overall accuracy of 58.15% and a Kappa coefficient of 0.41. These values indicate only moderate predictive agreement. The limitations of this performance should be emphasized more clearly, especially when interpreting detailed regional patterns and temporal transitions outside the reference basin.

AC5: We thank the commenter for noting that the interpretive scope of the national classification should be clarified given the overall accuracy of 58.15% and the Kappa coefficient of 0.41. The revised manuscript now explains that the model reproduces the main regional differences among trajectory archetypes within the reference basin, but confusion remains in locally complex transition zones. The framework is therefore more suitable for regional-scale comparison of trajectory archetypes than for definitive interpretation of fine spatial boundaries or individual grid-cell assignments.

We added diagnostics based on the maximum class probability and probability margin. Across the national mapping domain, the median maximum class probability was 0.461 and the median probability margin was 0.136, indicating substantial spatial and archetype-specific variation in assignment clarity. Their spatial distributions and comparisons between OOD and non-OOD areas are presented in Fig. S7. Type-specific statistics, including comparisons between environmentally supported and OOD Type 4 cells, are provided in Table S3(b). Environmental support further constrains the interpretation of the national results. Under the primary threshold of 7.667, 36.9% of the mapped permafrost grid cells were identified as OOD, and the entire northeastern domain defined in this study fell outside the predictor-space support of the reference basin. OOD coverage and spatial agreement under alternative thresholds are reported in Table S3(a). Archetype assignments in these areas represent statistical correspondence to the reference trajectory archetypes, while their local physical interpretation is limited by the environmental coverage of the reference domain.

Accordingly, the revised manuscript positions the national classification as a regional representation of broad trajectory-archetype patterns and adopts more cautious language for transition zones with small probability margins and for OOD areas. The relevant interpretations in the Results, Discussion, Abstract, and Conclusions have been revised to avoid inferences beyond the support of the model. Location of changes: Lines 46-53, 351-357, 455-471, 624-639, and 678-684; Fig.

S7; Table S3(b).

6. CC6: The out-of-distribution screening is an important component of the framework, but its implementation is not described in sufficient detail. The authors should specify the statistical distance, threshold, or predictor-space criterion used to identify OOD grid cells and provide a sensitivity analysis showing how the mapped applicability boundary changes with the selected threshold.

   AC6: We thank the commenter for noting that the original manuscript did not provide enough detail on the OOD screening algorithm, threshold selection, and threshold sensitivity. The revised manuscript now clarifies that OOD screening is performed in the environmental predictor space used by the regional transfer model. We have added the OOD score definition, the basis for the primary threshold, and a threshold sensitivity analysis. Figure 3 and the related results have been updated accordingly. OOD screening is used to indicate whether target conditions extend beyond the support represented by the reference basin. It does not change the trajectory archetype assigned by the GAM.

   For each time window, TPI, VD, TPD, DEM, latitude, and precipitation were standardized using scalers fitted to the reference basin data. For each national grid cell i, the predictor space OOD score was defined as

$$S_i = \max_{w,j} \left| z_{iwj} \right|$$

   where w denotes the eight time windows, j denotes the environmental predictors, and z_{iwj} is the standardized value calculated from the corresponding reference basin mean and standard deviation. The primary threshold was set to 7.667, which is the empirical maximum of S_i across all reference basin units. A target grid cell was labeled as OOD when $S_i>7.667$. This threshold represents the maximum standardized deviation across predictor-window combinations. A target cell exceeds this boundary when at least one predictor in one time window has a standardized deviation larger than the maximum observed in the reference domain.

   We also evaluated two stricter thresholds based on the 99th and 99.5th percentiles of the reference S_i distribution, with values of 4.122 and 4.755. Under the 99th percentile, 99.5th percentile, and reference maximum thresholds, the national OOD fractions were 66.2%, 59.1%, and 36.9%, respectively. The corresponding area weighted fractions were 65.0%, 57.7%, and 34.6%. Within the QTP, the OOD fractions were 57.3%, 48.4%, and 20.4%. The northeastern permafrost region was entirely classified as OOD under all three thresholds. Relative to the primary threshold, the agreement values for the 99th and 99.5th percentile thresholds were 0.708 and 0.778. The corresponding Jaccard indices were 0.558 and 0.625. The complete statistics are reported in Table S3(a), and the corresponding spatial boundaries under the three thresholds are shown in Figure C1.

These results show that the local OOD boundary within the QTP changes with the threshold, while the conclusion for the northeastern high latitude permafrost region remains stable. The revised manuscript therefore treats OOD as a diagnostic of predictor space support and states clearly that it does not alter the assigned trajectory archetype.

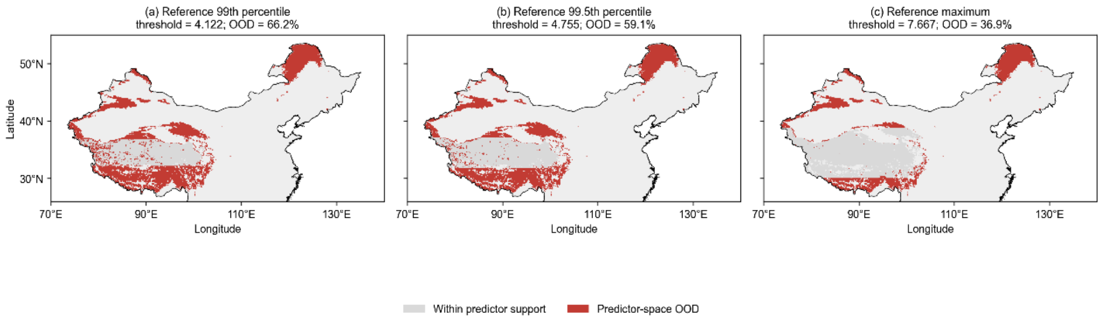

Figure C1. Sensitivity of predictor-space OOD boundaries to the reference 99th-percentile (4.122), 99.5th-percentile (4.755), and maximum (7.667) thresholds. Red indicates predictor-space OOD cells, and gray indicates cells within predictor support.

Location of change: Lines 359-373 and 449-485; Fig. 3; Table S3(a).

7. CC7: Some physical interpretations of regime transitions appear stronger than the available evidence supports. For example, transitions among the four statistical archetypes are interpreted as changes in active-layer storage, groundwater connectivity, and thaw-driven flow paths. These interpretations should be presented more cautiously unless they can be validated using independent observations of

active-layer thickness, groundwater, soil moisture, or baseflow.

AC7: We thank the commenter for pointing out that some of the physical interpretations of the trajectory archetypes and their transitions were stronger than supported by the available evidence. We have systematically reviewed and moderated the relevant statements in the revised manuscript.

Meanwhile, to maintain consistency between Figure 4 and the revised OOD definition, we updated the calculation and presentation logic of Figure 4 accordingly. As in Figure 3, OOD is used only as an overlay indicating environmental support and does not participate in archetype assignment or transition calculation.

In the Discussion, we removed or moderated statements that directly attributed changes in DCD-C to active-layer thickening, changes in storage capacity, enhanced groundwater connectivity, or the development of thaw-driven flow paths. These mechanisms are now presented only as possible interpretations consistent with previous permafrost hydrology studies, using conditional wording such as “may reflect”, “is consistent with”, and “could be associated with”.

Accordingly, Figure 4 presents statistical changes in trajectory-archetype assignments between the early and late periods, rather than direct observations or causal identification of specific subsurface hydrological transitions.

Location of change: Lines 324-327 and 506-621; Fig. 4 and its caption.