

S1 Informal Literature Review on Calibration Metric Choice

Further details about the informal literature review presented in the introduction of the paper can be found in Table S4 (uploaded as separate .xlsx file). Specifically, it lists, describes and categorizes the reasoning of choosing objective functions for 60 hydrological modelling studies.

S2 Further Details on Methodology and Analysis

S2.1 List of MARRMoT Models

Table S1: MARRMoT v2.1 models (1/2)

Number	Name	Storages	Params
01	collie1	1	1
02	wetland	1	4
03	collie2	1	4
04	newzealand1	1	6
05	ihacres	1	7
06	alpine1	2	4
07	gr4j	2	4
08	us1	2	5
09	susannah1	2	6
10	susannah2	2	6
11	collie3	2	6
12	alpine2	2	6
13	hillslope	2	7
14	topmodel	2	7
15	plateau	2	8
16	newzealand2	2	8
17	penman	3	4
18	simhyd	3	7
19	australia	3	8
20	gsfb	3	8
21	flexb	3	9
22	vic	3	10
23	lascam	3	24
24	mopex1	4	5

Table S1: MARRMoT v2.1 models (2/2)

Number	Name	Storages	Params
25	tcm	4	6
26	flexi	4	10
27	tank	4	12
28	xinanjiang	4	12
29	hymod	5	5
30	mopex2	5	7
31	mopex3	5	8
32	mopex4	5	10
33	sacramento	5	11
34	flexis	5	12
35	mopex5	5	12
36	modhydrolog	5	15
37	hbv	5	15
38	tank2	5	16
39	mcrm	5	16
40	smar	6	8
41	nam	6	10
42	hycymodel	6	12
43	gsmsocont	6	12
44	echo	6	16
45	prms	7	18
46	classic	8	12
47	IHM19	4	16

S2.2 Catchment Clustering

To select these basins, we categorized the 2901 available CARAVAN catchments by using a climate indices approach as applied in Knoben et al. (2018). It is based on a monthly moisture index (Willmott and Feddema, 1992) which is defined as follows:

$$MI = \begin{cases} 1 - \frac{E_p(t)}{P(t)}, & P(t) > E_p(t) \\ 0, & P(t) = E_p(t) \\ \frac{P(t)}{E_p(t)} - 1, & P(t) < E_p(t) \end{cases} \quad (1)$$

Where $E_p(t)$ is monthly values of potential evapotranspiration and $P(t)$ is monthly precipitation.

Based on this index, the dimensionless indices of aridity (I_M) and seasonality ($I_{M,r}$) can be derived:

$$I_M[-1, 1] = \frac{1}{12} \sum_{t=1}^{t=12} MI(t) \quad (2)$$

The moisture index I_M indicates the mean aridity of the location, which negative values indicating water limitation and positive values energy limitation.

$$I_{M,r}[0, 2] = \max(MI) - \min(MI) \quad (3)$$

The seasonality index $I_{M,r}$ shows the variance in moisture availability, with higher values indicating more variance.

Finally, the fraction of snow f_s is calculated using a threshold temperature of $T_0 = 0^\circ C$ in our case:

$$f_s = \frac{\sum_{t=1}^{t=12} P(T(t) \leq T_0)}{\sum_{t=1}^{t=12} P(t)} \quad (4)$$

f_s adds additional information regarding the fraction of precipitation that falls as snow and ranges from 0 (none) to 1 (all).

We used a k-means clustering algorithm to assign each of the 2901 catchments to one of ten clusters according to their aridity, seasonality and fraction of snow. We then picked one catchment closest to the centre of each cluster. Figure S2 shows the ten obtained clusters as well as the cluster location of the catchments that were selected (purple dots). When selecting the catchments we considered two things. First, we limited the catchment size from 100 to 1000 km^2 to increase the probability of an adequate signature representation through the calibrated models. Smaller catchments are less likely to dampen catchment responses and therefore better contain scale-dependent effects that also affect signature representation. Second, we ensured that the catchments had an data length of at least 20 years. To ensure the reliability of the identified clusters we repeated the clustering ten times with random seeds and could not identify large changes in the location of the centroids.

K-means clustered CARAVAN catchments in climate index space

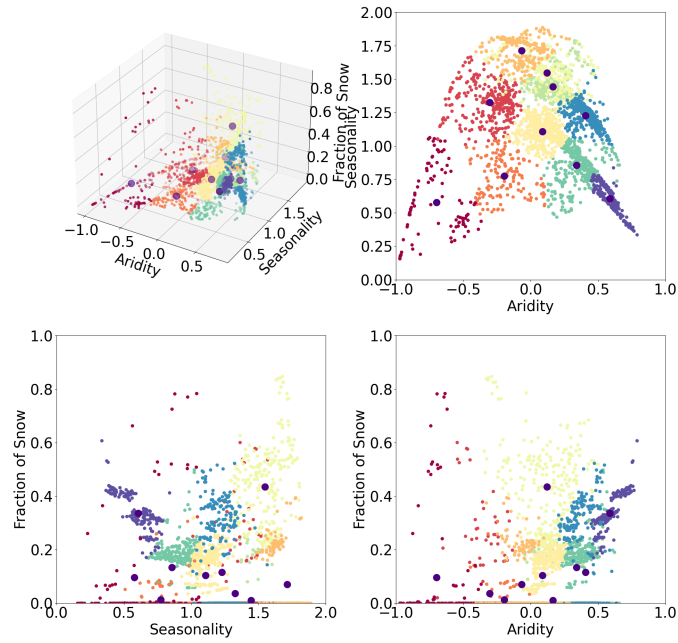


Figure S2. Results of the k-means clustering of 2901 CARAVAN catchments for their aridity, seasonality and fraction of snow. Each colored point cloud refers to one cluster and the thicker purple dots are the catchments selected for this study.

S2.3 Testing Differences in Forcing Data

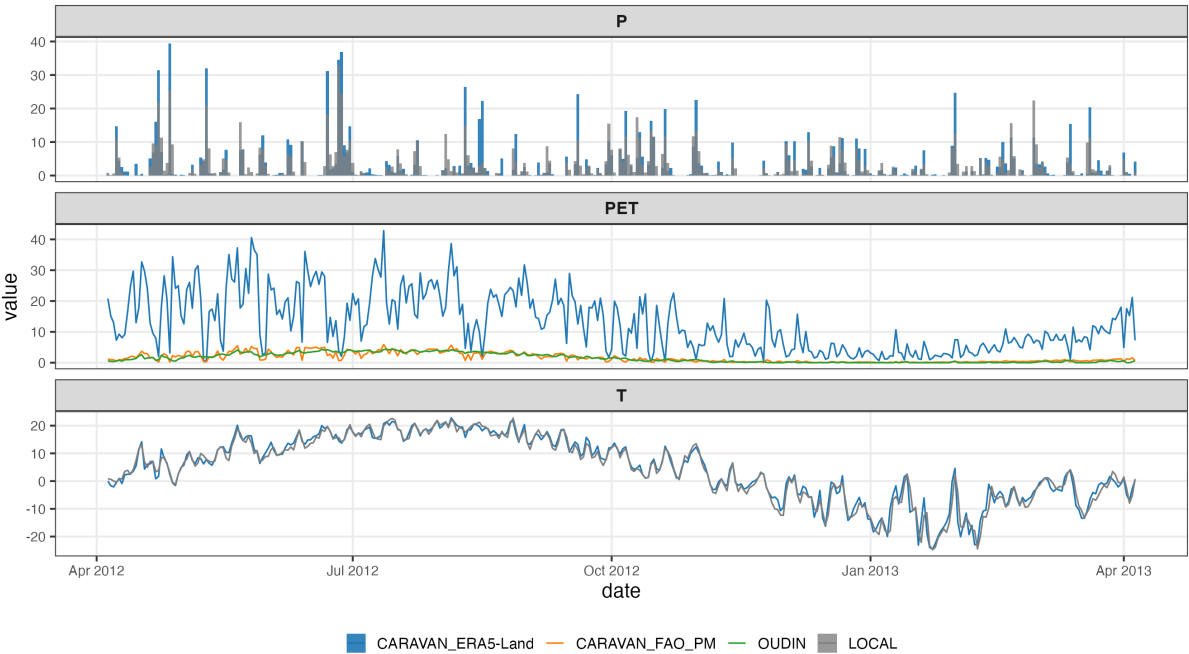


Figure S3. Comparison of Different forcing data for a selected year at HYS

Figure S3 shows the comparison of different available forcing choices for temperature, potential evapotranspiration and precipitation. In agreement with (Clerc-Schwarzenbach et al., 2024), we chose to replace the default CARAVAN PET (ERA5-Land) with a provided alternative (Penman-Monteith), but retained temperature and precipitation data.

Below are comparable plots for all other catchments (depending on data availability of the specific data sets). We compare to all of the local datasets in this study (Newman et al., 2015; Addor et al., 2017; Arsenault et al., 2020; Klingler et al., 2021; Fowler et al., 2021; Chagas et al., 2020; Coxon et al., 2020).

Catchment C02:

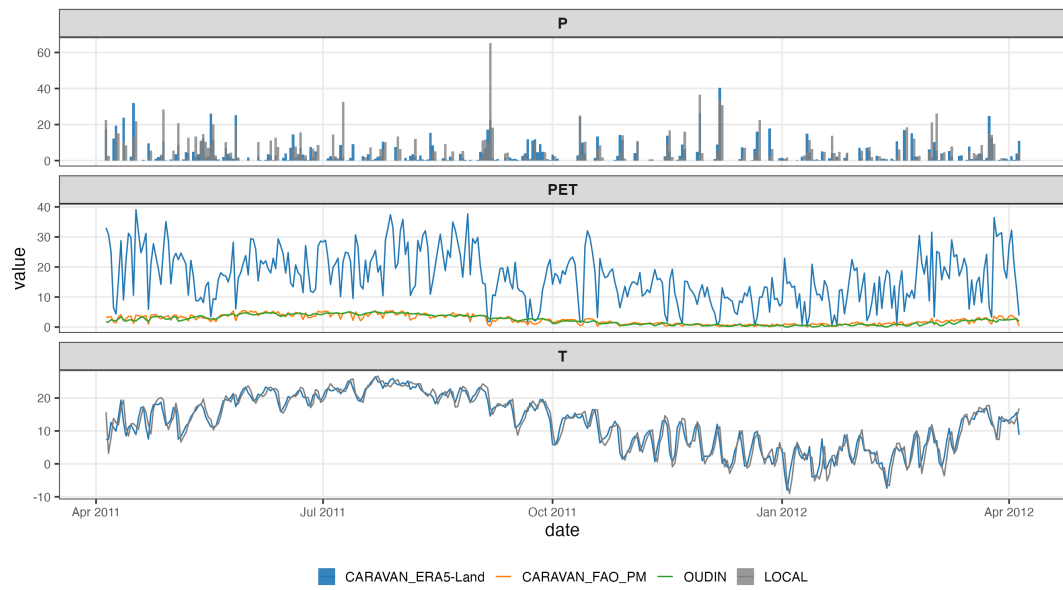


Figure S4. Comparison of Different forcing data for a selected year at C02

Catchment C03:

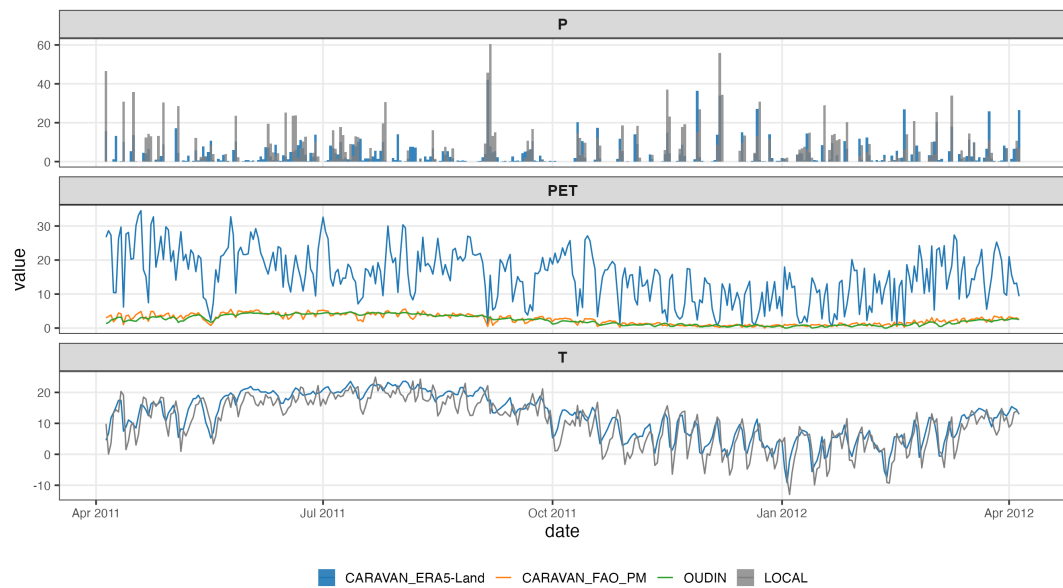


Figure S5. Comparison of Different forcing data for a selected year at C03

Catchment C12:

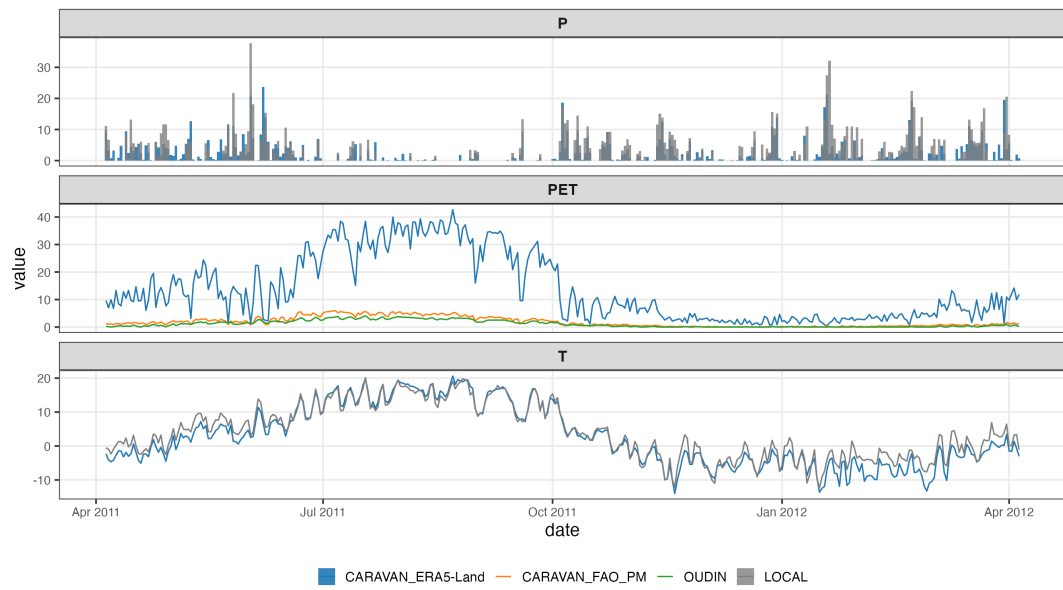


Figure S6. Comparison of Different forcing data for a selected year at C03

Catchment AUS1:

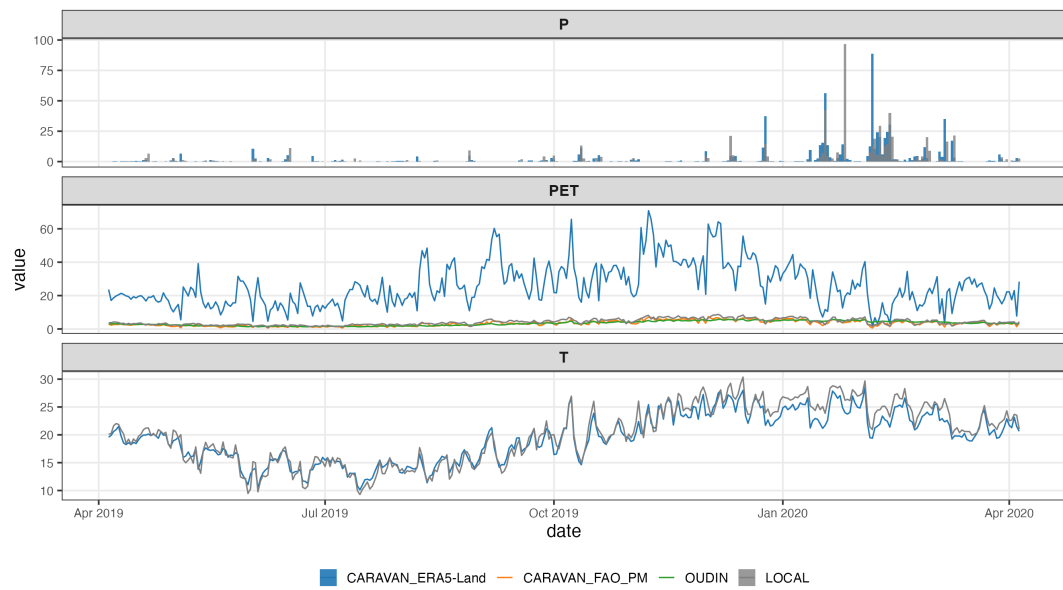


Figure S7. Comparison of Different forcing data for a selected year at C03

Catchment AUS6:

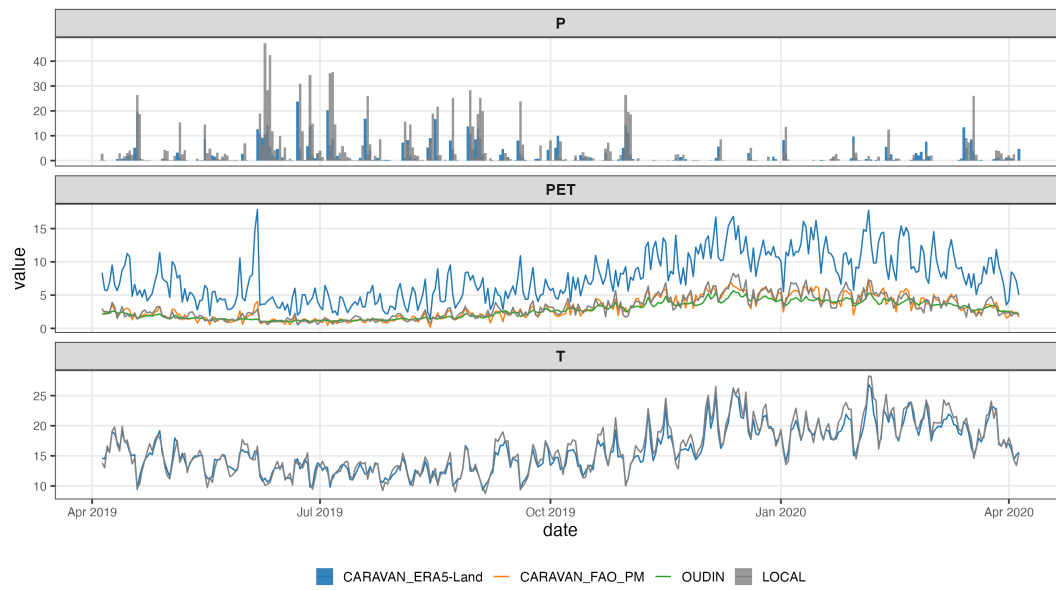


Figure S8. Comparison of Different forcing data for a selected year at AUS6

Catchment BR6:

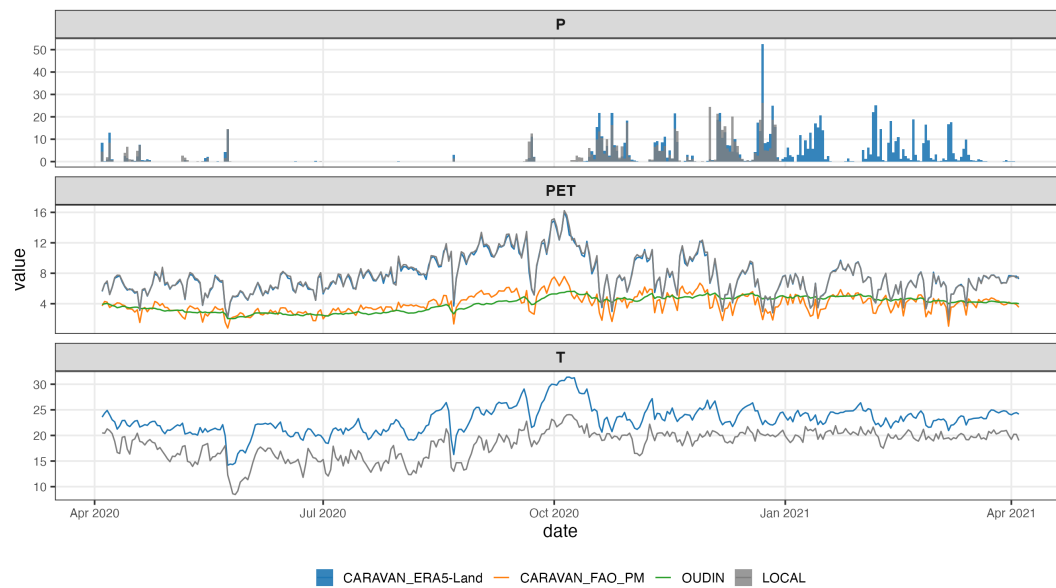


Figure S9. Comparison of Different forcing data for a selected year at BR6

Catchment GB2:

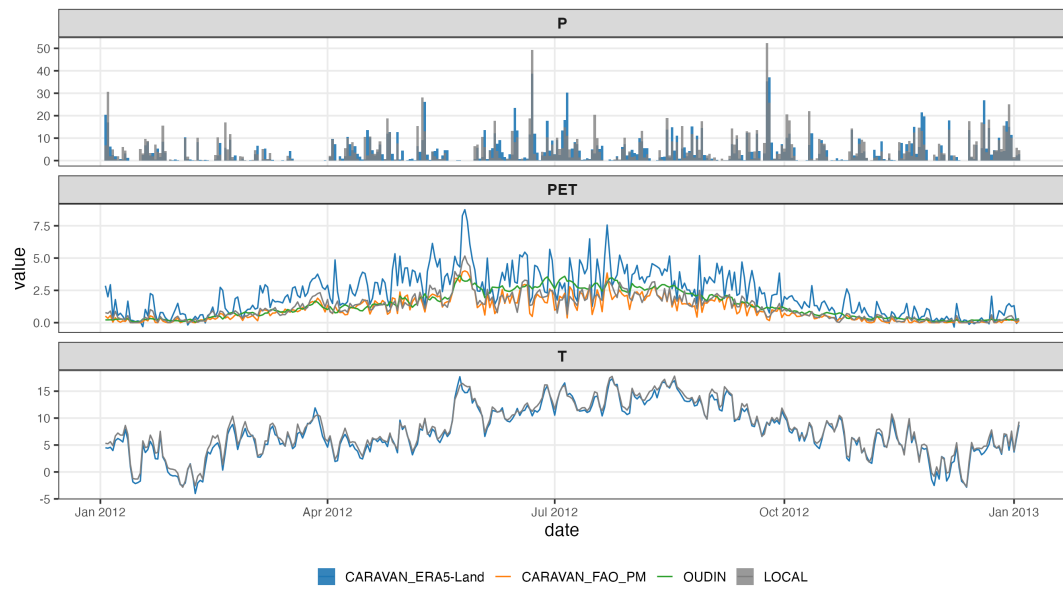


Figure S10. Comparison of Different forcing data for a selected year at GB2

Catchment GB3:

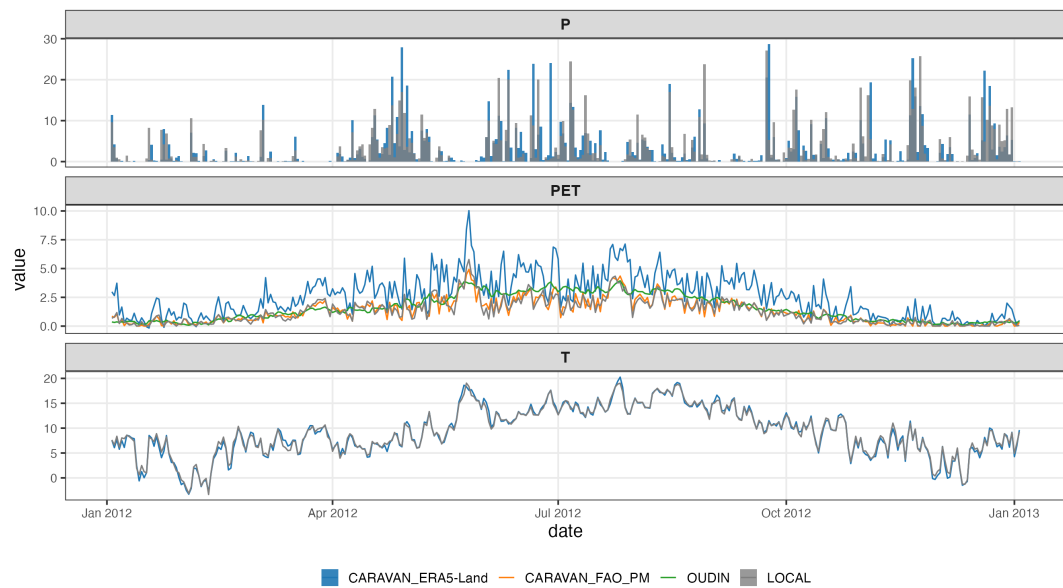


Figure S11. Comparison of Different forcing data for a selected year at GB3

Catchment LAM:

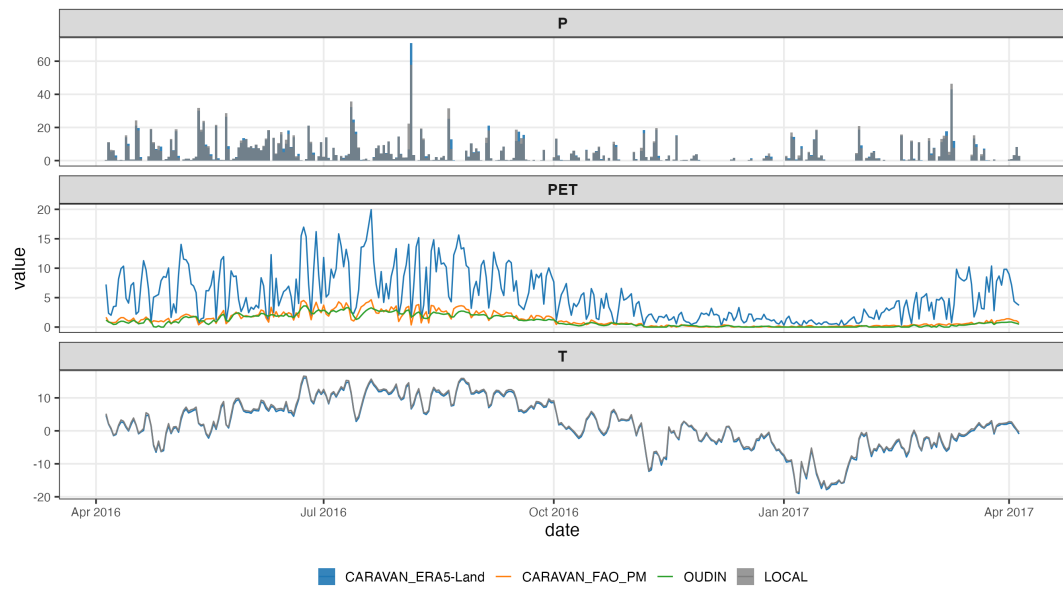


Figure S12. Comparison of Different forcing data for a selected year at LAM

S2.4 Difference Calibration and Validation

Observed signature differences:

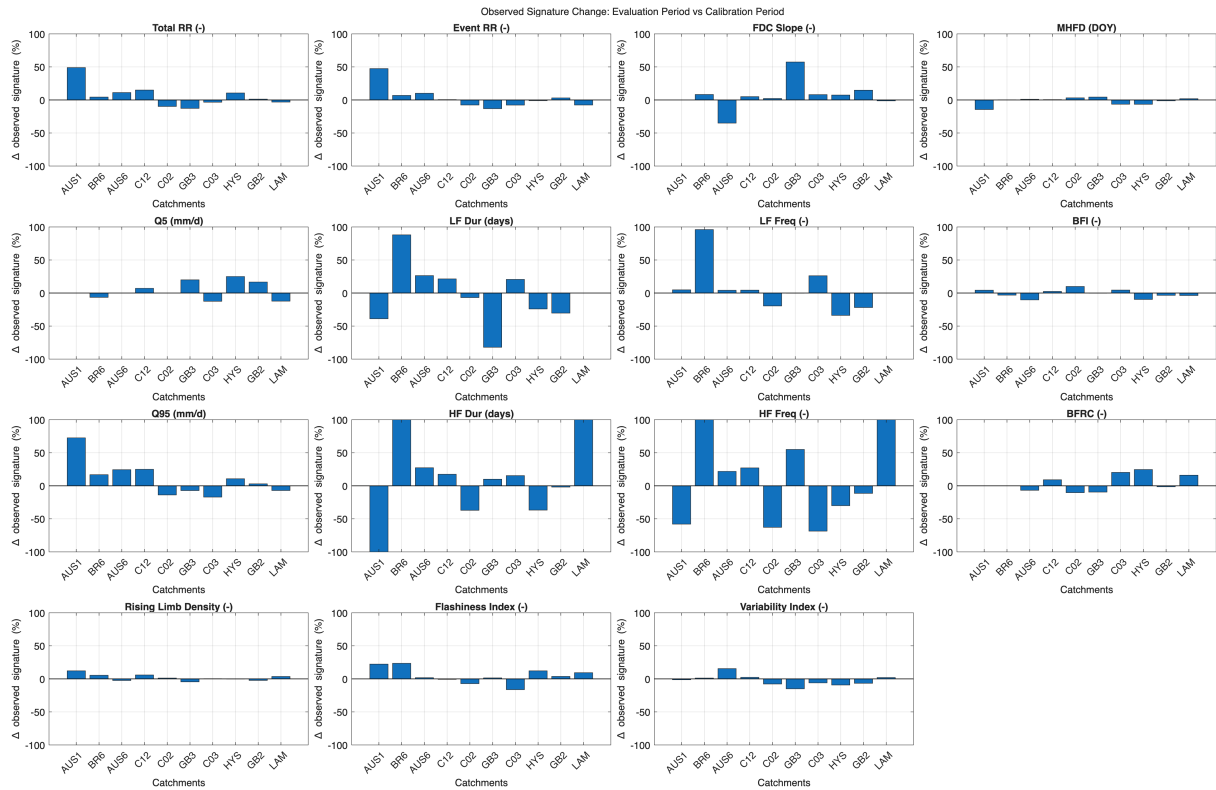


Figure S13. Differences in observed signature value for calibration vs. evaluation period.

Comparing calibration and evaluation period, we find that observed differences depend on (1) the catchment, such that AUS1 often has a considerable shift in signature representation, while other catchments are more consistent and (2) the signature, where TRR, ERR, MHFD, BFI, BRFC, RLD stay rather consistent across periods, while extreme flow duration and frequency are more susceptible to change. This change is represented in our model results as well, indicating adequate representation across both modelling periods.

Model differences for best runs in both time periods:

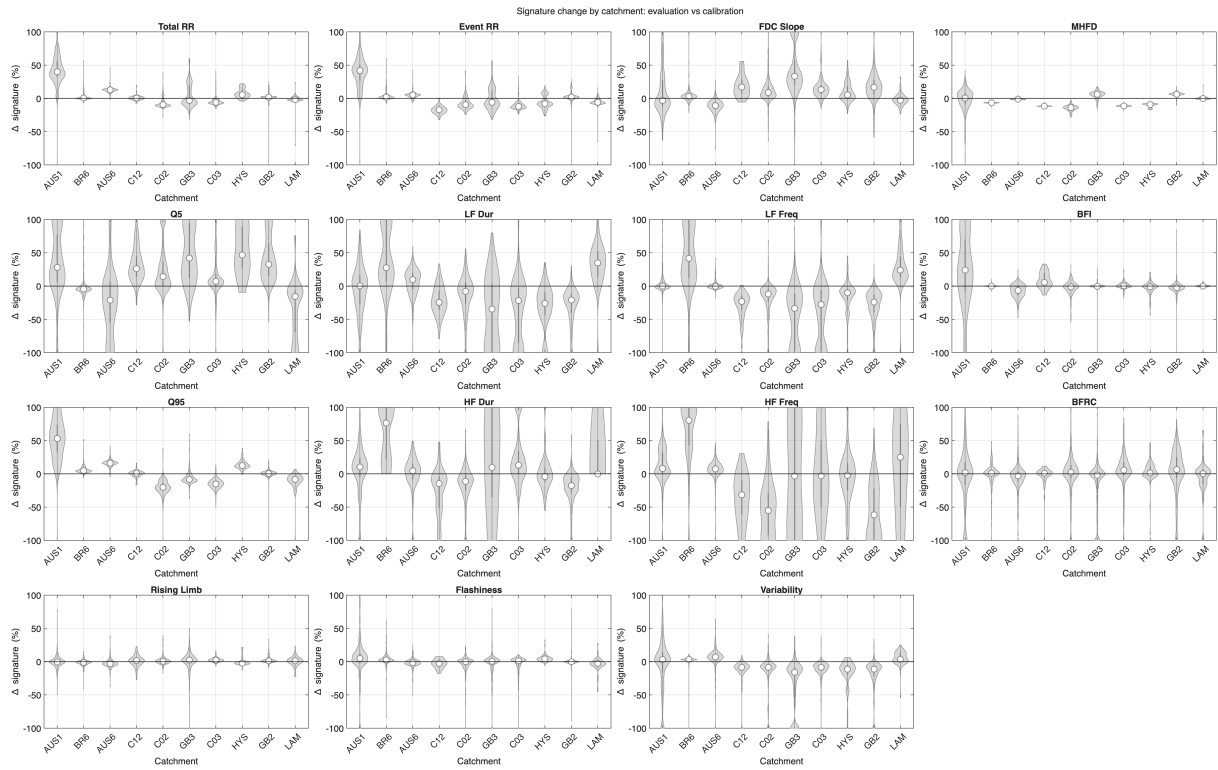


Figure S14. Differences in modelled signature value for calibration vs. evaluation period.

S2.5 Signature Definitions

All 15 streamflow signatures were computed with the TOSSH toolbox (Gnann et al., 2021) using default parameter settings unless stated otherwise. Throughout, $Q(t)$ is daily discharge [mm d^{-1}], $P(t)$ daily precipitation [mm d^{-1}], N the number of time steps, $\overline{(\cdot)}$ the temporal mean, and $\tilde{Q}$ the median daily discharge.

S2.5.1 Total Runoff Ratio (Total RR).

The fraction of precipitation leaving the catchment as streamflow,

$$\text{TotalRR} = \frac{\overline{Q}}{\overline{P}}. \quad (5)$$

S2.5.2 Event Runoff Ratio (Event RR).

The mean ratio of event runoff to event precipitation, averaged over rainfall–runoff events extracted from the precipitation series. Events are delineated using the TOSSH defaults: minimum intensities of 2 mm h^{-1} and 10 mm d^{-1} , and up to one recession day assigned to the preceding event,

$$\text{EventRR} = \frac{1}{N_e} \sum_{n=1}^{N_e} \frac{\sum_{t \in \text{event}_n} Q(t)}{\sum_{t \in \text{event}_n} P(t)}, \quad (6)$$

where N_e is the number of events.

S2.5.3 Mean Half-Flow Date (MHFD).

The mean number of days since the start of the water year (1 October) on which cumulative discharge first exceeds half of the annual total.,

$$\text{MHFD} = \left\{ \tau_y : \sum_{t=t_0^y}^{\tau_y} Q(t) \geq \frac{1}{2} \sum_{t=t_0^y}^{t_0^y + T_y} Q(t) \right\}, \quad (7)$$

averaged over years y , with t_0^y the start of water year y .

S2.5.4 High-flow percentile (Q95).

The flow not exceeded 95% of the time (i.e. the 95th non-exceedance percentile),

$$Q_{95} = F_Q^{-1}(0.95), \quad (8)$$

where F_Q^{-1} is the quantile function of daily discharge.

S2.5.5 Low-flow percentile (Q5).

The flow not exceeded 5% of the time (the 5th non-exceedance percentile),

$$Q_5 = F_Q^{-1}(0.05). \quad (9)$$

S2.5.6 High-Flow Frequency (HF Freq).

The mean number of high-flow days per year, where a high-flow day exceeds nine times the median daily discharge,

$$\text{HFF} = \frac{1}{N_{\text{yr}}} |\{t : Q(t) > 9\tilde{Q}\}|. \quad (10)$$

S2.5.7 High-Flow Duration (HF Dur).

The mean duration (in days) of high-flow events, where an event is a consecutive run of days with $Q(t) > 9\tilde{Q}$,

$$\text{HFD} = \overline{D}_{\text{high}}. \quad (11)$$

S2.5.8 Low-Flow Frequency (LF Freq).

The mean number of low-flow days per year, where a low-flow day falls below 20% of the mean daily discharge (Olden and Poff, 2003),

$$\text{LFF} = \frac{1}{N_{\text{yr}}} |\{t : Q(t) < 0.2\overline{Q}\}|. \quad (12)$$

S2.5.9 Low-Flow Duration (LF Dur).

The mean duration (in days) of low-flow events, where an event is a consecutive run of days with $Q(t) < 0.2\overline{Q}$,

$$\text{LFD} = \overline{D}_{\text{low}}. \quad (13)$$

S2.5.10 Baseflow Index (BFI).

The ratio of baseflow volume to total streamflow volume, with baseflow $Q_b(t)$ obtained by the UKIH smoothed-minima method using a block length of 5 days (the TOSSH default),

$$\text{BFI} = \frac{\sum_t Q_b(t)}{\sum_t Q(t)}. \quad (14)$$

S2.5.11 Baseflow Recession Constant (BFRC).

The recession constant k of an exponential master recession curve $Q(t) = Q_0 e^{-kt}$, constructed from recession segments of at least 15 days using the adapted matching-strip method; recession onset is identified with a Lyne–Hollick filter ($\alpha = 0.925$),

$$\text{BFRC} = k, \quad Q(t) = Q_0 e^{-kt} \quad [\text{d}^{-1}]. \quad (15)$$

S2.5.12 Flow-Duration-Curve Slope (FDC Slope).

The slope of the flow-duration curve between the 33rd and 66th exceedance percentiles, fitted in logarithmic space,

$$\text{FDCslope} = \frac{\ln Q_{33} - \ln Q_{66}}{0.66 - 0.33}, \quad (16)$$

where Q_{33} and Q_{66} are the flows exceeded 33% and 66% of the time.

S2.5.13 Flashiness Index (FI).

The Richards–Baker flashiness index, the sum of absolute day-to-day changes in discharge normalised by total discharge,

$$\text{FI} = \frac{\sum_{t=2}^N |Q(t) - Q(t-1)|}{\sum_{t=1}^N Q(t)}. \quad (17)$$

S2.5.14 Variability Index (VI).

The standard deviation of the common logarithm of discharge evaluated at the nine deciles (10%, 20%, ..., 90%) of the flow-duration curve,

$$\text{VI} = \text{std}(\{\log_{10} Q_p : p = 10\%, 20\%, \dots, 90\%\}). \quad (18)$$

S2.5.15 Rising-Limb Density (RLD).

The ratio of the number of rising limbs to the total time the hydrograph is rising, where a rising limb is an increasing-flow segment of at least one day (Sawicz et al., 2011),

$$\text{RLD} = \frac{N_{\text{RL}}}{\sum_t \mathbb{1}[Q(t) > Q(t-1)]} \quad [\text{d}^{-1}], \quad (19)$$

with N_{RL} the number of rising-limb segments.

S2.6 Test Cases Error Metrics

The motivation behind the used error metric was to quantify the signature error in a manner that is able to capture both the deviation from the observed value and the performance relative to the alternative objective functions while doing so for a large numbers of models. Additionally, we would like to be the metric representative when compared across signatures.

To achieve this, we first calculate the difference between simulated and observed value for all models, catchments, signatures and objective functions. In a first step of aggregation, we calculate the median over the models (which exceeded the benchmark), denoted through subscript i . This is a robust way to find a representative values. Next, we normalize by absolute of the maximum value found for this signature across all objective functions and per catchment. This lead to the worst performance being characterized by ± 1 for each catchment. This can further be aggregated by calculating the median over all catchments (denoted through j) leading to a single error metric per signature and objective function.

This following section clarifies the applied metrics using a generic example. Each Matrix represents the signature values depending on the catchment (columns) and model (rows) for a specific objective function.

In practice, each model is not evaluated once but is represented by a small ensemble of retained parameter sets (runs). The example therefore starts one level lower: for every combination of model, catchment, and objective function there are several runs, each producing its own signature value. As a first step, these runs are collapsed to a single representative value per model by taking the median over runs (denoted through subscript r). To illustrate, consider model 1 calibrated with Objective Function 1, for which three runs were retained:

$$\text{Runs (Model 1, Objective Function 1): } \begin{bmatrix} 0.78 & 0.38 & 0.58 \\ 0.80 & 0.40 & 0.60 \\ 0.83 & 0.43 & 0.63 \end{bmatrix}$$

Taking the median over these runs (i.e. over the rows) yields the single value $[0.80 \quad 0.40 \quad 0.60]$, which is the entry shown in the first row of Matrix 1 below. The same run-level reduction is applied to every model cell, so that the three matrices below already represent the per-model median over runs. From here on, the aggregation proceeds over models (i) and then catchments (j) as described.

$$\text{Matrix 1 (Objective Function 1): } \begin{bmatrix} 0.80 & 0.40 & 0.60 \\ 0.75 & 0.38 & 0.62 \\ 0.70 & 0.35 & 0.58 \end{bmatrix}$$

$$\text{Matrix 2 (Objective Function 2): } \begin{bmatrix} 0.60 & 0.30 & 0.52 \\ 0.55 & 0.28 & 0.51 \\ 0.50 & 0.32 & 0.50 \end{bmatrix}$$

$$\text{Matrix 3 (Objective Function 3): } \begin{bmatrix} 0.70 & 0.36 & 0.44 \\ 0.66 & 0.34 & 0.45 \\ 0.61 & 0.35 & 0.43 \end{bmatrix}$$

Error Metric:

$$EM_k = med_j (EM_{jk}) = med_j \left(\frac{med_i (S_{ijk} - y_j)}{|max (med_i (S_{ijk} - y_j))|} \right) \quad (20)$$

Observed Values:

$$[0.75 \quad 0.37 \quad 0.61]$$

For the applied error metric, we calculate the error in signature representation for all of them:

$$\text{Error Matrix 1: } \begin{bmatrix} 0.05 & 0.03 & -0.01 \\ 0.00 & 0.01 & 0.01 \\ -0.05 & -0.02 & -0.03 \end{bmatrix}$$

$$\text{Error Matrix 2: } \begin{bmatrix} -0.15 & -0.07 & -0.09 \\ -0.20 & -0.09 & -0.10 \\ -0.25 & -0.05 & -0.11 \end{bmatrix}$$

$$\text{Error Matrix 3: } \begin{bmatrix} -0.05 & -0.01 & -0.17 \\ -0.09 & -0.03 & -0.16 \\ -0.14 & -0.02 & -0.18 \end{bmatrix}$$

Next, we can calculate the median over the all the modelled values:

$$\text{Median Error Matrix 1: } [0.00 \quad 0.01 \quad -0.01]$$

$$\text{Median Error Matrix 2: } [-0.20 \quad -0.07 \quad -0.10]$$

$$\text{Median Error Matrix 3: } [-0.09 \quad -0.02 \quad -0.17]$$

To normalize, mainly to make the results comparable for multiple signatures with different ranges, we divide the error by the absolute of the maximum median error for each catchment. That means the range will be limited from -1 to +1 with an value of ± 1 indicating that the largest error was found for this objective function in this catchment. A value of 0 indicates that the median signature over all models equals the observed value.

$$\text{Absolute Maximum Median Error } [0.20 \quad 0.07 \quad 0.17]$$

Based on this, we can calculate EM_{jk} for each catchment:

$$EM_{j1}: [0.00 \quad 0.14 \quad -0.06]$$

$$EM_{j2}: [-1.00 \quad -1.00 \quad -0.59]$$

$$EM_{j3}: [-0.45 \quad -0.29 \quad -1.00]$$

For this test case, OF2 is worst metric regarding signature representation for catchments 1 and 2 while OF3 has worst representation in catchment 3. OF1 is the best metric for all of the catchments.

Finally, we can derive EM_k , the median normalized error for each objective function for this signature.

$$EM_1 = [0.00]$$

$$EM_2 = [-1.00]$$

$$EM_3 = [-0.45]$$

This median is only useful for larger amounts of catchments, but the applied sample size (10) should be sufficient. Alternatively, one can use the mean as the metric is strictly contained in range.

S2.7 Violinplots for Signature Error

We can plot the derived errors for selected signatures to evaluate the impact of the objective function on the signature representation.

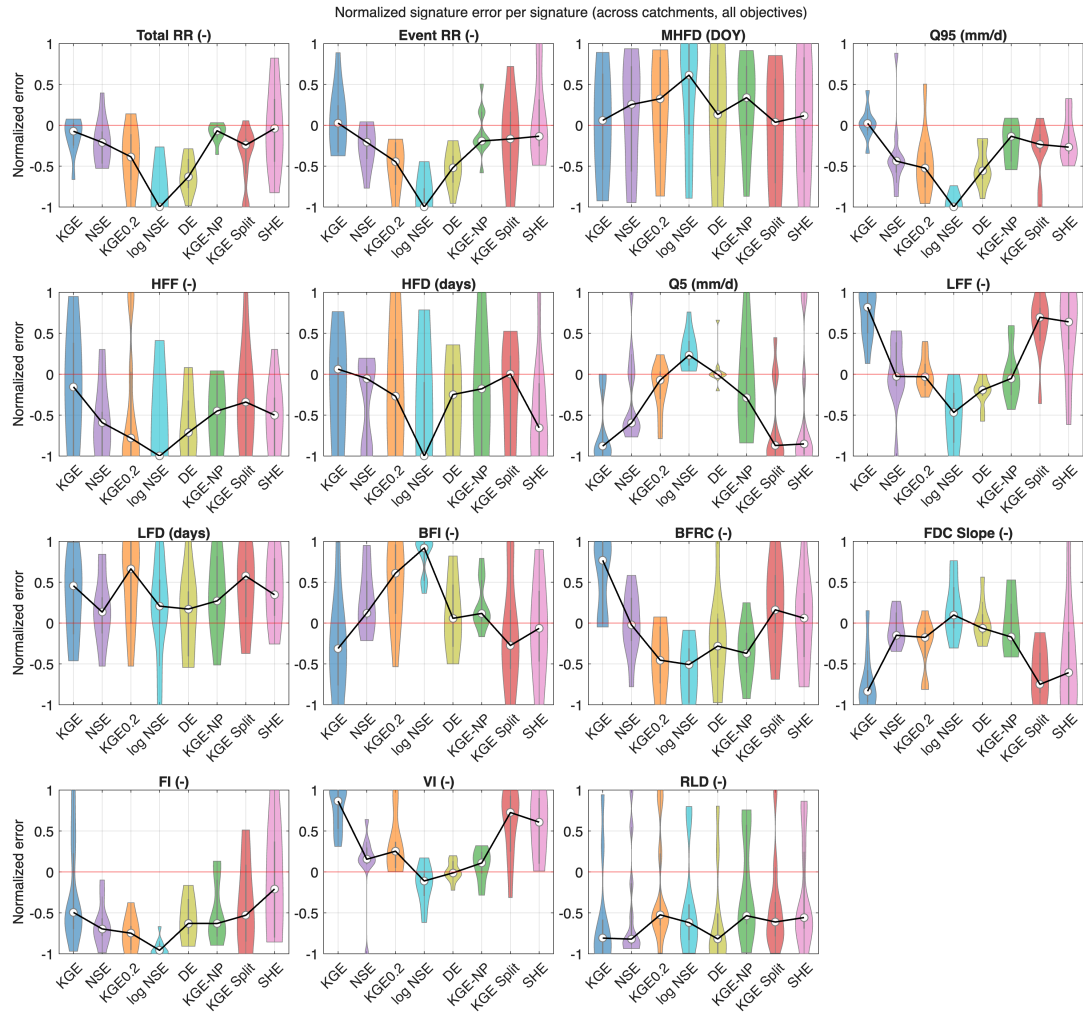


Figure S15. Variance-normalized Error over all Catchments and Models for 9 Signatures. The violinplots show the distribution of the median error over all models and parameter sets for each objective function and signature. This distribution represents the distribution for the objective function with each dot representing a catchment.

S2.8 Table with Normalized Error Values

Table S2. Medians and standard deviations of normalized signature errors for each objective function across 15 hydrological signatures.

Objective	Total RR (-)	Event RR (-)	MHFD (DOY)	FDC Slope (-)	Q5 (mm/d)	LF Dur (days)	LF Freq (-)	BFI (-)	Q95 (mm/d)	HF Dur (days)	HF Freq (-)	BFRC (-)	Rising Limb Density (-)	Flashiness Index (-)	Variability Index (-)
KGE median	-0.073	0.026	0.060	-0.835	-0.878	0.457	0.819	-0.311	0.023	0.061	-0.155	0.768	-0.808	-0.494	0.866
KGE sd	0.224	0.400	0.671	0.381	0.427	0.505	0.307	0.616	0.201	0.662	0.697	0.467	0.661	0.641	0.269
NSE median	-0.209	-0.209	0.254	-0.150	-0.596	0.134	-0.027	0.122	-0.438	-0.053	-0.592	-0.023	-0.820	-0.696	0.155
NSE sd	0.282	0.259	0.685	0.204	0.568	0.422	0.441	0.408	0.477	0.535	0.465	0.404	0.688	0.268	0.426
KGE0.2 median	-0.386	-0.445	0.325	-0.176	-0.073	0.663	-0.031	0.611	-0.525	-0.269	-0.781	-0.454	-0.524	-0.748	0.253
KGE0.2 sd	0.362	0.282	0.671	0.325	0.307	0.535	0.210	0.514	0.412	0.801	0.793	0.388	0.685	0.207	0.354
log NSE median	-1.000	-1.000	0.613	0.099	0.232	0.207	-0.467	0.918	-1.000	-1.000	-1.000	-0.508	-0.619	-0.957	-0.110
log NSE sd	0.259	0.187	0.710	0.358	0.228	0.600	0.357	0.244	0.082	0.697	0.513	0.361	0.636	0.105	0.242
DE median	-0.631	-0.518	0.133	-0.064	-0.005	0.171	-0.193	0.058	-0.557	-0.249	-0.713	-0.282	-0.819	-0.628	-0.011
DE sd	0.214	0.238	0.787	0.261	0.227	0.503	0.174	0.441	0.245	0.548	0.417	0.589	0.622	0.256	0.124
KGE-NP median	-0.067	-0.193	0.338	-0.171	-0.288	0.271	-0.053	0.118	-0.134	-0.179	-0.450	-0.371	-0.534	-0.630	0.109
KGE-NP sd	0.115	0.298	0.582	0.353	0.598	0.491	0.324	0.308	0.238	0.714	0.424	0.357	0.638	0.391	0.197
KGE Split median	-0.243	-0.167	0.036	-0.750	-0.871	0.575	0.695	-0.274	-0.235	0.003	-0.341	0.162	-0.611	-0.529	0.726
KGE Split sd	0.364	0.525	0.708	0.320	0.536	0.480	0.385	0.695	0.339	0.599	0.631	0.551	0.584	0.531	0.433
SHE median	-0.039	-0.134	0.116	-0.607	-0.852	0.346	0.639	-0.065	-0.268	-0.655	-0.500	0.059	-0.557	-0.209	0.608
SHE sd	0.525	0.543	0.773	0.681	0.797	0.445	0.556	0.620	0.279	0.635	0.432	0.572	0.665	0.683	0.388

S2.9 Additional Plots for Detailed Analysis

Event RR:

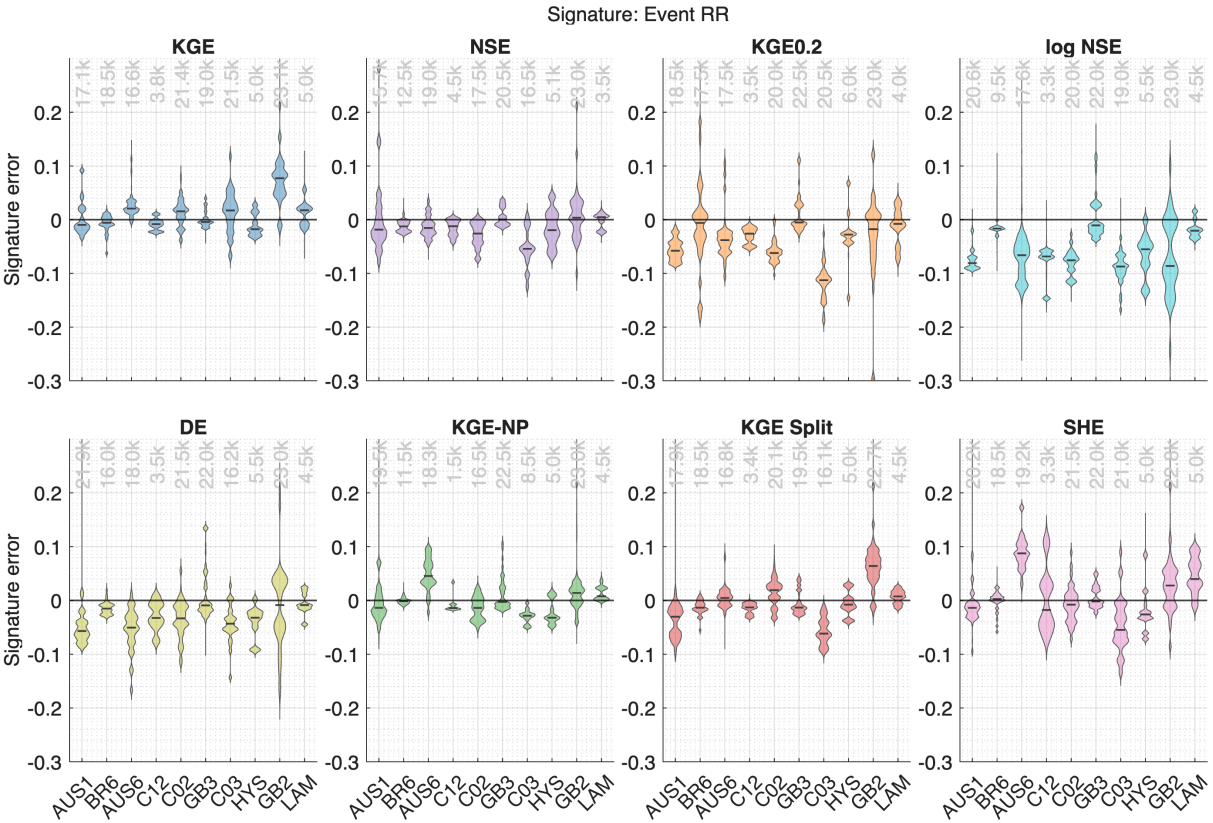


Figure S16. ERR for all OFs

Mean Half Flow Date:

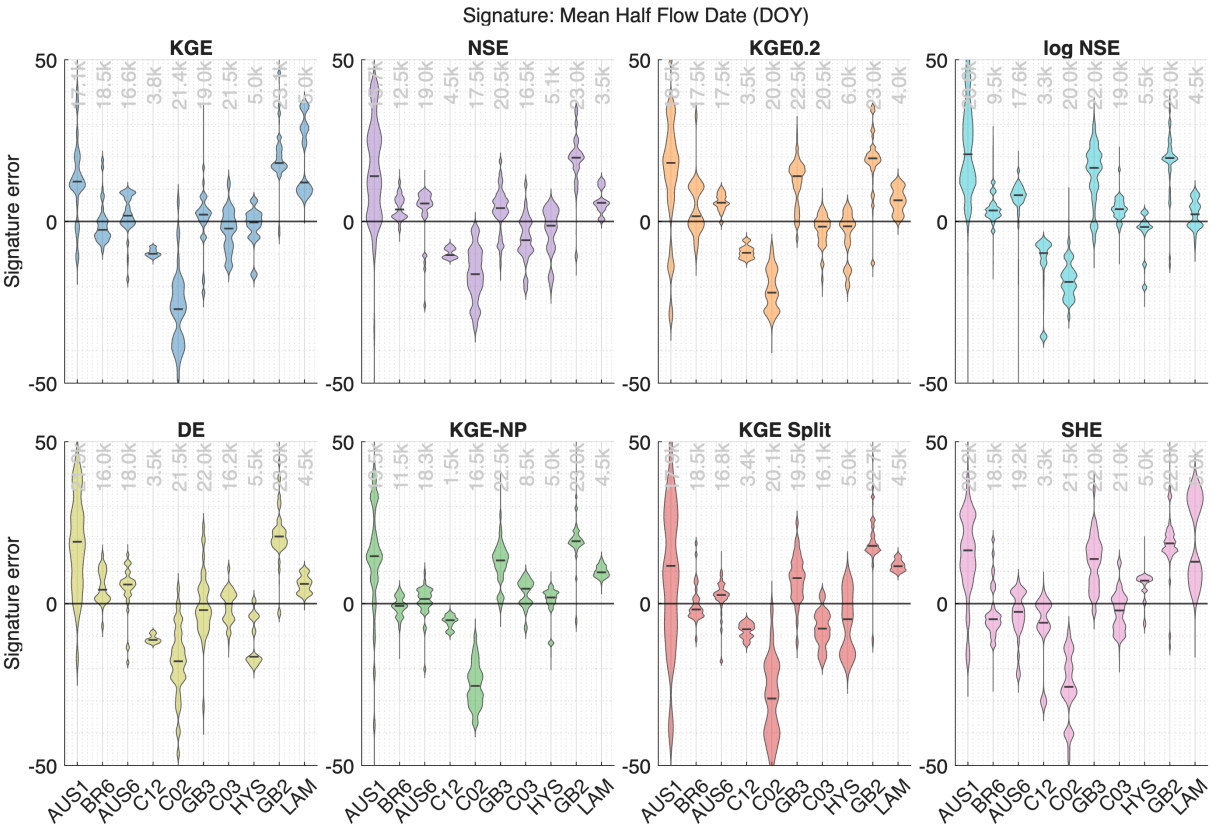


Figure S17. MHFD for all OFs

High Flow Percentile:

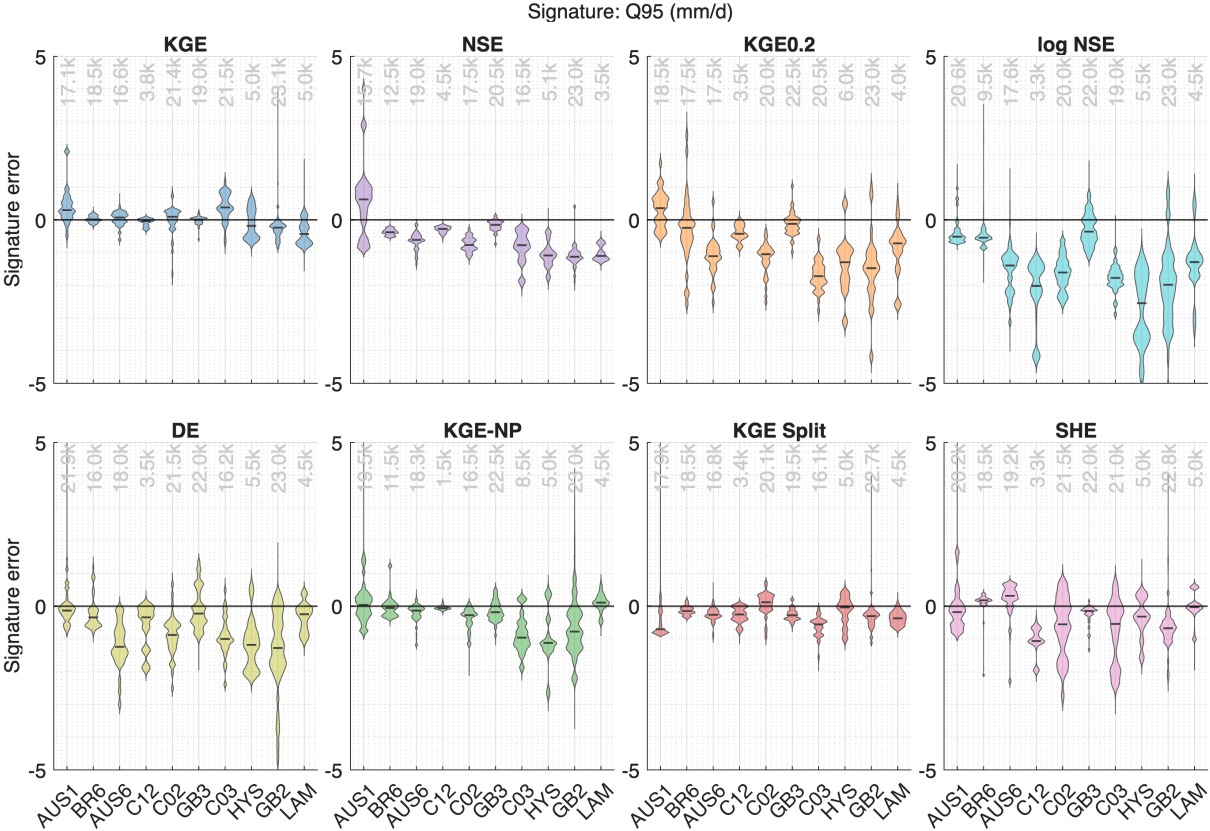


Figure S18. High Flow Percentile for all OFs

High Flow Frequency:

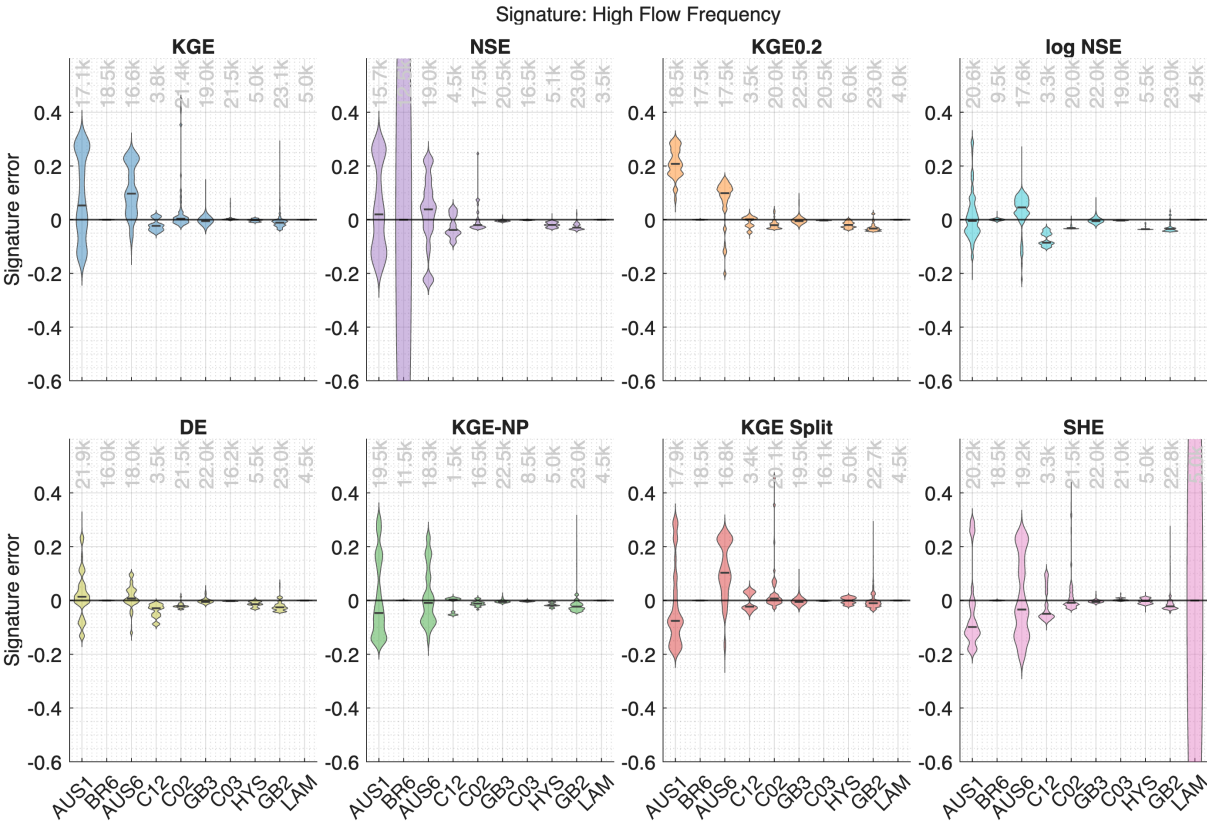


Figure S19. HFF for all OFs

High Flow Duration:

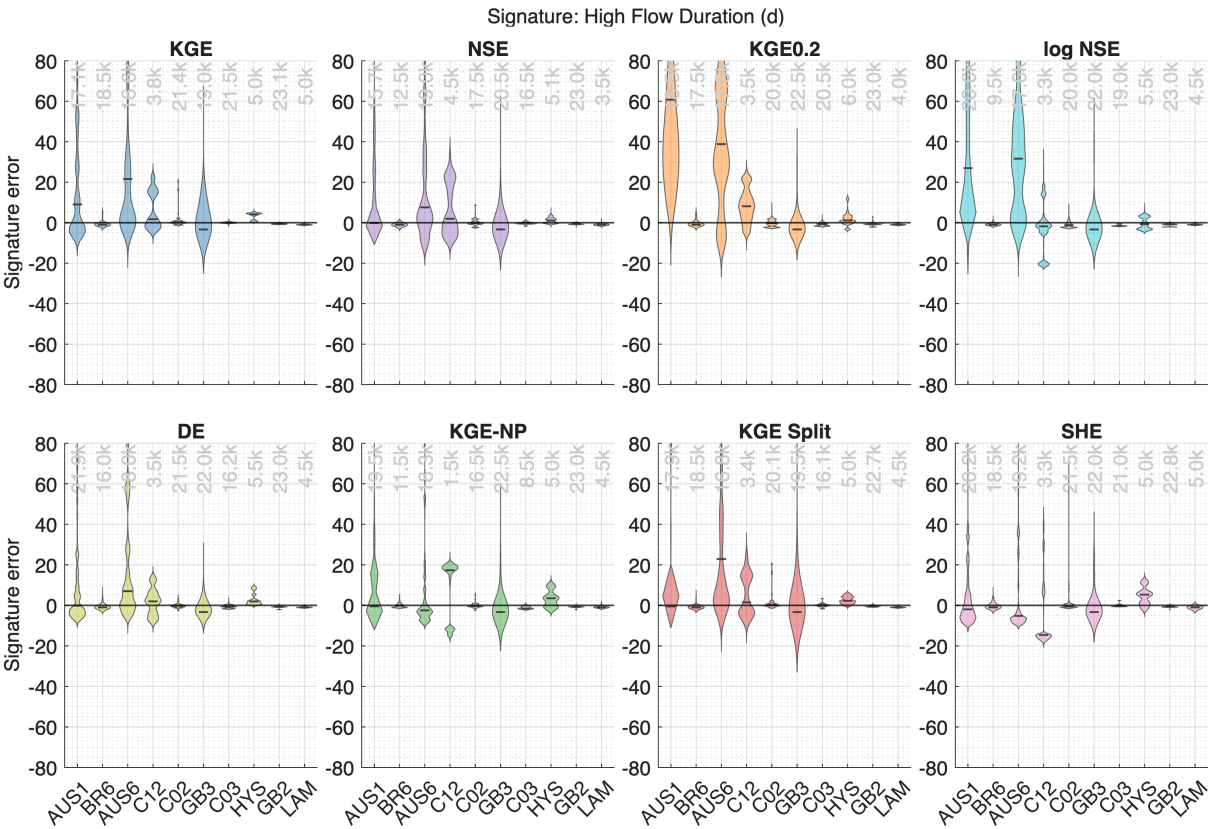


Figure S20. HFD for all OFs

Low Flow Percentile:

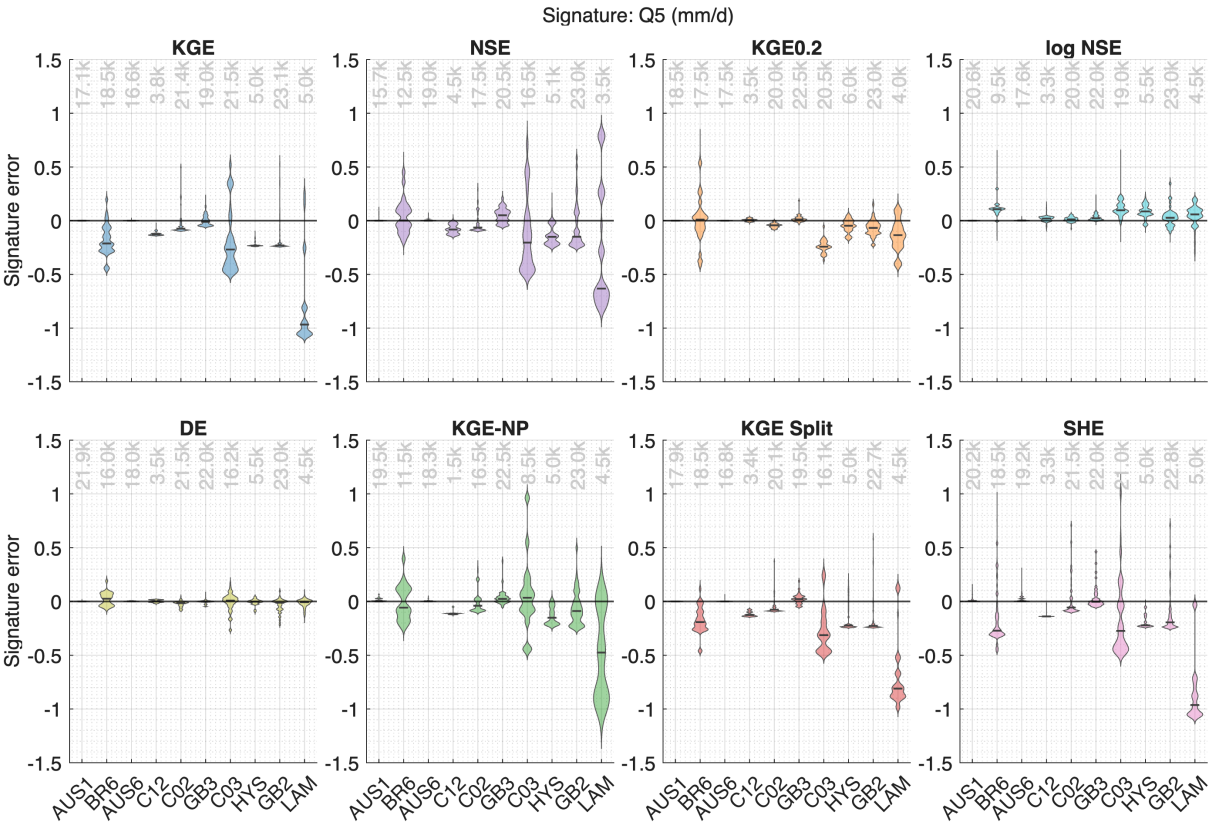


Figure S21. Details on Low Flow Percentile

Low Flow Frequency:

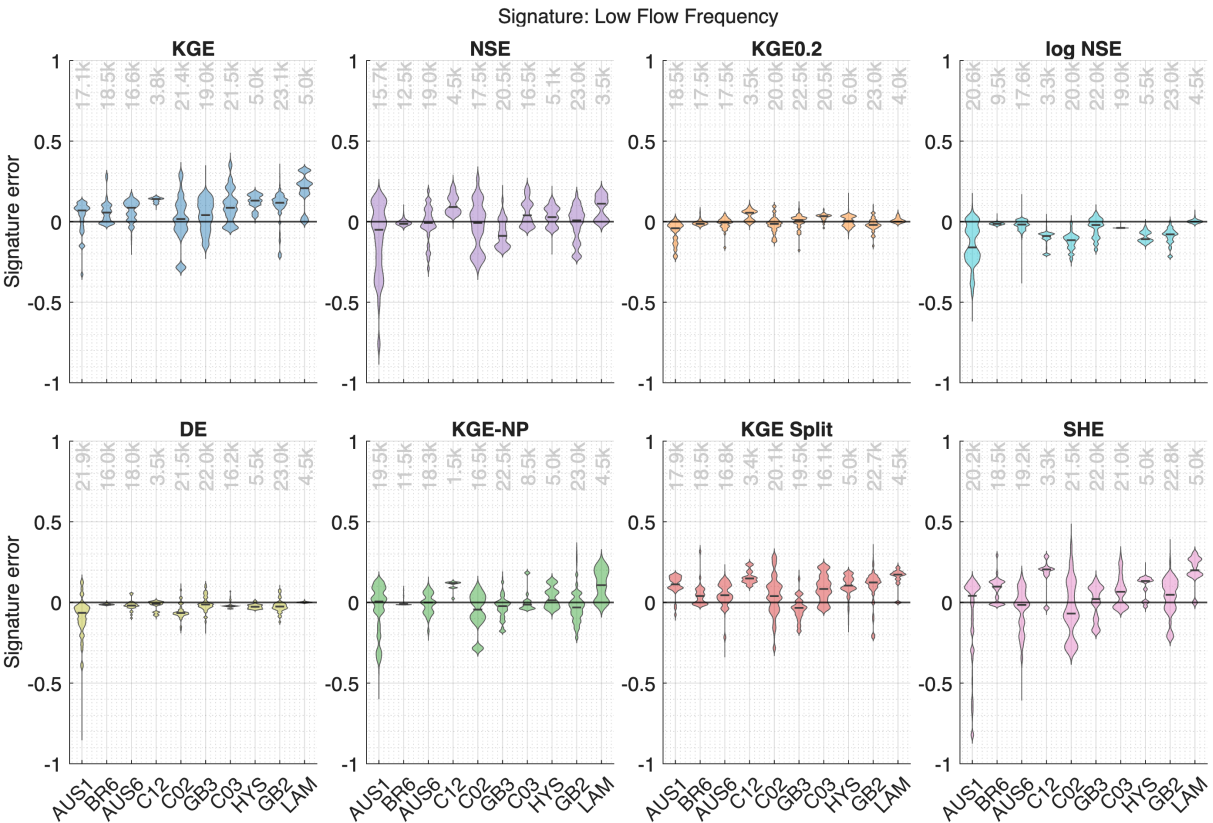


Figure S22. LFF for all OFs

Low Flow Duration:

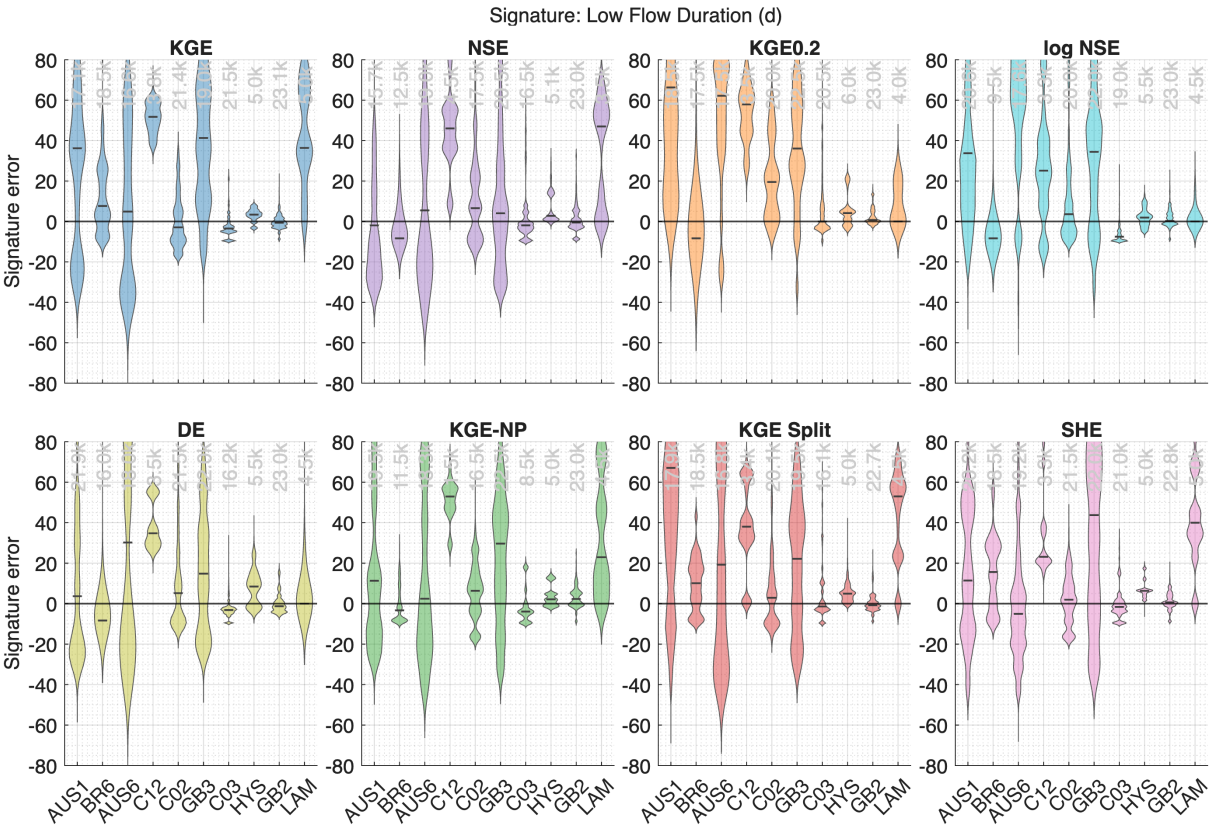


Figure S23. LFD for all OFs

Baseflow Index:

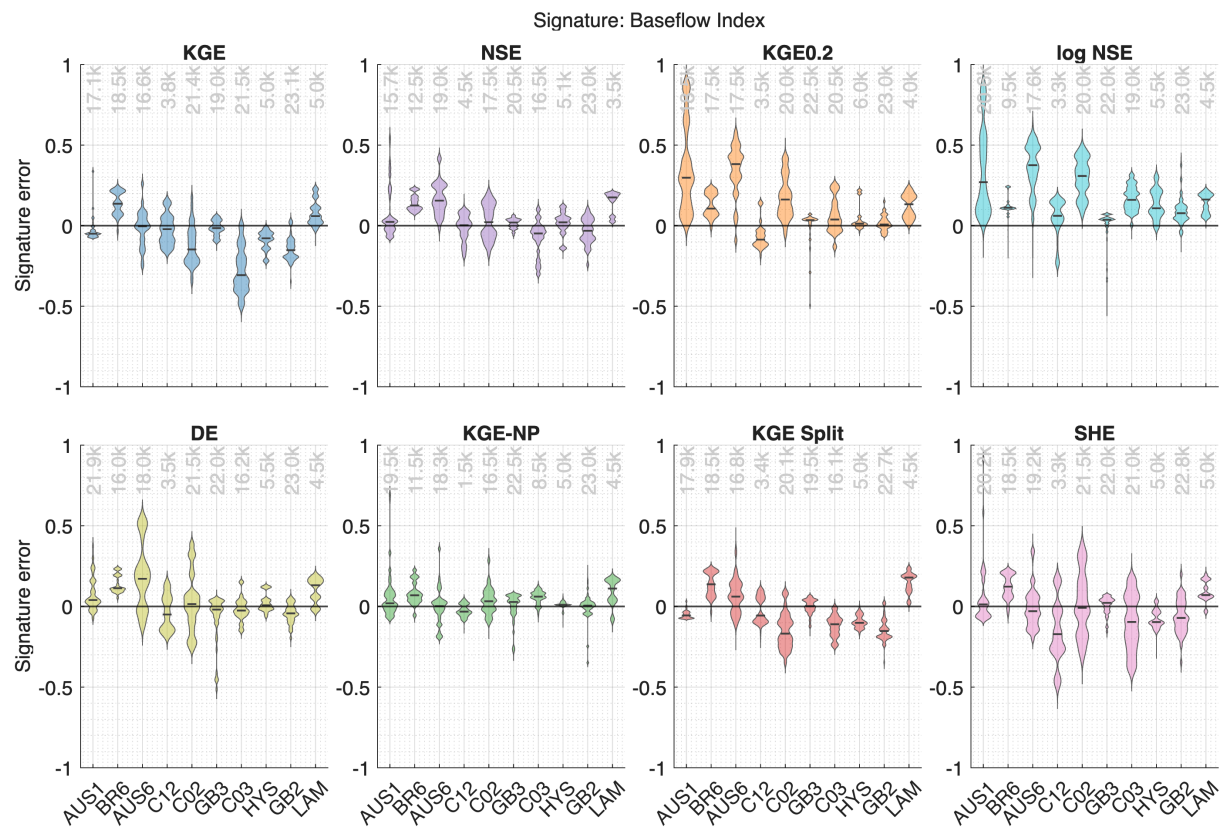


Figure S24. Baseflow Index for all OFs

Baseflow Recession Coefficient:

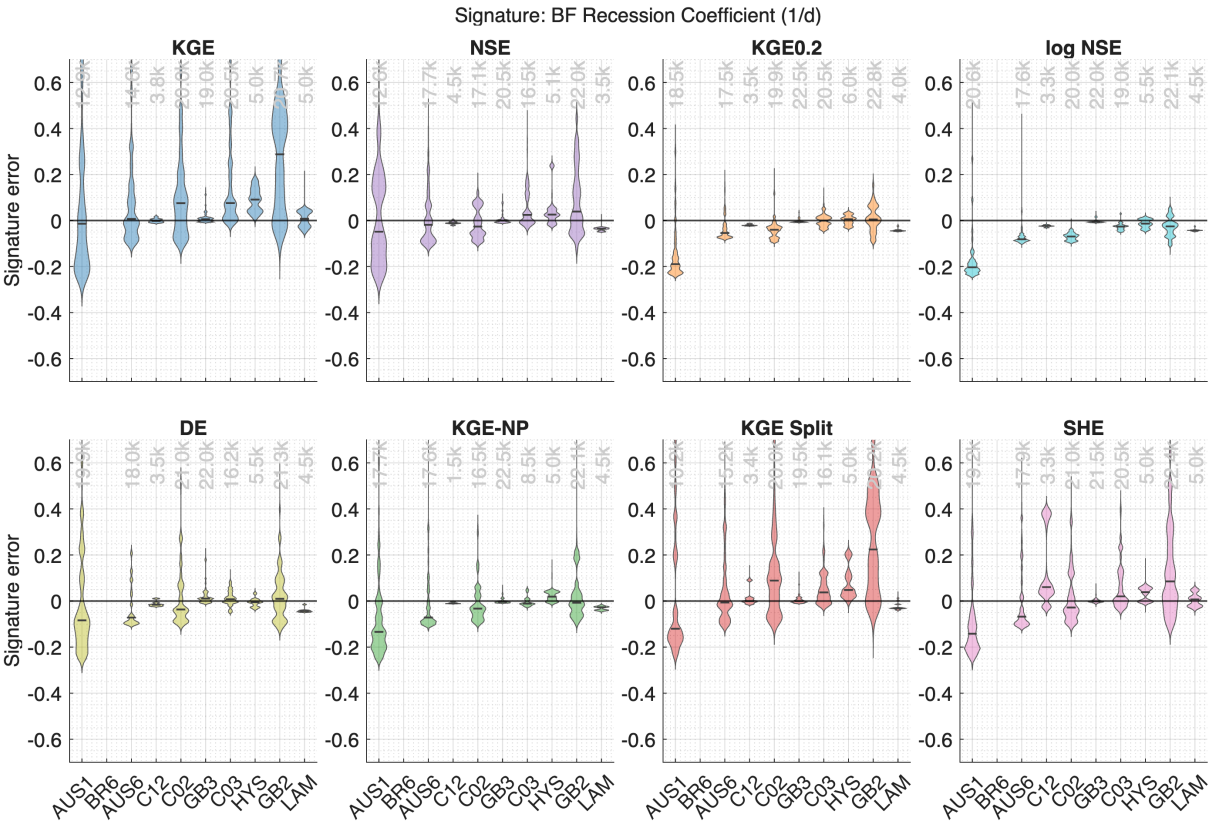


Figure S25. BRFC for all OFs

FDC Slope:

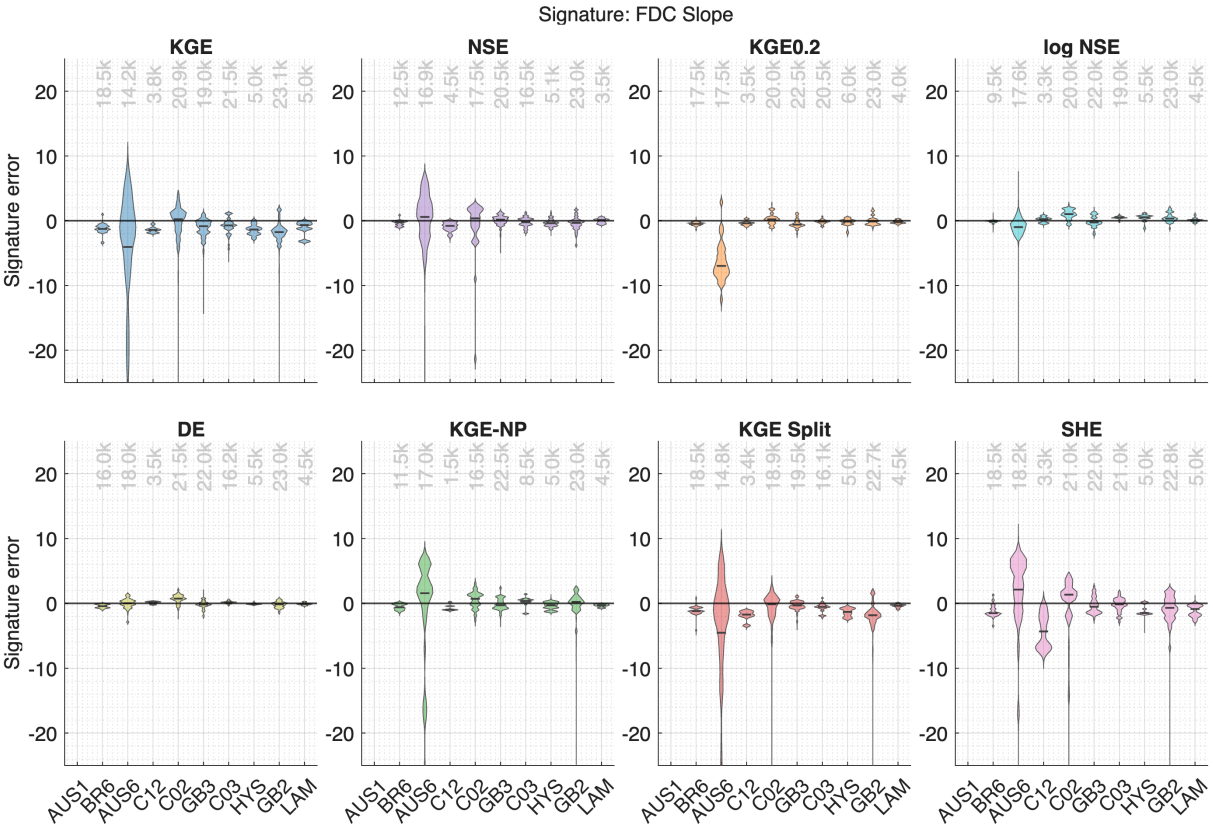


Figure S26. FDC Slope for all OFs

Flashiness Index:

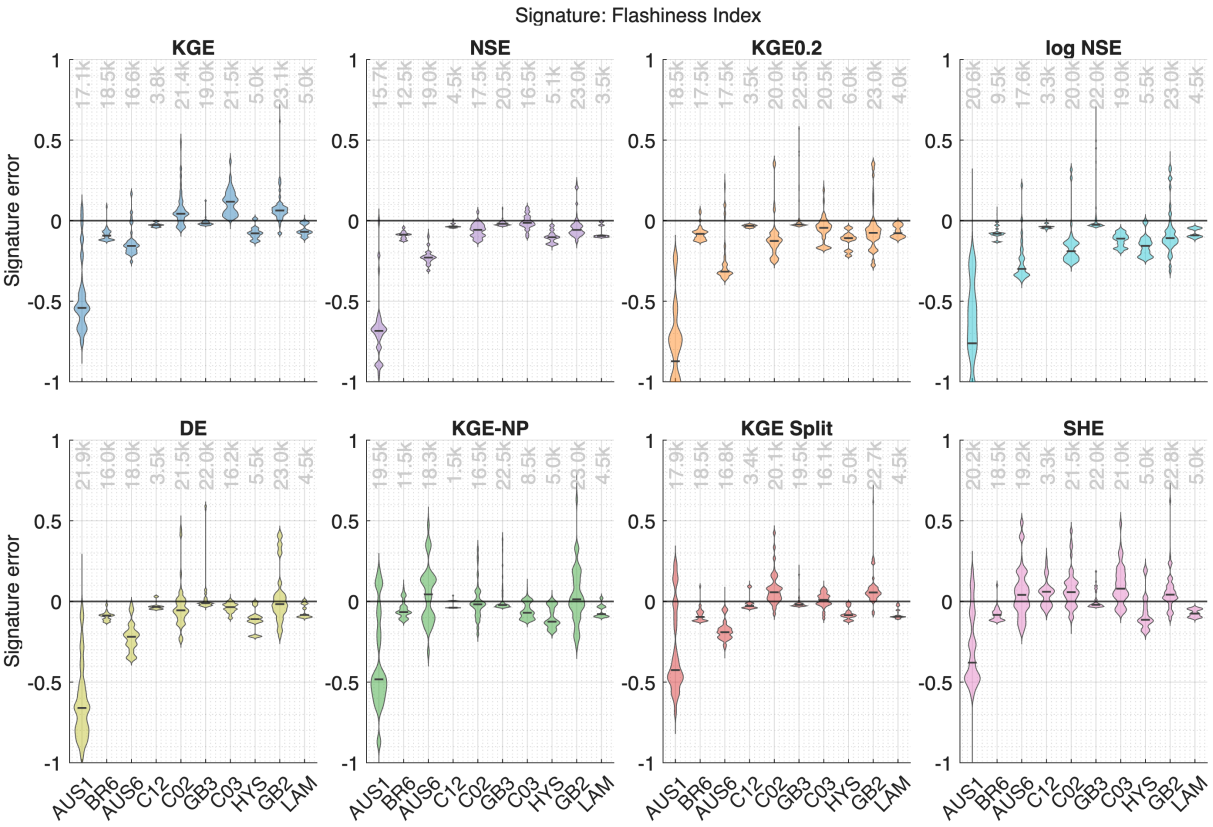


Figure S27. FI for all OFs

Variability Index:

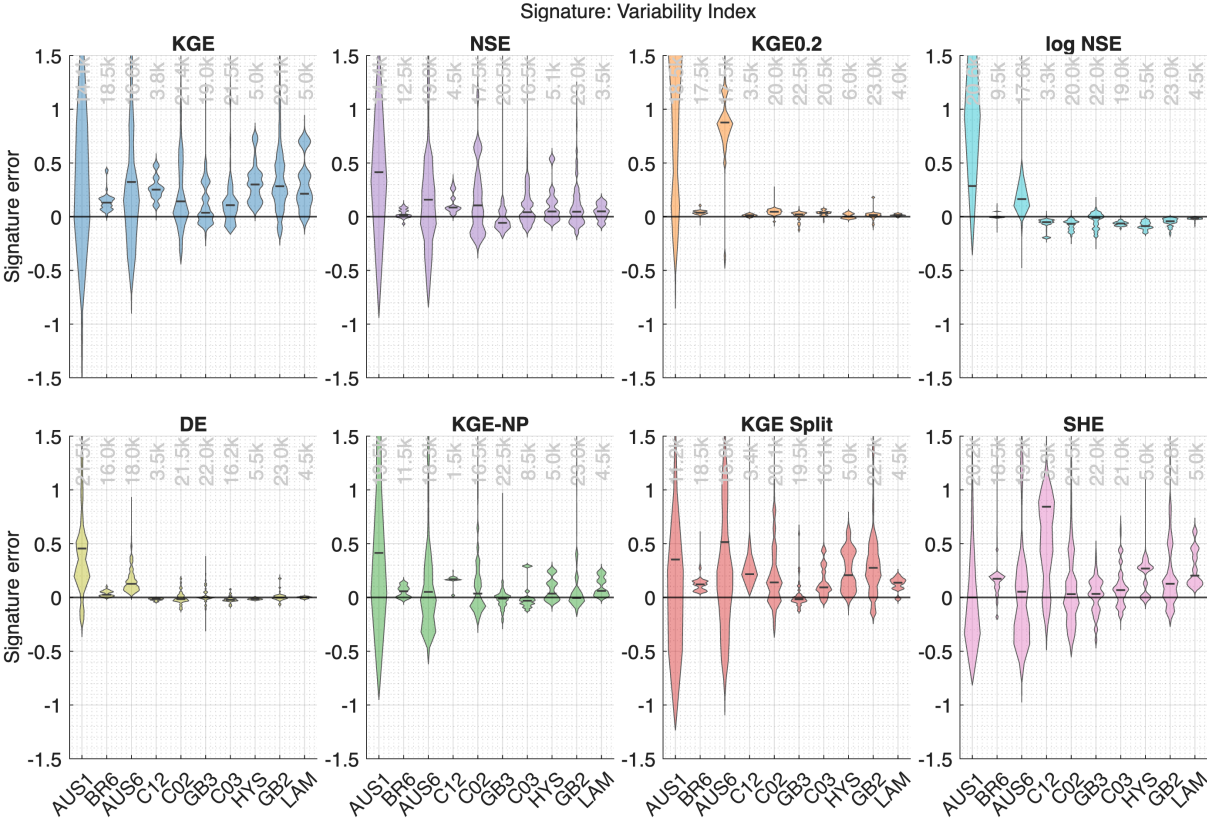


Figure S28. Variability Index for all OFs

Rising Limb Density:

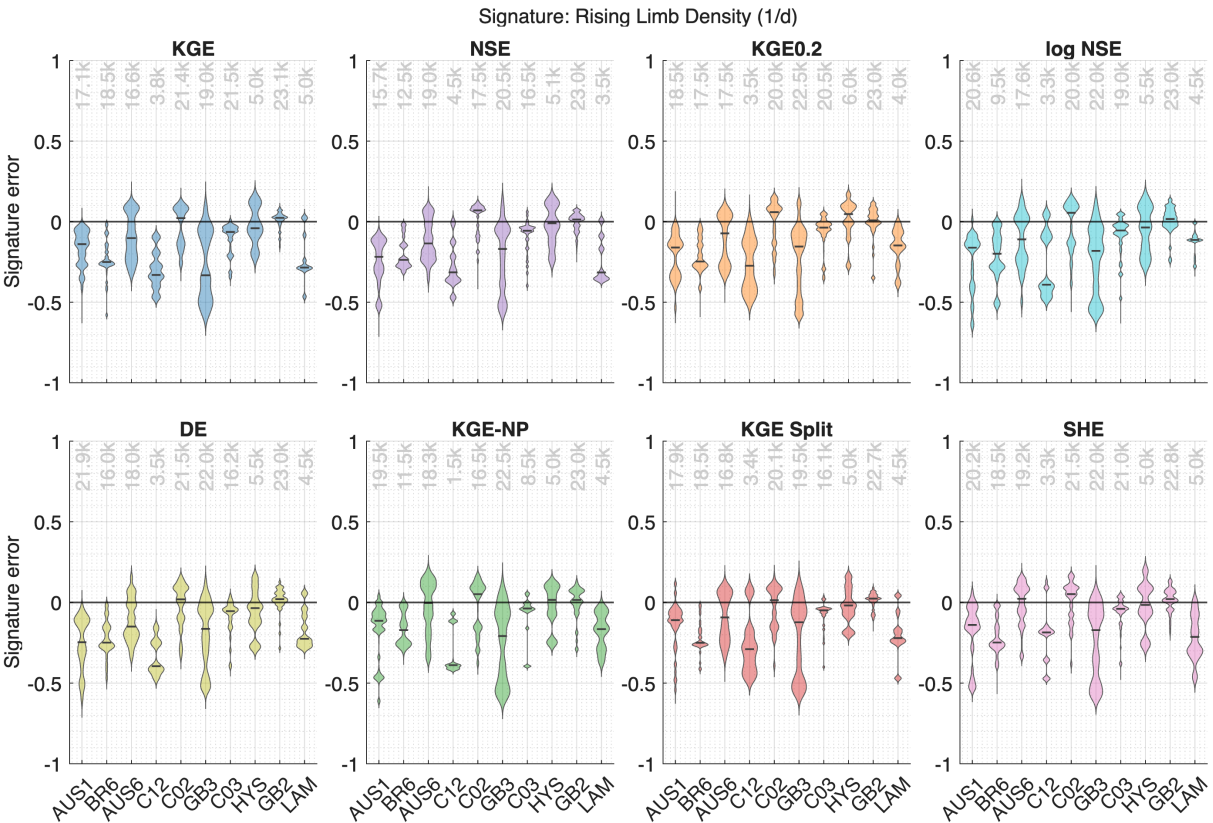


Figure S29. RLD for all OFs

S2.10 Additional Plots for Smaller and Larger Number of Runs

Considering 25 runs:

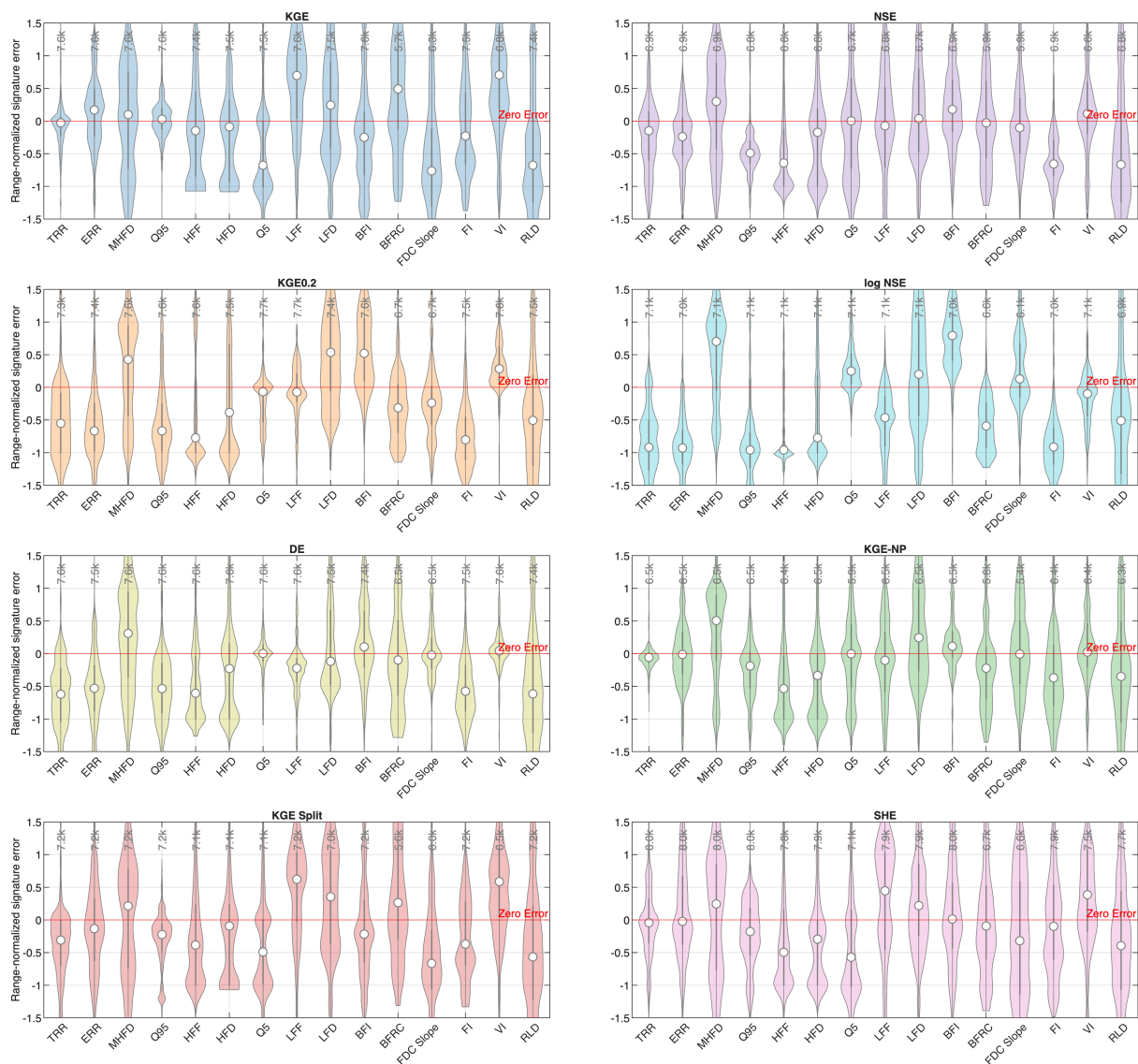


Figure S30. Violin Errors with 25 runs considered

Considering 100 runs:

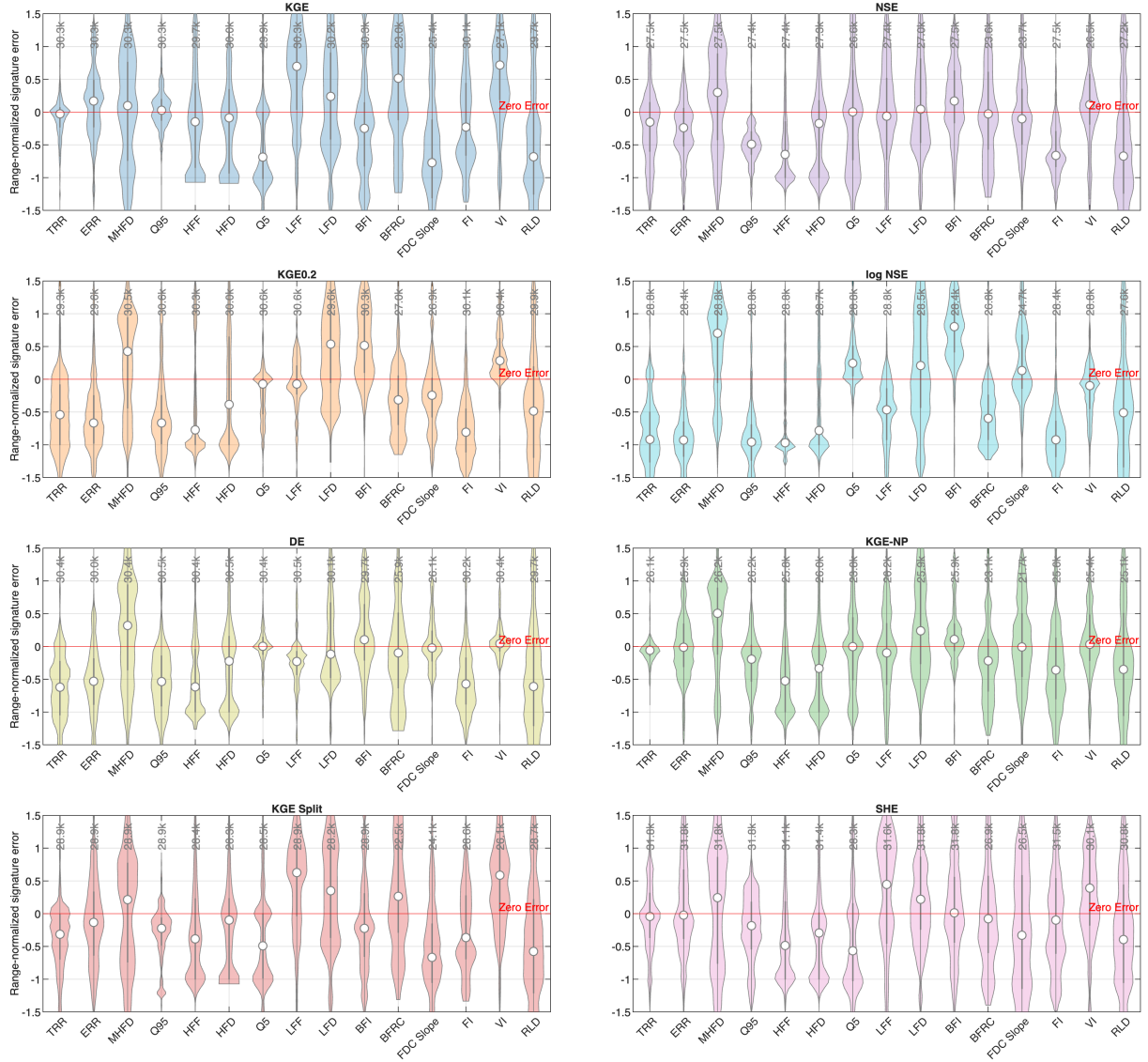


Figure S31. Violin Errors with 100 runs considered

To assess the role of parameter uncertainty, we retained the highest-performing parameter sets across all five seeds rather than a single optimum. The resulting spread in signature values was generally narrow relative to the spread induced by model structure and catchment, indicating that our conclusions are robust to parameter uncertainty within the retained ensemble. In Figure S32, we are collapsing into a median across the top 500 runs and 47 models (sequentially), so only 1 value retained per catchment, following the approach for the overall error metric. This highlights differences between the objective functions, but disregards spread for models and runs. Overall, strengths and weakness of specific OFs align with previous plots.

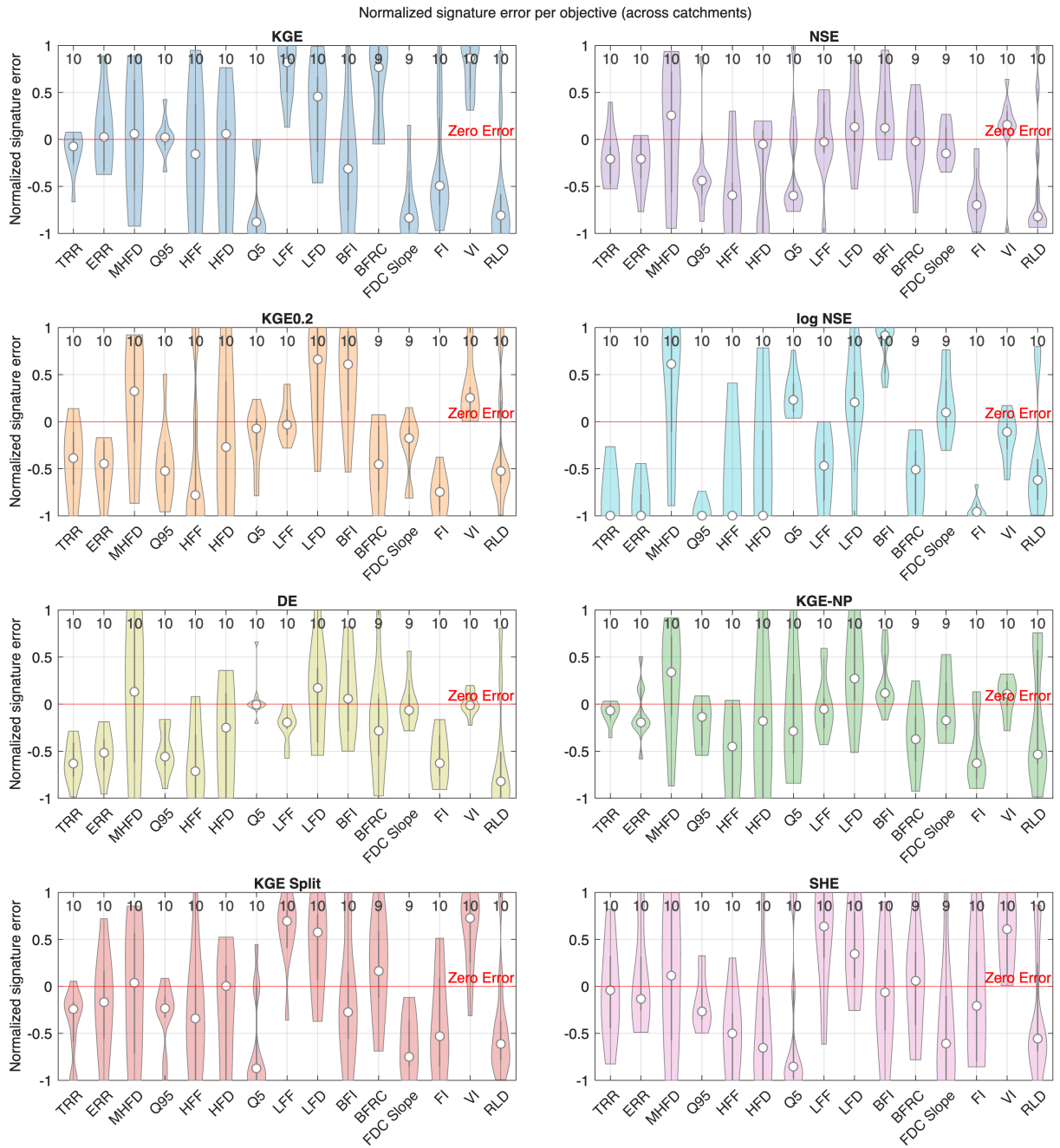


Figure S32. Violin Errors with median across runs

S2.11 Table with various p-values for statistical test

Table S3. Median p -values and fraction of tests with $p < 0.10$ and $p < 0.05$ across all objective function pairs for each hydrological signature.

Signature	Median p	$\#(p < 0.10)$ [%]	$\#(p < 0.05)$ [%]
Total RR (-)	0.053	67.9	42.9
Event RR (-)	0.017	75.0	64.3
Q95 (mm/d)	0.057	67.9	42.9
Q5 (mm/d)	0.042	78.6	60.7
BFI (-)	0.022	67.9	67.9
BFRC (-)	0.049	64.3	53.6
Flashiness Index (-)	0.057	71.4	42.9
LF Freq (-)	0.005	78.6	78.6
HF Freq (-)	0.258	14.3	7.1
FDC Slope (-)	0.141	32.1	21.4
Variability Index (-)	0.163	39.3	32.1
LF Dur (days)	0.383	3.6	3.6
HF Dur (days)	0.245	3.6	0.0
MHFD (DOY)	0.375	14.3	7.1
Rising Limb Density (-)	0.226	35.7	25.0

S2.12 Comparison of Calibration and Evaluation Results

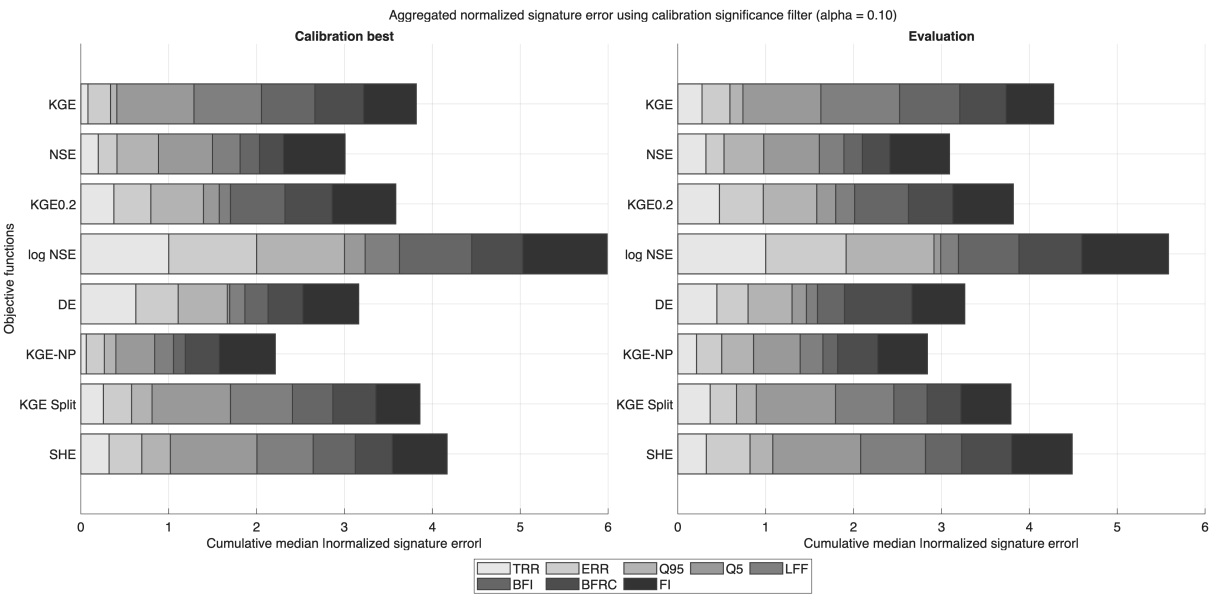


Figure S33. Comparison of Calibration and Evaluation Aggregated Error Metric

S2.13 Random Forest for all Signatures

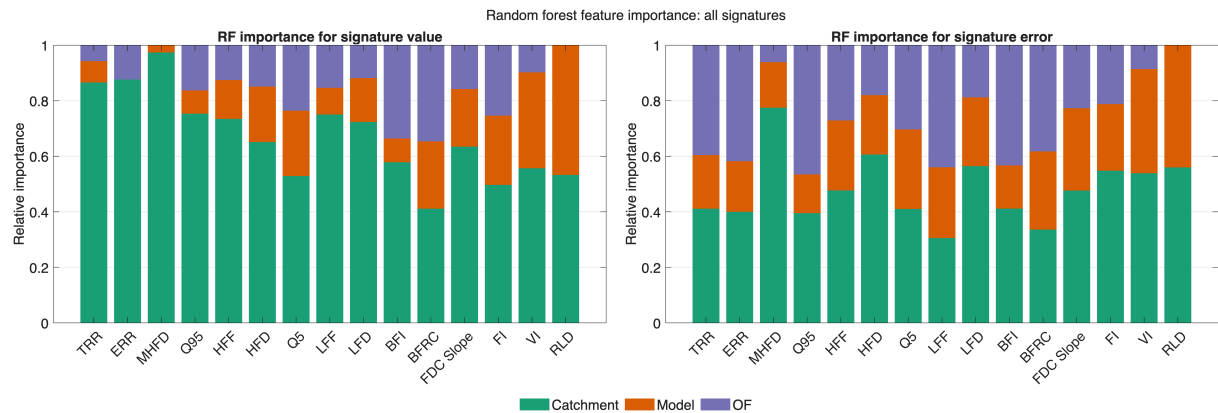


Figure S34. Random Forest including all 15 Signatures

S2.14 Correlation for Models

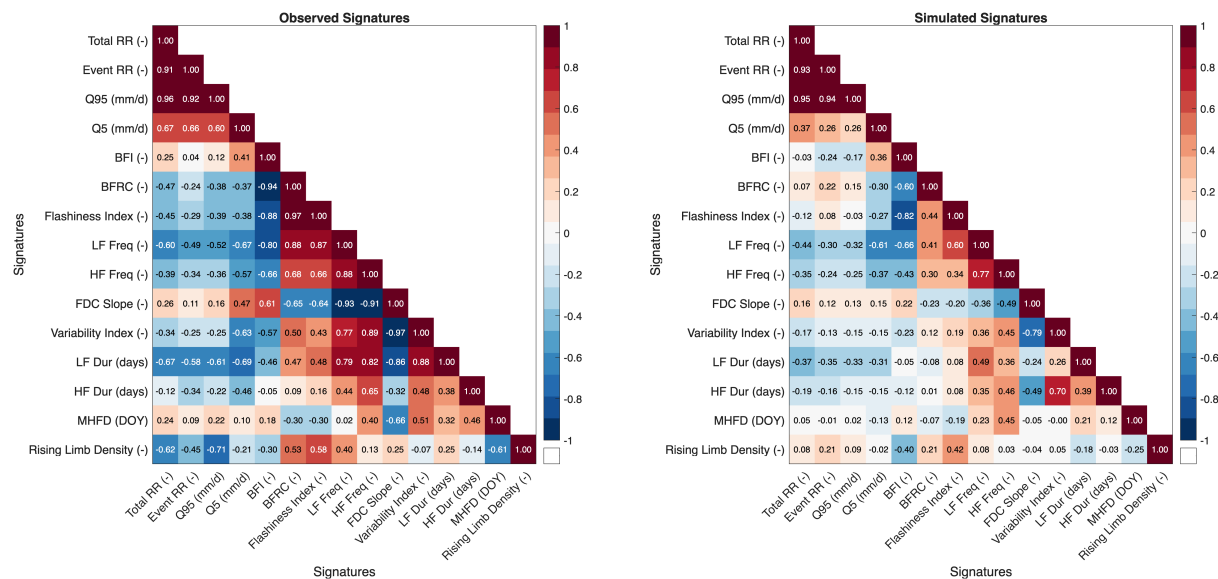


Figure S35. Signature Correlation Model-based

References

- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, *Hydrology and Earth System Sciences*, 21, 5293–5313, <https://doi.org/10.5194/hess-21-5293-2017>, 2017.
- Arsenault, R., Brisette, F., Martel, J.-L., Troin, M., Lévesque, G., Davidson-Chaput, J., Gonzalez, M. C., Ameli, A., and Poulin, A.: A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds, *Scientific Data*, 7, 243, <https://doi.org/10.1038/s41597-020-00583-2>, 2020.
- Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., and Siqueira, V. A.: CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil, *Earth System Science Data*, 12, 2075–2096, <https://doi.org/10.5194/essd-12-2075-2020>, 2020.
- Clerc-Schwarzenbach, F., Selleri, G., Neri, M., Toth, E., Van Meerveld, I., and Seibert, J.: Large-sample hydrology – a few camels or a whole caravan?, *Hydrology and Earth System Sciences*, 28, 4219–4237, <https://doi.org/10.5194/hess-28-4219-2024>, 2024.
- Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, *Earth System Science Data*, 12, 2459–2483, <https://doi.org/10.5194/essd-12-2459-2020>, 2020.
- Fowler, K. J. A., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: hydrometeorological time series and landscape attributes for 222 catchments in Australia, *Earth System Science Data*, 13, 3847–3867, <https://doi.org/10.5194/essd-13-3847-2021>, 2021.
- Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J., and McMillan, H. K.: TOSSH: A Toolbox for Streamflow Signatures in Hydrology, *Environmental Modelling and Software*, 138, 104983, <https://doi.org/10.1016/j.envsoft.2021.104983>, 2021.
- Klingler, C., Schulz, K., and Hernegger, M.: LamaH-CE: LARge-SaMple DATA for Hydrology and Environmental Sciences for Central Europe, *Earth System Science Data*, 13, 4529–4565, <https://doi.org/10.5194/essd-13-4529-2021>, 2021.
- Knoben, W. J. M., Woods, R. A., and Freer, J. E.: A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data, *Water Resources Research*, 54, 5088–5109, <https://doi.org/10.1029/2018WR022913>, 2018.
- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, *Hydrol. Earth Syst. Sci.*, 19, 209–223, <https://doi.org/10.5194/hess-19-209-2015>, 2015.
- Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, *River Research and Applications*, 19, 101–121, <https://doi.org/10.1002/rra.700>, 2003.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrology and Earth System Sciences*, 15, 2895–2911, <https://doi.org/10.5194/hess-15-2895-2011>, 2011.
- Willmott, C. J. and Feddema, J. J.: A More Rational Climatic Moisture Index*, *The Professional Geographer*, 44, 84–88, <https://doi.org/10.1111/j.0033-0124.1992.00084.x>, 1992.