

# Review of Song and Li for publication in *Hydrology and Earth System Sciences*

The manuscript develops a reference basin framework for transferring permafrost hydrological coupling regimes from the Source Region of the Yellow River (SRYR) to other permafrost regions of China. The approach constructs a two dimensional process space using dominant control depth (DCD) and integrated coupling strength ( $C$ ), identifies four trajectory based regimes from moving window analyses, and then predicts these regimes at the national scale using a generalized additive model (GAM) with climatic, topographic, and soil thermal predictors. The inclusion of out of distribution (OOD) screening is a useful attempt to identify regions where transfer from the reference basin is less well supported.

The regionalization problem addressed here is important, and the idea of transferring trajectory based information rather than individual model parameters is interesting. My main concern is that several steps between the reference basin analysis and the national interpretation require greater methodological clarity. In particular, the spatial meaning of the DCD and  $C$  fields, the validation of the national transfer, the definition of the OOD criterion, and the temporal transition analysis need to be clarified before the national patterns can be interpreted with confidence. I therefore recommend **major revision**.

## Major Comments

### **Q1: Spatial interpretation of correlations with basin outlet runoff.**

The spatial meaning of the DCD and  $C$  fields within the reference basin is not sufficiently clear. Soil ice is evaluated for individual spatial units at 5 km resolution, but the hydrological response variable is the same runoff record from the Tangnaihai basin outlet. Consequently, the analysis does not directly measure a local runoff response at each grid cell. A high soil ice and runoff correlation at a particular location may indicate that soil ice variability there covaries with the integrated basin response, but it does not by itself demonstrate that this location or soil depth exerts dominant hydrological control on outlet runoff. This distinction is important because the manuscript subsequently interprets the spatial patterns in terms of shallow and deep hydrological control. The authors should state explicitly what can and cannot be inferred when spatially distributed soil ice series are related to a single basin integrated discharge series. The terminology in the Results and Discussion should also distinguish statistical association with basin runoff from local hydrological control.

### **Q2: Validation of the national transfer model.**

The reported overall accuracy of 58.15% and Kappa coefficient of 0.41 indicate moderate agreement, but the procedure used to obtain these statistics is not described clearly enough to assess the performance of the national transfer model. It is not clear whether these values are calculated from the observations used to fit the GAM, from a withheld subset, from cross validation, or from a spatially independent validation. This distinction is particularly important for gridded environmental data because neighboring cells are spatially correlated, and model performance estimated

from randomly divided spatial samples may differ substantially from performance in geographically independent regions. The manuscript should state exactly how the predictions used to calculate accuracy and Kappa were generated. Class specific performance or a confusion matrix should also be reported because overall accuracy alone cannot show whether some archetypes are predicted substantially more reliably than others. Since the GAM produces class probabilities, information on the confidence of individual assignments would also help distinguish well supported regime cores from uncertain transition areas. With the reported performance, the national classification should be interpreted primarily at the regional scale rather than as a precise grid cell representation of subsurface hydrological state.

**Q3: OOD screening and the applicability boundary.**

The OOD analysis is central to the proposed applicability control, but the mathematical procedure used to define OOD conditions is not provided. The manuscript states that cells outside the environmental support represented by the reference basin are identified as OOD, but it does not specify how predictor space is standardized, what distance or range criterion defines environmental support, whether the predictors are considered independently or jointly, what numerical threshold is used, or how that threshold is selected. These details are necessary for reproducibility because the extent of the OOD domain depends directly on the definition of environmental support. I therefore recommend that the authors provide the complete OOD criterion and show how the mapped applicability boundary responds to reasonable changes in the selected threshold. The fraction of each regional domain classified as OOD would also be useful. The interpretation of predictions within OOD regions should remain clearly separated from predictions within the environmental support of the reference basin, since a numerical class can still be produced by the GAM even when the target conditions fall outside the predictor space represented by the training data.

**Q4: Temporal transition analysis in Figure 4.**

The procedure used to construct the early and late period classifications in Section 3.3 needs to be explained explicitly in the Methods. The four archetypes are initially defined by clustering complete DCD and  $C$  trajectories constructed from eight successive time windows. Figure 4 then assigns regime types separately to the first three and last three windows and interprets differences between these classifications as transitions such as Type 2 to Type 4 and Type 4 to Type 3. It is not clear how an archetype defined from a complete eight window trajectory is assigned to a shorter early or late subset. It is also unclear whether the period specific classifications are obtained directly from DCD and  $C$ , from the GAM using period specific environmental predictors, or through another procedure. These alternatives have different meanings. In particular, a change in statistically predicted class resulting from changes in environmental predictors is not equivalent to directly observing a transition between independently identified DCD and  $C$  trajectories. The exact procedure used to generate the early and late classifications and the transition percentages in Figure 4 should therefore be stated clearly, and the interpretation of these changes should match the actual classification procedure.

## Minor Comments

- The manuscript should provide the exact soil layer depths used to calculate  $C$  and DCD. Since  $C$  is calculated by summing contributions over the soil profile, the vertical discretization is relevant to the numerical value of the metric.
- Figure 2 labels the horizontal axis as Time Window (0 to 7). Please provide the corresponding calendar periods in the caption or axis labels.

- Please state the exact calendar periods included in the early and late periods shown in Figure 4.
- The terms “type,” “regime,” “coupling regime,” and “trajectory archetype” are used interchangeably. I suggest selecting one term and using it consistently.
- A compact table reporting the characteristic DCD,  $C$ , temporal behavior, and spatial proportion of the four reference basin archetypes would make the classifications easier to compare quantitatively.