

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*}

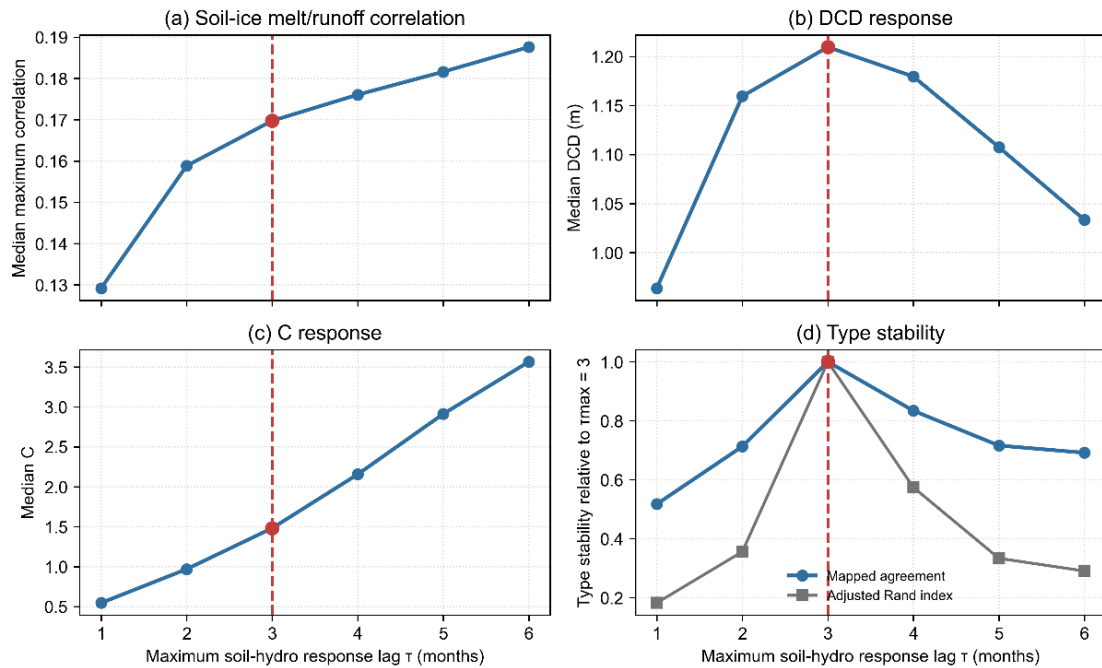
¹ College of Hydrology and Water Resources, Hohai University, Nanjing, China

² College of Oceanography, Hohai University, Nanjing, China

Contents of this file

Figures S1-S4

Sensitivity of DCD-C Trajectories to Maximum Response-Lag Window



The red dashed line indicates the selected value $r_{max} = 3$ months.

Figure S1. Sensitivity of DCD-C trajectories to the maximum response-lag window. The maximum response lag τ defines the longest monthly lag considered when calculating the correlation between soil-ice melt and runoff response. (a) shows the median maximum soil-ice melt/runoff correlation under different τ values. (b) shows the median DCD derived from the corresponding depth-wise coupling structure. (c) shows the median cumulative coupling strength C. (d) compares the regime assignments obtained under each τ with those obtained under $\tau = 3$ months, using mapped agreement and the adjusted Rand index. The red dashed line indicates the selected value $\tau = 3$ months.

Diagnostic Evaluation of the Number of DCD-C Trajectory Regimes

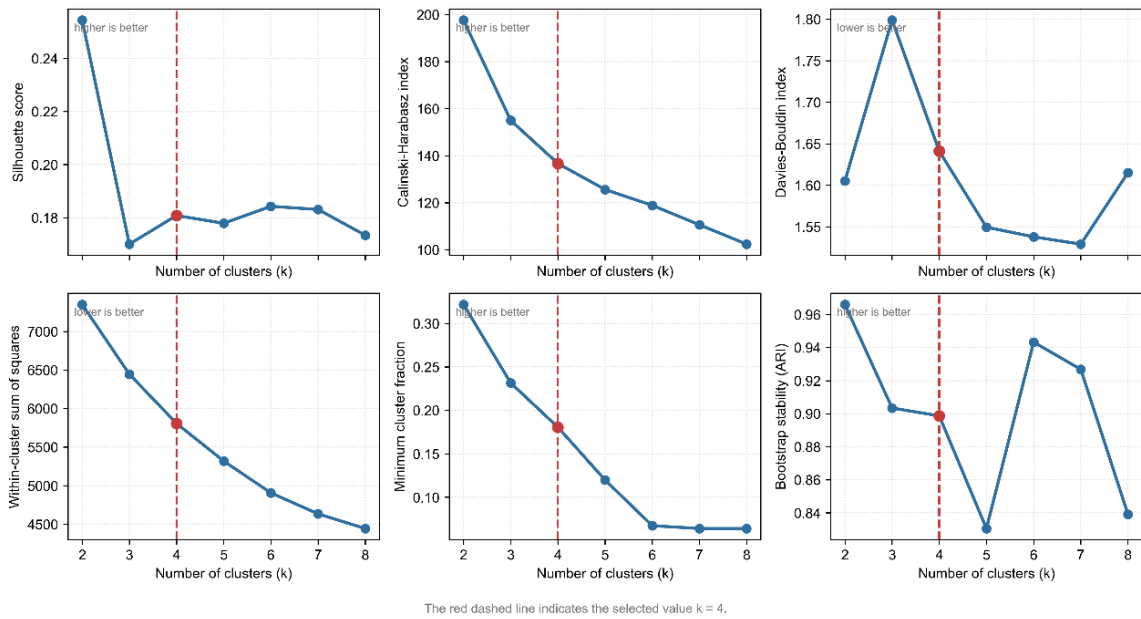


Figure S2. Diagnostic evaluation of candidate numbers of DCD-C trajectory regimes. Candidate cluster numbers from $k = 2$ to $k = 8$ were evaluated using multiple clustering diagnostics: (a) silhouette score, (b) Calinski-Harabasz index, (c) Davies-Bouldin index, (d) within-cluster sum of squares, (e) minimum cluster fraction, and (f) bootstrap stability. The red dashed line indicates the retained four-cluster solution. The diagnostics were used as complementary evidence together with cluster-size balance and physical interpretability, rather than as a single-metric optimization criterion.

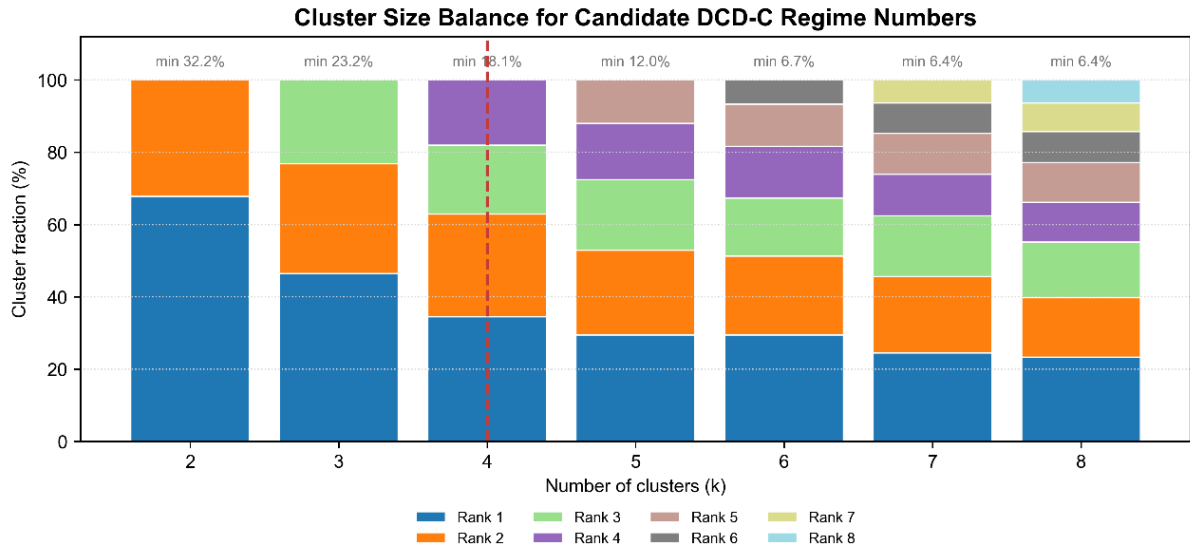


Figure S3. Cluster-size balance for candidate DCD-C regime numbers. The stacked bars show the relative fraction of samples assigned to each cluster under candidate values from $k = 2$ to $k = 8$. The four-cluster solution retains relatively balanced regime sizes, whereas larger k values increasingly introduce smaller subgroups. The red dashed line indicates the retained four-cluster solution.

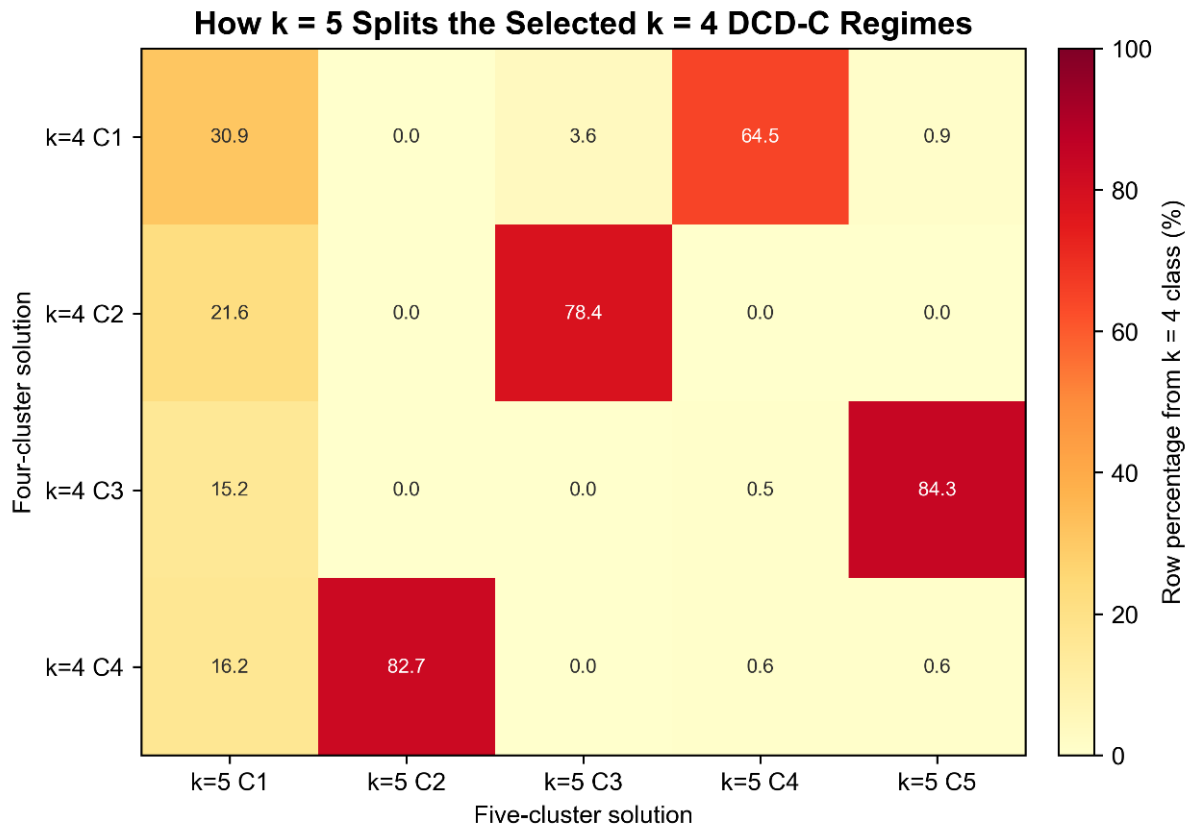


Figure S4. Relationship between the retained k = 4 solution and the alternative k = 5 solution. The heatmap shows how each regime in the four-cluster solution is partitioned into the five-cluster solution. Percentages indicate the fraction of each k = 4 regime assigned to each k = 5 regime. The comparison shows that the additional k = 5 class mainly increases classification granularity by producing mixed or locally split subgroups, rather than forming a clearly independent DCD-C trajectory regime. This supports the use of k = 4 as a parsimonious and physically interpretable regime classification.