



Supplement of

Technical note: Benchmarking large-domain model performance under sampling uncertainty

Gaby J. Gründemann et al.

Correspondence to: Wouter J. M. Knoben (wouter.knoben@ucalgary.ca)

The copyright of individual parts of the supplement might differ from the article licence.

Contents of this file

1. Figures S1 to S9

- (a) Figures S1-S3: a set of figures comparable to those in the main manuscript but with the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) as model performance metric.
- 5 (b) Figure S4: map showing which benchmark category results in the best KGE validation scores in each basin.
- (c) Figures S5-S7: similar maps as shown in the main paper for each of the three benchmark groups.
- (d) Figures S8, S9: a separation of the benchmark KGE scores into its three bias, variability and correlation components, for calibration and regionalization stations.

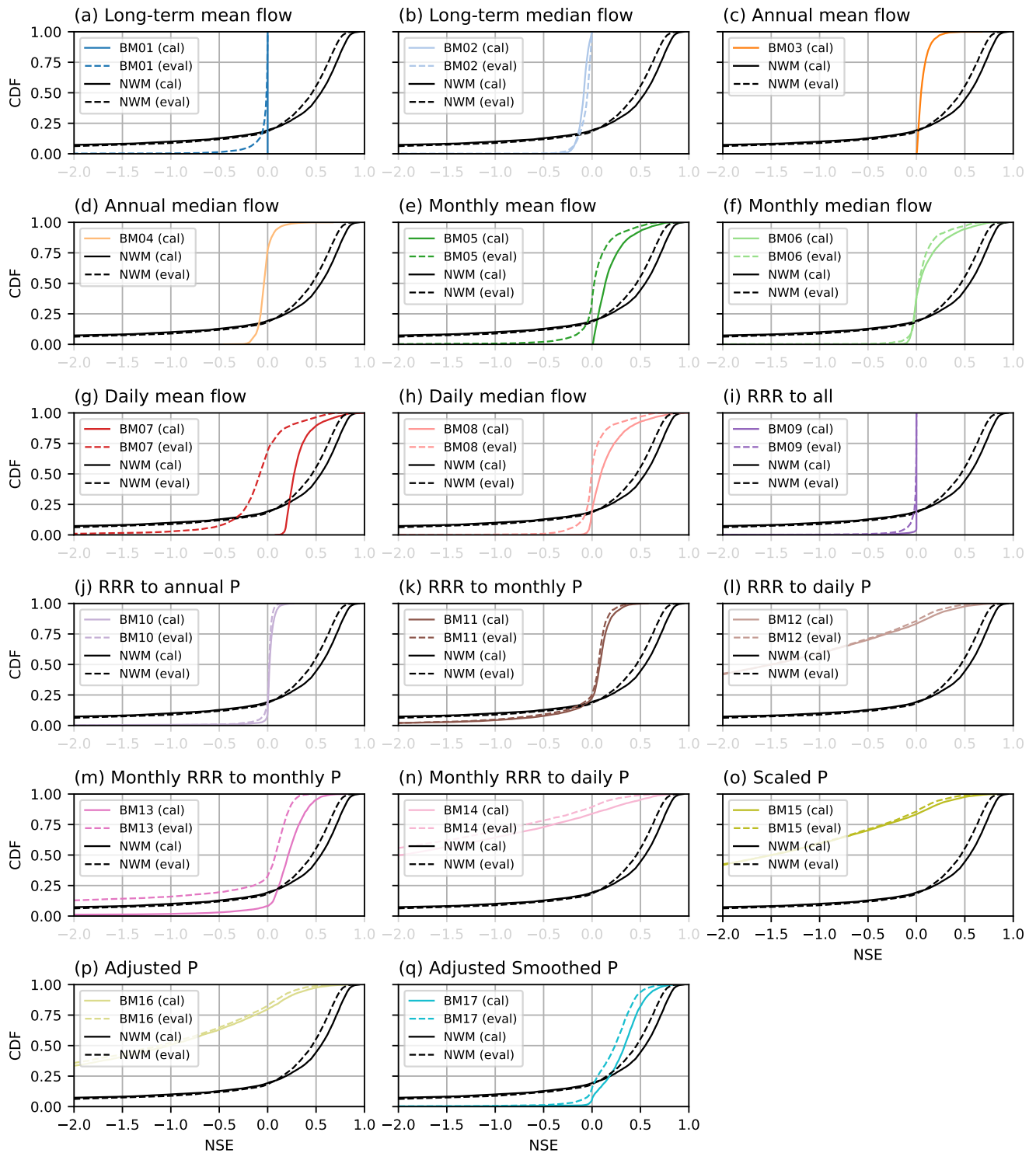


Figure S1. Cumulative Distribution Function (CDF) plot of the Nash–Sutcliffe Efficiency (NSE) scores for the National Water Model v3.0 and 17 simple benchmarks. For benchmarks 11, 12, 13 and 14, *RRR* stands for Rainfall Runoff Ratio. *P* (benchmarks 11–16) stands for precipitation.

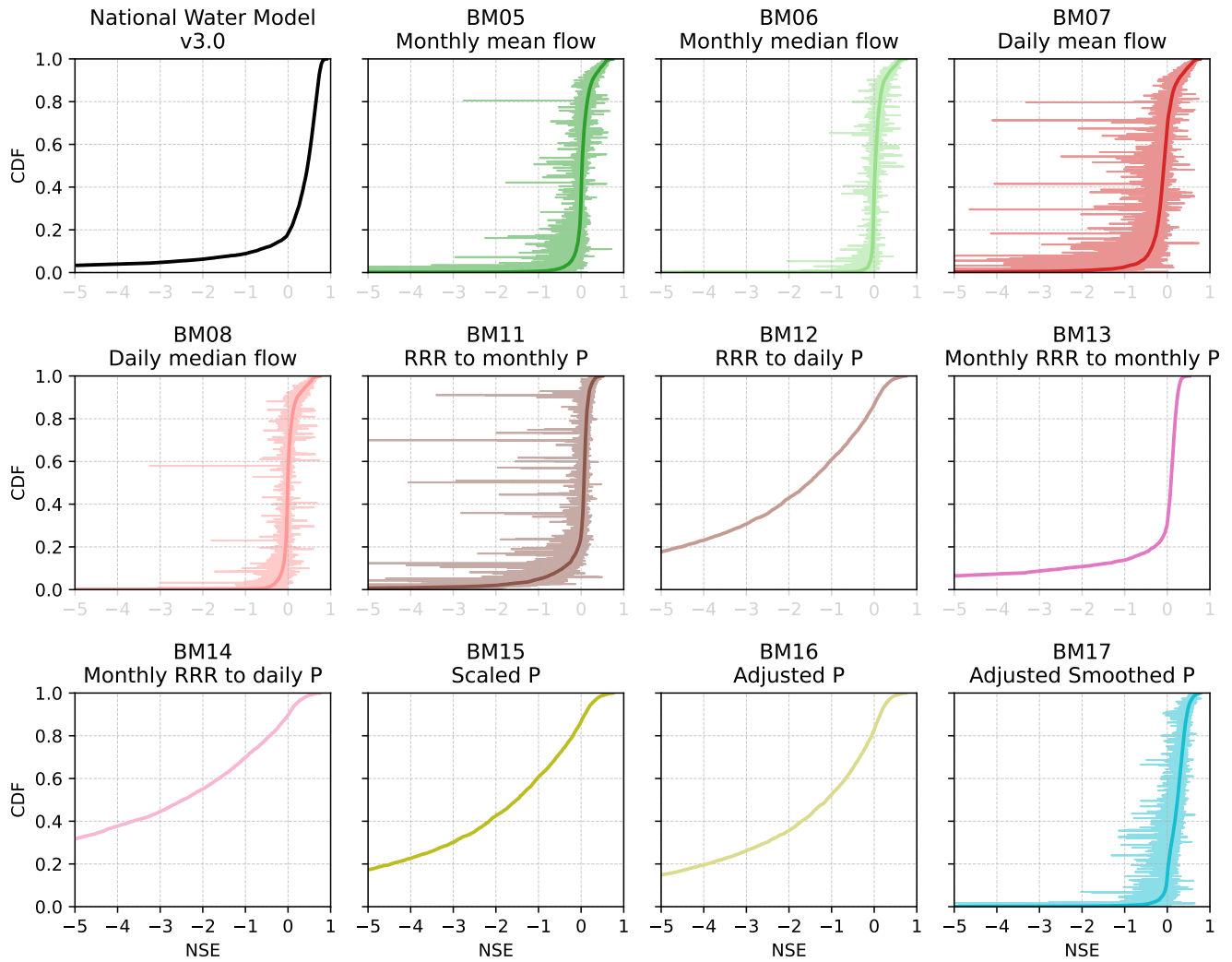


Figure S2. Cumulative Distribution Function (CDF) plot of the Nash–Sutcliffe Efficiency (NSE) scores of the evaluation period. The National Water Model v3.0 KGE scores are in black, and the NSE scores for the simple benchmarks in colors. Sampling uncertainty (defined as the difference between the 5th and 95th percentile NSE estimate) in lighter colors. For benchmarks 11, 12, 13 and 14, *RRR* stands for Rainfall Runoff Ratio. *P* (benchmarks 11–16) stands for precipitation.

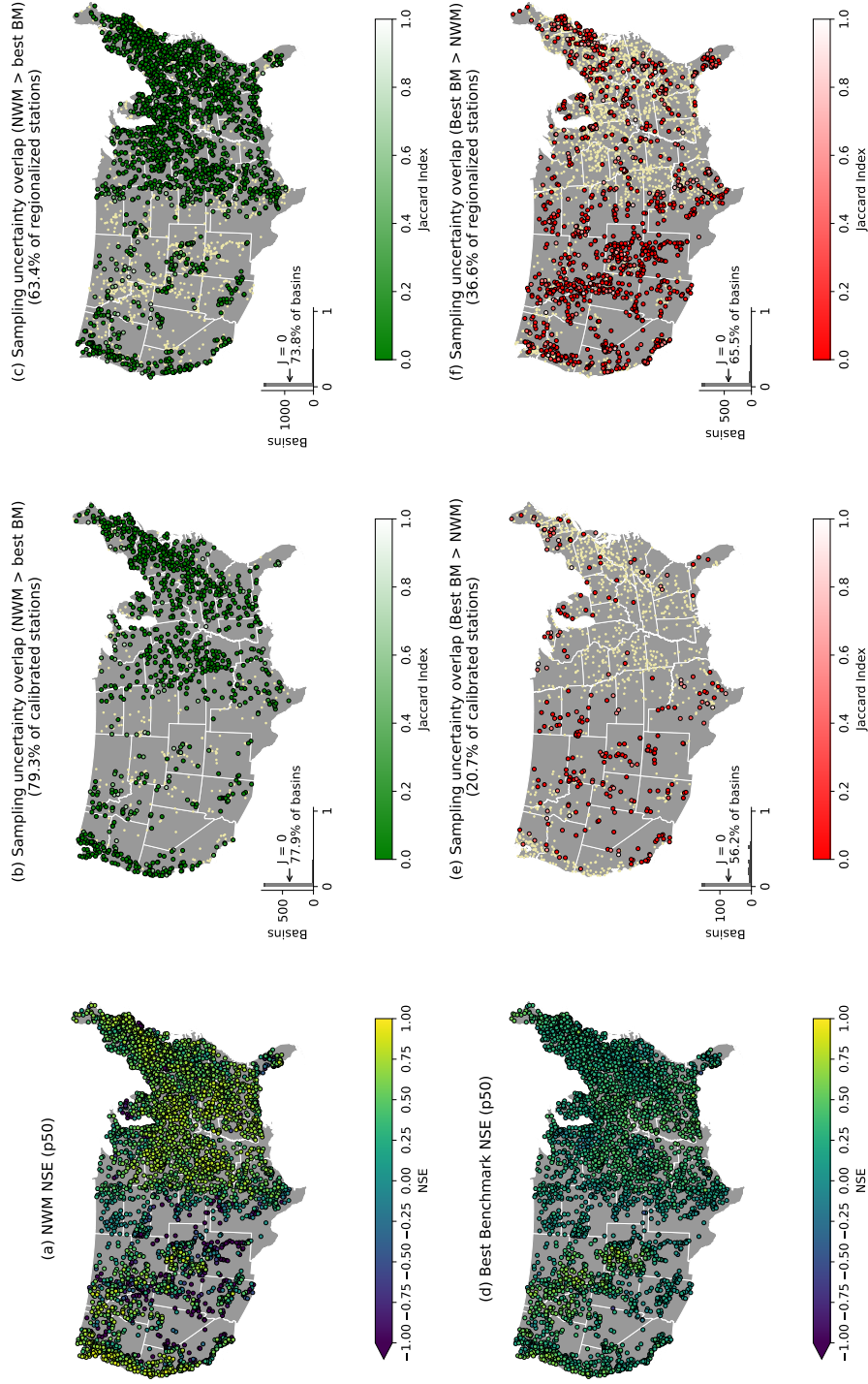


Figure S3. Overview of spatial patterns in model and benchmark performance during the evaluation period, using **NSE** and **all benchmarks simultaneously**. (a,d) Estimated 50th percentile NSE score for NWM and best benchmark respectively. (b,e) Jaccard index showing overlap between sampling uncertainty intervals where the 50th percentile NSE score for NWM > benchmark, and NWM < benchmark, respectively, for gauges used for model calibration. (c,f) Jaccard index showing overlap between sampling uncertainty intervals where the 50th percentile NSE score for NWM > benchmark, and NWM < benchmark, respectively, for gauges used for model regionalization. Histograms show Jaccard index distributions and specifically call out the number of $J = 0$ cases, where the estimated metric distributions have no overlap. Borders obtained from Commission for Environmental Cooperation (CEC) (2022).

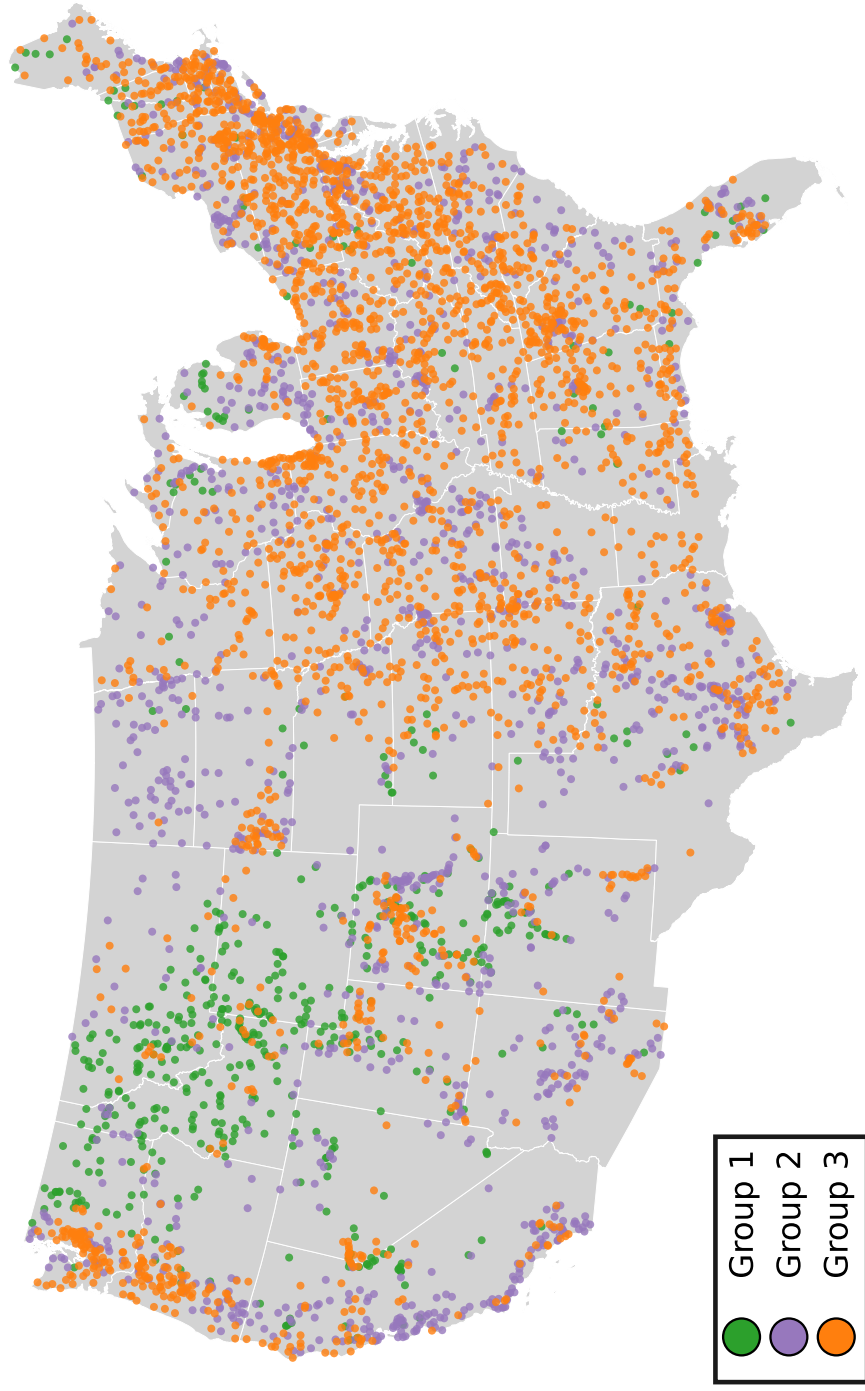


Figure S4. Composite map of best benchmark, summarizing benchmarks by category: (1) benchmarks derived from streamflow data (BM01-08); (2) benchmarks derived from rainfall-runoff ratios (BM09-15), and (3) simple model benchmarks (BM16-17). Borders obtained from Commission for Environmental Cooperation (CEC) (2022).

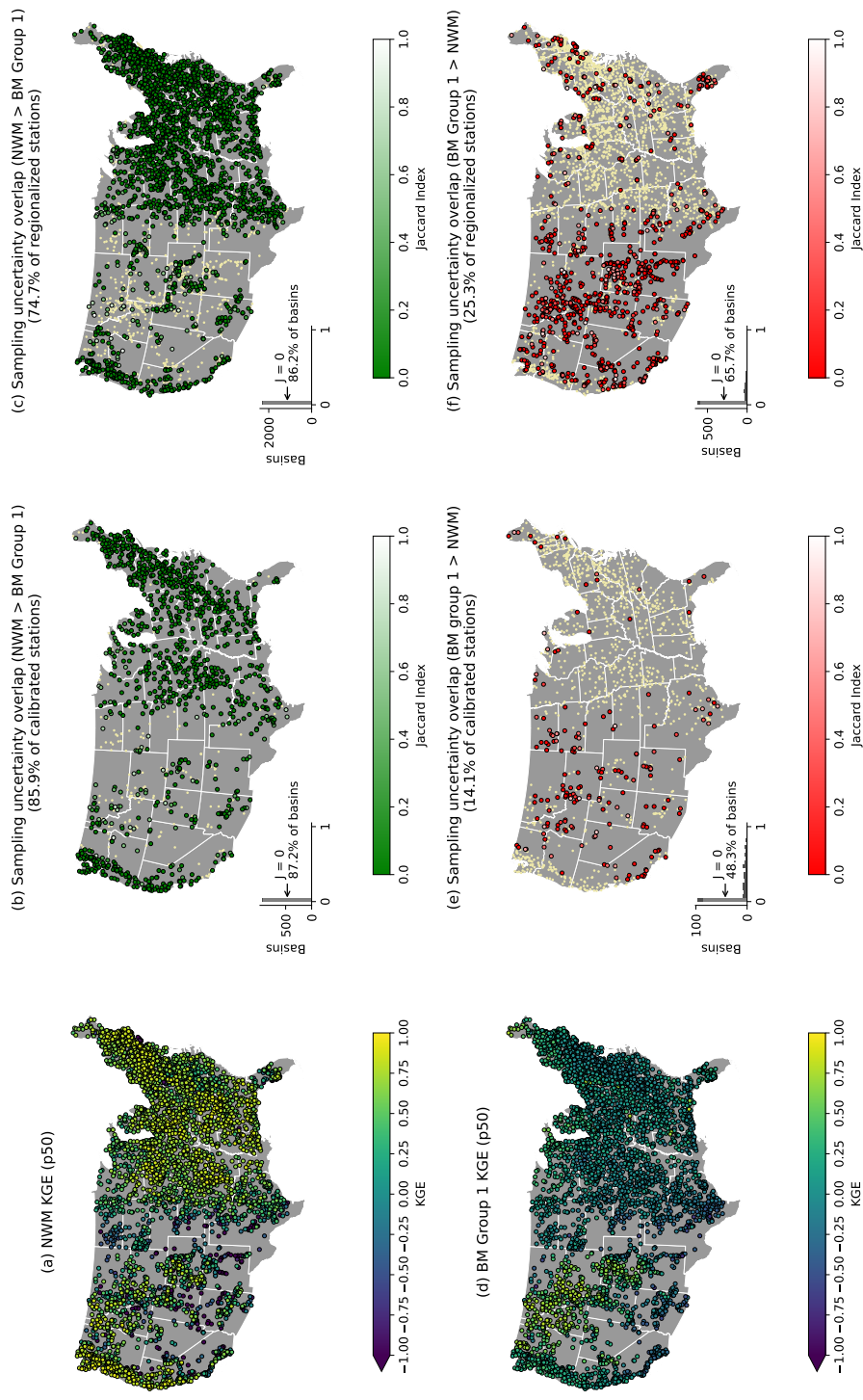


Figure S5. Overview of spatial patterns in model and benchmark performance, and overlap using **KGE** and **streamflow-only benchmarks**. (a,d) Estimated 50th percentile KGE score for NWM and best benchmark respectively. (b,e) Jaccard index showing overlap between sampling uncertainty intervals where the 50th percentile KGE score for NWM >benchmark, and NWM <benchmark, respectively, for gauges used for model calibration. (c,f) Jaccard index showing overlap between sampling uncertainty intervals where the 50th percentile KGE score for NWM >benchmark, and NWM <benchmark, respectively, for gauges used for model regionalization. Histograms show Jaccard index distributions and specifically call out the number of $J = 0$ cases, where the estimated metric distributions have no overlap. Borders obtained from Commission for Environmental Cooperation (CEC) (2022).

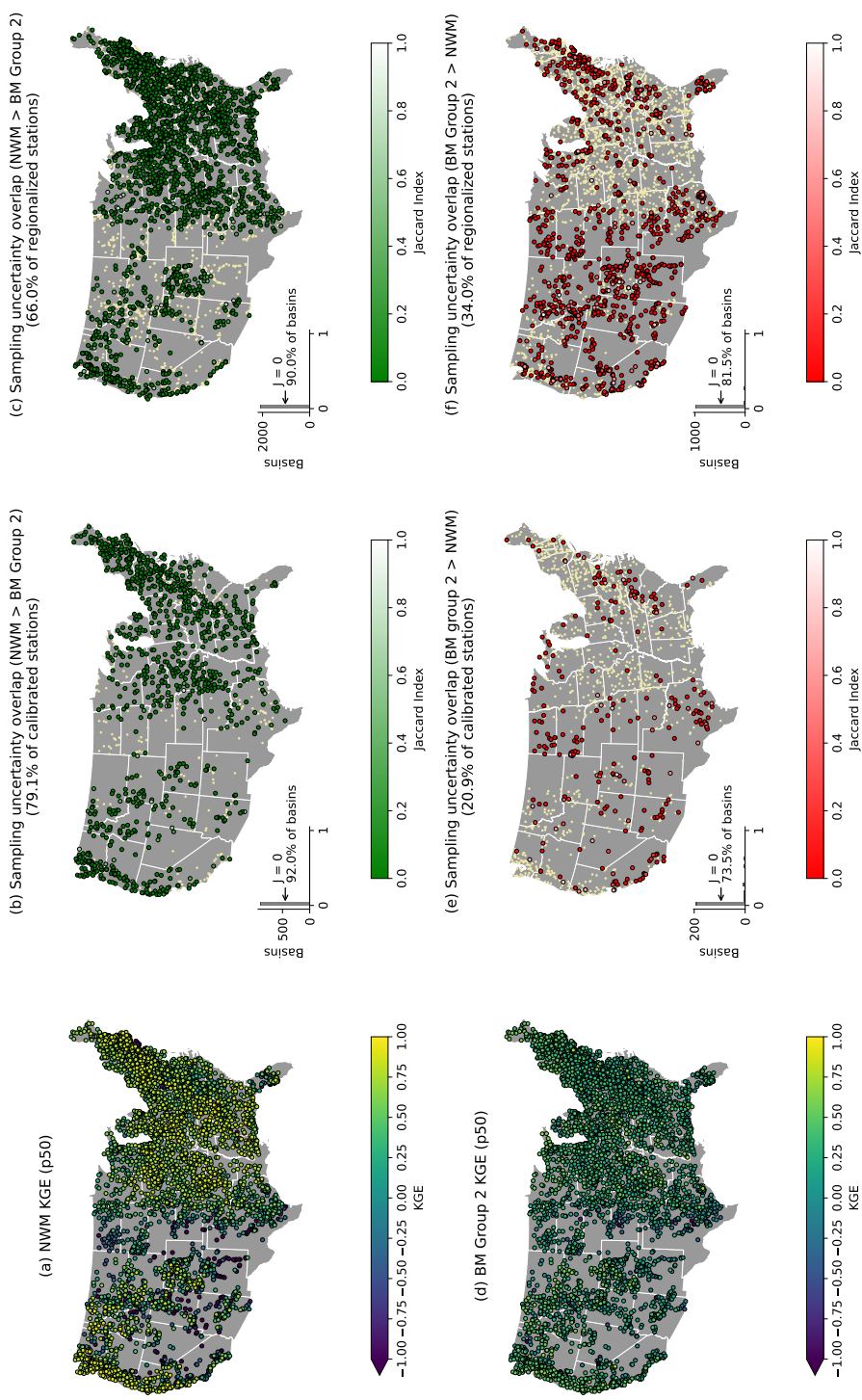


Figure S6. Overview of spatial patterns in model and benchmark performance, and overlap using **KGE** and **rainfall-runoff ratio benchmarks only**. (a,d) Estimated 50th percentile KGE score for NWM and best benchmark respectively. (b,e) Jaccard index showing overlap between sampling uncertainty intervals where the 50th percentile KGE score for NWM >benchmark, and NWM <benchmark, respectively, for gauges used for model calibration. (c,f) Jaccard index showing overlap between sampling uncertainty intervals where the 50th percentile KGE score for NWM >benchmark, and NWM <benchmark, respectively, for gauges used for model regionalization. Histograms show Jaccard index distributions and specifically call out the number of $J = 0$ cases, where the estimated metric distributions have no overlap. Borders obtained from Commission for Environmental Cooperation (CEC) (2022).

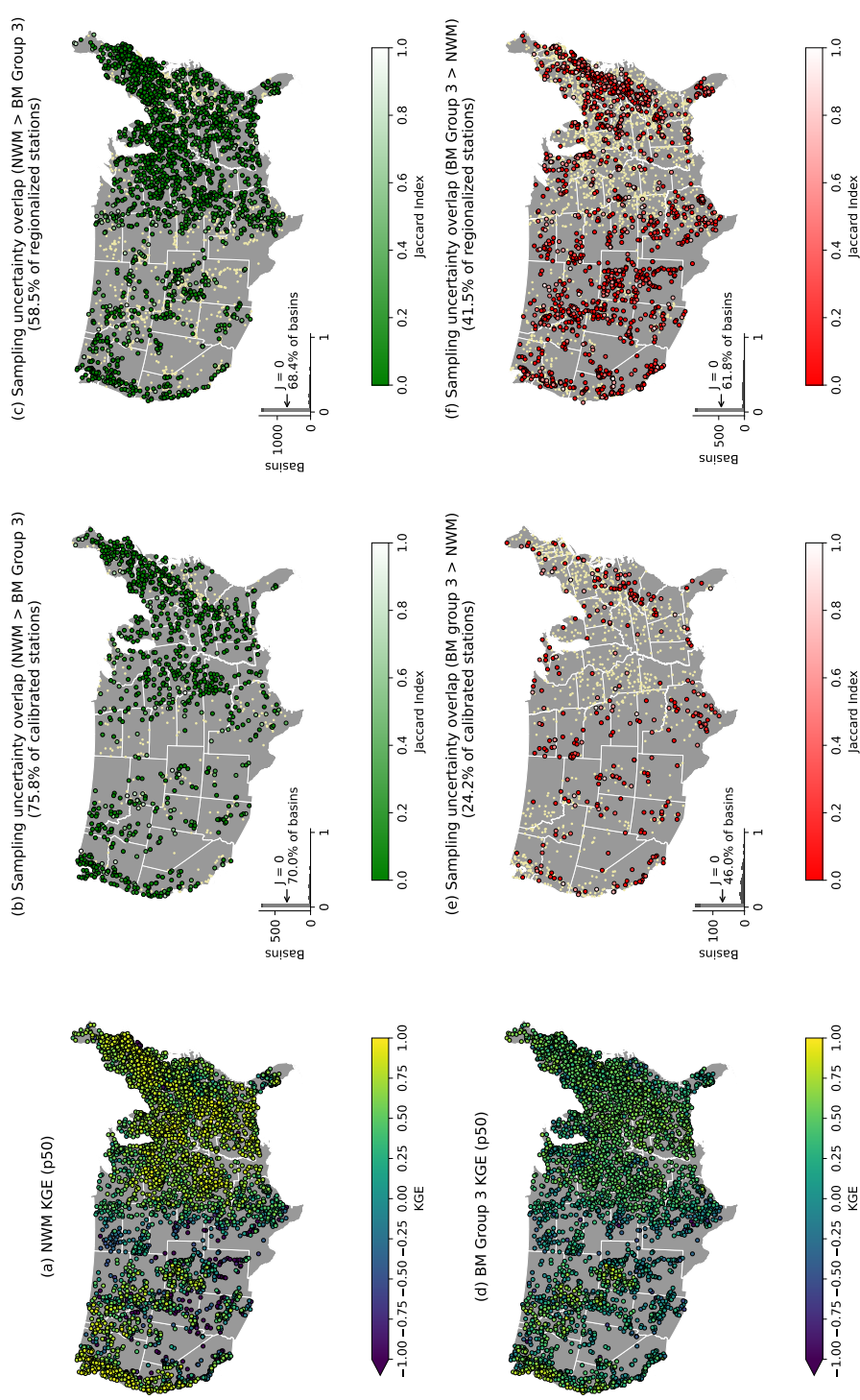


Figure S7. Overview of spatial patterns in model and benchmark performance, and overlap using **KGE** and **simple model benchmarks**. (a,d) Estimated 50th percentile KGE score for NWM and best benchmark respectively. (b,e) Jaccard index showing overlap between sampling uncertainty intervals where the 50th percentile KGE score for NWM >benchmark, and NWM <benchmark, respectively, for gauges used for model calibration. (c,f) Jaccard index showing overlap between sampling uncertainty intervals where the 50th percentile KGE score for NWM >benchmark, and NWM <benchmark, respectively, for gauges used for model regionalization. Histograms show Jaccard index distributions and specifically call out the number of $J = 0$ cases, where the estimated metric distributions have no overlap. Borders obtained from Commission for Environmental Cooperation (CEC) (2022).



Figure S8. KGE components for National Water Model simulations and all benchmarks, for stations used for parameter calibration.

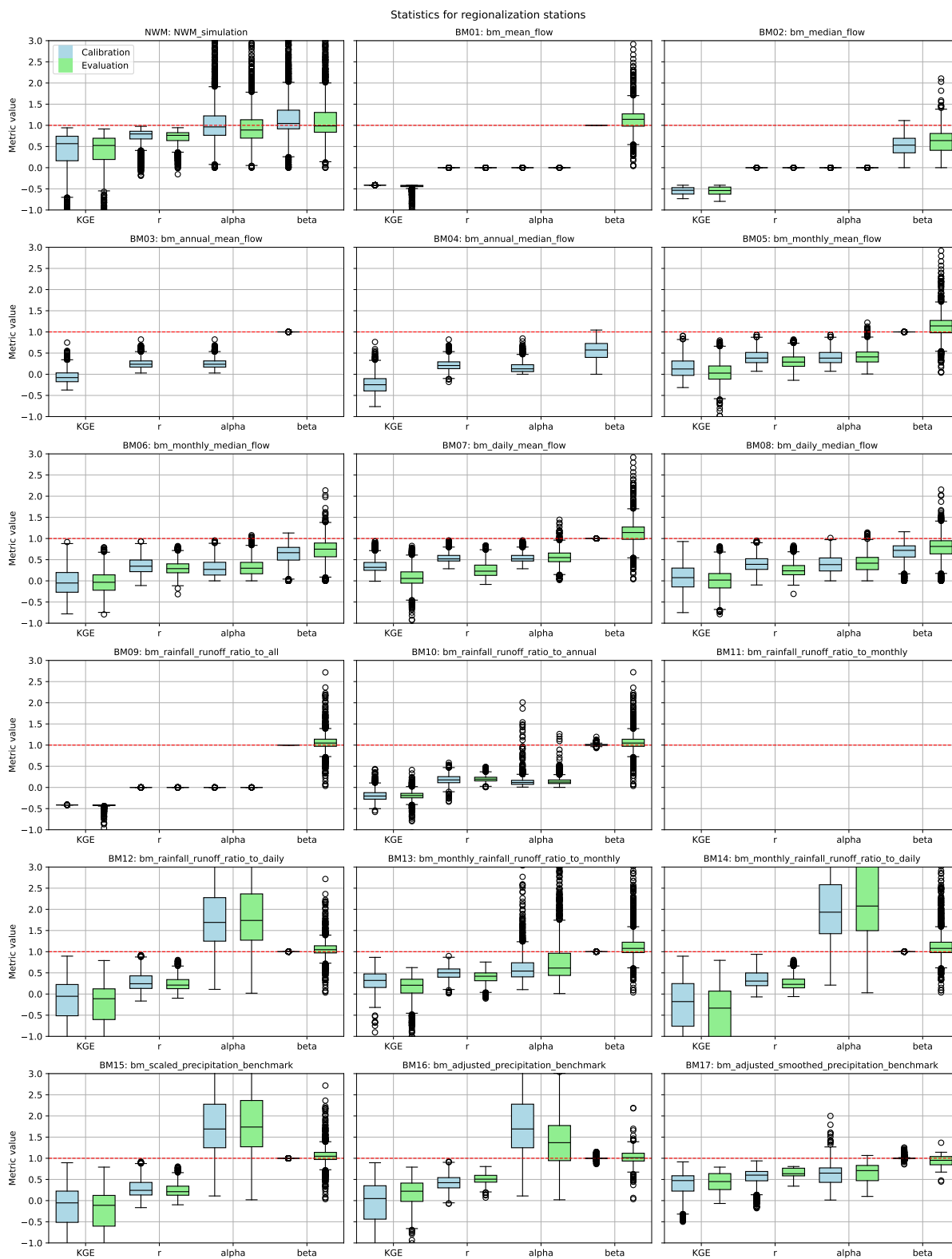


Figure S9. KGE components for National Water Model simulations and all benchmarks, for stations used for parameter regionalization.

References

- 10 Commission for Environmental Cooperation (CEC): North American Atlas – Political Boundaries, <http://www.cec.org/north-american-environmental-atlas/political-boundaries-2021/>, statistics Canada, United States Census Bureau, Instituto Nacional de Estadística y Geografía (INEGI), 2022.
- Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.