Supplementary Materials for

A hybrid Kolmogorov-Arnold networks-based model with attention for predicting Arctic River streamflow

Renjie Zhou [1, *], Shiqi Liu[2]

[1]: Department of Environmental and Geosciences, Sam Houston State University, Huntsville, TX 77340, USA; Email: renjie.zhou@shsu.edu

[2]: Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China; Email: liusq@igsnrr.ac.cn

Correspondence to: Renjie Zhou (renjie.zhou@shsu.edu)

S1. Hyperparameter Selection for Loss Weighting

The combined loss function balances data fidelity (MSE) and physical consistency (physics constraint) through weighting parameters:

$$\mathcal{L}_{combined} = \alpha \mathcal{L}_{MSE}(\hat{Q}, Q_{obs}) + \beta \mathcal{L}_{phys}, \tag{S1}$$

where α and β are weighting coefficients that control the relative importance of the data-driven loss (MSE) and physics-informed constraint terms in the combined loss function. The summation of α and β is 1. It aims at balancing two objectives: 1) fitting observed data through THE traditional data-driven loss function, and 2) respecting physical constraints through domain-specific penalty terms. The relative weighting of these objectives can influence model performance. Excessive data weighting may lead to physically implausible predictions that overfit noise, while excessive physics weighting can overconstrain the model and prevent learning of complex real-world behaviors not captured by simplified physical formulations.

A systematic grid search over $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ (corresponding to $\beta \in \{0.9, 0.7, 0.5, 0.3, 0.1\}$) is conducted to determine optimal weighting. Each configuration was trained across 1 to 12 months horizons. The mean values of their evaluation metrics are calculated and plotted in Fig. S1. It reveals that the optimum is obtained at $\alpha = 0.7$, $\beta = 0.3$, achieving NSE of 0.83, RMSE of 8.05 mm, and KGE' of 0.79. Lower α values produce degraded performance (NSE: 0.81-0.82), which indicates the physics constraint is overly restrictive when weighted too heavily to capture processes, and excessive weighting prevents the model from learning data-driven corrections. Conversely, a higher α value also reduces performance (NSE: 0.82). The optimal $\alpha = 0.7$ balances strong enough data-driven learning strategy to capture complex Arctic hydrology while maintaining physical consistency through meaningful but not dominant physics constraints. Therefore, $\alpha = 0.7$ and $\beta = 0.3$ are adopted in the manuscript unless otherwise stated.

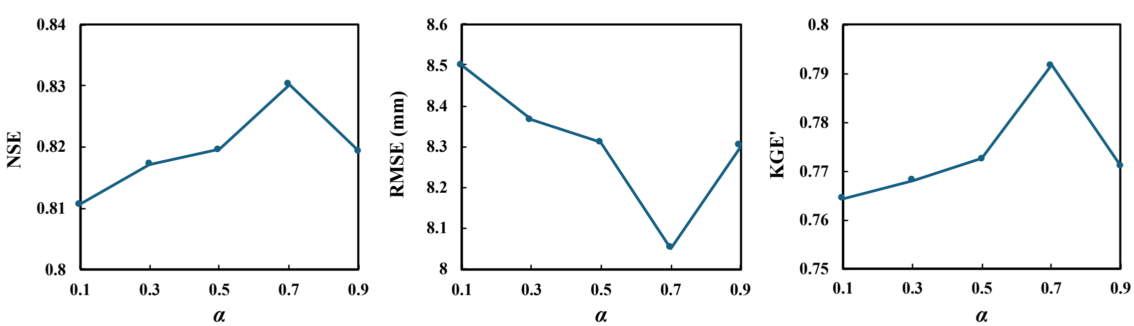

Figure S1: Sensitivity of RCPIKLA performance to the physics-informed loss weighting parameter α. Left: NSE; Middle: RMSE; Right: KGE'. Points denote the mean over 1 to 12 month horizons for each α.

S2. Evaluation metrics across temporal aggregation windows (10-run uncertainty)

The model performance across different time steps (1-12 months) reveals variations in predictive capabilities among the models tested. To ensure stable results, each model is run 10 times at each time step, and the evaluation metrics are averaged. Prediction ensemble means and variability across 10 independent training runs at each forecasting horizon are reported in Table S1–S3 of Supplementary Materials.

Table S1: NSE across temporal aggregation windows (10-run uncertainty). Values are reported as the ensemble mean across 10 independent runs, with the minimum and maximum values across runs for each forecasting horizon (1–12 months ahead).

	RNN			**GRU**			**LSTM**			**KAN-LSTM**			**RCPIKLA**		
Time step	NSE mean	NSE min	NSE max	NSE mean	NSE min	NSE max	NSE mean	NSE min	NSE max	NSE mean	NSE min	NSE max	NSE mean	NSE min	NSE max
1	0.72	0.72	0.74	0.75	0.74	0.76	0.69	0.69	0.70	0.72	0.72	0.74	0.82	0.78	0.83
2	0.76	0.75	0.76	0.77	0.76	0.77	0.74	0.74	0.75	0.78	0.76	0.78	0.83	0.81	0.84
3	0.72	0.71	0.74	0.75	0.75	0.76	0.72	0.71	0.72	0.78	0.78	0.79	0.83	0.78	0.85
4	0.70	0.67	0.72	0.73	0.73	0.75	0.69	0.68	0.70	0.78	0.77	0.79	0.82	0.80	0.84
5	0.66	0.64	0.69	0.71	0.69	0.73	0.66	0.64	0.68	0.77	0.76	0.79	0.81	0.76	0.83
6	0.66	0.64	0.69	0.70	0.68	0.72	0.66	0.65	0.68	0.73	0.70	0.77	0.83	0.78	0.87
7	0.66	0.64	0.69	0.72	0.71	0.74	0.68	0.65	0.70	0.78	0.74	0.80	0.85	0.81	0.89
8	0.65	0.63	0.69	0.72	0.69	0.74	0.68	0.65	0.70	0.78	0.75	0.80	0.85	0.80	0.90
9	0.69	0.63	0.71	0.75	0.74	0.76	0.69	0.64	0.72	0.79	0.78	0.81	0.86	0.77	0.90
10	0.68	0.65	0.71	0.75	0.73	0.77	0.69	0.67	0.72	0.80	0.76	0.82	0.83	0.79	0.90
11	0.68	0.63	0.71	0.74	0.72	0.75	0.68	0.65	0.70	0.78	0.76	0.81	0.85	0.81	0.88
12	0.69	0.66	0.70	0.74	0.72	0.75	0.69	0.67	0.72	0.76	0.74	0.78	0.83	0.78	0.86

Table S2: RMSE across temporal aggregation windows (10-run uncertainty). Values are reported as the ensemble mean across 10 independent runs, with the minimum and maximum values across runs for each forecasting horizon (1–12 months ahead).

	RNN			**GRU**			**LSTM**			**KAN-LSTM**			**RCPIKLA**		
Time step	RMSE mean	RMSE min	RMSE max	RMSE Mean	RMSE min	RMSE max	RMSE mean	RMSE min	RMSE max	RMSE mean	RMSE min	RMSE max	RMSE mean	RMSE min	RMSE max
1	10.50	10.18	10.61	9.90	9.70	10.13	11.04	10.9	11.15	10.44	10.19	10.56	8.53	8.26	9.35
2	9.87	9.69	10.02	9.62	9.50	9.69	10.08	9.92	10.2	9.46	9.26	9.68	8.32	8.04	8.66
3	10.54	10.16	10.84	9.99	9.87	10.09	10.66	10.54	10.82	9.30	9.13	9.40	8.32	7.74	9.29
4	11.07	10.67	11.53	10.35	10.08	10.50	11.19	10.97	11.31	9.37	9.26	9.55	8.48	7.99	9.10
5	11.70	11.20	12.16	10.82	10.49	11.15	11.72	11.36	12.11	9.59	9.28	9.85	8.73	8.24	9.79
6	11.78	11.20	12.15	11.05	10.76	11.51	11.74	11.45	11.99	10.51	9.61	11.00	8.39	7.34	9.60
7	11.79	11.22	12.17	10.70	10.42	10.95	11.54	11.18	11.95	9.45	9.09	10.26	7.87	6.86	8.93
8	11.97	11.38	12.44	10.74	10.38	11.29	11.58	11.14	11.99	9.64	9.09	10.26	7.86	6.31	9.08
9	10.49	10.17	11.39	9.42	9.16	9.67	10.49	9.99	11.32	8.58	8.16	8.83	7.08	6.08	9.13
10	10.51	9.99	10.99	9.40	8.91	9.75	10.35	9.91	10.68	8.39	7.92	9.12	7.62	6.00	8.49
11	10.47	10.02	11.21	9.54	9.33	9.78	10.55	10.18	10.93	8.62	8.08	9.00	7.23	6.32	8.14
12	10.47	10.21	10.85	9.55	9.28	9.93	10.46	9.95	10.75	9.07	8.73	9.48	7.77	6.93	8.79

Table S3: KGE' across temporal aggregation windows (10-run uncertainty). Values are reported as the ensemble mean across 10 independent runs, with the minimum and maximum values across runs for each forecasting horizon (1–12 months ahead).

	RNN			**GRU**			**LSTM**			**KAN-LSTM**			**RCPIKLA**		
Time step	KGE Mean	KGE min	KGE max	KGE Mean	KGE min	KGE max	KGE Mean	KGE min	KGE max	KGE Mean	KGE min	KGE max	KGE Mean	KGE min	KGE max
1	0.67	0.67	0.68	0.71	0.71	0.72	0.68	0.67	0.68	0.70	0.69	0.70	0.80	0.76	0.81
2	0.72	0.70	0.72	0.73	0.72	0.73	0.71	0.70	0.72	0.76	0.75	0.77	0.79	0.76	0.82
3	0.69	0.69	0.71	0.72	0.70	0.73	0.68	0.68	0.69	0.77	0.75	0.78	0.80	0.78	0.83
4	0.68	0.65	0.70	0.71	0.70	0.72	0.68	0.66	0.68	0.77	0.76	0.79	0.81	0.77	0.85
5	0.65	0.63	0.67	0.69	0.67	0.71	0.65	0.63	0.67	0.76	0.74	0.78	0.79	0.73	0.82
6	0.64	0.62	0.68	0.67	0.65	0.69	0.65	0.63	0.67	0.72	0.70	0.75	0.79	0.76	0.82
7	0.64	0.62	0.68	0.69	0.68	0.71	0.65	0.61	0.67	0.75	0.71	0.77	0.81	0.75	0.85
8	0.64	0.61	0.67	0.68	0.66	0.70	0.65	0.63	0.67	0.75	0.71	0.78	0.82	0.77	0.87
9	0.69	0.63	0.72	0.72	0.70	0.74	0.69	0.65	0.72	0.78	0.74	0.81	0.80	0.73	0.84
10	0.68	0.64	0.71	0.71	0.69	0.74	0.67	0.64	0.70	0.78	0.74	0.81	0.77	0.73	0.81
11	0.68	0.65	0.70	0.71	0.70	0.72	0.65	0.61	0.68	0.76	0.73	0.78	0.76	0.71	0.85
12	0.67	0.64	0.69	0.71	0.69	0.74	0.66	0.64	0.69	0.76	0.74	0.78	0.74	0.69	0.79