Overall assessment

This manuscript develops a reference-basin-based framework for identifying and transferring permafrost–hydrological coupling regimes across China. Using the Source Region of the Yellow River as a process-constrained reference basin, the authors construct a two-dimensional DCD–C process space, identify four trajectory-based coupling archetypes, and transfer these archetypes to broader permafrost regions using a parsimonious set of climatic, topographic, and soil thermal proxy variables. The study addresses an important problem in cold-region hydrology: how to scale process understanding from relatively well-studied basins to broader data-sparse permafrost regions.

Overall, I find the study promising and within the scope of HESS. However, I recommend substantial revision before publication. Most of the required revisions concern conceptual clarification, methodological transparency, uncertainty assessment, and more careful interpretation, rather than a fundamental redesign of the proposed framework. Detailed comments follow.

Detailed comments

1. 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.
2. The conceptual framing and motivation of the proposed framework need to be strengthened. The Introduction summarizes several existing approaches for large-scale 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.

3. 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.

4. 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 authors

should consider whether their definitions would be more appropriate in a methods subsection rather than in the data description.

5. 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.

6. 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.

7. 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.

8. 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.

9. The Abstract and Conclusion should better balance framework description, key findings, and uncertainty. Both sections clearly present the proposed reference-basin 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.”