Department of Hydrosciences, Institute of Hydrology and Meteorology, Chair of Hydrology
TU Dresden, Germany

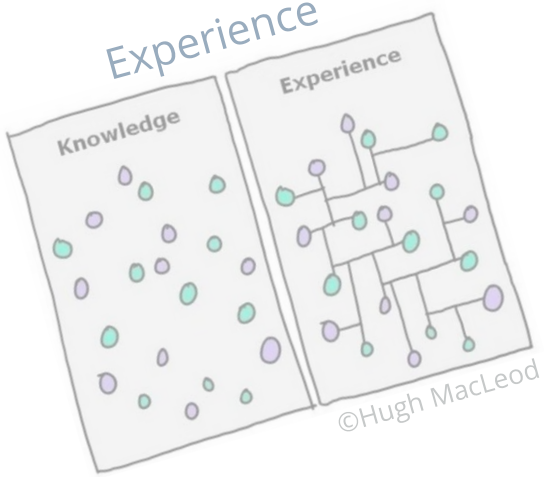
Benchmarking Automatically Identified Model Structures with a Large Model Ensemble

Diana Spieler, Kan Lei and Niels Schütze

diana.spieler@tu-dresden.de

Automatic Model Structure Identification

Experience



©Hugh MacLeod

Addor & Melsen, 2018

Typically done by …

Practicality

Intercomparison Studies

Fenicia et al., 2014

Modular Modelling Framework

Raven

UNIVERSITY OF WATERLOO

+

(Mixed Integer) Calibration Algorithms

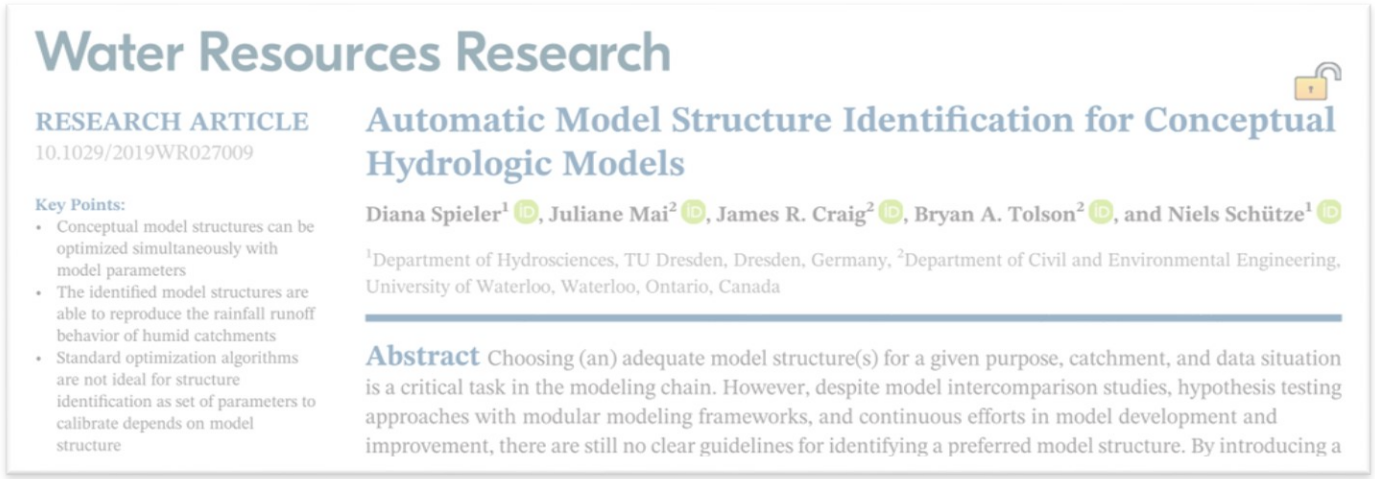
DDS

Tolson & Shoemaker, 2007

Automatic Model Structure Identification

(AMSI)

Method description and proof of concept:



Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR027009

Key Points:
- Conceptual model structures can be optimized simultaneously with model parameters
- The identified model structures are able to reproduce the rainfall runoff behavior of humid catchments
- Standard optimization algorithms are not ideal for structure identification as set of parameters to calibrate depends on model structure

Automatic Model Structure Identification for Conceptual Hydrologic Models

Diana Spieler[1], Juliane Mai[2], James R. Craig[2], Bryan A. Tolson[2], and Niels Schütze[1]

[1]Department of Hydrosciences, TU Dresden, Dresden, Germany, [2]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario, Canada

Abstract Choosing (an) adequate model structure(s) for a given purpose, catchment, and data situation is a critical task in the modeling chain. However, despite model intercomparison studies, hypothesis testing approaches with modular modeling frameworks, and continuous efforts in model development and improvement, there are still no clear guidelines for identifying a preferred model structure. By introducing a

https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019WR027009

→ ☑ Does AMSI work?

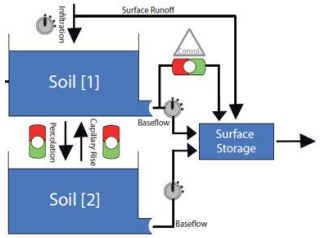
Research Question:

→ How well does AMSI perform* in regard to different benchmarks?

→ ☑ How well does it work?

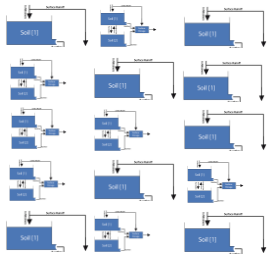
* Performance will be defined in KGE performance henceforth, even though the definition of when (a) model structure(s) perform well is a huge question in itself !!!

Automatic Model Structure Identification



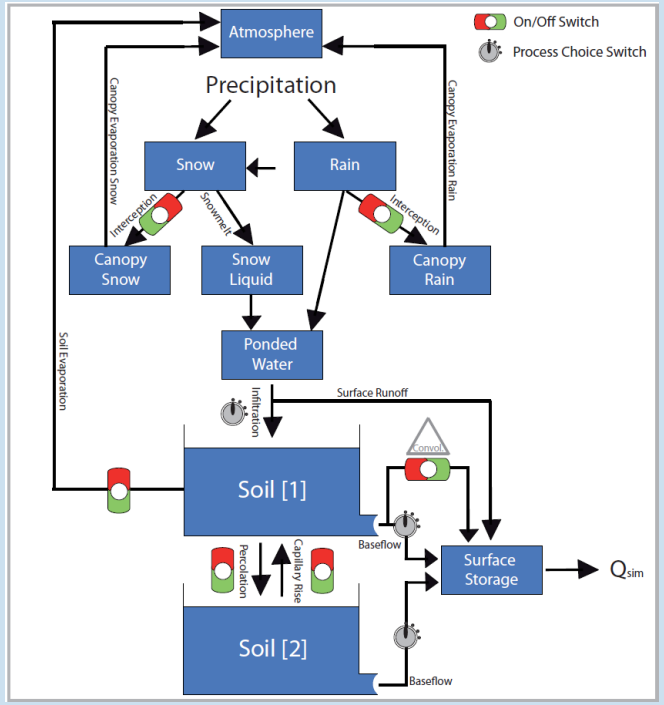
MMF combined with mixed integer optimization algorithm to simultaneously calibrate model structures (integer p.) and model parameters (continuous p.)

Brute Force Modelling

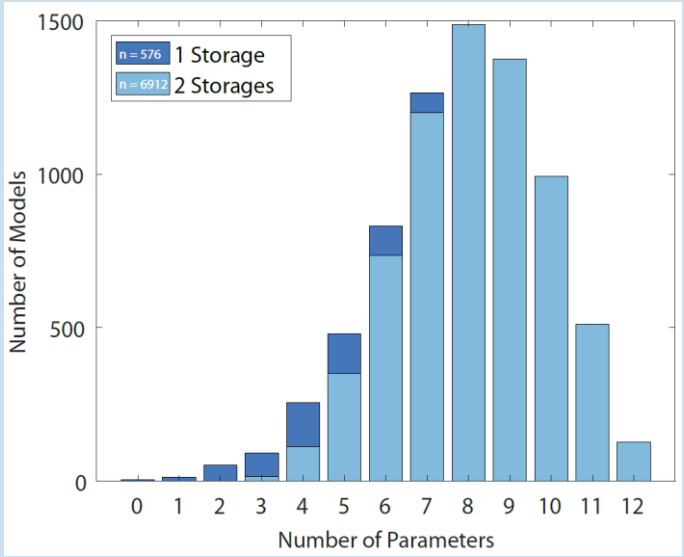


1 standard parameter calibration for 7488 fixed model structures build with MMF. Model structures are resembling the same model space AMSI is searching.

Identical Model Space with 13824 possible combinations

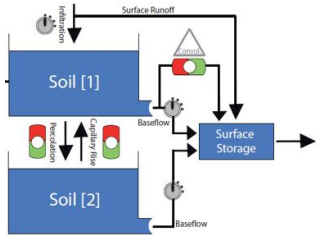


- The model space allows 1 or 2 soil storages
- 9 processes can be in- or excluded from the model structure
  - Interception (R/S) /Canopy Evaporation (R/S)
  - Infiltration
  - Soil Evaporation
  - Percolation
  - Capillary Rise
  - Baseflow 1
  - Baseflow 2
  - Convolution
- 3 of 9 processes also have several process options



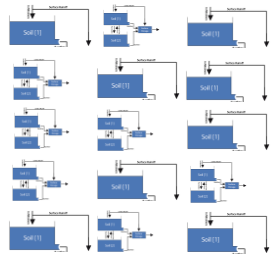
- 0-12 parameters depending on chosen model structure
  - **BUT** 29 parameters are constantly calibrated for AMSI to work
- 7488 most likely model structures are calibrated for the brute force modelling (BFM)

Automatic Model Structure Identification



MMF combined with mixed integer optimization algorithm to simultaneously calibrate model structures (integer p.) and model parameters (continuous p.)

Brute Force Modelling



1 standard parameter calibration for 7488 fixed model structures build with MMF. Model structures are resembling the same model space AMSI is searching.

MARRMoT Toolbox

1 standard parameter calibration for 45 literature based model structures that are structurally more divers than AMSI

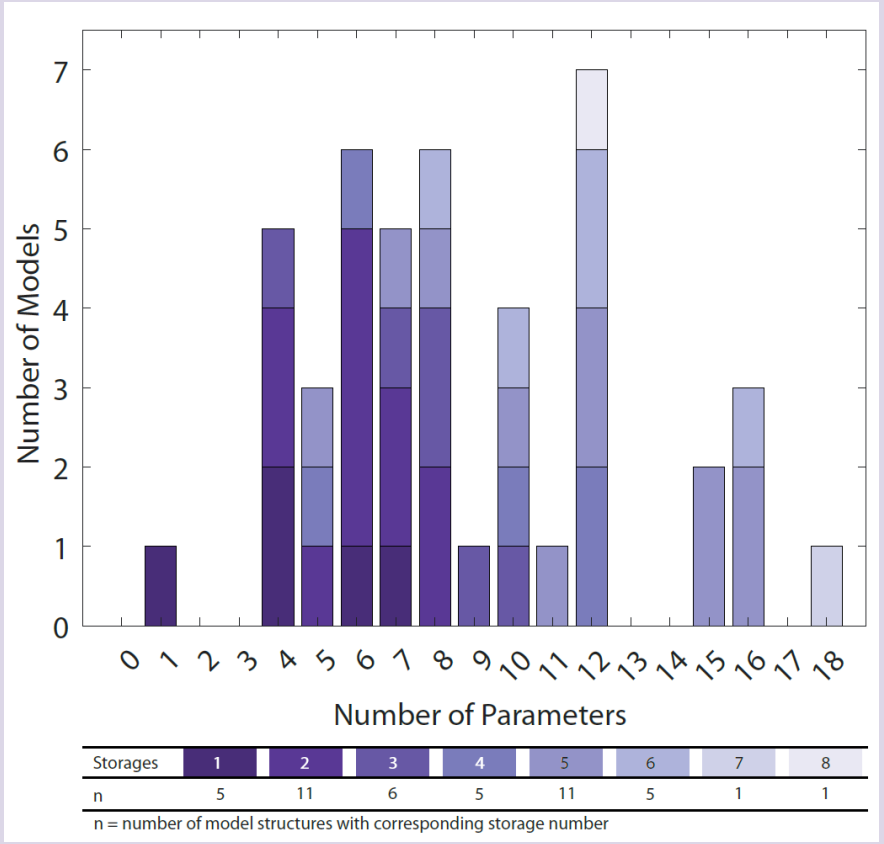
Knoben et al., 2019

Structurally more divers model ensemble with 45 fixed model structures

- MARRMoT is a modular open-source toolbox containing documentation and model code based on 47 existing conceptual hydrologic models. (Knoben et al., 2019)
- MATLAB/Octave based

- 1 – 8 storages
- 1 – 18 parameters
- known models taken from literature (e.g. GR4J, HBV, HyMOD)

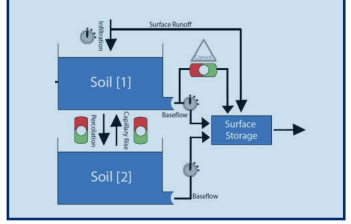
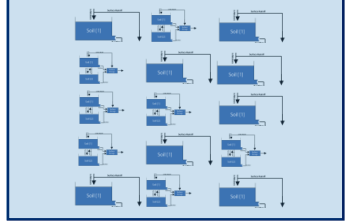
→ Results are structural benchmark for AMSI runs



Storages	1	2	3	4	5	6	7	8
n	5	11	6	5	11	5	1	1

n = number of model structures with corresponding storage number

Calibration Details:

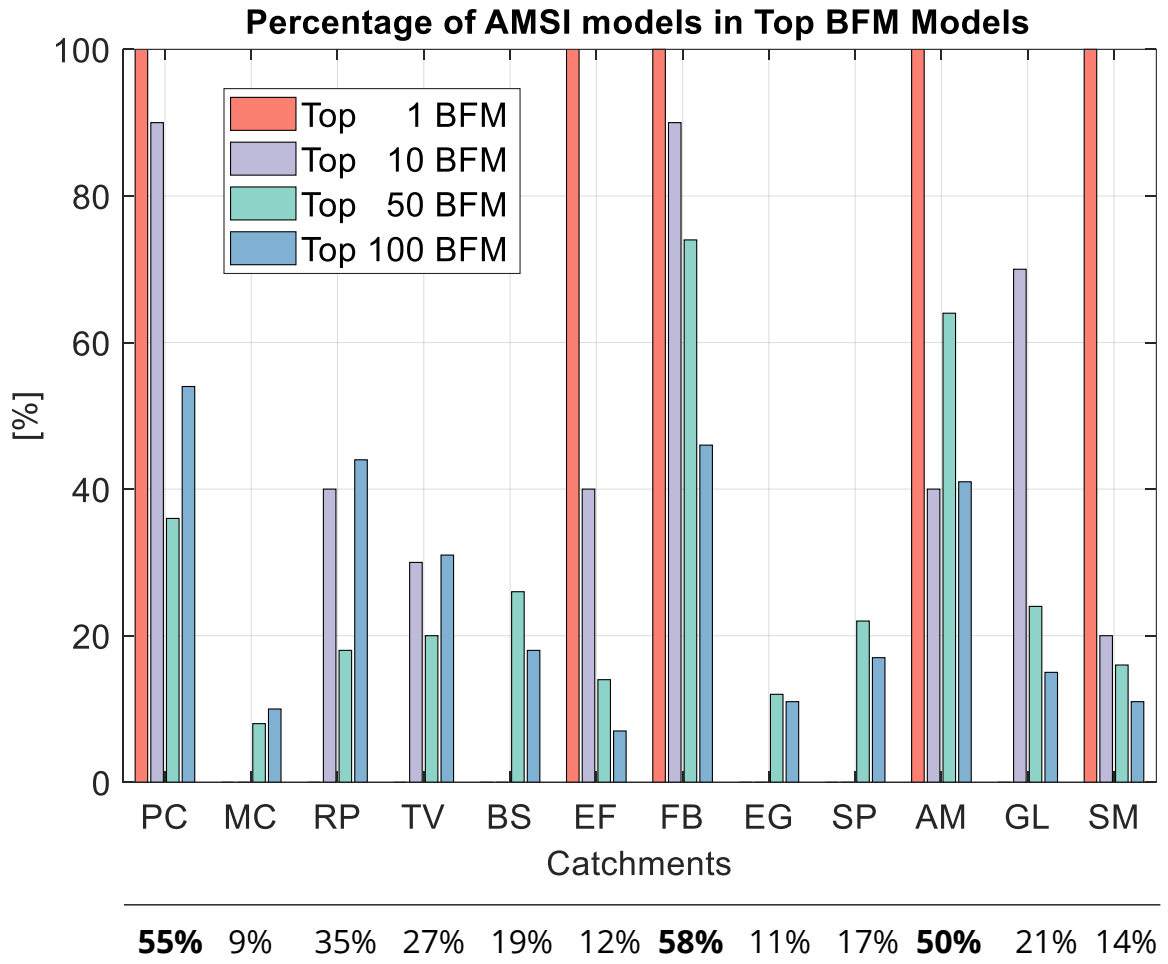
	AMSI	BFM	MARRMoT
			
Calibration runs	**100** (structure + parameter)	**1** (parameter)	
Optimizer	**DDS/** CMAES	**CMAES**	
Objective Fun.	**KGE**		
Budget	**max of 25.000 iterations**		
Period	**Calibration:** 1975 - 2000 **Evaluation:** 1950 - 1975		

Study Area:

- 12 MOPEX catchments (Duan et al., 2006)
- Hydro-climatically divers and unregulated

Do we find the same MS that are ranked as the “top” ones in BFM?

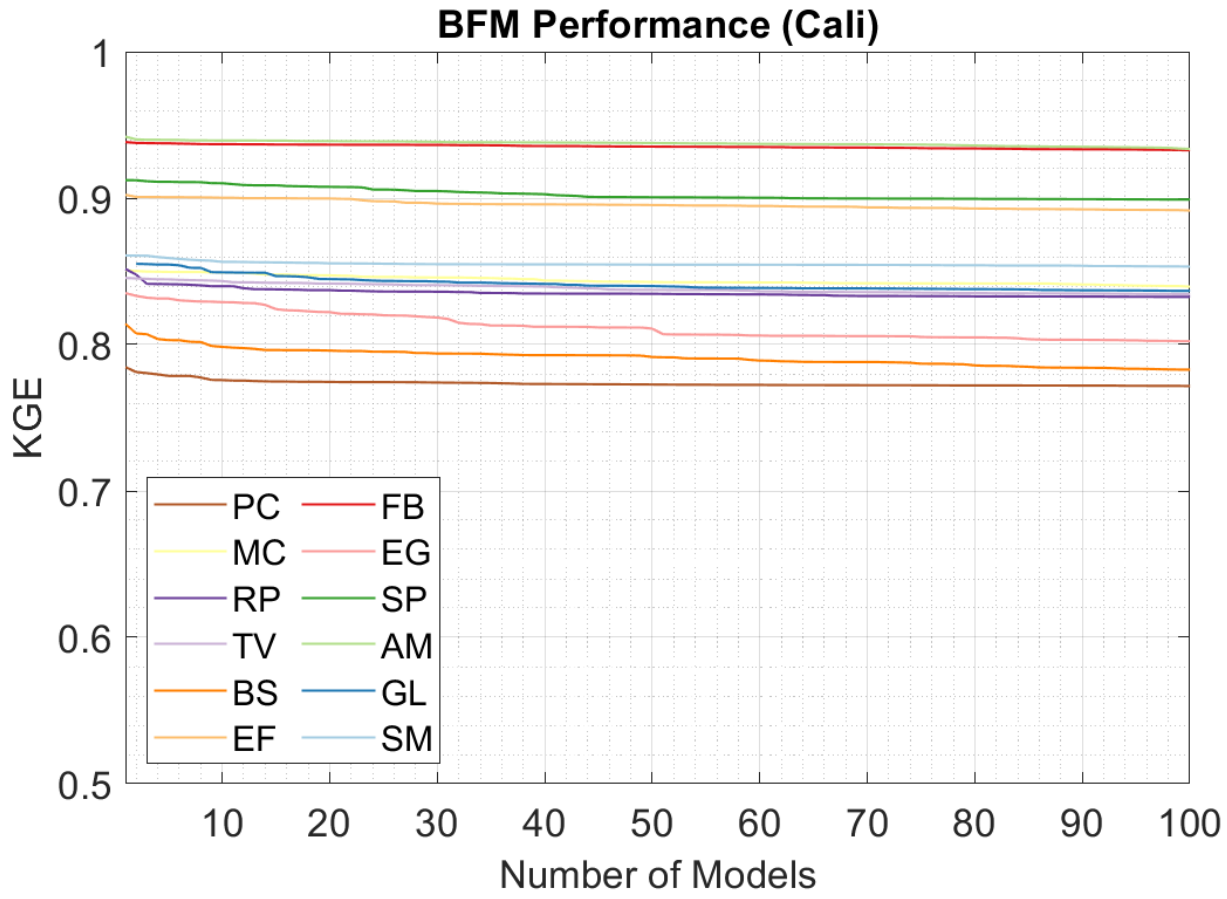
- Okay. This looks somewhat disappointing …
- But is it really?
- DDS was developed for the purpose of finding good global solutions (as opposed to globally optimal solutions)
- How different do the top 100 Models actually perform? So how easy would it be for DDS to find the “Top Model Structures”?



Do we find the same MS that are ranked as the “top” ones in BFM?

- How different do the top 100 Models actually perform?

Catchment	ΔKGE Top 100
PC	0.013
MC	0.011
RP	0.019
TV	0.011
BS	0.031
EF	0.011
FB	0.006
EG	0.033
SP	0.013
AM	0.008
GL	0.018
SM	0.008
Max ΔKGE	0.033

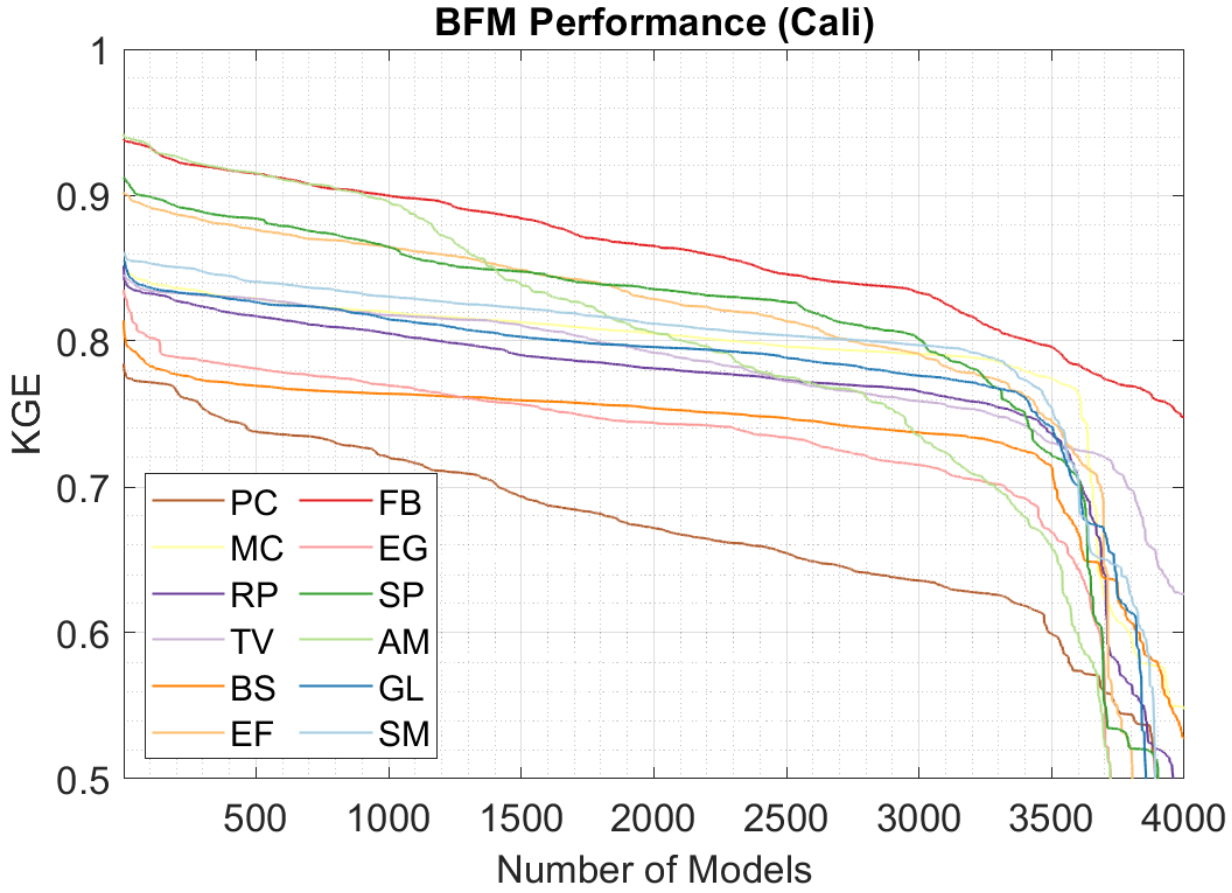


- Where does Equifinality start? ΔKGE = 0.01 / 0.05 / 0.10?

Do we find the same MS that are ranked as the "top" ones in BFM?

- How different do the top 100 Models actually perform?

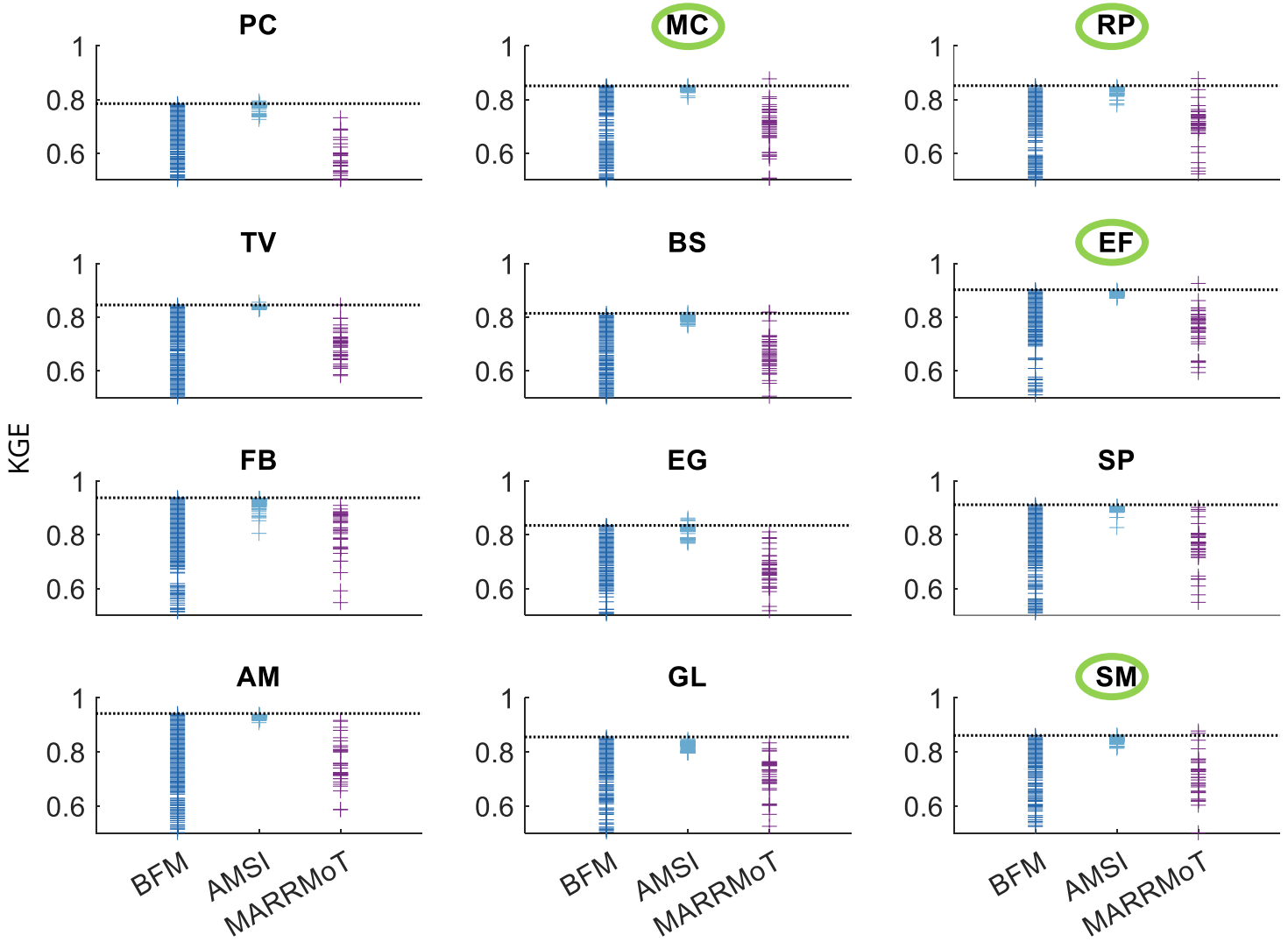
Catchment	ΔKGE Top 100	ΔKGE Top 1000	ΔKGE Top 3500
PC	0.013	0.064	0.185
MC	0.011	0.031	0.076
RP	0.019	0.047	0.115
TV	0.011	0.028	0.115
BS	0.031	0.050	0.101
EF	0.011	0.038	0.157
FB	0.006	0.039	0.142
EG	0.033	0.066	0.166
SP	0.013	0.048	0.191
AM	0.008	0.047	0.284
GL	0.018	0.040	0.114
SM	0.008	0.030	0.107
Max ΔKGE	0.033	0.066	0.284



- Where does Equifinality start? ΔKGE = 0.01 / 0.05 / 0.10?

How does AMSI compare to MARRMoT (Structural Benchmark)

- **AMSI** can compete with the structurally more divers **MARRMoT** in most catchments
- For **4** out of 12 catchments MARRMoT offers a better model structure
- For catchments MC, RP & EF that's always HBV with 5 storages and 15 parameters (according to MARRMoT) but we could also consider it a model with 3 soil storages
- For the arid SM it's a 2 storage structure with 8 parameters (m16 - newzealand2) that could also have been reproduced in AMSI, but might have a better parameter calibration



Conclusions

- We need more than an integral performance metric like KGE to discriminate between suitable model structures.
- 1 or 2 storage models perform well for the tested catchments. Only 3 out of 12 catchments benefit from one additional soil storages.

Questions

- Where does equifinality start?
- How can we identify the catchments that need more storages a priori? (e.g. P.C. David et al., 2022)
- If we consider MARRMoT to be the collected empirical knowledge on suitable conceptual model structures of the last decades. What have we actually learned?

Thank you for your attention!