

Hydrology and Earth System Sciences

Supplementary Information for

**A reference-basin based framework for identifying and transferring permafrost
hydrological coupling regimes across China**

Lei Song^{1*}, Yanping Li^{2*}

¹ College of Hydrology and Water Resources, Hohai University, Nanjing, China

² College of Oceanography, Hohai University, Nanjing, China

Contents of this file

Figures S1-S7 and Tables S1-S4

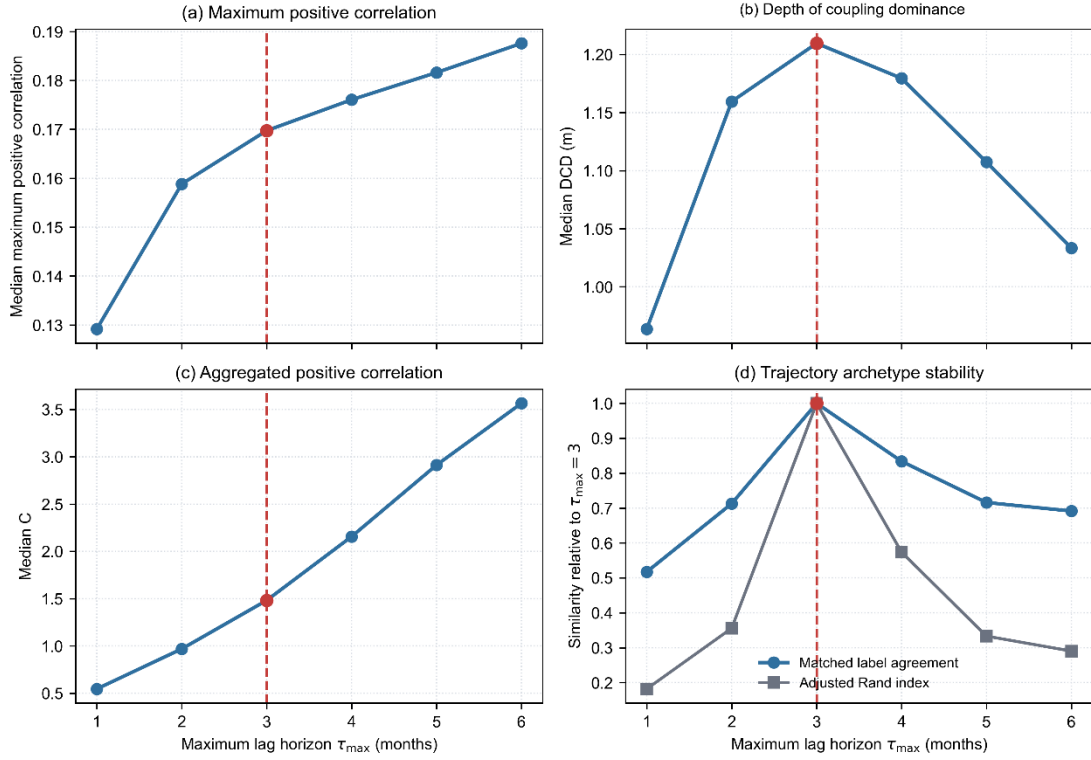


Figure S1. Sensitivity of DCD-C trajectories to the maximum lag horizon. The maximum lag horizon, $\tau_{\max}$, defines the longest monthly lag included when calculating positive correlations between soil ice loss and precipitation-adjusted runoff at the native 5 km reference-basin scale. (a) Median maximum positive correlation. (b) Median dominant control depth (DCD). (c) Median aggregated positive correlation metric C. (d) Matched label agreement and adjusted Rand index (ARI) between the trajectory archetypes obtained under each lag horizon and those obtained under $\tau_{\max} = 3$ months. The red dashed line indicates the primary setting of $\tau_{\max} = 3$ months.

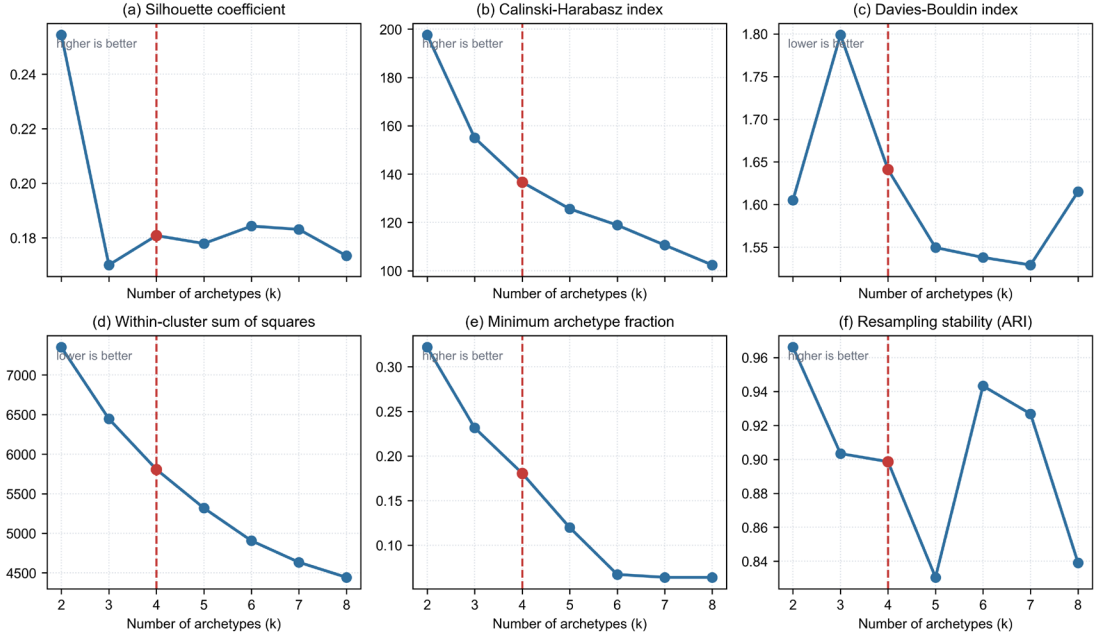


Figure S2. Diagnostic evaluation of candidate numbers of DCD-C trajectory archetypes. Candidate solutions from $k = 2$ to $k = 8$ were compared using (a) the silhouette coefficient, (b) the Calinski-Harabasz index, (c) the Davies-Bouldin index, (d) the within-cluster sum of squares, (e) the minimum archetype fraction, and (f) resampling stability measured by the adjusted Rand index. The red dashed line indicates the retained four-archetype solution. No single metric uniquely selected $k = 4$; the retained solution represents a compromise among separation, class balance, resampling stability, and trajectory interpretability.

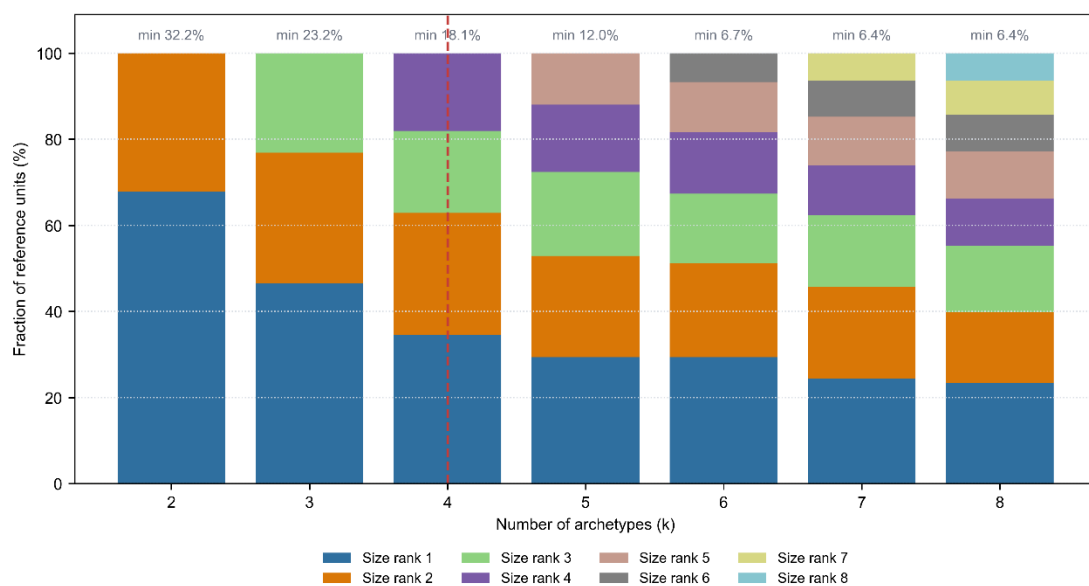


Figure S3. Size balance of candidate DCD-C trajectory-archetype solutions.

Stacked bars show the fractions of reference units assigned to each size-ranked archetype for $k = 2$ to $k = 8$. Labels above the bars report the minimum archetype fraction. The red dashed line indicates the retained $k = 4$ solution.

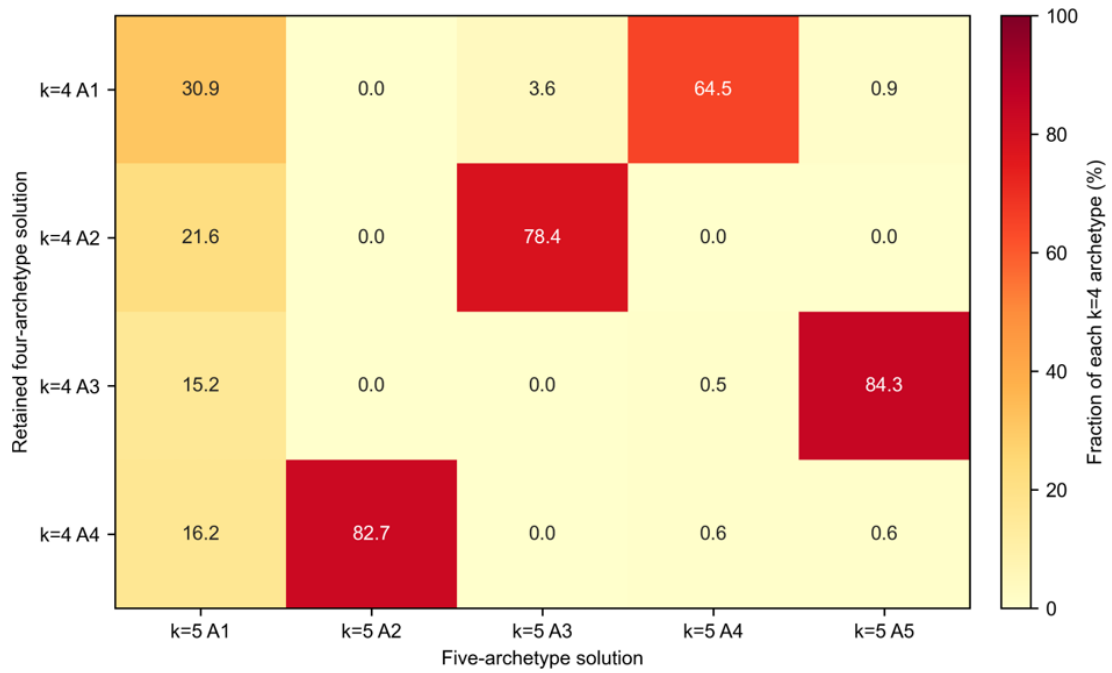


Figure S4. Relationship between the retained $k = 4$ solution and the alternative $k = 5$ solution. The heat map shows how each trajectory archetype in the retained four-archetype solution is partitioned under $k = 5$. Values give the percentage of each $k = 4$ archetype assigned to each $k = 5$ archetype.

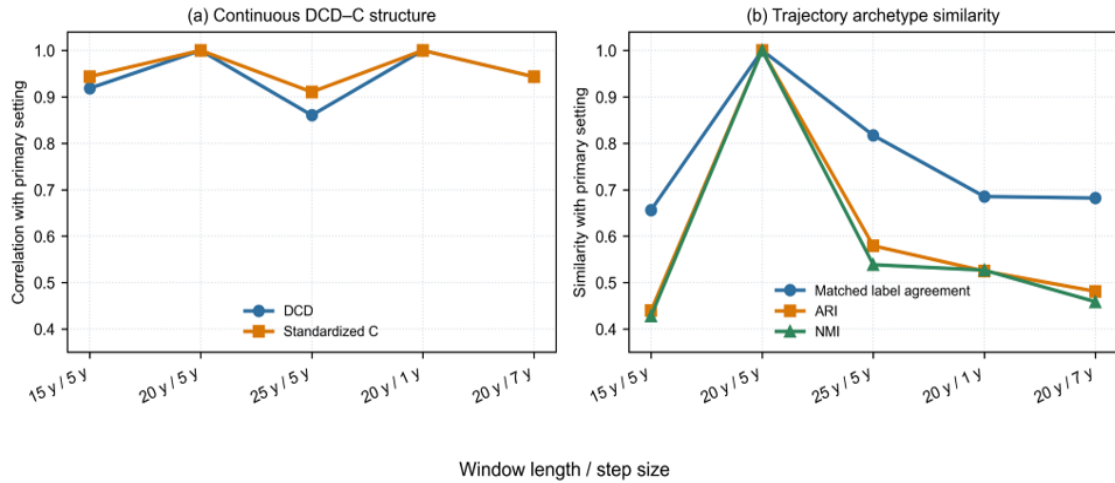


Figure S5. Sensitivity of DCD-C trajectories to window length and step size. The primary setting uses a 20-year window and a 5-year step. Alternative settings use 15- and 25-year windows with a 5-year step, and a 20-year window with 1- and 7-year steps. (a) Pearson correlations of the DCD and standardized C fields with the primary setting after alignment to the primary window-center years. (b) Matched label agreement, adjusted Rand index (ARI), and normalized mutual information (NMI) between independently clustered trajectory archetypes and the primary solution. For the 20-year window with a 1-year step, the continuous fields are identical at window centers shared with the primary setting; however, the denser trajectory contains additional intermediate windows and therefore yields a different independently clustered solution.

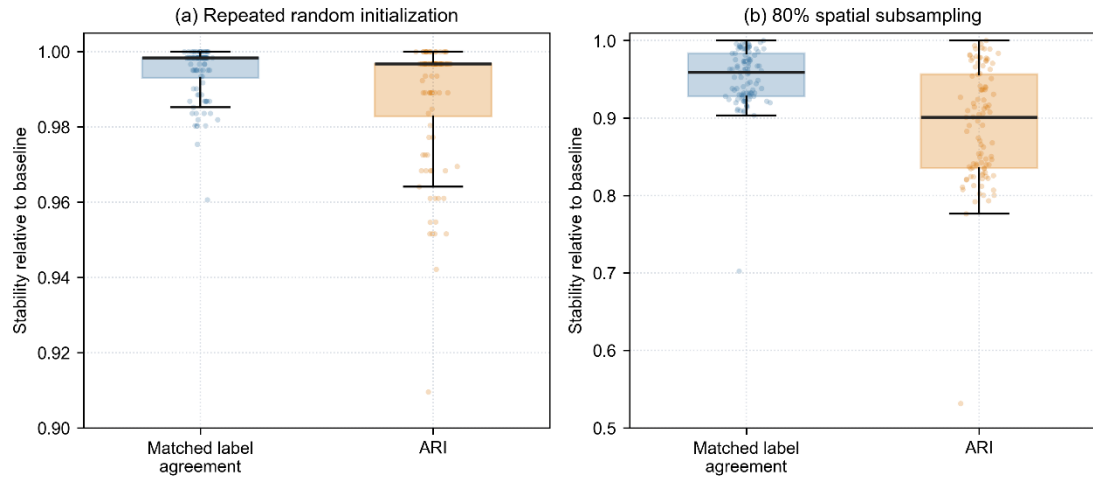


Figure S6. Stability of the retained four-archetype solution. (a) Distribution of matched label agreement and adjusted Rand index (ARI) across 100 random-seed trials, with 30 initializations per fit. (b) Distribution of the same metrics across 100 independent 80% spatial subsamples. Different y-axis ranges are used to show the narrow spread in panel (a) while retaining the full distribution in panel (b). Labels were aligned through optimal matching before matched label agreement was calculated; ARI is invariant to label permutation.

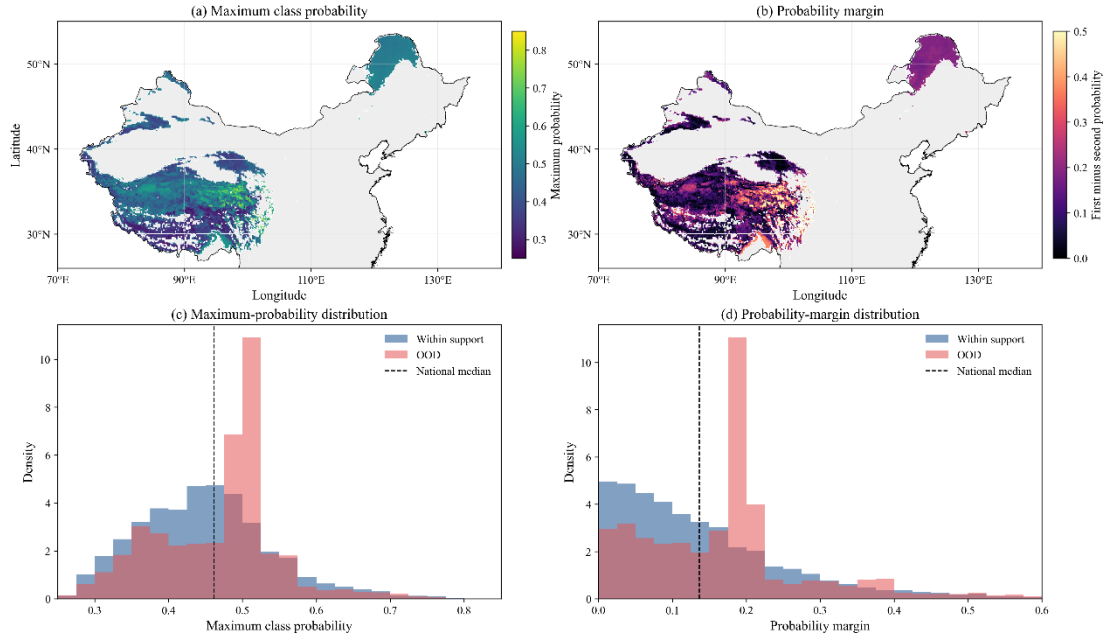


Figure S7. Spatial distribution and frequency distributions of classification probability diagnostics for the national archetype transfer. (a) Maximum class probability. (b) Probability margin between the largest and second-largest class probabilities. (c) Distribution of maximum class probability within and outside predictor support. (d) Distribution of probability margin within and outside predictor support. Larger values indicate a clearer model preference for the selected archetype and greater separation from the closest alternative, respectively.

Table S1. Stability of the retained four-archetype solution under repeated random initialization and 80% spatial subsampling. The random-initialization analysis used 100 random seeds, with 30 initializations per fit. The spatial analysis used 100 independent subsamples containing 80% of the reference units. Labels were aligned through optimal matching before matched label agreement was calculated; the adjusted Rand index (ARI) and normalized mutual information (NMI) are invariant to label permutation.

Analysis	Repetitions	Metric	Minimum	Median	Mean	Maximum
Random initialization	100	Matched label agreement	0.961	0.998	0.995	1.000
Random initialization	100	ARI	0.910	0.997	0.987	1.000
Random initialization	100	NMI	0.880	0.993	0.980	1.000
80% spatial subsampling	100	Matched label agreement	0.702	0.959	0.953	1.000
80% spatial subsampling	100	ARI	0.531	0.901	0.892	1.000
80% spatial subsampling	100	NMI	0.548	0.875	0.873	1.000

Table S2. Sensitivity of the continuous DCD-C structure and independently clustered trajectory archetypes to window length and step size. The primary setting is a 20-year window with a 5-year step. Continuous DCD and standardized C fields were compared after alignment to the primary window-center years. Matched label agreement, adjusted Rand index (ARI), and normalized mutual information (NMI) compare the independently clustered solutions with the primary solution. For the 20-year window with a 1-year step, the continuous fields are identical at window centers shared with the primary setting; however, the denser trajectory contains additional intermediate windows and therefore yields a different independently clustered solution.

Window / step	Windows	DCD r	Standardized C r	Matched label agreement	ARI	NMI
15 y / 5 y	9	0.918	0.944	0.656	0.440	0.428
20 y / 5 y (primary)	8	1.000	1.000	1.000	1.000	1.000
25 y / 5 y	7	0.861	0.910	0.817	0.579	0.538
20 y / 1 y	36	1.000	1.000	0.685	0.525	0.527
20 y / 7 y	6	0.943	0.943	0.682	0.481	0.459

Table S3. Predictor-space OOD threshold sensitivity and classification probability diagnostics.

(a) Sensitivity of predictor-space OOD coverage to the reference-support threshold

Threshold basis	Threshold	National OOD (%)	Area-weighted OOD (%)	QTP OOD (%)	NE OOD (%)	Agreement	Jaccard
99th percentile	4.122	66.2	65.0	57.3	100.0	0.708	0.558
99.5th percentile	4.755	59.1	57.7	48.4	100.0	0.778	0.625
Reference maximum	7.667	36.9	34.6	20.4	100.0	1.000	1.000

(b) Classification probability diagnostics by archetype and predictor support status.

Group	n	pmax median [IQR]	Probability margin median [IQR]
Type 1	4791	0.444 [0.379, 0.492]	0.075 [0.038, 0.126]
Type 2	7424	0.449 [0.381, 0.500]	0.116 [0.049, 0.211]
Type 3	2609	0.381 [0.346, 0.440]	0.126 [0.068, 0.201]
Type 4	9459	0.496 [0.436, 0.515]	0.186 [0.111, 0.215]
Type 4 within predictor support	1966	0.444 [0.385, 0.514]	0.138 [0.060, 0.254]
Type 4 OOD	7493	0.500 [0.459, 0.515]	0.188 [0.135, 0.210]

Note: The reference maximum (7.667) is used as the main OOD threshold. Agreement and Jaccard are calculated relative to this main mask. The probability margin equals the largest minus the second-largest class probability. Type 4 contains 1966 cells within predictor support, equivalent to 12.8% of all supported cells and 20.8% of all Type 4 cells.

Table S4. Confusion matrix and class-specific performance of the reference-basin classification evaluation.

(a) Confusion matrix. Rows indicate the reference trajectory archetypes, and columns indicate the out of fold archetype assignments.

Reference archetype	Classified as Type 1	Classified as Type 2	Classified as Type 3	Classified as Type 4	Support
Type 1	38	54	19	5	116
Type 2	31	146	20	12	209
Type 3	9	27	125	11	172
Type 4	3	37	26	44	110
Classified total	81	264	190	72	607

(b) Class-specific precision, recall, and F1 score.

Archetype	Precision	Recall	F1 score	Support
Type 1	0.469	0.328	0.386	116
Type 2	0.553	0.699	0.617	209
Type 3	0.658	0.727	0.691	172
Type 4	0.611	0.400	0.484	110

Note: The evaluation used out of fold archetype assignments obtained from five fold spatial cross validation for 607 spatial units in the reference basin, with folds constructed from 0.5° spatial groups. Each unit was classified only when its spatial group was withheld. Overall accuracy was 58.15%, and Cohen’s kappa was 0.41. Precision is the fraction of assignments to an archetype that were correct; recall is the fraction of reference units of that archetype that were correctly assigned.

