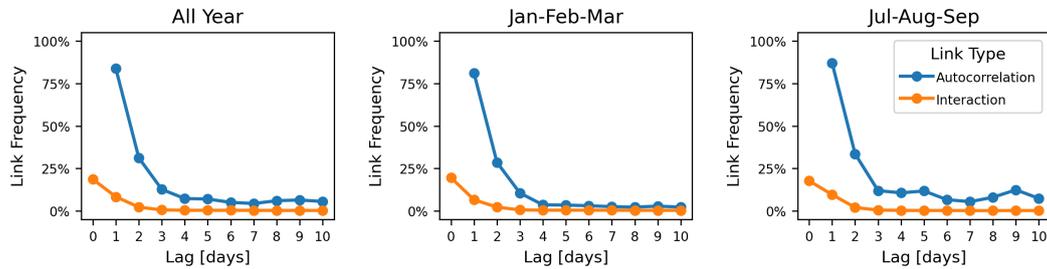
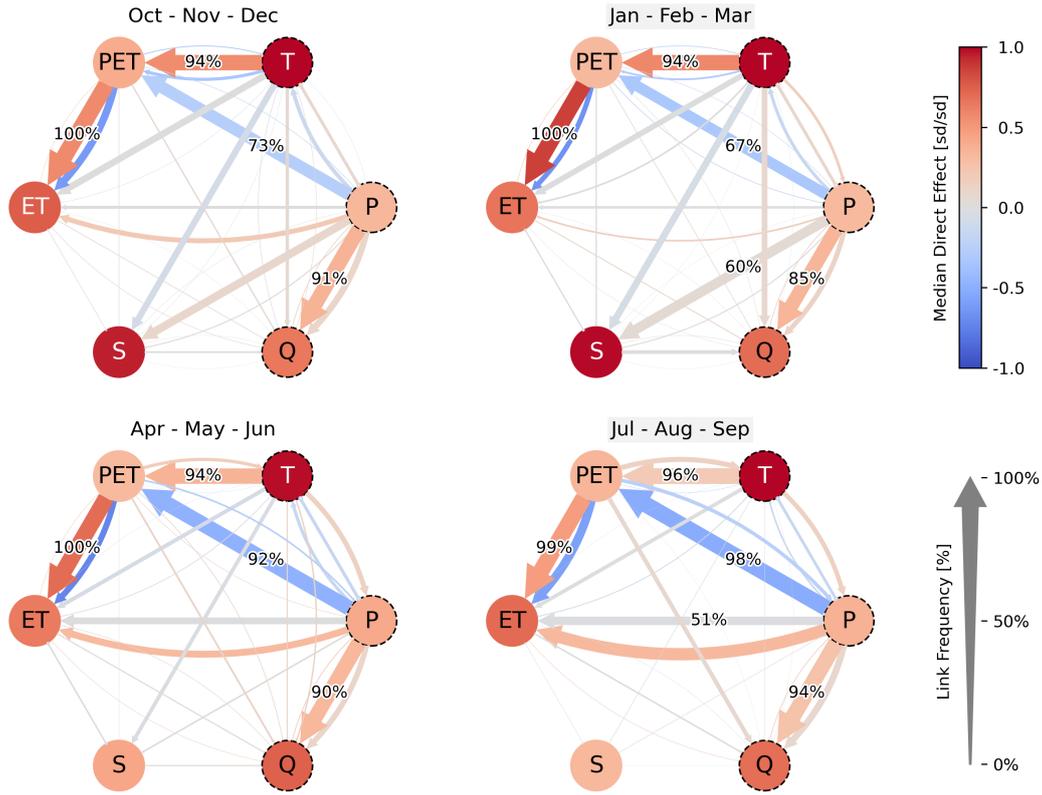


Supplementary Material: Causal Analysis of Streamflow, Evapotranspiration and Snow Dynamics in Large Sample Hydrology

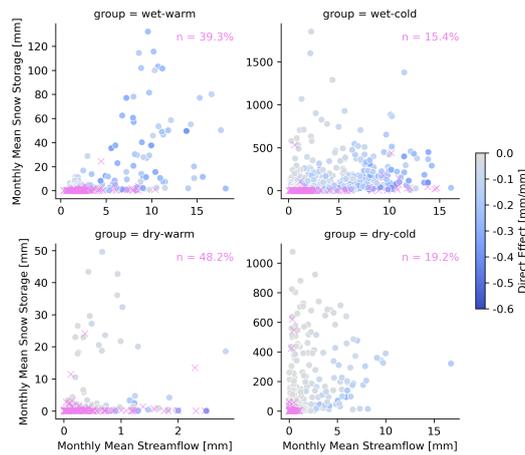
The supplementary material contains additional figures and tables that add further information on methodological details and relevant results. In the main text, they are both referenced in text and in captions.



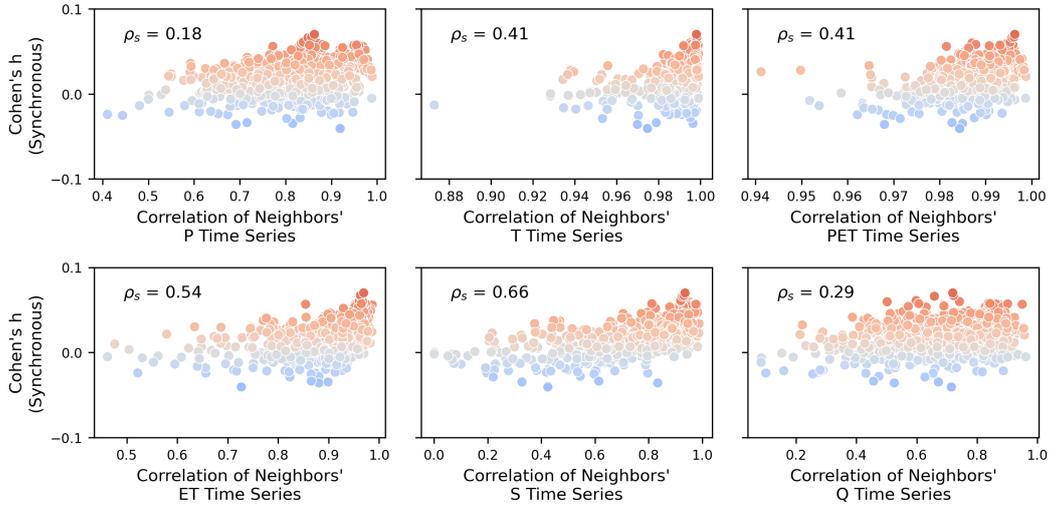
Supplementary Figure 1. Most links between variables are contemporaneous and have lag zero. With the available time resolution, the temporal order of cause and effect can not be identified, and collider-based orientation rules have to be used. Many contemporaneous links could not be oriented for the graphs in Fig. 1, making directionality assumptions necessary for a causal effect estimation. Autocorrelational links point from a lagged version of a variable to itself and can therefore only be found at positive lags. Link frequency is taken over all catchments and months. The maximum time lag of ten days is a hyperparameter chosen for the analysis.



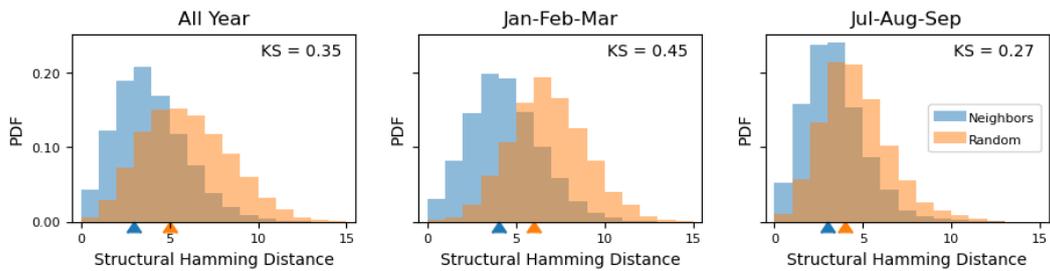
Supplementary Figure 2. As Fig. 1, but using snow storages (S) instead of derivatives (dS). Snow interactions such as the $P \rightarrow S$, $T \rightarrow S$, and $S \rightarrow Q$ links are detected less frequently and with smaller causal effects when storages are used instead of derivatives.



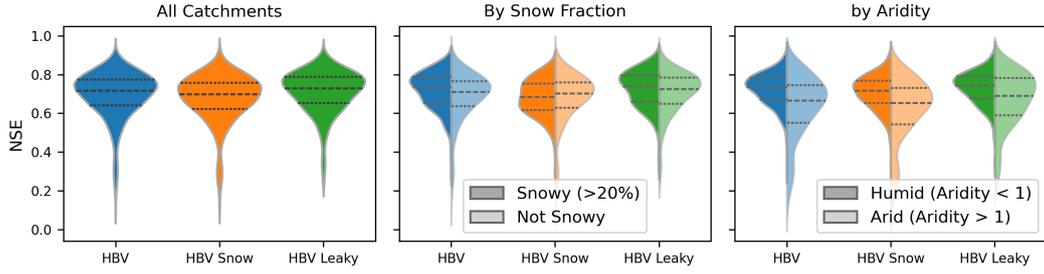
Supplementary Figure 3. On average 24% of $dS \rightarrow Q$ links have unreasonably large or positive regression slopes (violet). We treat slopes below -0.27 (5th percentile) and above zero as outliers. They are primarily found in catchments with low snow storage overall, where regression estimates are dominated by infrequent snowmelt events.



Supplementary Figure 4. Synchronous link changes between neighboring catchments are most strongly associated with spatial patterns in the correlation of neighboring snow time series. Rank correlation is measured using Spearman's ρ_s . Spatial patterns of Cohen's h for synchronous link changes are shown in Fig. 3B.



Supplementary Figure 5. Neighboring catchments have more similar causal graphs than random pairings. The Structural Hamming Distance (SHD) is calculated by counting how many adjacencies differ between causal graphs, irrespective of their lag. Similar to Fig. 3, we use the five closest neighboring catchments. Distributional differences are measured in terms of the Kolmogorov-Smirnoff distance. Graphs are most dissimilar in winter, when differences in energy and snow availability change snow-related adjacencies. Triangular marks indicate median values of the SHD distributions.



Supplementary Figure 6. HBV model variants result in very similar NSE distributions across catchments, illustrating the problem of equifinality. Model variants are independently calibrated using a Differential Evolution algorithm. Horizontal lines in violin plots indicate lower quartile, median, and upper quartile from bottom to top. Snowy catchments have more than 20% of precipitation falling as snow. Humid and arid catchments are defined based on their aridity.

	in Tigramite	Value
Preprocessing	-	deseasonlization, detrending, outlier removal
Masking	mask	monthly
Causal Discovery ↔ Significance Threshold ↔ Maximum Lag ↔ Link Assumptions	run_pcmciplus(...) pc_alpha tau_max link_assumptions	PCMCi+ 0.01 10 days precipitation, temperature and potential evapotranspiration causes, streamflow effect
Independence Test ↔ Mask Type	RobustParCorr(...) mask_type	RobustParCorr 'xy' (cause and effect)
Causal Effect Estimation ↔ Preprocessing ↔ Causal Parents ↔ Bootstrap Samples ↔ Block Length	LinearMediation data_transform all_parents boot_samples boot_blocklength	- none for physical units, standardization for standardized effects from causal discovery 50 10 days

Supplementary Table 1. Setup and Hyperparameters for Causal Discovery and Effect Estimation. In addition to monthly masks, outliers are removed by masking out the 0.1% largest absolute values for each variable across the whole available time period. While outliers are automatically addressed by the independence test, they are problematic for the causal effect estimation.