Metrics that Matter:

Calibration Choices and Their Impact on Signature Representation in Conceptual Hydrological Models

Peter Wagener[1*] // Diana Spieler[2] // Niels Schütze[2]

Motivation

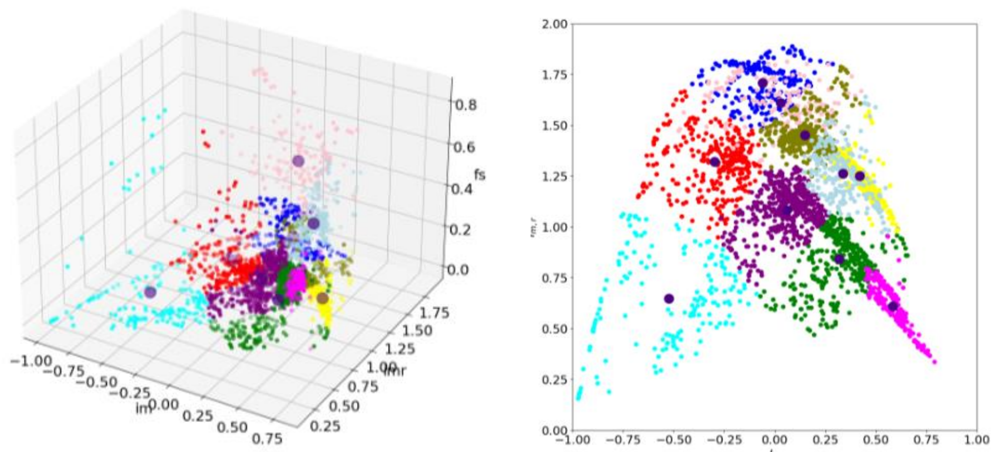
- Calibration is a key part of the hydrological modeling chain and typically guided by an objective function employing a performance metric.
- Calibration determines model parameters that govern model performance but also the implicit representation of hydrological processes.
- The impact the choice of metric may have on process representation is not well understood. This hampers the informed selection of suitable calibration metrics for specific modeling purposes.

This study investigates the impact of 8 objective functions on 15 different signatures for 11 climatically diverse catchments and 47 conceptual, lumped hydrological models.

Data and Methods

Catchments:

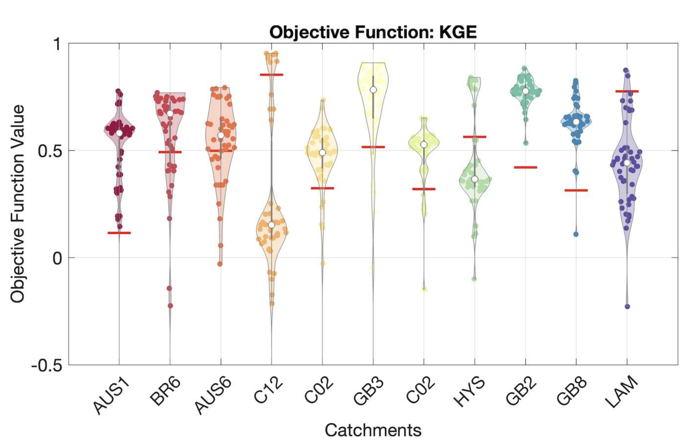
CARAVAN



- The 11 catchments are a subset of the CARAVAN dataset [g].
- Catchments were selected using a k-means clustering algorithm using aridity, seasonality, and snow fraction [f].

Models:

- Models from the MARRMoT-Toolbox [e, k] include a variety of widely used models (e.g., GR4J, HBV, VIC, HYMOD, TOPMODEL).
- 47 models were calibrated (using CMA-ES) and evaluated for a 10-year time-period with a 1-year warm-up period.
- The interannual mean was used as a benchmark to select models for the analysis.



driest (left) to wettest (right)

Metrics:

- The evaluated metrics are KGE [c], NSE [h] and the logarithmic versions of both.
- Additionally, DE [j], KGE-NP [i], KGE-Split [a] and SHE [d] are evaluated.

Signatures

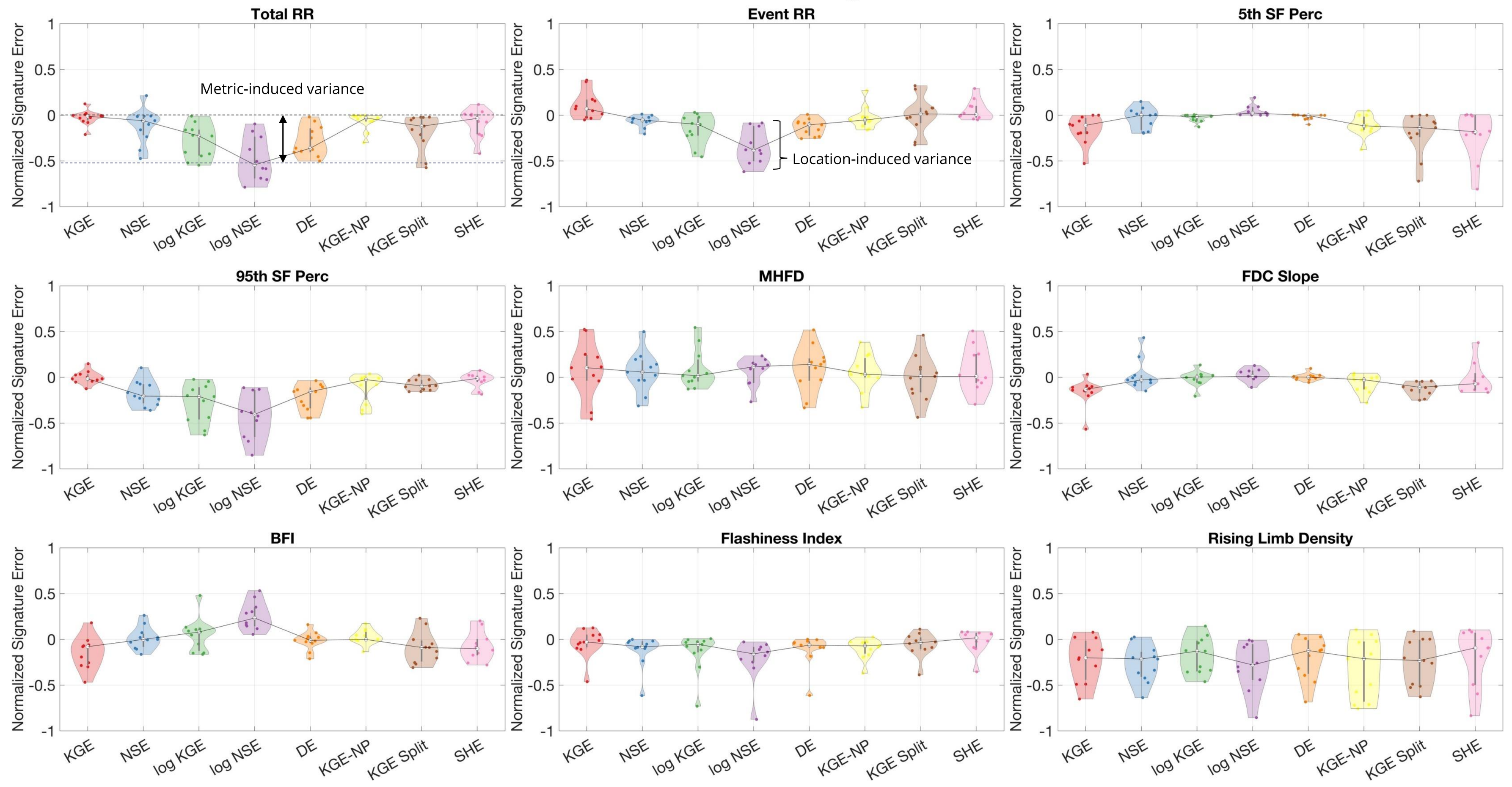
- The signatures used are a subset of the signatures in the TOSSH-Toolbox [b].

Toolbox for Streamflow Signatures in Hydrology

Category/Processes	Signatures
Streamflow	Slope of Flow Duration Curve, 5th and 95th SF Percentile, High/Low Flow Frequency and Duration, Mean Half Flow Date
Water Balance	Total Runoff Ratio
Partitioning/Connectivity	Event Runoff Ratio
Baseflow	Baseflow Index, Baseflow Recession Coefficient
Water Storage	Flashiness Index, Variability Index
Channel Processes	Rising Limb Density

Results

Metric- and location-induced Variance:



Each violinplot shows the distribution of the median signature error for the 11 catchments (median calculated over all models surpassing the benchmark)

	KGE	NSE	logKGE	logNSE	DE	KGE-NP	KGE-Split	SHE	OF Impact
TRR	+	+	-	- -	-	+ +	+	+ +	**+ +**
ERR	+	+	-	- -	+	+ +	+	+ +	**+**
5SF%	-	+	+ +	+ +	+ +	+	- -	- -	**+**
95SF%	+ +	-	-	- -	-	+	+	+ +	**+ +**
MHFD	-	-	+	+ +	+	+ +	+	+ +	**+**
FDC-Slope	-	+	+ +	+ +	+ +	+	+	+	**+**
BFI	-	+	-	- -	+ +	+ +	+	+	**+**
FI	+ +	+	+	+	+ +	+	+ +	+ +	**-**
RLD	-	-	+	-	+	-	-	+	**- -**

SIGNATURES: + + very good, + good, - bad, - - very bad,
OF IMPACT: + + very high, + high, - low, - - very low

Impacts of Objective Functions (OF):

- KGE: good on Runoff Ratios and high flows, underestimated low flows, solid performance on additional metrics
- NSE: solid throughout, better than KGE on low flows
- log KGE: good on low flows, better than log NSE overall
- log NSE: good on low flows, very high ET and BFI, underestimates high flow drastically
- DE: good on low flows, better than log KGE/NSE on BFI and Event RR
- KGE-NP: best performance overall, slight underestimate on low flows, improves standard KGE in multiple aspects
- KGE-Split: underestimated runoff, partial improvements over KGE
- SHE: very good overall, slight weakness on low-flows

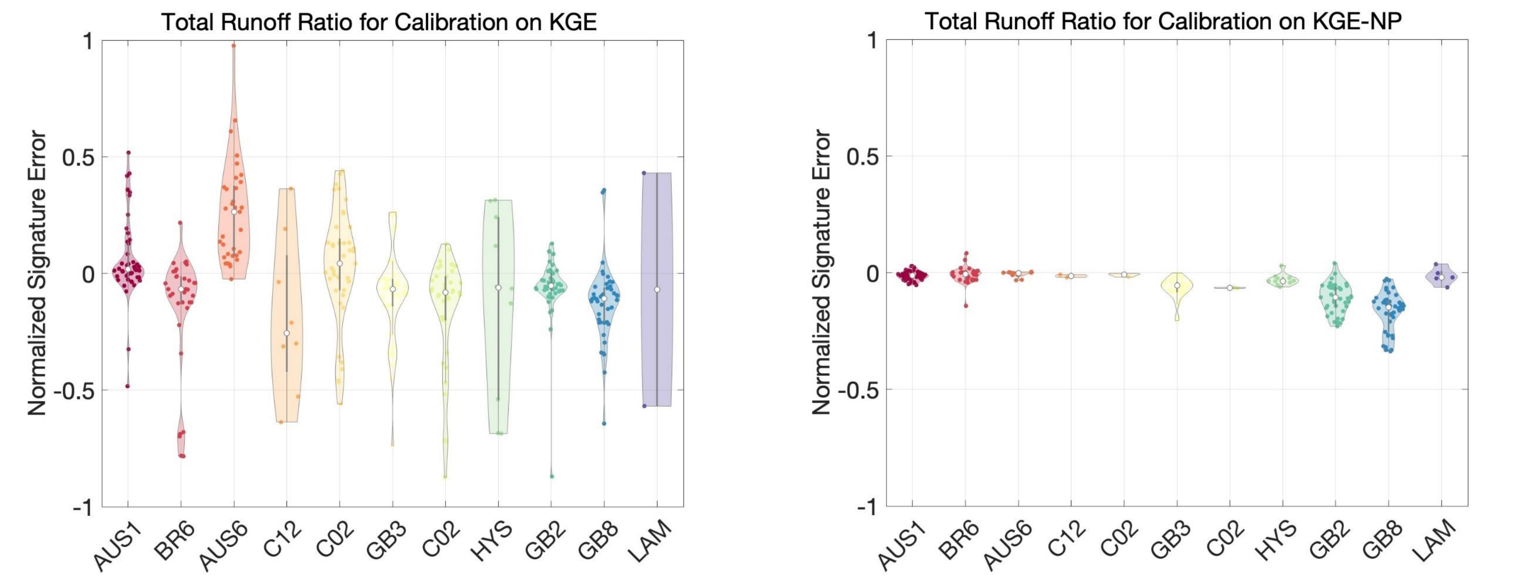
KEY MESSAGE

The choice of calibration metric can strongly impact the signature representation in conceptual hydrologic modeling!

Discussion

- The influence of the calibration metric on the signature representation depends largely on the signature, catchment, and model
- **KGE-NP** and **SHE** are the most accurate estimates for signature representation
- General patterns show that every OF has specific abilities and shortcomings

→ We recommend selecting a signature that accurately represents the aspect of streamflow of interest when applying conceptual models



New Objective Functions:

- The two new versions of the KGE (non-parametric version [i] and KGE-Split [a]) have improvements for some signatures (less variance, better mean), thus showing improvement compared to the KGE (see Figure above).
- The diagnostic efficiency (DE) performs similarly to the logarithmic version of the KGE/NSE on low flows but shows improvement regarding other signatures.
- The SHE emerged as one of the best OFs in this study and implies further potential by incorporating specific signatures within it.

Main Limitation:

- The number of catchments is too low and does neither account for extreme climates nor incorporate the non-climatic properties.
- Expanding this study could allow us to disentangle the impacts of objective functions, locations, signatures , and models.
- The selection of objective functions was subjective and could be extended.

TECHNISCHE UNIVERSITÄT DRESDEN

Contact: peter.wagener@ucalgary.ca

[1] Schulich School of Engineering, University of Calgary
[2] Institute of Hydrology and Meteorology, TUD Dresden University of Technology
* research primarily conducted at TUD, now at UofC

References

[a] Fowler, K., Peel, M., Western, A., & Zhang, L. (2018). Improved rainfall-runoff calibration for drying climate: Choice of objective function. Water Resources Research,54 (5), 3392–3408. https://doi.org/10.1029/2017WR022466
[b] Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J., & McMillan, H. K. (2021). TOSSH: A toolbox for streamflow signatures in hydrology. Environmental Modelling & Software, 138, 104983. https://doi.org/10.1016/j.envsoft.2021.104983
[c] Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377 (1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003
[d] Kiraz, M., Coxon, G., & Wagener, T. (2023). A signature-based hydrologic efficiency metric for model calibration and evaluation in gauged and ungauged catchments. Water Resources Research, 59 (11), e2023WR035321. https://doi.org/10.1029/2023WR035321
[e] Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., & Woods, R. A. (2019). Modular assessment of rainfall–runoff models toolbox (MARRMoT) v1.2: An open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. Geoscientific Model Development, 12 (6), 2463–2480. https://doi.org/10.5194/gmd-12-2463-2019
[f] Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. Water Resources Research, 54 (7), 5088–5109. https://doi.org/10.1029/2018WR022913
[g] Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., & Matias, Y. (2023). Caravan - a global community dataset for large-sample hydrology. Scientific Data, 10 (1), 61. https://doi.org/10.1038/s41597-023-01975-w
[h] Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part i — a discussion of principles. Journal of Hydrology, 10 (3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6
[i] Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: Towards a non-parametric variant of the kling-gupta efficiency. Hydrological Sciences Journal, 63 (13), 1941–1953. https://doi.org/10.1080/02626667.2018.1552002
[j] Schwemmle, R., Demand, D., & Weiler, M. (2021). Technical note: Diagnostic efficiency – specific evaluation of model performance. Hydrology and Earth System Sciences, 25 (4), 2187–2198. https://doi.org/10.5194/hess-25-2187-2021
[k] Trotter, L., Knoben, W. J. M., Fowler, K. J. A., Saft, M., & Peel, M. C. (2022). Modular assessment of rainfall–runoff models toolbox (MARRMoT) v2.1: An object-oriented implementation of 47 established hydrological models for improved speed and readability. Geoscientific Model Development, 15 (16), 6359–6369. https://doi.org/10.5194/gmd-15-6359-2022

IHM