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

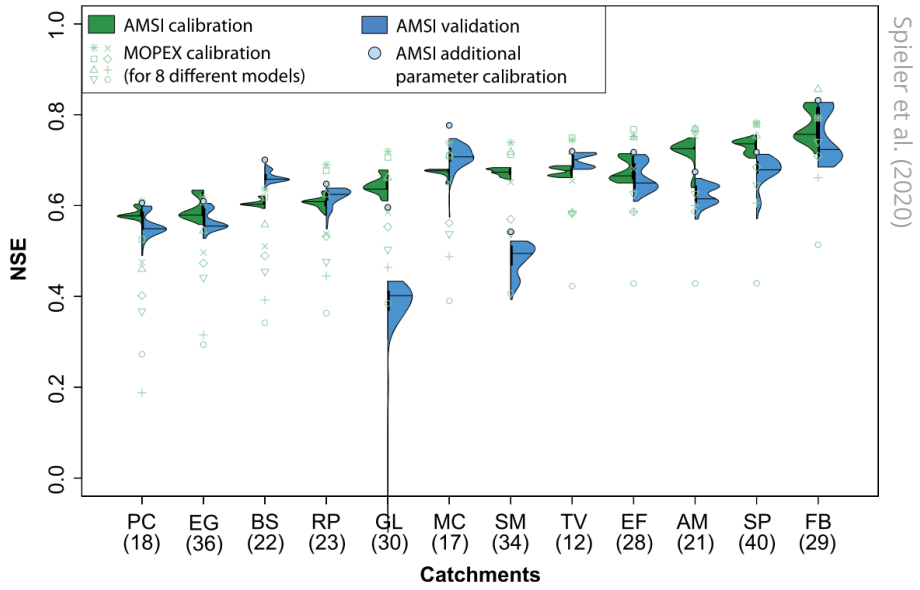
How good does Automatic Model Structure Identification work? A Benchmark Study with 6912 Model Structures.

Diana Spieler and Niels Schütze

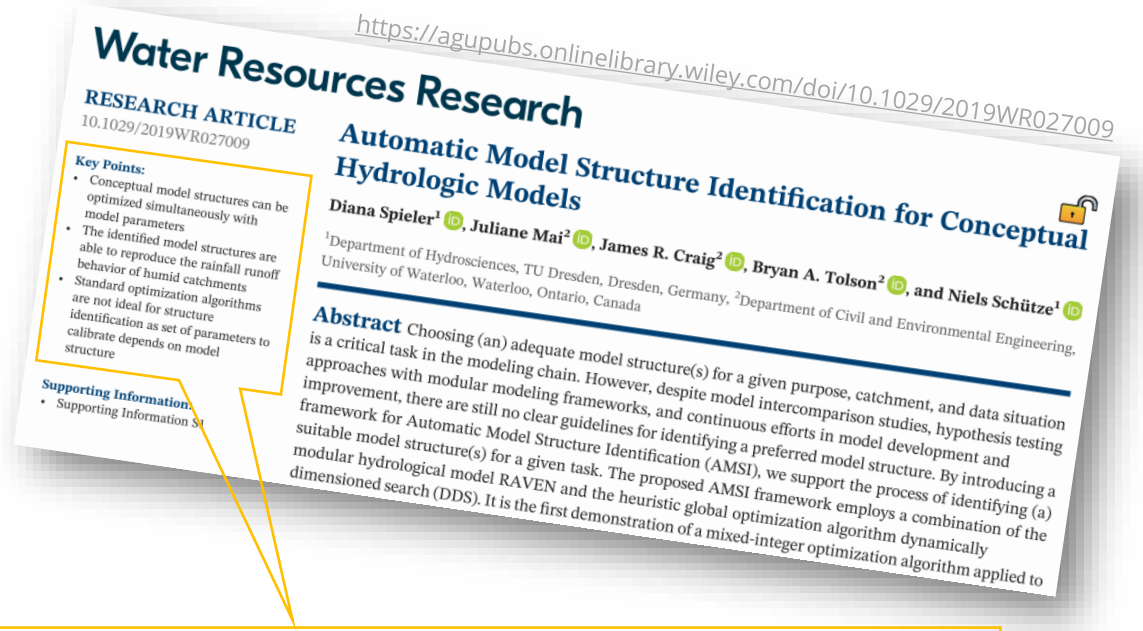
diana.spieler@tu-dresden.de

I previously developed a tool for Automatic Model Structure Identification (AMSI) allowing to calibrate conceptual model structures simultaneously with model parameters.

I tested it on 12 hydro-climatically differing MOPEX catchments (Duan,2006) and the identified model structures worked well.



Spieler et al. (2020)


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

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

Supporting Information:
- Supporting Information S1

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 framework for Automatic Model Structure Identification (AMSI), we support the process of identifying (a) suitable model structure(s) for a given task. The proposed AMSI framework employs a combination of the modular hydrological model RAVEN and the heuristic global optimization algorithm dynamically dimensioned search (DDS). It is the first demonstration of a mixed-integer optimization algorithm applied to

- Conceptual model structures can be optimized simultaneously with model parameters
- The identified model structures can 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

BUT – we don't know how well AMSI really works.

Potential problems with AMSI:

- AMSI is computationally challenging as different model structures may use a different number of parameters.
- Some parameters may be shared between model structures, others might be relevant for only a few structures.
- Shared model parameters might cause different effects in different model structures, causing their optimal values to differ across structures.

Research Questions

- **How do two "of the shelf" mixed-integer optimization algorithms perform, when having to handle these peculiarities during AMSI?**
- **Do we find the "best"* available model structure(s) out of a given model space?**

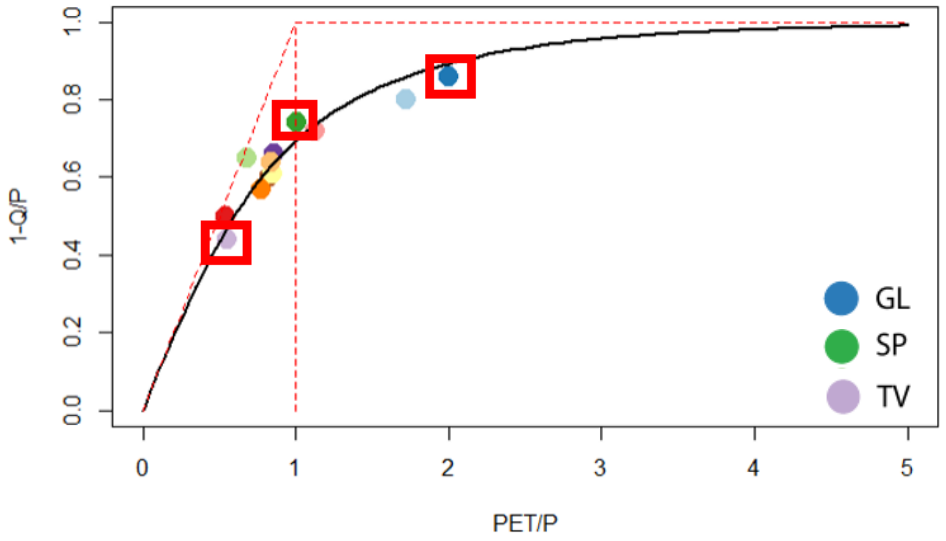
* "best" will be defined in KGE performance henceforth, even though the definition of (a) "best" model structure(s) is a question in itself

Experimental Design

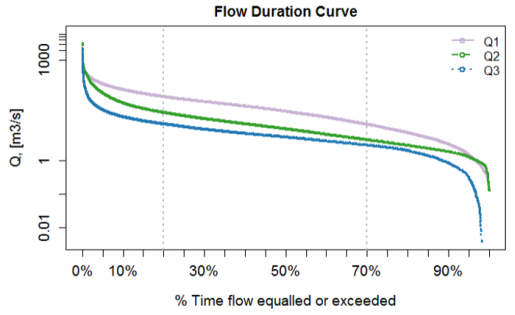
We employ two "of the shelf" mixed-integer optimization algorithms with AMSI.

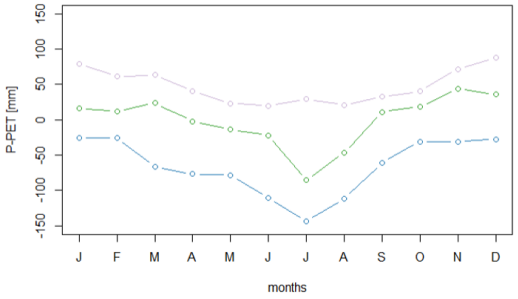
Dynamically Dimensioned Search Tolson et al. (2009)	DDS	CMAES	Covariance Matrix Adaptation Evolution Strategy Hansen et al. (2011)

And pick 3 catchments from the 12 previously used MOPEX catchments



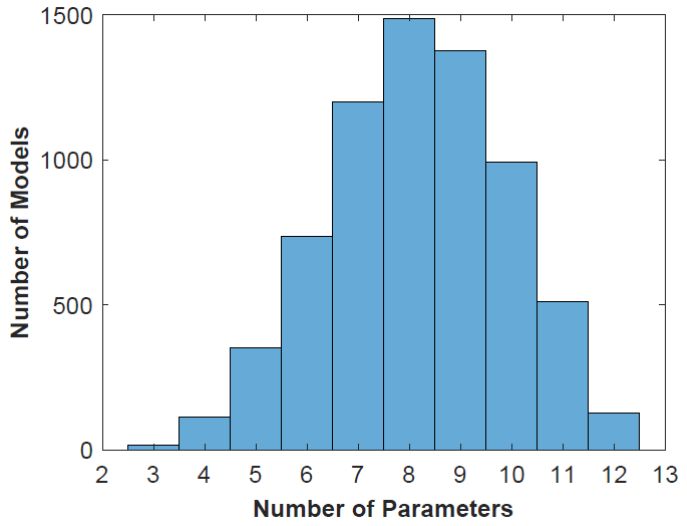
Label	Catchment Name	Size [km²]	Elevation [masl]	P [mm]	PET [mm]	Q [mm]	P/PET	Q/P	Forest Cover [%]
GL	Guadalupe River, TX	3406	289	767	1529	104	0.50	0.14	11
SP	Spring River, MI	3015	254	1082	1095	285	0.99	0.26	3
TV	Tygart Valley River, WV	2372	390	1301	711	729	1.83	0.56	78

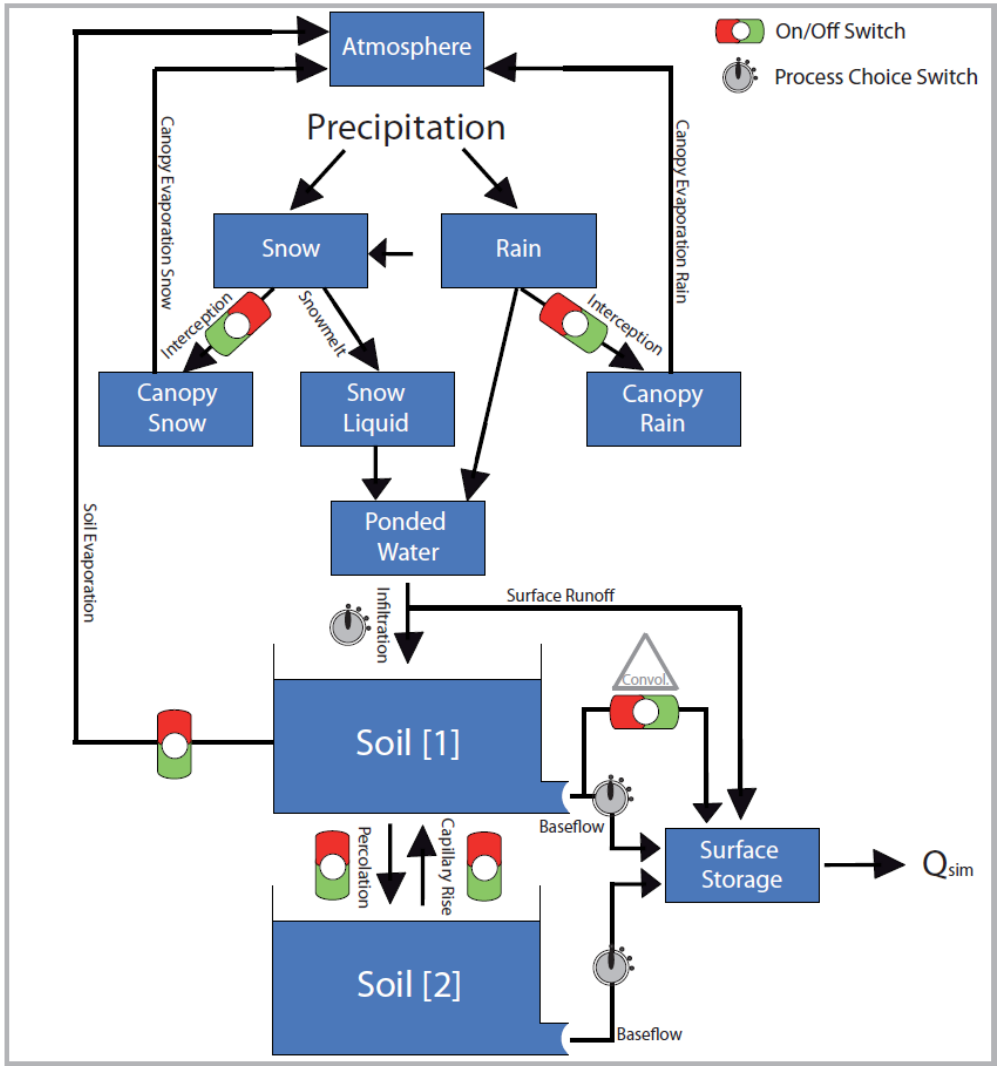




We created a maximum model space which allows for 13824 structure combinations

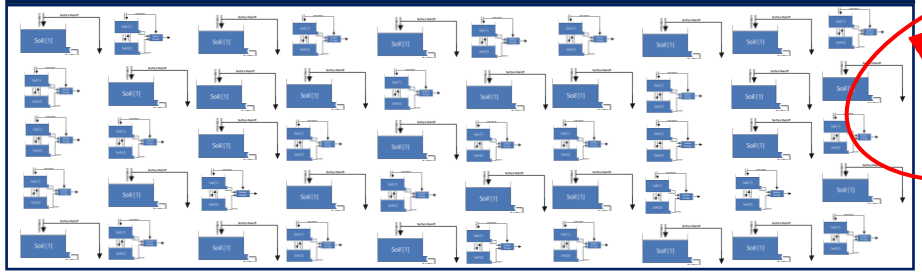
- The model space allows 1 or 2 soil storages
- 9 processes can be in- or excluded from the model structure
- 3 of 9 processes also have several process options
- A rain-snow routine is fixed
- Model structures have between 3 and 12 parameters depending on the included processes/process options





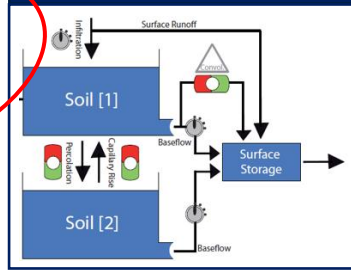
We compare the "Brute Force Calibration" of 6912 Model Structures with the AMSI approach

Parameter calibration of all possible model structures out of a given model space



VS

Automatic Model Structure Identification



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

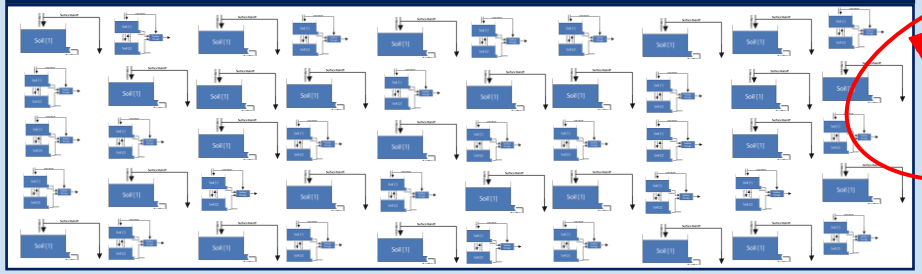
But why 6912 instead of 13824??

- Some possible combinations within the chosen model space might not make much sense
  - E.g., 960 times the lower soil storage may be active but doesn't allow any outflow
- Similarly, percolation into the lower soil storage is turned off 6912 times (making all combinations connected to the 2nd soil storage practically useless)

  → these combinations are excluded for the brute force calibration but remain possible during the AMSI calibration

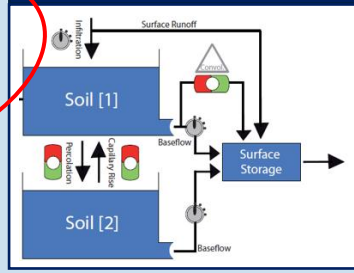
We compare the "Brute Force Calibration" of 6912 Model Structures with the AMSI approach

Brute Force Calibration
(6912 combinations)



VS

Automatic Model Structure Identification
(13824 possibilities)



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

Identical Model Space with 13824 possible combinations

- 1 standard parameter calibrations for 6912 models
- Performed for 3 catchments
- Calibration: CMAES, KGE, 1975-2000
- Max. budget of 25.000 iterations
- Validation: 1950-1975
- 3 to 12 parameters depending on model structure

→ **Results are Benchmark for AMSI runs**

- 100 AMSI runs (multiple starts)
- Performed for 3 catchments
- Calibration: CMAES+DDS, KGE, 1975-2000
- Max. budget of 25.000 iterations
- Validation: 1950-1975
- 3 to 12 parameters depending on model structure
  - **BUT** 29 parameters are constantly calibrated for AMSI (9 integer parameters for structural choices + 20 continuous parameters for potentially necessary process parameters)

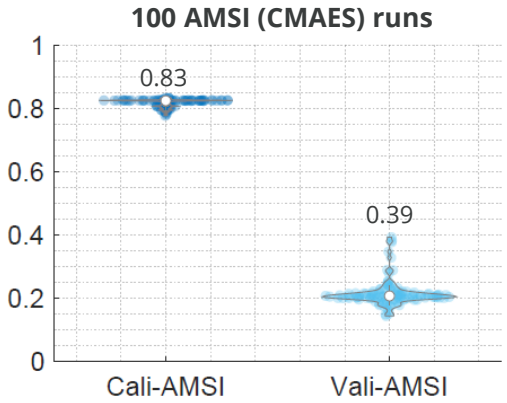
Calibration and Validation Results for AMSI approaches and "Brute Force Calibration"

Guadalupe Catchment

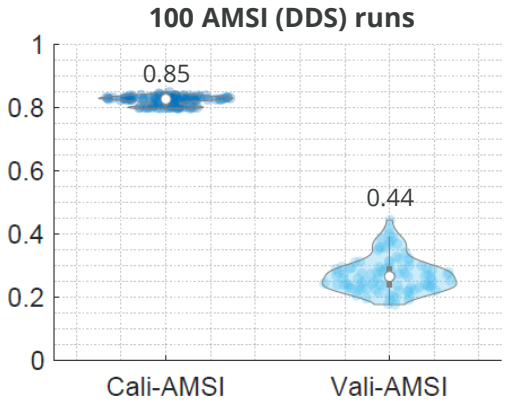
arid catchment = challenging example

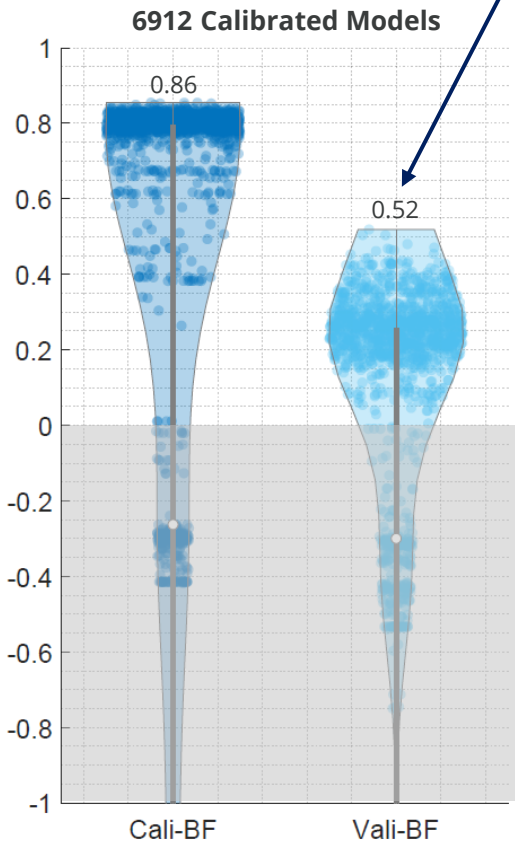
Max KGE values

Cali/Vali results for 100 AMSI runs with CMAES



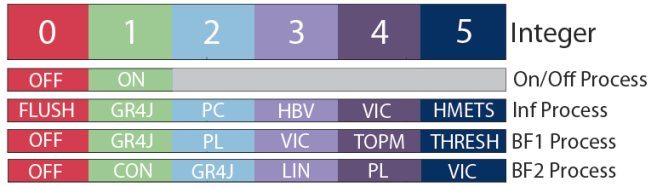
Cali/Vali results for 100 AMSI runs with DDS





Cali/Vali results for 6912 standard calibrations

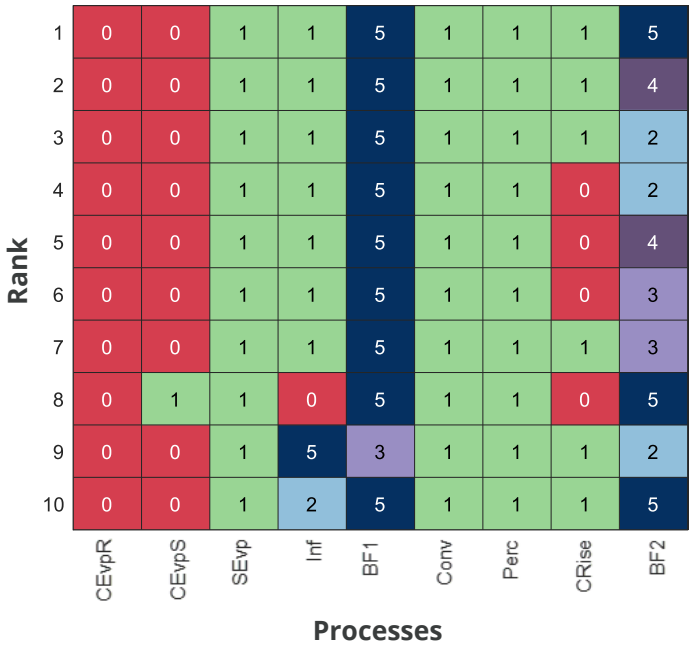
Structural Choices of the 10 "Best" Model Structures in Calibration



Guadalupe Catchment

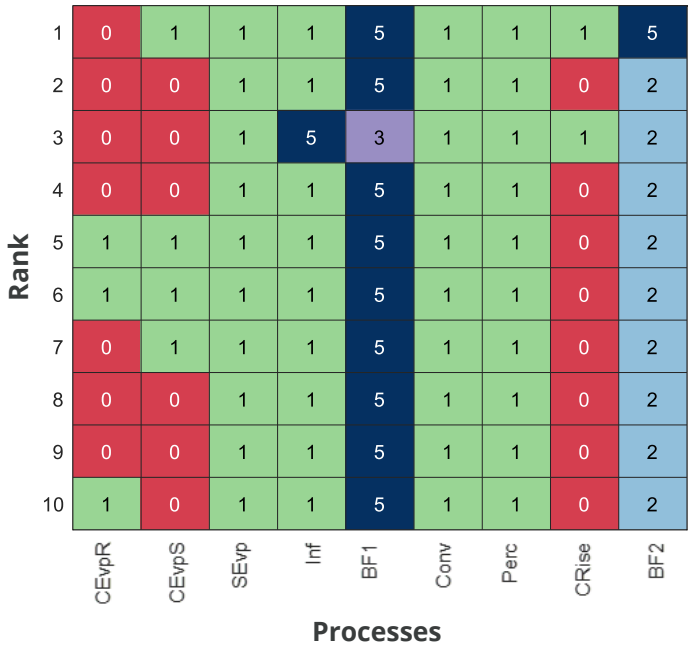
"Brute Force" (Benchmark)

10 "best" out of 6912 model structures



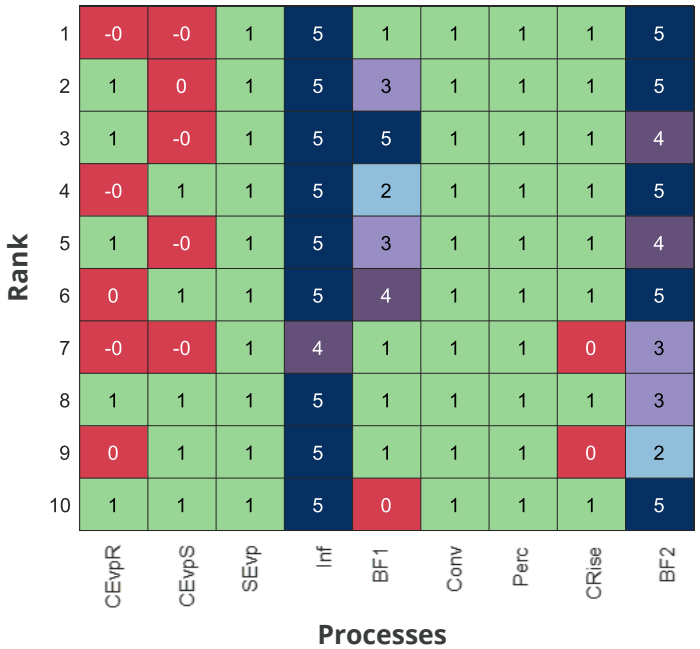
AMSI - DDS

10 "best" model structures identified with AMSI-DDS

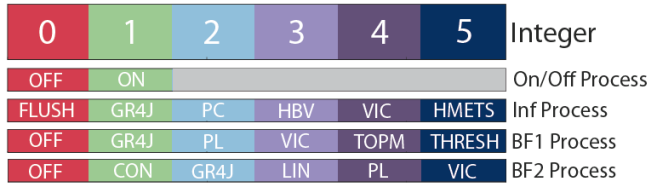


AMSI - CMAES

10 "best" model structures identified with AMSI-CMAES



Do the same models perform well in Validation?



Guadalupe Catchment

"Brute Force"
Calibration

"Brute Force"
Validation

- During calibration it seems the process description for the baseflow of the lower soil storage (BF2) can be chosen arbitrarily and there is a clear favorite for the flow out of the upper soil storage (BF1) and Infiltration (Inf)

10 "best" out of 6912 model structures

	CEvpR	CEvpS	SEvp	Inf	BF1	Conv	Perc	CRise	BF2
1	0	0	1	1	5	1	1	1	5
2	0	0	1	1	5	1	1	1	4
3	0	0	1	1	5	1	1	1	2
4	0	0	1	1	5	1	1	0	2
5	0	0	1	1	5	1	1	0	4
6	0	0	1	1	5	1	1	0	3
7	0	0	1	1	5	1	1	1	3
8	0	1	1	0	5	1	1	0	5
9	0	0	1	5	3	1	1	1	2
10	0	0	1	2	5	1	1	1	5

10 "best" out of 6912 model structures

	CEvpR	CEvpS	SEvp	Inf	BF1	Conv	Perc	CRise	BF2
1	0	1	1	0	1	1	1	1	2
2	0	0	1	0	3	0	1	1	2
3	0	1	1	0	3	0	1	1	2
4	0	1	1	0	4	1	1	1	2
5	0	1	1	3	2	1	1	1	2
6	0	0	1	5	3	1	1	1	2
7	0	1	1	3	3	1	1	1	2
8	0	0	1	4	2	1	1	1	2
9	0	0	1	2	3	1	1	1	2
10	0	0	1	0	2	0	1	1	2

- Validation shows that models using a GR4J-like baseflow approach* in the lower soil storage and no Threshold based approach* in the upper soil layer tend to perform better.

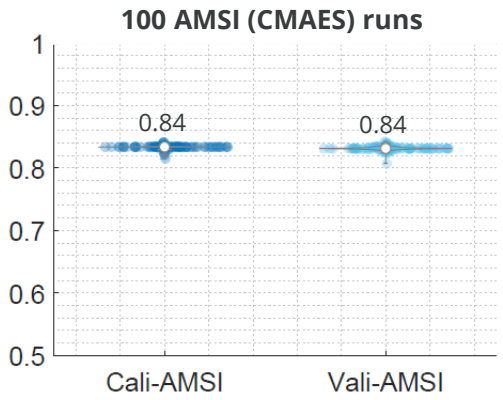
* Please refer to the RAVEN documentation for more information on the different process algorithms → RAVEN

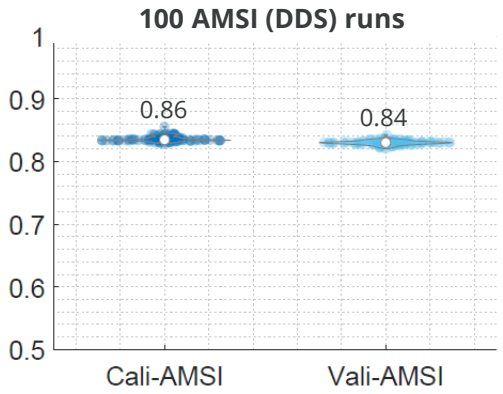
What about the results for the other 2 catchments?

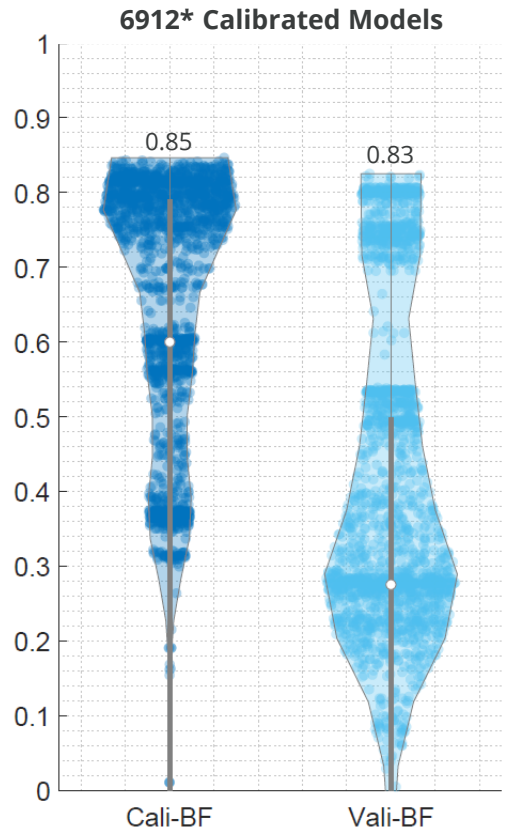
I made a stupid mistake and then the supercomputer and I hit a rough patch in our relationship

Tygart Valley Catchment (preliminary)

- Unfortunately, not all results were available as planned.
- Thus, no reliable conclusion about the performance of AMSI can be drawn just yet.







*1230/6912 structures evaluated

Mind the different axes limits

What do we know (so far)

- Be very careful when handling files that are the foundation of your currently running ~20.000 calibration jobs.
- Use open-source software – license problems are a pain.
- Okay, okay but what about the research questions?
    - **How do two "of the shelf" mixed-integer optimization algorithms perform, when having to handle these peculiarities during AMSI?**
        - Better than (I) expected.
        - However, CMAES has a variance-based optimization strategy which does not seem to be quite as suitable for structural calibration
    - **Do we find the "best"* available model structure(s) out of a given model space?**
        - We get very close in finding the same structures that perform well in calibration; at least for DDS it seems to be a matter of parameter fine tuning
        - However, this does not necessarily imply those same structures work well in validation.

⇨ In order to pursue AMSI a lot of thought needs to be put into the calibration setup (objective function, optimization algorithm, multi-criteria calibration, several datasets, etc.)

* "best" will be defined in KGE performance henceforth, even though the definition of (a) "best" model structure(s) is a question in itself

Thank you for your interest in my work!

Feel free to reach out in case of any questions or remarks!

diana.spieler@tu-dresden.de