Responses to the referee's comments

1. RC1: The three-way classification of existing extrapolation pathways in the Introduction is somewhat confusing to me. The distinction among data-driven transfer, remote sensing/data assimilation, and physically based regionalization is not entirely clear, as these approaches can overlap substantially in practice. In particular, data assimilation is more of a state-constraining approach than an extrapolation pathway in the strict sense, while physically based large-scale modelling could be positioned more explicitly within this classification. I suggest that the authors clarify the logic behind this classification and explain whether these pathways are intended to be mutually exclusive categories or complementary and partially overlapping approaches.

   AR1: We thank the reviewer for pointing out that the original manuscript did not organize the existing approaches to regional-scale extension with sufficient clarity. In the original version, data-driven transfer, remote sensing/data assimilation, and physically based regionalization were presented as three parallel extrapolation pathways. This formulation did not adequately reflect their distinct functions in practice or the ways in which they are interconnected. In particular, data assimilation primarily serves to update and constrain model states using observational information, rather than constituting an independent extrapolation pathway in the strict sense.

   In the revised Introduction, we have therefore replaced the original three-part classification with a functional organization based on the roles that different approaches play in regional-scale extension. Data driven and machine learning methods primarily use observations and environmental covariates to establish spatial predictive relationships. Physically based models support regional process simulation and may be extended to data sparse regions through parameter or model regionalization. Remote sensing provides spatially extensive environmental and state information, while data assimilation incorporates such information into land surface or hydrological models to update and constrain simulated states. We have also expanded the discussion of large scale physically based modelling and added a recent example demonstrating how remote sensing, data assimilation, and physically based hydrological modelling can be integrated within a single analytical framework. The placement of relevant references has also been adjusted to better match the specific points they support.

On this basis, the revised manuscript further clarifies that the present study does not propose a new extrapolation algorithm parallel to these existing approaches. Instead, it first uses simulated soil ice information and long-term runoff records from a reference basin to define comparable archetypes of vertical soil profile structure, and then uses publicly accessible regional datasets with continuous spatial coverage to represent target regions relative to these reference archetypes.

Location of change: Lines 75-90.

2. RC2: The conceptual framing and motivation of the proposed framework need to be strengthened. The Introduction summarizes several existing approaches for largescale assessment of permafrost hydrological effects. However, the manuscript does not clearly explain what these approaches have already achieved, what specific limitations remain unresolved, and how these unresolved limitations lead to the need for the proposed reference-basin and trajectory-based archetype framework. As a result, the transition from the literature review to the proposed method appears somewhat incomplete. Related to this issue, several key concepts are introduced before they are sufficiently defined, including "permafrost-hydrological coupling", "process space," "temporal trajectories," and "coupling archetypes." The introduction of many new terms within a short section makes it difficult for the reader to fully understand the structure and rationale of the proposed framework. I suggest that the authors revise the Introduction to provide a more explicit logical chain: first, define permafrost-hydrological coupling as used in this study; second, explain why existing large-scale approaches cannot adequately represent or transfer such subsurface hydrothermal coupling structures; third, introduce the rationale for using the Source Region of the Yellow River as a process-constrained reference

basin; and finally, briefly explain what is meant by the DCD-C process space, temporal trajectories, and coupling archetypes. This would make the motivation and structure of the proposed framework much easier to follow.

AR2: We thank the reviewer for this constructive suggestion. We have reorganized the Introduction accordingly and clarified the conceptual basis and motivation of the proposed framework.

The revised Introduction defines permafrost-hydrological coupling as a structural relationship between changes in permafrost conditions and basin scale hydrological responses, characterized by vertical organization within the soil profile and temporal evolution. We further use the depth-dependent statistical association between a monthly soil-ice loss signal and runoff response as its operational representation. The soil-ice loss signal is derived from the positive month to month decrease in simulated soil-ice content. Its calculation and the preprocessing of runoff are described in greater detail in the revised Methods section.

The revised manuscript also clarifies that existing data-driven prediction, remote sensing, data assimilation, and physically based modelling approaches already support regional state estimation, observational constraint, and process simulation. Physically based models can represent freeze-thaw processes, hydrothermal transport, and their hydrological responses. However, such results generally depend on specific model structures, parameterizations, and data conditions, and therefore do not automatically provide a unified structural representation that can be directly compared and transferred across regions. Existing regional products are commonly organized around individual states, fluxes, model parameters, or hydrological responses, whereas vertical soil profile structure and its evolution across successive time windows are less often treated as a unified object of regional representation. In addition, long-term soil ice, vertical thermal state, and runoff data required to identify such structures remain difficult to obtain continuously across extensive permafrost regions. Regional extension therefore cannot assume that all target regions possess process information comparable to that available in the reference basin.

The revised manuscript accordingly positions the Source Region of the Yellow River as a process-informed reference domain. To avoid repetition and to address the related concerns raised in RC3 and RC4, the detailed rationale for selecting the SRYR, its frozen-ground and hydrothermal gradients, and its environmental relationship with the national transfer domain have been consolidated in the expanded Study area section (Sect. 2.1). And the revised Introduction now also briefly defines the DCD-C space, temporal trajectories, trajectory archetypes, and OOD screening.

At the regional representation stage, target regions are not required to have soil-ice profiles or long-term runoff records comparable to those available in the reference basin. Instead, publicly available climatic, topographic, and soil-thermal proxy variables with continuous spatial coverage are used to establish statistical relationships between target environments and the reference trajectory archetypes. This procedure does not directly reconstruct DCD, C, or soil-ice profiles in the target regions; rather, it represents each target unit relative to the trajectory archetypes defined in the reference domain. OOD screening is then used to identify where target environmental conditions extend beyond the support represented by the reference domain, thereby delimiting the interpretive boundary of the regional transfer.

Location of change: Lines 70-74, 91-147, and 149-177.

3. RC3: The current study-area description is somewhat too focused on justifying the use of the SRYR as a reference basin. This rationale is useful but may be more appropriate for the Introduction. The Study area section should provide a more detailed physical description of the SRYR and the national transfer domain, including geography, climate, elevation, permafrost distribution, and hydrological characteristics. This would make the regional context clearer before the datasets and methods are introduced.?

AR3: Thank you for pointing out that the original manuscript provided insufficient information on the physical geographical background of the study areas. In response, we retained the main rationale for selecting the Source Region of the Yellow River (SRYR) as the reference basin in the Introduction and expanded the Study area section.

The revised manuscript now describes the geographical location, elevation, climate, precipitation pattern, and runoff contribution of the SRYR. We also added information on the coexistence and internal spatial gradients of continuous permafrost, sporadic permafrost, and seasonally frozen ground within the basin. In addition, we clarified that the national transfer domain includes the major permafrost regions of the Qinghai-Tibetan Plateau and the northeastern cold region, covering contrasting high-elevation and high-latitude permafrost environments that differ substantially in latitude, elevation, climate, topography, and soil thermal conditions. These revisions more clearly establish the geographical context of the reference basin and the environmental settings to which the trajectory archetypes are regionally represented, while clarifying that the SRYR serves as a process-informed reference domain rather than a statistically representative sample of all permafrost regions in China.

Location of change: Lines 149-177; References currently at lines 784-786, 799-801, and 855-857.

4. RC4: The organization of Section 2.1 could be improved. The current subsection combines study-area description, dataset information, and the definition of derived thermal indicators, which makes the structure somewhat unclear. I suggest separating the study-area description and the datasets into different subsections. In addition, because soil ice content from previous WEB-DHM-pf simulations is a key input for constructing the DCD-C process space, the authors should briefly explain the reliability and validation of these model outputs. Finally, TPI, TPD, and VD are not simply datasets but derived process-relevant indicators; therefore, the authorsshould consider whether their definitions would be more appropriate in a methods subsection rather than in the data description.

AR4: We thank the reviewer for pointing out that the original Sect. 2.1 combined the study area description, data sources, and derived indicators, which made the organization of the section and the relationship between datasets and subsequent analyses unclear. We have therefore reorganized the Data and Methods section.

The revised manuscript separates the original "Study area and datasets" subsection into two independent subsections, "Study area" and "Datasets". Sect. 2.1 now focuses only on the geographic, climatic, and permafrost background of the SRYR and the national transfer domain. Sect. 2.2 describes the data sources, spatiotemporal resolutions, and specific applications of the datasets used for the reference basin analysis and national-scale representation. We have also added a brief description of previous WEB-DHM-pf validation studies in Sect. 2.2. These studies evaluated the model using observations of runoff, evapotranspiration, soil temperature, soil moisture, seasonal freezing depth, and active layer thickness. The revised manuscript also clarifies that soil ice content from WEB-DHM-pf is used as simulated input data rather than direct observations.

In addition, TPI, TPD, and VD are derived indicators calculated from soil temperature data rather than original datasets. Therefore, we moved their definitions, calculation methods, and physical interpretations out of the data description and introduced a separate subsection entitled "Derived soil-thermal indicators". The subsequent procedures, including DCD-C construction, trajectory clustering, and regional representation, are also organized into independent subsections. This revised structure clearly separates the study-area background, original datasets, derived variables, and subsequent analytical procedures.

Location of change: Lines 149-236 and 328; References currently at lines 816-826.

5. RC5: The construction of the DCD-C process space and the trajectory clustering procedure require further clarification. The coupling strength C is derived from the absolute value of the lagged correlation between soil ice content and outlet runoff. This is a useful diagnostic, but it should be described more cautiously as a correlation-based proxy rather than a direct measure of physical coupling. The

authors should clarify whether the monthly time series were detrended, deseasonalized, or standardized before the correlation analysis, and whether temporal autocorrelation and common climatic forcing were considered. The choice of a fixed lag of $\tau = 3$ months also needs stronger justification. Since DCD, C, the temporal trajectories, and the subsequent clustering all depend on this lag, the authors should summarize the sensitivity of the results to alternative lag choices in the main text, rather than only referring readers to the Supplement. The trajectory clustering procedure should be described more transparently. The manuscript should specify how the DCD-C trajectories were represented as inputs to k-means, whether DCD and C were standardized, and how cluster stability was assessed. In addition, the equations should be numbered?

AR5: We thank the reviewer for noting that the construction of the DCD-C space and the trajectory clustering procedure required further clarification. Following this comment, we rechecked the implementation against the original Methods. The original text oversimplified the input signal, correlation treatment, and clustering input and therefore did not fully document the workflow used to generate the reported results. The revised manuscript now provides the corrected definitions, equations, temporal treatment, and clustering settings. These revisions document the implemented analysis more accurately; the underlying analysis used to generate the reported results has not been replaced after review.

The analysis first constructs a monthly soil ice loss signal from the positive decrease in simulated soil ice content between adjacent months:

$$M(t,z) = \max[0, VIC(t-1,z) - VIC(t,z)]$$

where t denotes the ending month of the soil ice decrease and z denotes soil depth. This variable is a diagnostic signal derived from changes in simulated soil ice content. It is not the absolute soil ice content or a directly simulated melt flux.

Within each 20 year sliding window, outlet runoff is first residualized using a linear regression on concurrent and one month lagged basin mean precipitation:

$$Q_{res}(t) = residual\{Q(t) \sim 1 + P(t) + P(t-1)\}$$

Pearson correlations are then calculated between the soil ice loss signal and precipitation adjusted runoff for each soil layer and for lags from 0 to 3 months:

$$r(z,\tau) = corr[M(t,z), Q_{res}(t+\tau)], \tau = 0,1,2,3$$

Only the positive component of the correlation is used:

$$r^{+}(z,\tau) = \max[r(z,\tau), 0]$$

The metric C is defined as the discrete sum of positive correlations across all soil layers and lag values:

$$C = \sum_{z}\sum_{\tau=0}^{3} r^{+}(z,\tau)$$

and DCD is defined as the soil depth centroid weighted by these positive correlation contributions:

$$DCD = \frac{\sum_{z} z \sum_{\tau=0}^{\tau=3} r^{+}(z,\tau)}{\sum_{z}\sum_{\tau=0}^{\tau=3} r^{+}(z,\tau)}$$

Accordingly, C is interpreted as an aggregated positive correlation metric over the soil profile and lag range. DCD represents the weighted depth centroid of these positive contributions. The revised manuscript no longer describes C as a direct measurement of physical coupling strength.

We have also clarified the temporal preprocessing. Before the correlation analysis, the monthly series were not additionally deseasonalized, linearly detrended, or standardized using a z score. Pearson correlation performs mean centering and variance normalization during calculation. No explicit correction for temporal autocorrelation was applied. Runoff residualization reduces the direct influence of concurrent and one month lagged precipitation covariation. It does not remove temperature, snow, or other shared climatic controls.

The correlation coefficients were not used for significance testing. Given the absence of an explicit autocorrelation correction and the remaining shared climatic influences, DCD and C are used as descriptive correlation based diagnostics. They are not used to infer causality or an independent effect of permafrost degradation.

We further clarify that the 3 months setting represents a maximum lag range,

$\tau_{max} = 3$. Correlations are calculated at 0, 1, 2, and 3 months. DCD and C summarize the positive contributions across this full lag range. This setting provides a common intraseasonal range for considering both concurrent and short delay responses. It does not assign a uniform 3 month response time to all locations.

Following the reviewer's suggestion, we summarized the main lag-sensitivity results in the revised manuscript. The sensitivity analysis was conducted at the native 5 km reference-basin scale, where the depth-lag correlation structure and the corresponding DCD and C metrics were initially calculated. For each candidate τ_{max} from 1 to 6 months, we recalculated DCD and C, reconstructed the 16-dimensional trajectories, standardized the trajectory features, and repeated the clustering independently. Cluster labels were aligned through optimal label matching before matched label agreement was calculated.

The median DCD values obtained with τ_{max}=2, 3, and 4 months were 1.159, 1.210, and 1.179 m, respectively. Because C is defined as a discrete sum over the included soil layers and lag values, its median value increased from 0.969 at τ_{max}=2 to 1.482 at τ_{max}=3 and 2.156 at τ_{max}=4 as additional lag values were included. Relative to τ_{max}=3, the independently derived trajectory-archetype classifications for τ_{max}=2 and 4 had matched label agreement rates of 71.3% and 83.4%, with adjusted Rand index values of 0.356 and 0.574, respectively.

These results indicate that the characteristic DCD remains similar within the neighboring 2-4 month range, whereas the magnitude of C and discrete archetype membership are more sensitive to the selected maximum lag horizon. We therefore retained τ_{max}=3 months as a common intraseasonal comparison horizon. We do not interpret it as a uniquely optimal or universally applicable response lag. Complete results for τ_{max}=1-6 months are reported in Fig. S1.

The revised manuscript also specifies the trajectory representation used for clustering. For each spatial unit i, the DCD and C values from eight successive windows form a 16-dimensional trajectory vector:

$$x = [DCD_0, \dots, DCD_7, C_0, \dots, C_7]$$

Within each window, the depth-lag correlation structure is first summarized into one DCD value and one C value; the individual lag correlations are not included as separate clustering features. Each of the 16 trajectory features is standardized separately across all reference-basin spatial units using StandardScaler. K-means clustering is then applied with k = 4, n_init = 30, and a fixed random seed of 42.

We further added random-initialization and spatial-subsampling stability analyses. Across 100 random-seed trials with n_init = 30, the median matched label agreement was 0.998, with a minimum of 0.961, and the median adjusted Rand index was 0.997, with a minimum of 0.910. Across 100 independent trials using 80% spatial subsamples, the median matched label agreement was 0.959 and the median adjusted Rand index was 0.901, with minimum values of 0.702 and 0.531, respectively. The complete distributions and summary statistics are reported in Fig. S6 and Table S1.

These results show that the retained four-archetype solution is highly stable to random initialization and maintains high overall consistency under spatial-sampling perturbations, although some boundary units can still change class. The revised manuscript therefore describes the four groups as trajectory archetypes with transitional boundaries rather than as strictly separated natural classes.

All core equations are now numbered. The revised Methods document the implemented DCD-C construction, temporal treatment, trajectory representation, feature standardization, and clustering procedure. Sect. 2.4 summarizes the lag-sensitivity results, while the complete lag and clustering-stability analyses are reported in the Supplement.

Location of change: Lines 237-323 and 341-342; Figs. S1 and S6; Table S1.

6. RC6: The four coupling regimes are described mainly in qualitative terms. A quantitative summary table showing the mean/range of DCD and C, temporal trends, and the spatial proportion of each type would make the regime characteristics more transparent and easier to compare.

AR6: We thank the reviewer for suggesting a quantitative summary of the four trajectory archetypes. In response to this comment, we have added Table 1 in the revised Results section to provide a unified comparison of the spatial proportion, overall DCD and C characteristics, temporal window ranges, and changes between the first and last windows for the four archetypes.

As shown in Table 1, Type 2 had the largest area weighted proportion in the reference basin (34.3%), followed by Type 3 (28.5%), Type 1 (19.1%), and Type 4 (18.1%). The mean DCD values of the four archetypes were 1.057, 1.162, 1.276, and 1.205 m, respectively, with Type 3 showing the deepest characteristic depth. The corresponding mean C values were 1.261, 1.695, 1.740, and 1.357, respectively, with the highest value in Type 3 and the lowest value in Type 1.

Regarding temporal evolution, Type 1 maintained a relatively shallow DCD, with partial recovery of C after an initial decline. Type 2 showed a shallower DCD and reduced C during the middle period, followed by recovery in both metrics during the later period. Type 3 consistently exhibited the deepest DCD across the temporal windows. Its C declined substantially and recovered slightly in the later period, but remained lower than the initial value. Type 4 showed pronounced DCD reorganization, while C declined and remained at a relatively low level. Between the first and last windows, C changed by -0.380, -0.694, -1.096, and -0.922 for Types 1-4, respectively, with the largest decreases occurring in Type 3 and Type 4.

The added table enables direct comparison of the four archetypes in terms of spatial proportion, characteristic depth, aggregated correlation magnitude, and temporal evolution. The revised Sect. 3.1 has also been adjusted accordingly to incorporate the quantitative comparison provided by Table 1.

Location of change: Lines 386-445; Table 1.

7. RC7: The uncertainty of the probabilistic classification should be better reflected in the Results. The manuscript reports an overall accuracy of 58.15% and a Kappa coefficient of 0.41, indicating moderate agreement, but it does not show how confident the model is in assigning each grid cell to a specific regime type. Since each grid cell is assigned to the class with the highest predicted probability, it would

be useful to know whether the winning class is clearly separated from the alternatives. I suggest that the authors report the maximum class probability, the probability margin between the first- and second-ranked classes, or classification entropy. This would help distinguish robust regime cores from uncertain transition zones and would support a more cautious interpretation of the national classification map.

AR7: We thank the reviewer for pointing out that the national archetype map lacked information on classification probability uncertainty. We have added two diagnostics to the revised manuscript: the maximum class probability and the probability margin. For each target unit, the classifier provides the probabilities P_k of belonging to the four reference trajectory archetypes, where k=1,…,4. The two diagnostics are defined as

$$p_{max} = \max_k p_k$$

$$\Delta p = p_{(1)} - p_{(2)}$$

where $p_{(1)}$ and $p_{(2)}$ are the largest and second-largest class probabilities, respectively. A larger p_{max} indicates greater within-model decisiveness for the selected archetype, while a larger Δp indicates clearer separation between the two leading candidate archetypes.

Across all valid cells in the national mapping domain, the median p_{max} was 0.461, with an interquartile range of 0.390-0.509. The median probability margin was 0.136, with an interquartile range of 0.059-0.200. Clear differences were observed among archetypes. Type 1 had the lowest median probability margin, at 0.075, indicating a relatively less distinct boundary from neighboring archetypes. Type 4 had a median probability margin of 0.186, indicating relatively clearer class separation.

The revised manuscript now presents the spatial distributions of the maximum class probability and probability margin, together with their distributions within and outside the predictor support domain (Fig. S7). Table S3(b) summarizes the maximum class probability and probability margin statistics for each archetype.

Location of change: Lines 351-357 and 464-471; Fig. S7; Table S3(b).

8. RC8: From Fig. 3, there appears to be a strong spatial overlap between OOD areas and Type 4 transitional decoupled regimes, especially in northeastern China. I do not see a clear explanation of whether this overlap is coincidental, physically meaningful, or partly induced by the transfer/classification procedure. Since OOD labels do not change the assigned type but indicate weaker environmental support, the authors should quantify and discuss the overlap between OOD areas and Type 4. This is important for determining whether Type 4 can be interpreted as a robust transitional decoupled regime, or whether its spatial pattern partly reflects extrapolation uncertainty.

AR8: We thank the reviewer for pointing out the clear spatial overlap between OOD areas and Type 4. In response, we quantified the relationship between the trajectory archetypes and predictor-space OOD and updated Figure 3 and the related discussion.

As shown in Table R1, under the primary threshold of 7.667, 36.9% of the nationally mapped permafrost cells were labeled as OOD. Among all OOD cells, 83.6% were assigned to Type 4. Conversely, 79.2% of all Type 4 cells were located in OOD areas. The corresponding OOD fractions for Types 1, 2, and 3 were 8.6%, 8.6%, and 16.2%, respectively. These results show a strong and systematic spatial association between Type 4 and OOD.

Table R1. Cross-statistics between the nationally mapped trajectory archetypes and predictor-space OOD under the primary threshold of 7.667. The table reports the total number of cells assigned to each archetype, the number of OOD cells, the proportion of each archetype located in OOD areas, and the composition of the OOD domain by archetype.

Type	Total cells, n (%)	OOD cells, n	P(OOD | Type) (%)	P(Type | OOD) (%)
Type 1	4791 (19.7)	412	8.6	4.6
Type 2	7424 (30.6)	637	8.6	7.1
Type 3	2609 (10.7)	423	16.2	4.7
Type 4	9459 (39.0)	7493	79.2	83.6

The overlap is most pronounced in the northeastern permafrost region (revised Figure 3 and Fig. R2). The entire region is classified as OOD under the primary threshold, and 97.4% of its cells are assigned to Type 4. Type 4 in this region therefore represents the closest statistical match to the Type 4 reference archetype under environmental conditions that extend beyond the predictor support of the reference basin. Local physical interpretation in this region is consequently limited by the coverage of the reference domain.

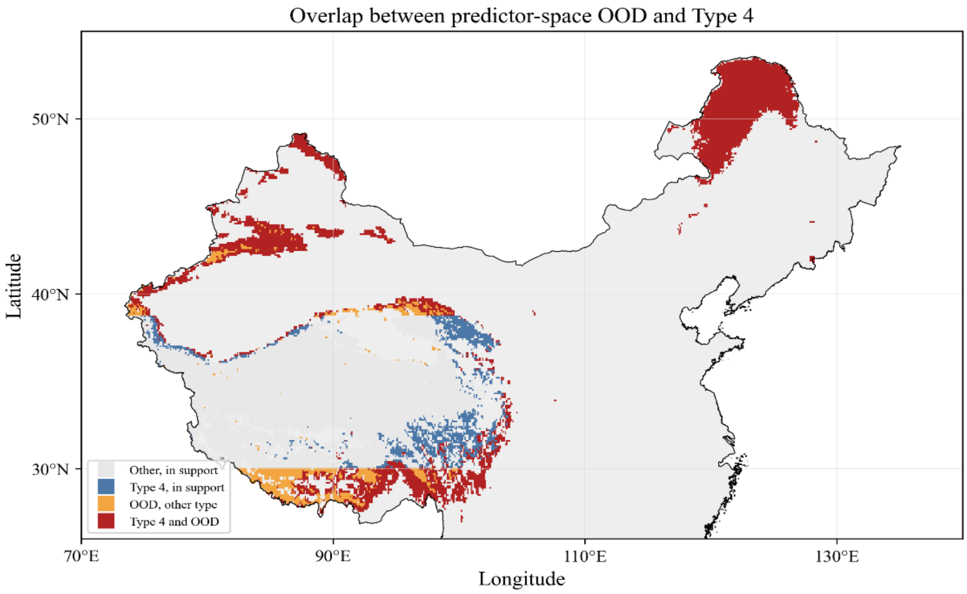

Figure R2. Spatial overlap between Type 4 and predictor-space OOD under the primary threshold of 7.667. Gray indicates other archetypes within predictor support, blue indicates Type 4 within predictor support, orange indicates OOD cells assigned to other archetypes, and red indicates cells that are both Type 4 and OOD.

Type 4 and OOD are not identical. Within the QTP, 56.7% of Type 4 cells are labeled as OOD. A total of 1,966 Type 4 cells remain within predictor support. These cells account for 12.8% of all supported cells and 20.8% of all Type 4 cells. Type 4 within predictor support can therefore still be discussed as an assignment within predictor support, while Type 4 in OOD areas requires more cautious interpretation.

The revised Results report these overlap statistics, and the Discussion now distinguishes Type 4 within predictor support from Type 4 in OOD areas. The updated Figure 3 displays both the assigned trajectory archetypes and the predictor-space OOD labels. Figure R2 and Table R1 provide the spatial overlap and complete cross-statistics for direct inspection.

We also recalculated the early- and late-period classifications and their transitions. OOD screening was not used to establish the period-specific types or determine transition categories. It was overlaid only as an environmental support label. The transitions in the revised Figure 4 are therefore derived directly from the early- and late-period class assignments.

Location of change: Lines 449-533, 599-603, and 615-621; Figs. 3 and 4.

9. RC9: The Abstract and Conclusion should better balance framework description, key findings, and uncertainty. Both sections clearly present the proposed referencebasin framework, but they remain somewhat general and method-oriented. The Abstract in particular provides limited information on the main results, while the Conclusion emphasizes the usefulness of the framework without sufficiently reflecting the moderate classification performance and the uncertainty associated with national-scale transfer. I suggest that the authors revise both sections to include more concrete findings, such as the dominant spatial patterns of the four regimes, the main regions identified as OOD, and the implications of the applicability boundary. In addition, since the manuscript already includes a separate Section 5 titled "Conclusion," Section 4 should be titled simply "Discussion" rather than "Discussion and conclusions."

AR9: We thank the reviewer for pointing out that the original Abstract and Conclusion placed excessive emphasis on describing the framework, while providing insufficient synthesis of the main results, classification performance, and uncertainties associated with the regional transfer. In response, we have rebalanced the presentation of the methods, results, and applicability boundaries in both sections.

The revised Abstract now includes the main differences among the four trajectory archetypes within the reference basin, the dominant spatial patterns at the national scale, and the results of the classification performance and OOD screening. The revised Conclusion further quantifies the type proportions and DCD-C characteristics within the reference basin, summarizes the national distribution of the archetypes and the changes in archetype assignment between the early and late periods, and explicitly clarifies the interpretive scale associated with the overall accuracy of 58.15% and the Kappa coefficient of 0.41. Accordingly, the national classification map is positioned as a regional-scale representation of the reference archetypes, rather than as a deterministic reconstruction of subsurface processes or fine-scale type boundaries at individual grid cells.

The revised manuscript also clarifies that, under the primary threshold, 36.9% of the nationally mapped permafrost grid cells were labeled as OOD, including 20.4% of the Qinghai-Tibetan Plateau domain, while the entire northeastern permafrost region fell outside the environmental support represented by the reference basin. Numerical classifications can still be obtained for these areas, but their local physical interpretation is constrained by the environmental coverage of the reference domain.

Following the reviewer’s suggestion, the title of the original Section 4, “Discussion and conclusions” has been revised to “Discussion” while Section 5 remains as a separate “Conclusion” section.

Location of change: Lines 23-53, 533, and 663-691.