Supplement of Hydrol. Earth Syst. Sci., 27, 4385–4408, 2023
https://doi.org/10.5194/hess-27-4385-2023-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Towards robust seasonal streamflow forecasts in mountainous catchments: impact of calibration metric selection in hydrological modeling

Diego Araya et al.

Correspondence to: Pablo A. Mendoza (pamendoz@uchile.cl)

Table S1. Catchment attributes: mean elevation (m a.s.l.; **ME**), area (km^2, **A**), fraction of basin covered by forest (%, **FBCF**), fraction of basin covered by barren land (%, **FBCBL**), mean annual temperature (°C, **MAT**), aridity index (-, **AI**), precipitation seasonality (-, **PS**), mean annual precipitation (mm/yr, **MAP**), precipitation falling as snow (%, **PFS**), mean annual runoff (mm/yr, **MAR**), baseflow index (-, **BFI**), interannual runoff variability (-, **IRV**).

Catchment	ME	A	FBC	FBCFL	MAT	AI	PS	MAP	PFS	MAR	BI	IRV
Pulido en Vertedero	3566	2022	0.0	63	6.3	7.0	-0.96	126	36	20	0.83	0.45
Toro antes de junta con la	3905	467	0.0	87	4.3	5.6	-0.62	133	50	43	0.68	0.38
La Laguna en Salida	4275	558	0.0	78	2.6	4.3	0.71	156	71	104	0.77	0.65
Derecho en Alcohuaz	3544	338	0.1	69	4.6	3.2	-1.12	222	64	107	0.81	1.00
Hurtado en Angostura	3724	672	0.1	66	3.9	2.4	-1.04	251	70	128	0.86	0.81
Rapel en Junta	2661	821	0.6	34	8.7	3.0	-1.05	313	12	71	0.67	1.30
Grande en las Ramadas	3098	569	0.0	67	6.7	2.9	-1.12	284	34	202	0.78	0.89
Grande en Cuyano	2726	1287	0.2	49	8.8	3.6	-1.21	277	8	147	0.75	0.96
Choapa en Cuncumén	3142	1132	0.2	64	6.0	2.7	-1.18	318	53	236	0.75	0.74
Pedernal en Tajada	2454	81	0.1	27	9.6	2.8	-1.25	355	3	89	0.65	1.06
Sobrante en Piñadero	2610	241	0.2	40	8.3	2.3	-1.21	397	13	90	0.78	1.15
Alicahue en Colliguay	2403	348	0.2	37	9.1	2.5	-1.18	391	7	94	0.78	1.06
Aconcagua en Chacabuquito	3178	2113	0.3	66	5.3	1.9	-1.16	390	52	448	0.76	0.52
Pocuro en el Sifón	2006	181	0.2	20	11.1	2.5	-1.22	445	1	124	0.72	0.79
Arrayán en la Montosa	2513	216	0.4	50	8.1	1.6	-1.15	560	18	207	0.75	0.63
Yerba Loca antes de junta con	3424	110	0.2	70	3.3	1.2	-1.13	456	72	278	0.75	0.44
Maipo en el Manzano	3181	4839	0.2	69	5.0	0.9	-1.12	764	59	605	0.81	0.35
Claro en el Valle	1605	349	27.1	20	10.7	0.8	-1.16	1297	1	702	0.63	0.59
Claro en los Queñes	1857	354	14.6	36	8.9	0.6	-1.11	1568	8	1527	0.70	0.41
Colorado en junta con Palos	2288	878	11.5	57	7.1	0.5	-1.09	1650	23	1526	0.73	0.38
Palos en junta con Colorado	1973	490	16.7	55	8.1	0.5	-1.07	1743	14	1532	0.80	0.32
Sauces antes de junta con Ñuble	1683	607	9.8	20	8.3	0.6	-1.04	1752	10	1581	0.68	0.36

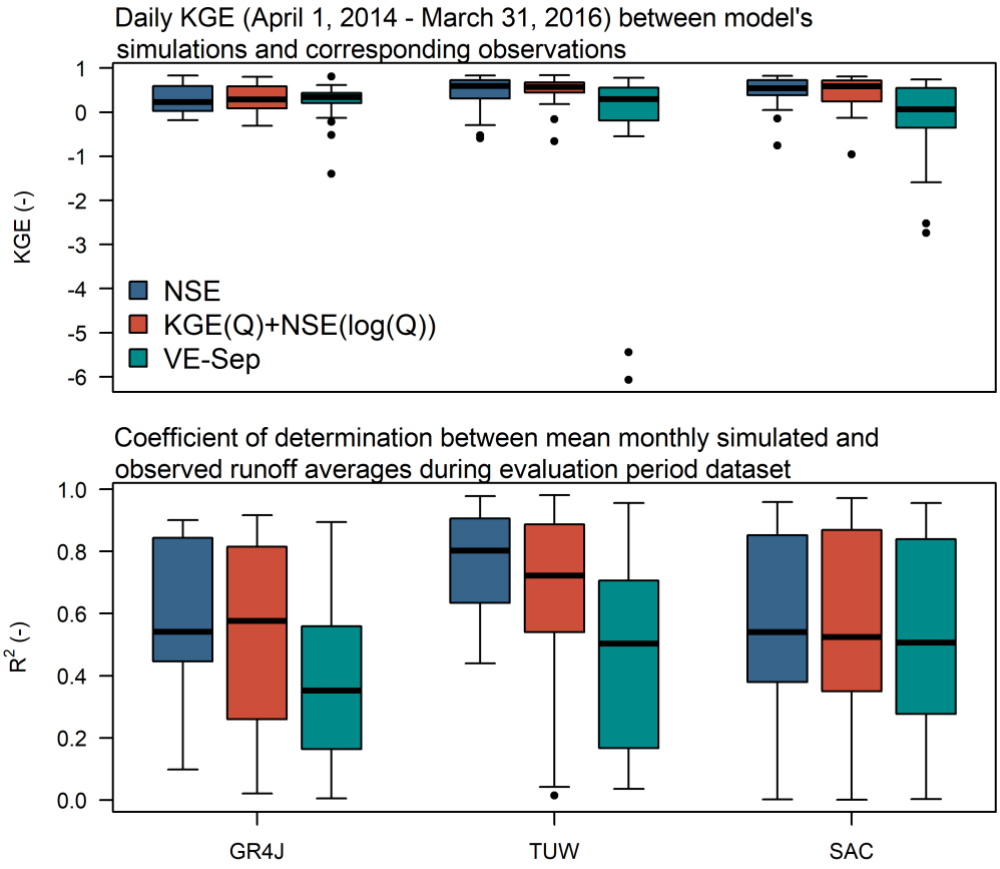

Figure S1. Kling-Gupta Efficiency (KGE) between simulated and observed daily streamflow for the period April/2014-March/2016, using parameter values obtained with different calibration objective functions and hydrological models (upper panel); and coefficient of determination (R^2) between mean monthly simulated and observed runoff averages for the evaluation dataset (April/1987 – March/1994 and April/2013 – March/2020, bottom panel). Each boxplot comprises results from the 22 case study basins. The boxes correspond to the interquartile range (IQR, i.e., 25th and 75th percentiles), the horizontal line in each box is the median, and the whiskers extend to the $\pm 1.5 \cdot IQR$ of the ensemble. The points correspond to outliers beyond the whiskers' range.

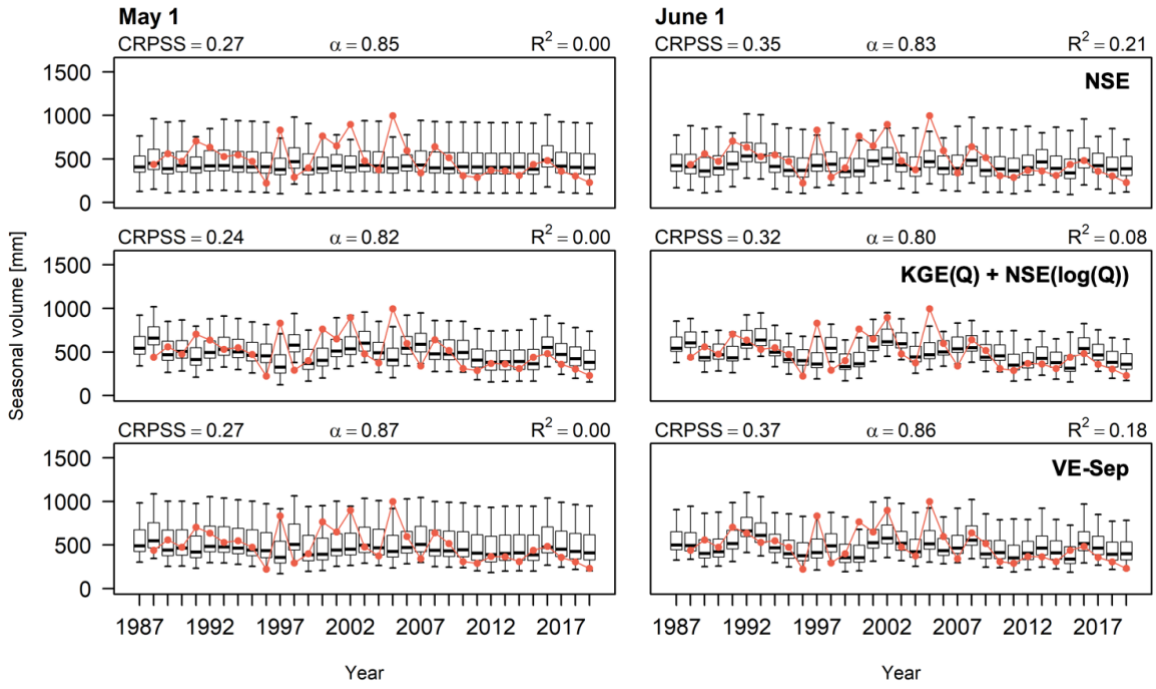

Figure S2. Time series with ESP seasonal hindcasts (i.e., September-March runoff) initialized on May 1 (left panels), and June 1 (right panels) for the Maipo at El Manzano basin. The boxes correspond to the interquartile range (IQR, i.e., 25th and 75th percentiles); the horizontal line in each box is the median, whiskers extend to the $\pm 1.5 \cdot IQR$ of the ensemble, and the red dots represent the observations. The results were produced with the TUW model, using parameters obtained from calibrations conducted with NSE, KGE(Q)+NSE(log(Q)) and VE-Sep. Each panel displays the CRPSS, the reliability index α, and the coefficient of determination R^2 (computed using the ensemble forecast median).

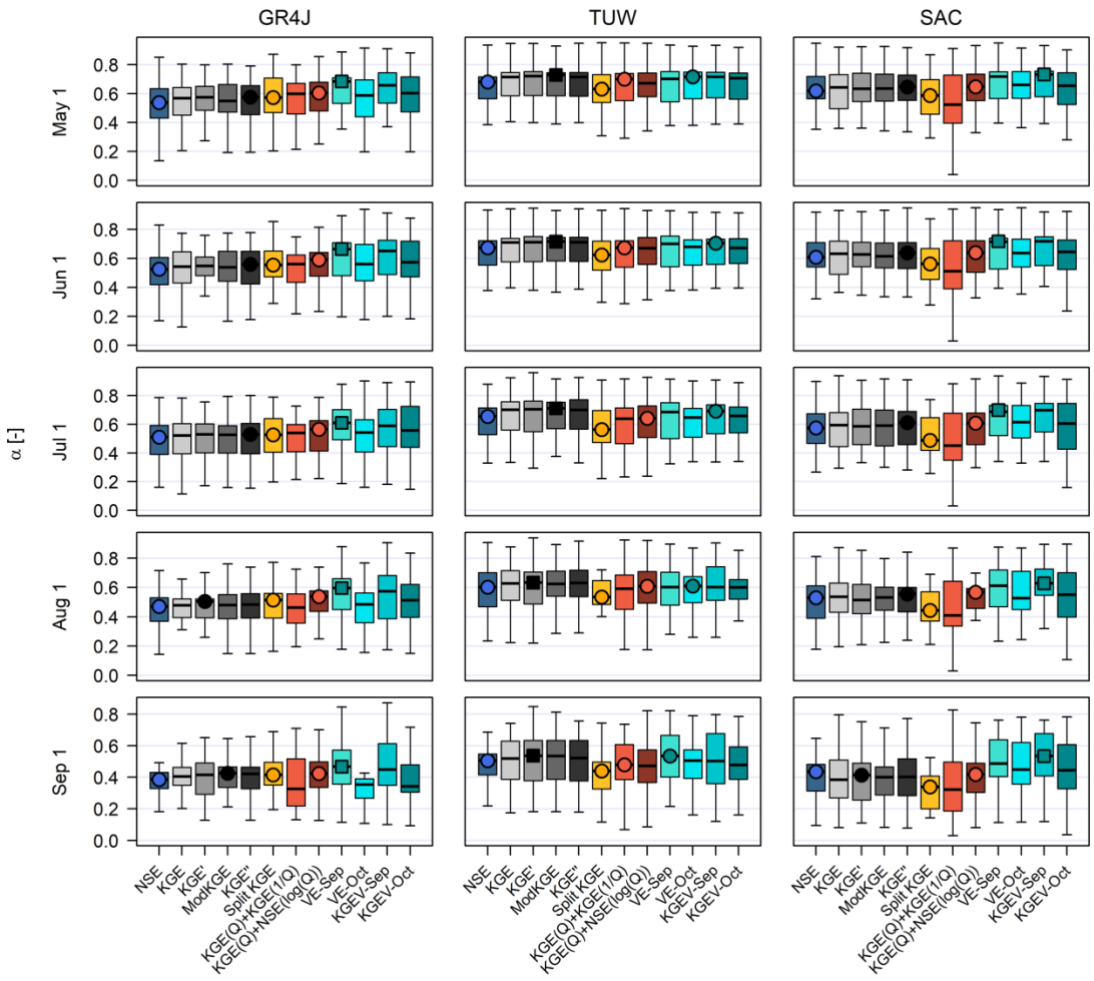

Figure S3. Comparison of the α reliability indices obtained with different calibration objective functions. Each panel contains results for a specific combination initialization time (rows) and hydrological model (columns), and each boxplot comprises results from the 22 case study basins. The boxes correspond to the interquartile range (IQR, i.e., 25th and 75th percentiles), the horizontal line in each box is the median, and whiskers extend to the $\pm 1.5 \cdot IQR$ of the ensemble. The circle indicates the objective function providing the highest median within each family of calibration metric (identified with different colors), and the square indicates the objective function that delivers the best set of metric values using a specific combination of initialization time and hydrological model.

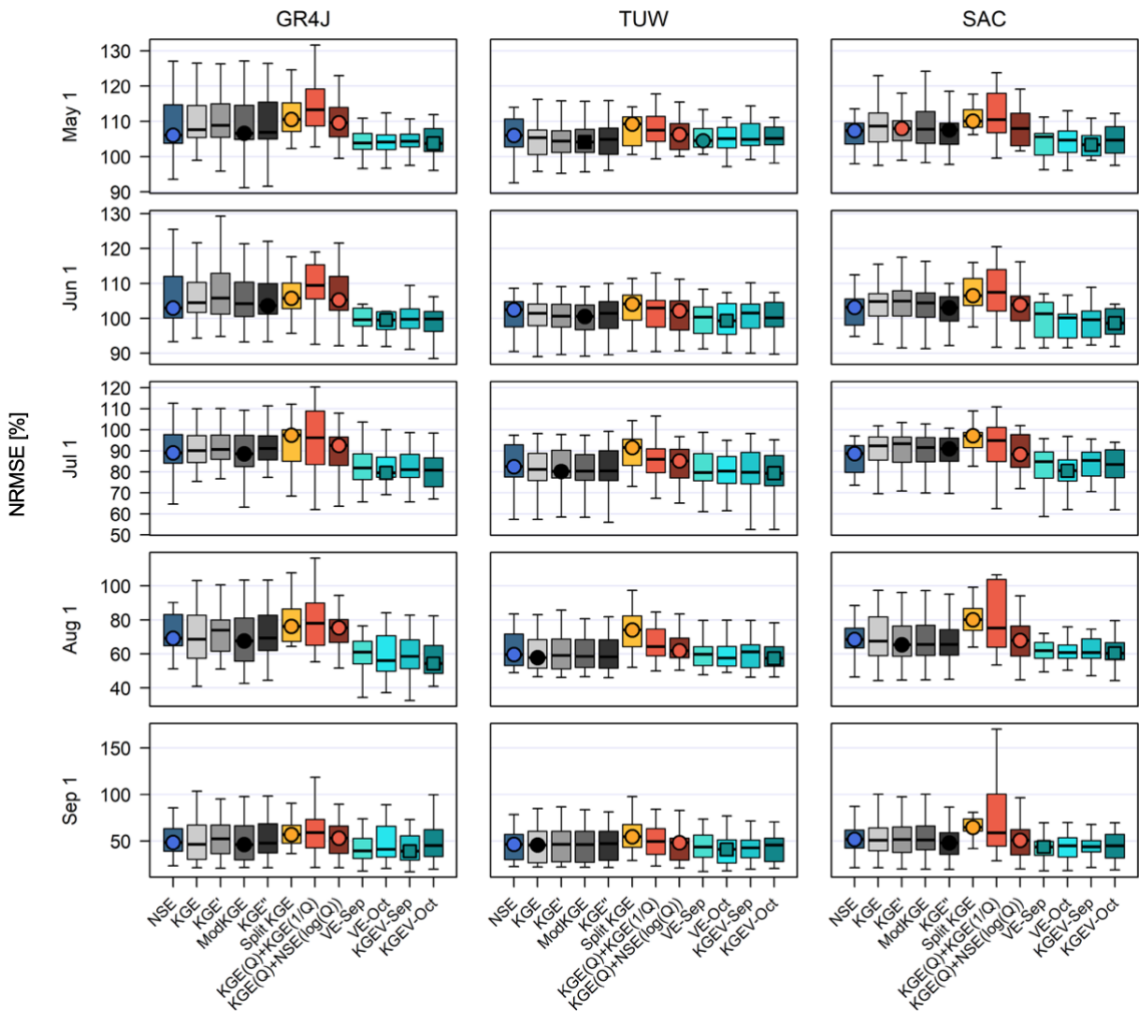

Figure S4. Same as in Figure S3, but for NRMSE.

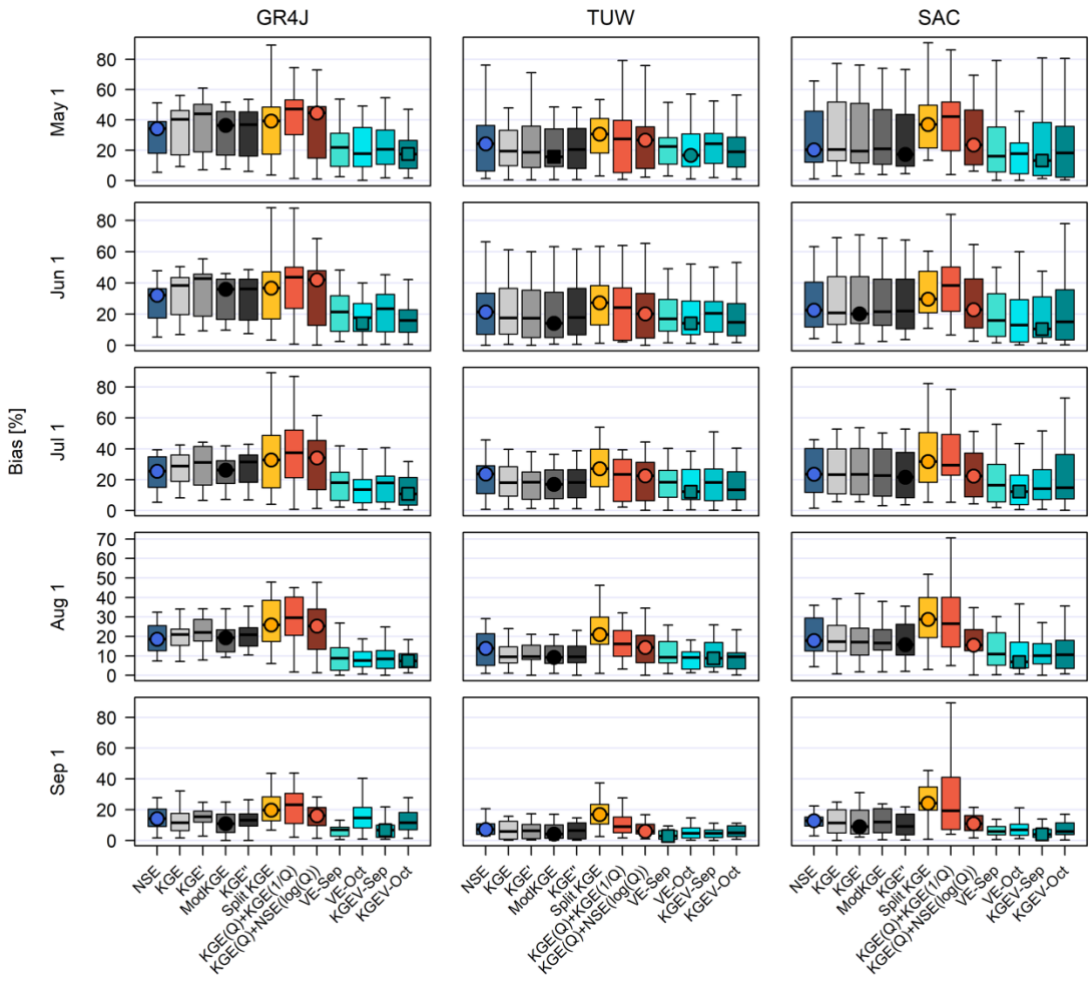

Figure S5. Same as in Figure S3, but for the percent bias.

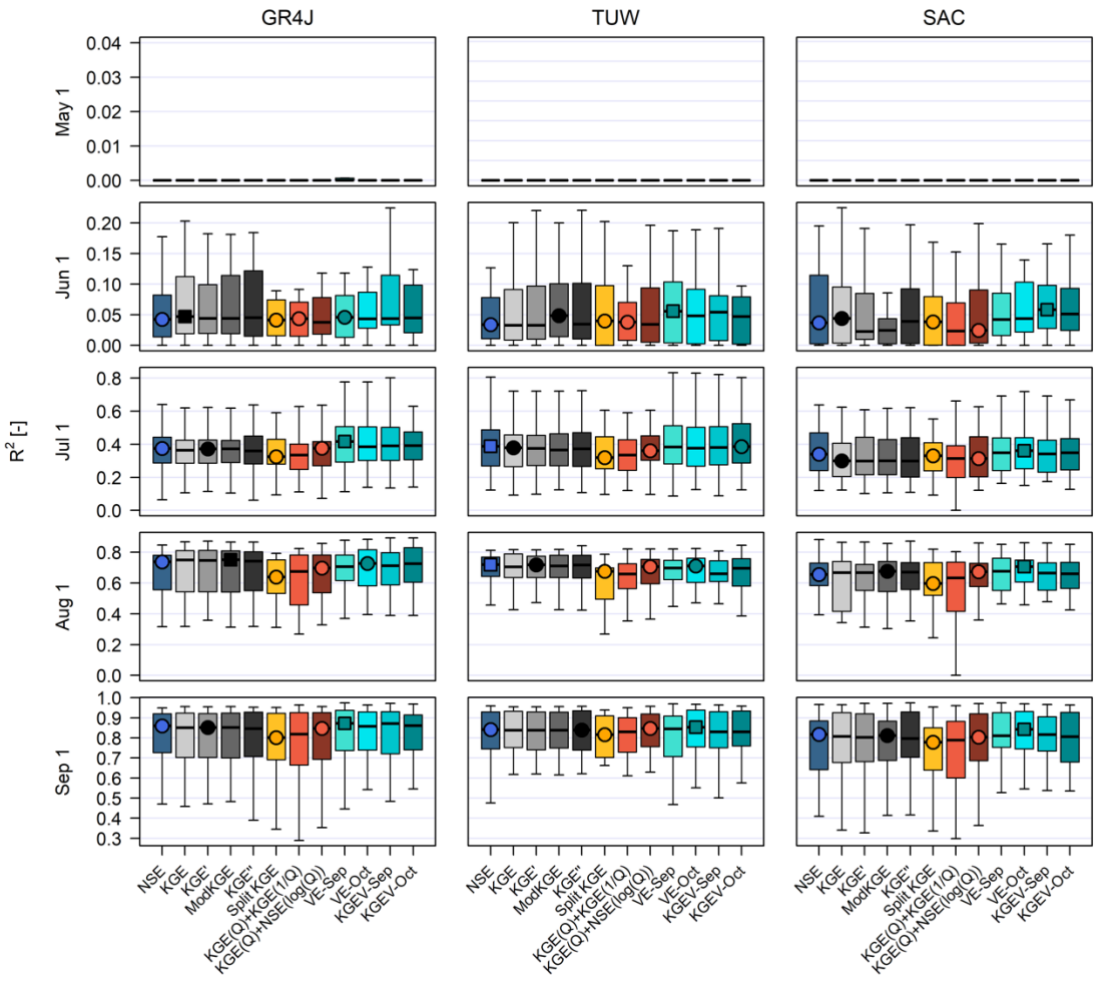

Figure S6. Same as in Figure S3, but for R^2.

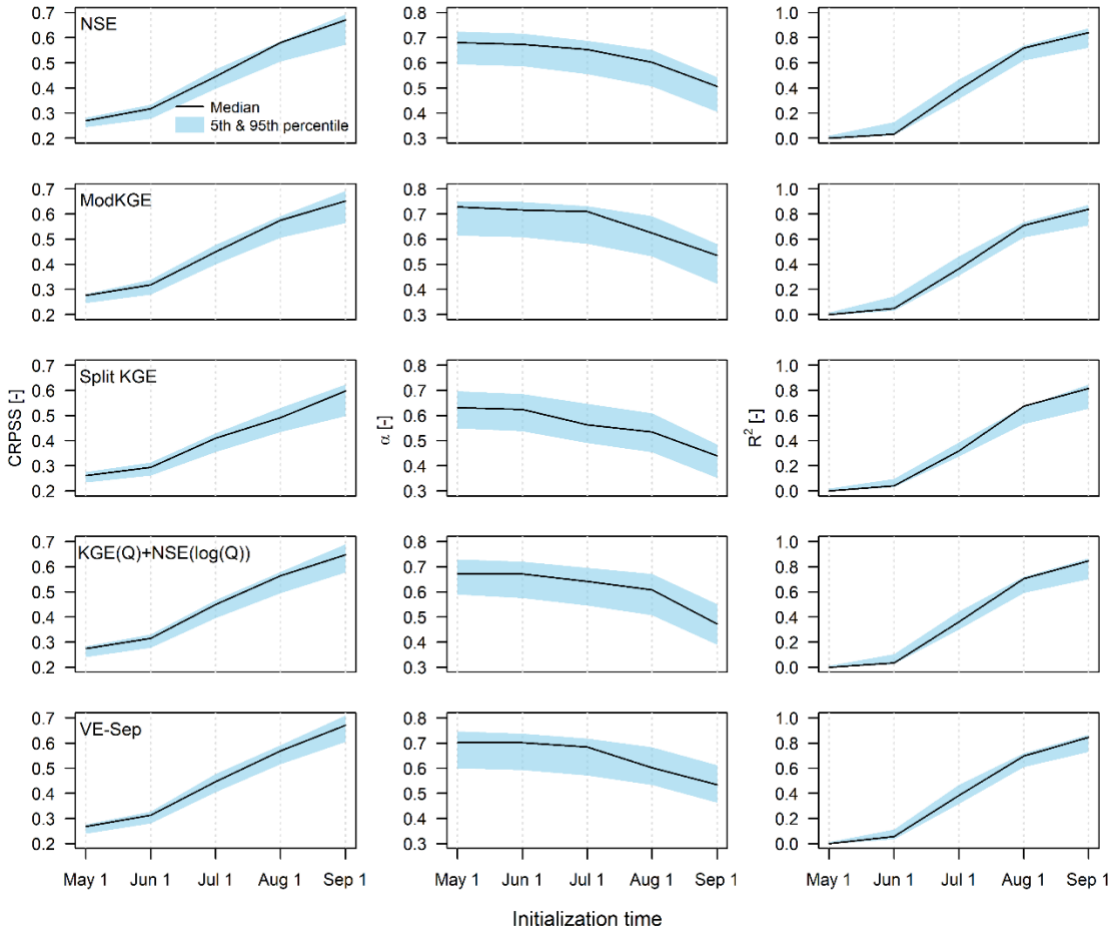

Figure S7. Impact of initialization time on hindcast performance using NSE, ModKGE, Split KGE, KGE(Q)+NSE(log(Q)) and VE-Sep as calibration metrics with the TUW model. The shades represent the 5th and 95th percentiles in each metric from the 22 case study catchments, and the solid line represents the median of each metric.

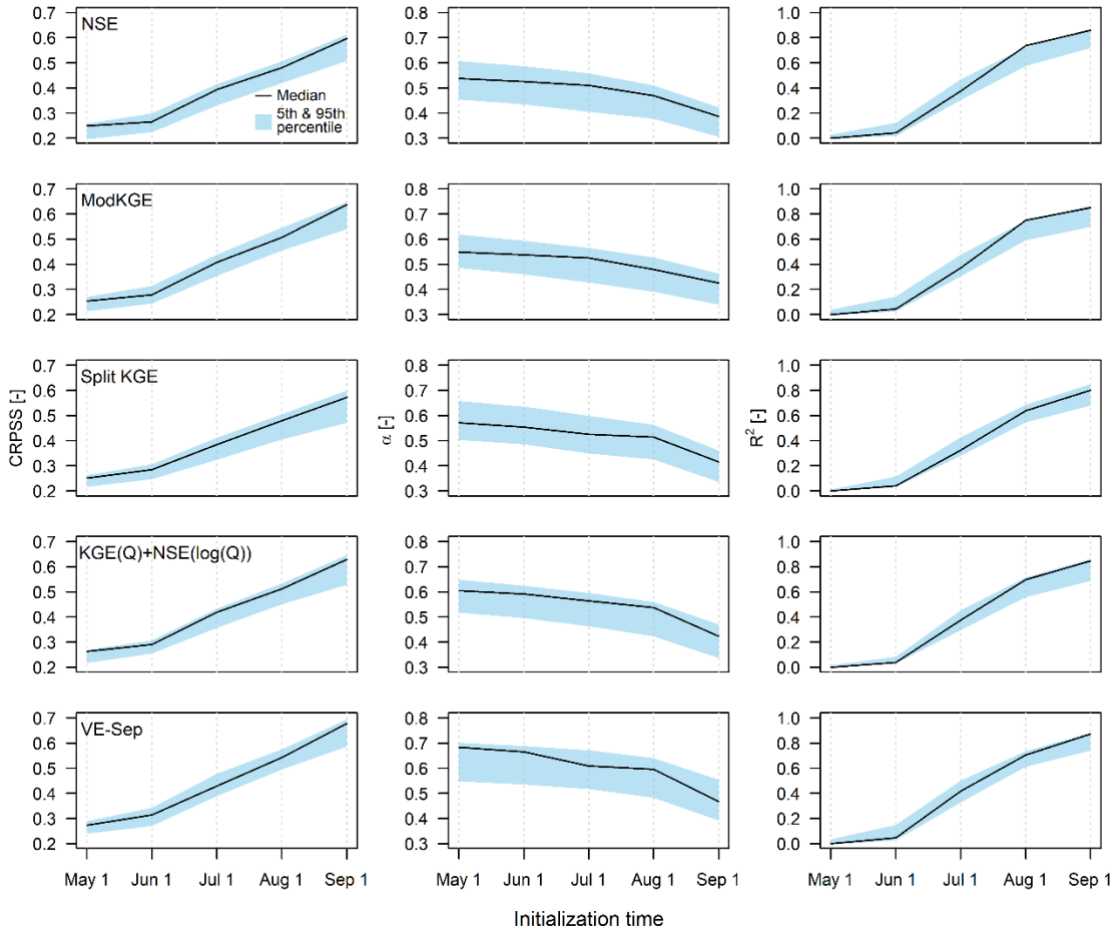

Figure S8. Same as in Figure S7, but for the GR4J model.

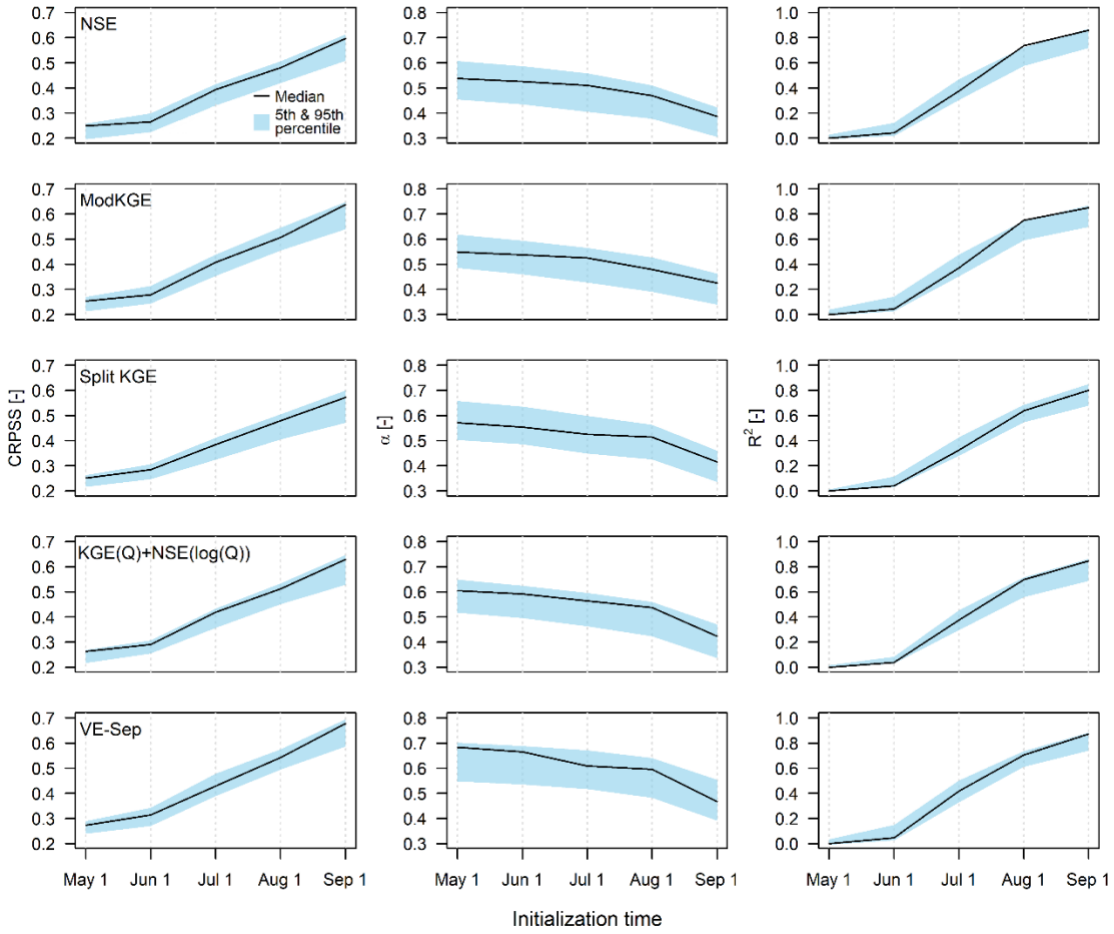

Figure S9. Same as in Figure S7, but for the SAC-SMA model.

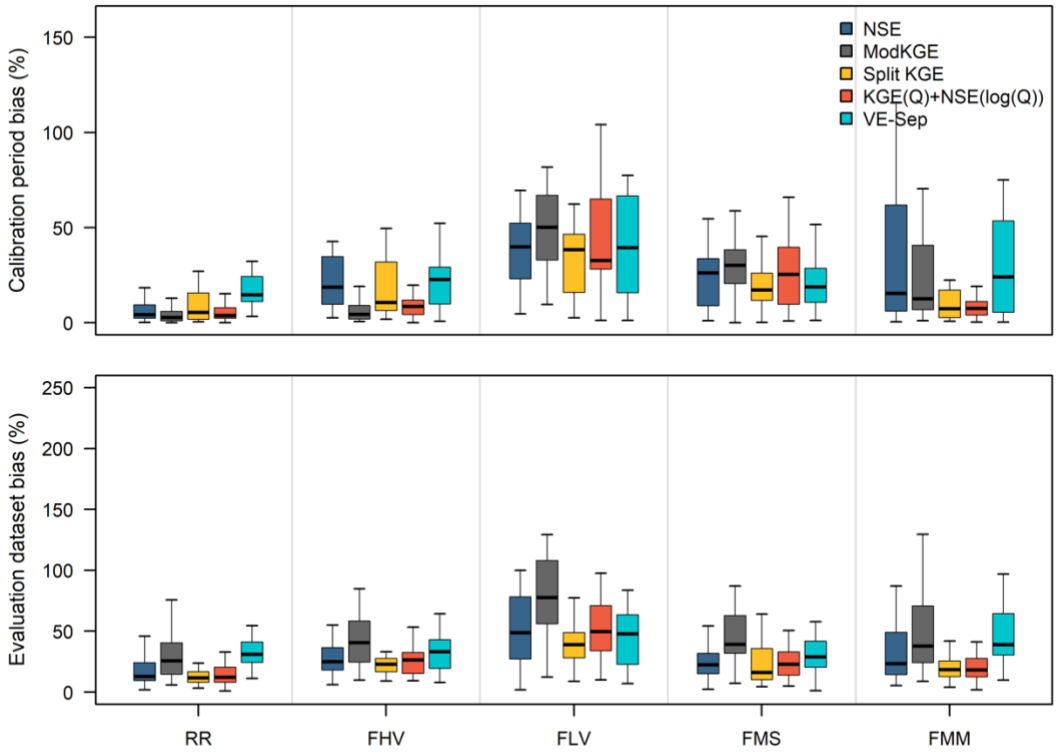

Figure S10. Percent biases (y-axis) in hydrological signatures (x-axis) obtained with the five representative objective functions and the GR4J model for the (upper panel) calibration (April/1994 – March/2013) and (bottom panel) evaluation dataset (April/1987 – March/1994 and April/2013 – March/2020). Each boxplot comprises results for our 22 case study basins. The boxes correspond to the interquartile range (IQR, i.e., 25^{th} and 75^{th} percentiles), the horizontal line in each box is the median, and whiskers extend to the $\pm 1.5 \cdot IQR$ of the ensemble.

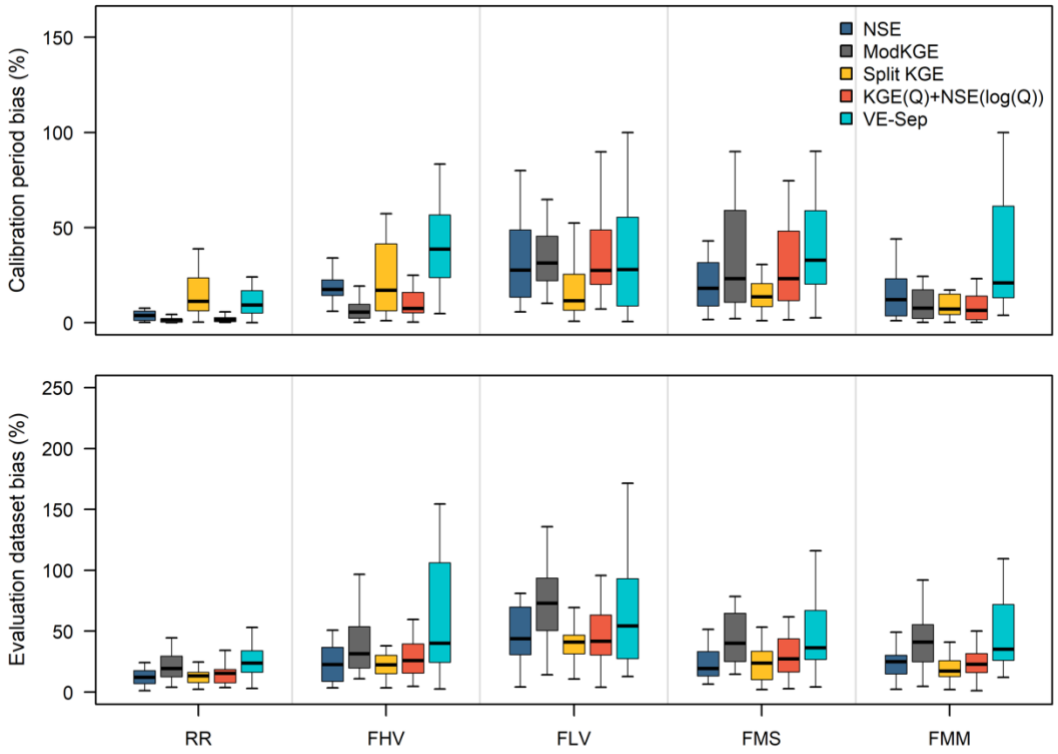

Figure S11. Same as in Figure S10, but for SAC-SMA model.

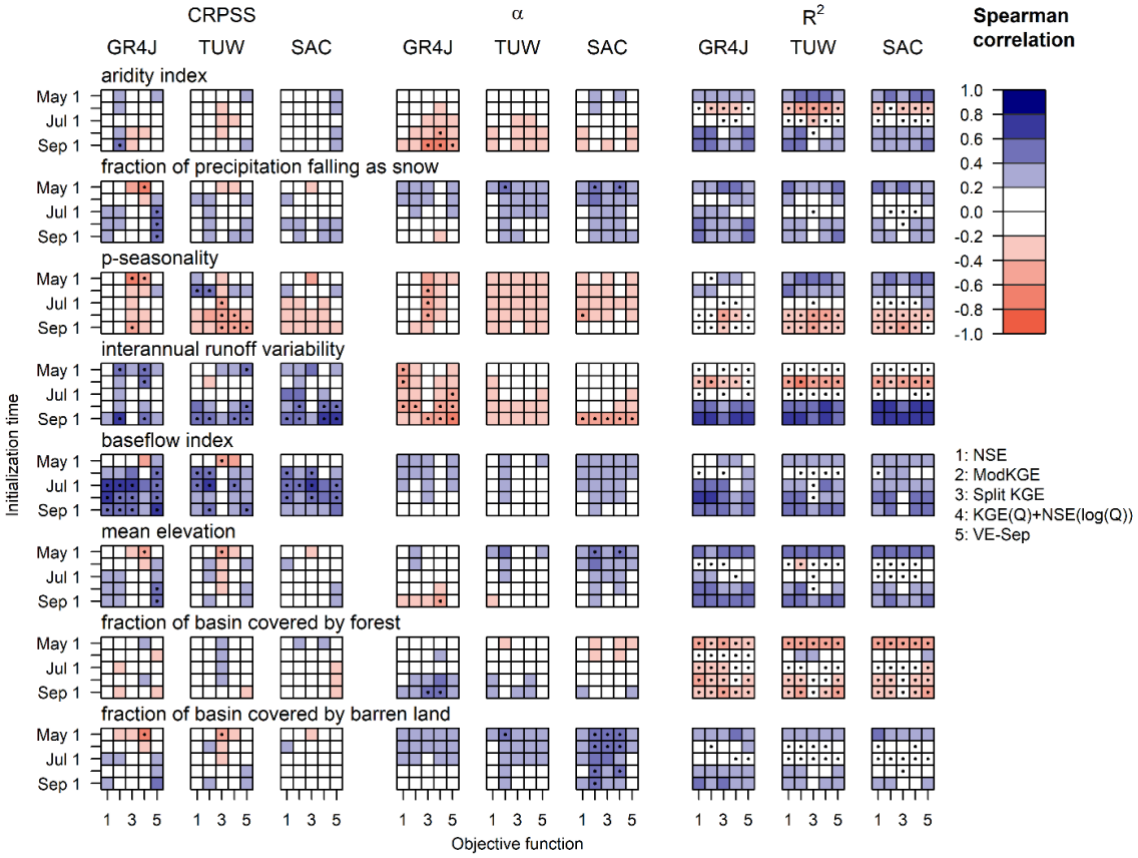

Figure S12. Spearman's rank correlation coefficients between catchment characteristics (shown in different rows) and the CRPSS (left), α reliability index (center), and the coefficient of determination R^2 (right) obtained for seasonal streamflow hindcasts (period April/1987 – March/2020), produced with the five representative objective functions (x-axis in each color matrix), different initialization times (y-axis in each color matrix), and the three hydrological models. Black dots indicate statistically significant ($p < 0.05$) correlations.