Testing the Hydrological Adequacy of 7533 KGE Calibrated Conceptual Model Structures

Diana Spieler [1] // Niels Schütze [1]

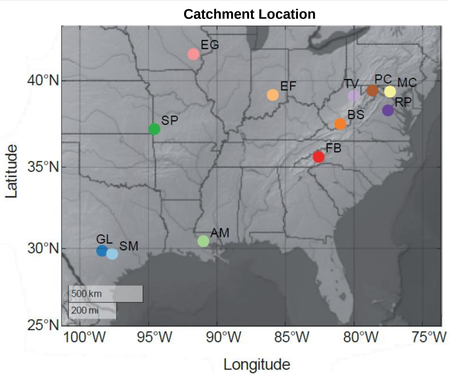
[1] Chair of Hydrology, TUD Dresden University of Technology

Motivation

- A considerable number of modelling studies calibrates their models on KGE. This raises the impression that calibration metrics, just like hydrological models themselves, are often chosen by legacy (e.g. convenience, experience, habit) rather than adequacy (e.g. fit for purpose, representing the variable of interest).
- We test with 7533 conceptual lumped model structures what this means for the representation of the hydrological flow regime as represented through signatures.

Data

- 12 hydro-climatically diverse US MOPEX catchments (Duan et al., 2006) are tested.
- They are semi-arid to humid, unregulated, have rainfall-runoff coefficients between 0.14 (GL) and 0.56 (TV), and BFI values between 0.32 (SP) and 0.70 (FB).
- A large variety in potential runoff generation processes is therefore expected.



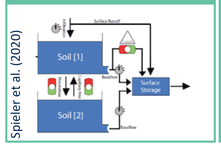
Models

- The 7533 conceptual model stem from 3 different modelling experiments:
  - 7488 models are variants of a 1 or 2 soil storage structure that can vary in the availability and description of 9 different processes (AMSI + BMF)
  - 45 models are commonly used literature-based models significantly varying in storage and process availability

The 3 Modelling Experiments

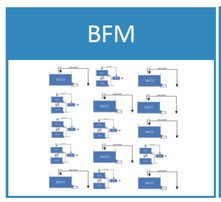
Using the same Model Hypothesis Space

AMSI — 100 Models



100 simultaneous structure and parameter calibrations of a flexible model structure with 2-6 storages and 0 to 12 parameters

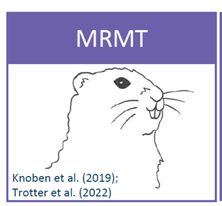
BFM — 7488 Models



Parameter calibration of 7488 individual fixed model structures with 2-6 storages from the same model space AMSI searches

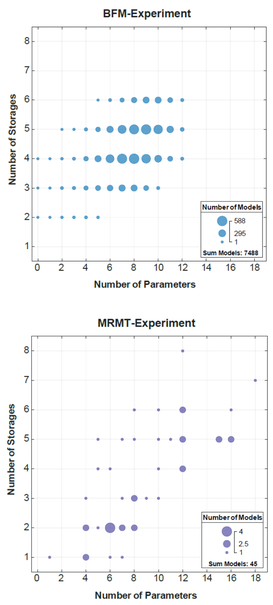
Literature Based

MRMT — 45 Models



Parameter calibration of 45 individual fixed and literature-based model structures with 1-8 storages and 1 to 18 parameters

Comparison Model Complexity



Hydrological Adequacy Requirements

- We consider a model to be accurate when it performs well in calibration (11/1975 – 10/2000) and validation (11/1950 – 10/1975).
- We consider all models that perform within a ΔKGE of 0.05 of the best KGE model for a catchment as accurate.
- We consider a model to be adequate when it additionally performs acceptable (less than 50% bias) in monthly and yearly mean flow representations and on 10 selected signatures.

Abbreviation	Unit	Description	Reasoning
RR	[-]	Runoff Ratio	General Water Balance
BFI	[-]	Baseflow Index	Magnitude Baseflow
Q5	[mm]	5th streamflow percentile	Magnitude Low Flows
Q95	[mm]	95th streamflow percentile	Magnitude High Flows
LFfreq	[-]	Low Flow Frequency	Frequency
HFfreq	[-]	High Flow Frequency	Frequency
LFdur	[d]	Low Flow Duration	Duration
HFdur	[d]	High Flow Duration	Duration
MHFD	[dowy]	Mean Half Flow Date	Timing
slopeFDC	[-]	slope of the FDC	Rate of Change

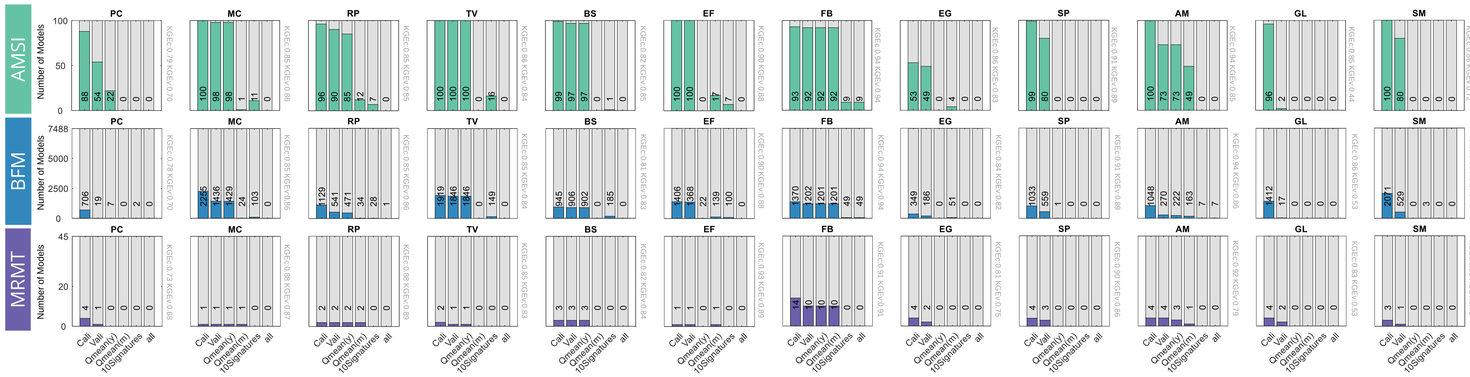
Main Message

In only 57 out of 90369 instances did the tested 7533 KGE calibrated models satisfy our adequacy requirements.

Model calibration and evaluation needs to move beyond single aggregated metric considerations!

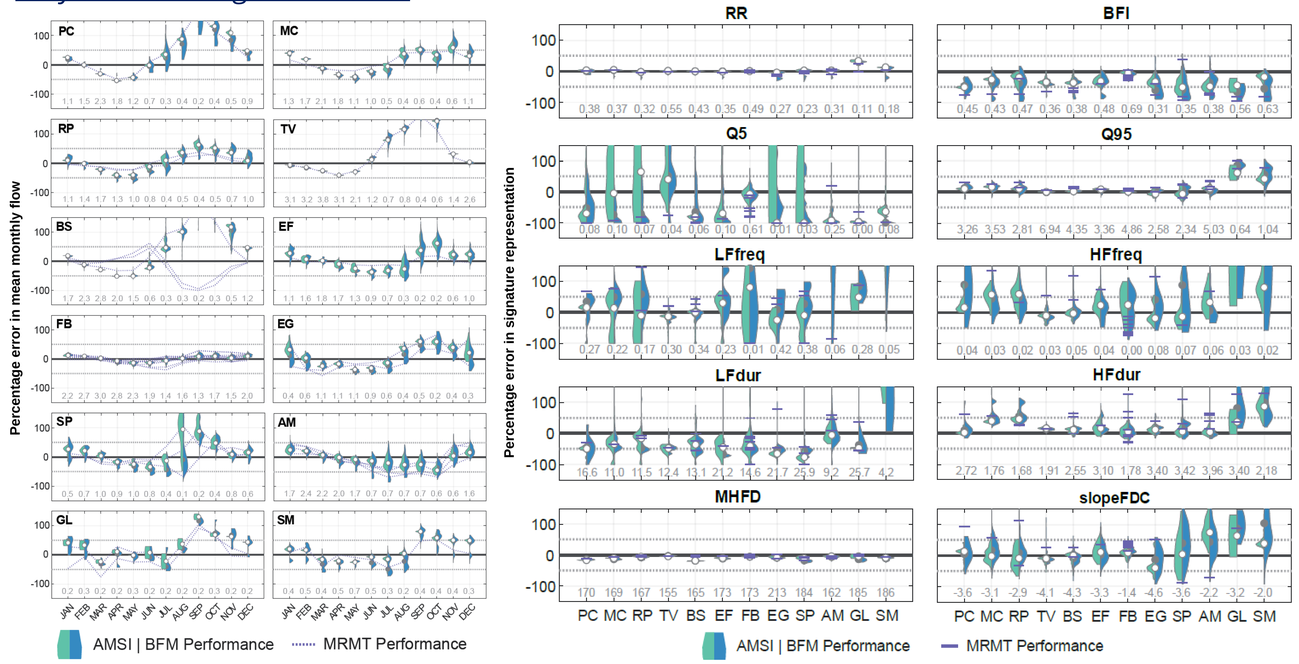
Results

Where are we losing the models?



- In nine out of twelve catchments, no model was considered adequate.
- In the remaining 3 catchments, only between 1 (0.1 %) and 49 (0.7 %) of all tested model structures are adequate.
- Most models are „lost" due to errors in their monthly Qmean or signature representation.

Why are we losing the models?



Conclusions

- Counterbalancing errors in KGE can lead to inadequate model structures (Q overestimation in autumn and underestimation in spring still lead to good Qmean over the entire time period)
- Signatures connected to KGE are represented well (RR, MHFD, Q95), BFI generally underestimated and large errors for frequency signatures, and low flows.
- The literature-based MARRMoT models do not show any immediate benefits compared to the other tested models.
- This indicates that as long as we constrain conceptual model structures through nothing but aggregated metrics, we do not benefit from any of the knowledge that might have gone into their development.

Contact:
diana.spieler@tu-dresden.de

OSPP Rating

References : Duan et al. (2006): Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.* **320**, 3–17.; Spieler et al. (2020): Automatic Model Structure Identification for Conceptual Hydrologic Models. *Water Resour Res* **56**, (2020).; Knoben et al. (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. *Geosci. Model. Dev.* **12**, 2463–2480.; Trotter et al. (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. *Geosci. Model. Dev.* **15**, 6359–6369.