W. Knoben[1], R. Woods[1], J. Freer[2], K. Fowler[3], M. Peel[3]
[1] Department of Civil Engineering, University of Bristol, UK
[2] School of Geographical Sciences, University of Bristol, UK
[3] Department of Infrastructure Engineering, University of Melbourne, AUS
w.j.m.knoben@bristol.ac.uk

What can we learn from applying 40+ models to 559 catchments?

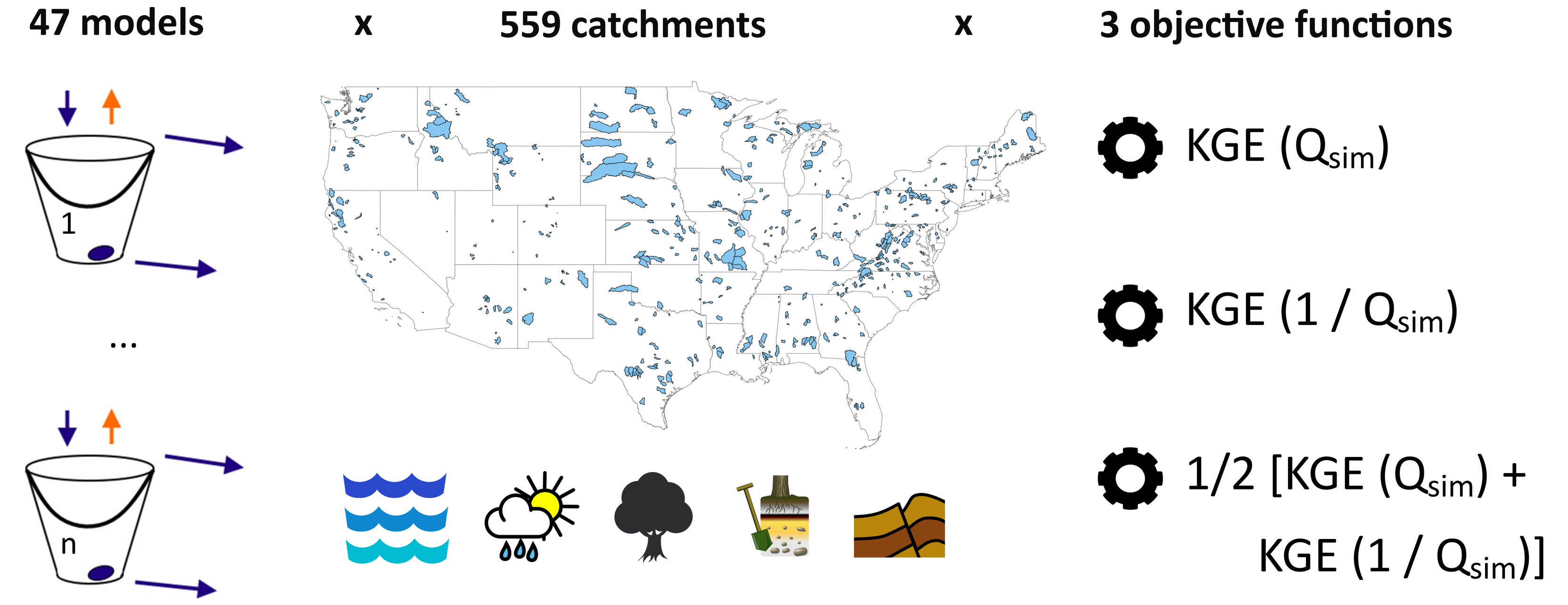
Relating differences in conceptual model accuracy to climatic forcing, catchment characteristics and model structure

EGU2018-3838

How do we choose an appropriate conceptual model structure for ungauged catchments? We need to formalize the model structure-catchment-performance relationship

Large-sample studies allow us to test hypotheses and come to conclusions that are less conditional on the choice of study catchment. This allows generalization of findings. The increase in breadth necessarily requires a sacrifice in depth[1], and there are thus trade-offs to be made:

- **Summary metrics** are necessary to make sense of the results
- Individual cases can not be examined in any great **detail**
- Experimental setup must make **concessions** to keep computational time feasible

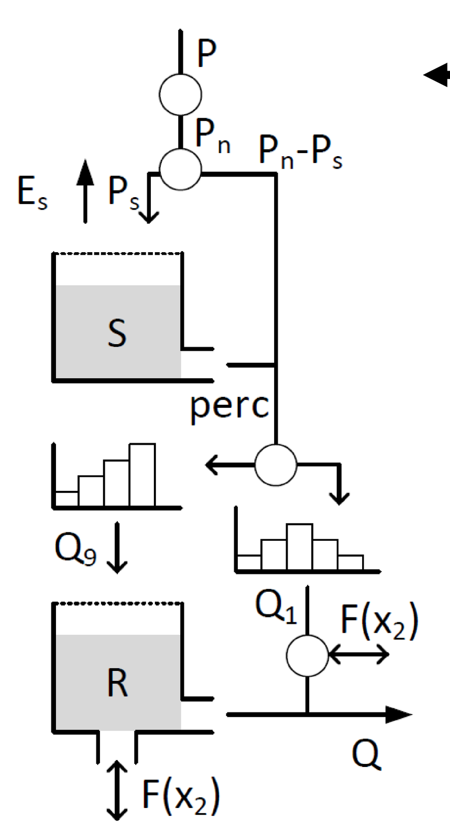


Models, at least in theory, represent a different selection of catchment processes

Catchments[2] cover a wide range of climates and are provided with catchment attributes covering climate, flow patterns, topography, land cover, soil types and geology

Objective functions show models' ability to simulate high / low / both types of flows

We show results for two 4-parameter models



Model 1[3]

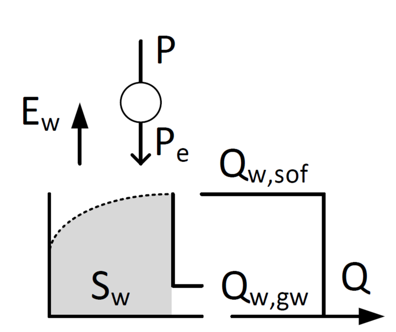
- Stylized representation of processes

Soil moisture maximum depth [mm]	[1,2000]
Water exchange [mm/d]	[-8,14]
Routing store depth [mm]	[1,300]
Flow delay parameter [d]	[0.5,5]

Model 2[4]

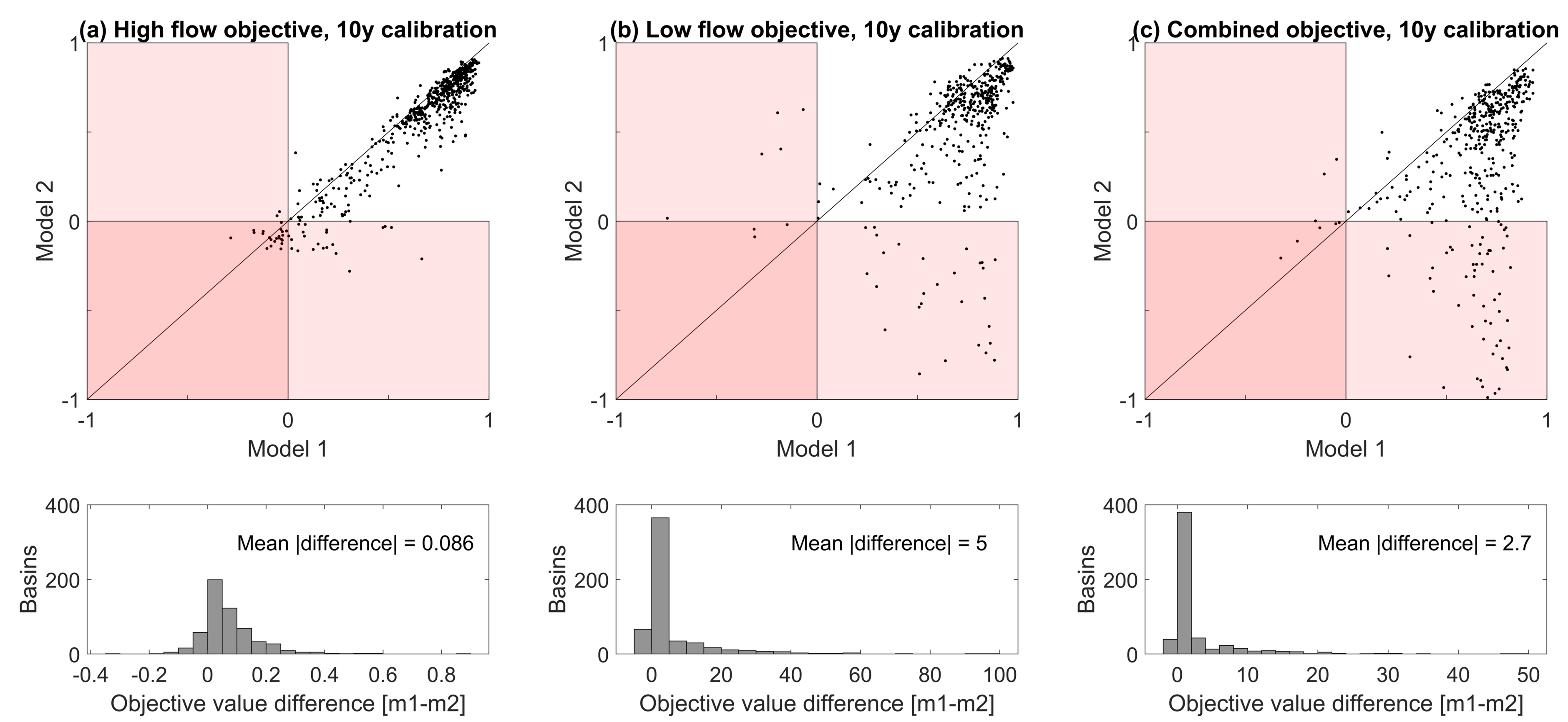
- Represents wetland processes, strong focus on saturation excess flow

Interception depth [mm]	[0,5]
Soil moisture distribution shape [-]	[0,10]
Soil moisture maximum depth [mm]	[1,2000]
Base flow time parameter [d-1]	[0,1]

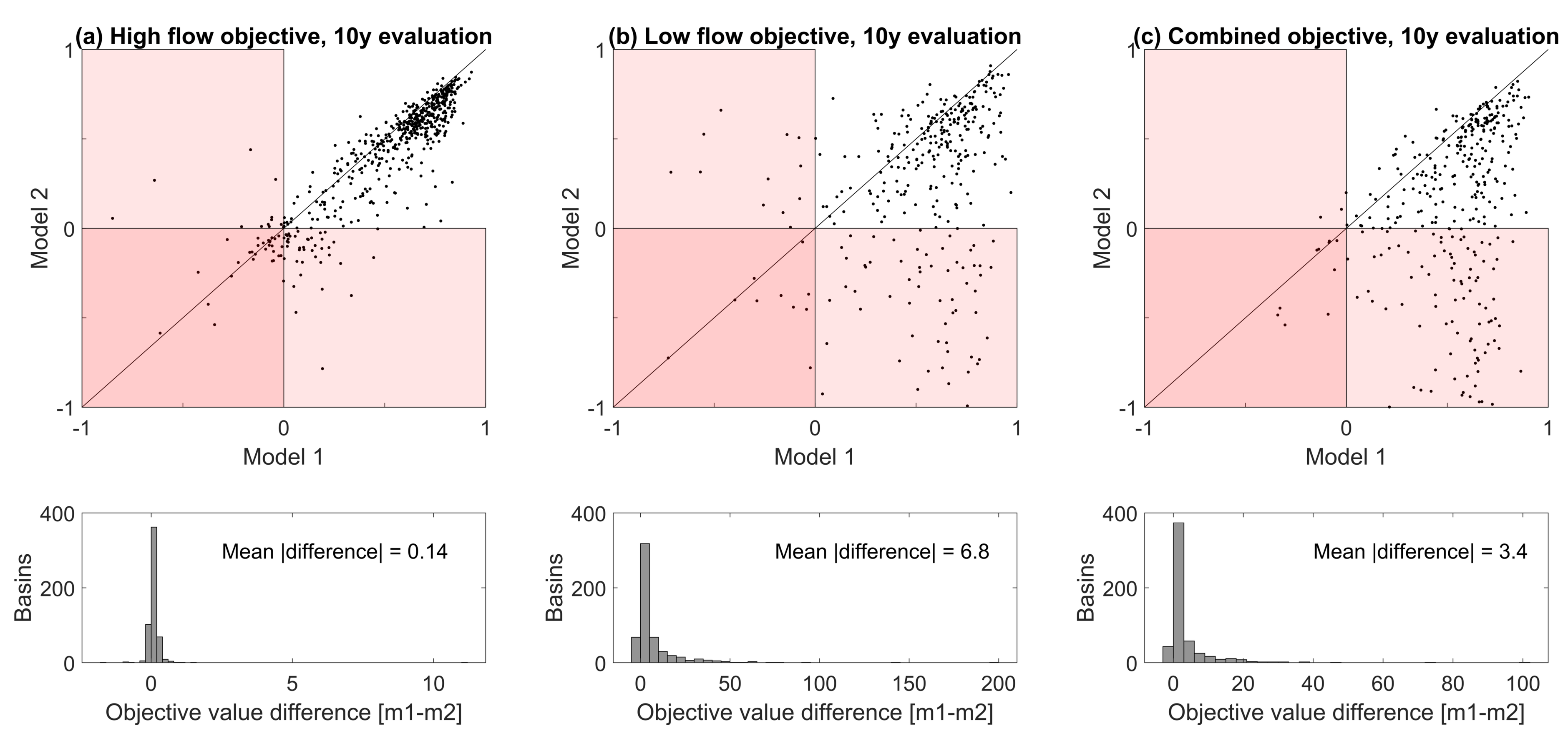


Neither model outperforms the other consistently, but model structure uncertainty is significantly larger when low flow simulation is included in the objective function (fig. 1, 2)

- Model structure uncertainty is defined as the difference between 'best' objective function values
- Model 2 struggles during low flow simulation, possibly because it is essentially a linear reservoir in these cases. Model 1 in contrast has high flexibility in flow delay and water import/export
- Still, there are cases where model 2 outperforms model 1 during evaluation (top-left quadrants in fig. 2a, b, c)
- ***Future goals:*** *quantify model structure uncertainty based on a larger sample of models and determine whether certain models are better at reproducing certain flow types*



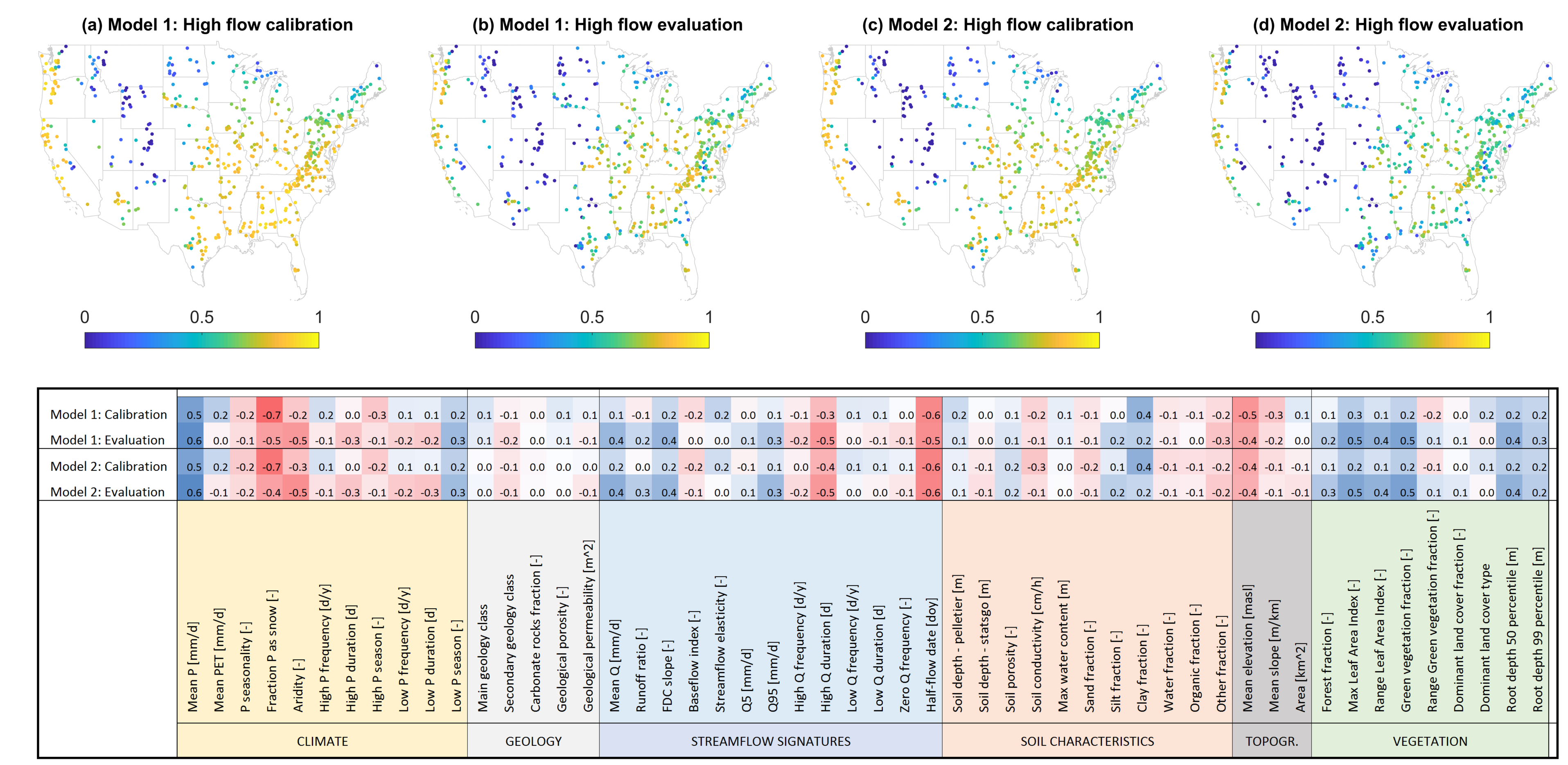
(fig. 1): Model structure uncertainty during high flow (a), low flow (b) and combined flow (c) calibration. Differences in objective function value during calibration are solely due to differences between model structures. Red shaded areas show regions in the objective space where performance is bad (KGE < 0) for one of the models (lighter shading) or both (darker shading). Histograms quantify the model structure uncertainty in terms of objective function values.



(fig. 2): Model structure uncertainty during high flow (a), low flow (b) and combined flow (c) evaluation.

Model performance is subject to strong spatial organization, which correlates more with climatic conditions than with catchment attributes (fig. 3)

- Model performance tends to be better in less arid catchments, in line with findings from earlier work
- Model performance is worse for catchments with more snow; neither model has a snow module
- Geological, topographical and most soil characteristics seem unrelated to model performance
- However, catchments with higher clay fraction result in better calibration performance. Why?
- More vegetation leads to better evaluation performance. Related to aridity perhaps?
- ***Future goal:*** *formalize the relationships between model performance and catchment attributes*

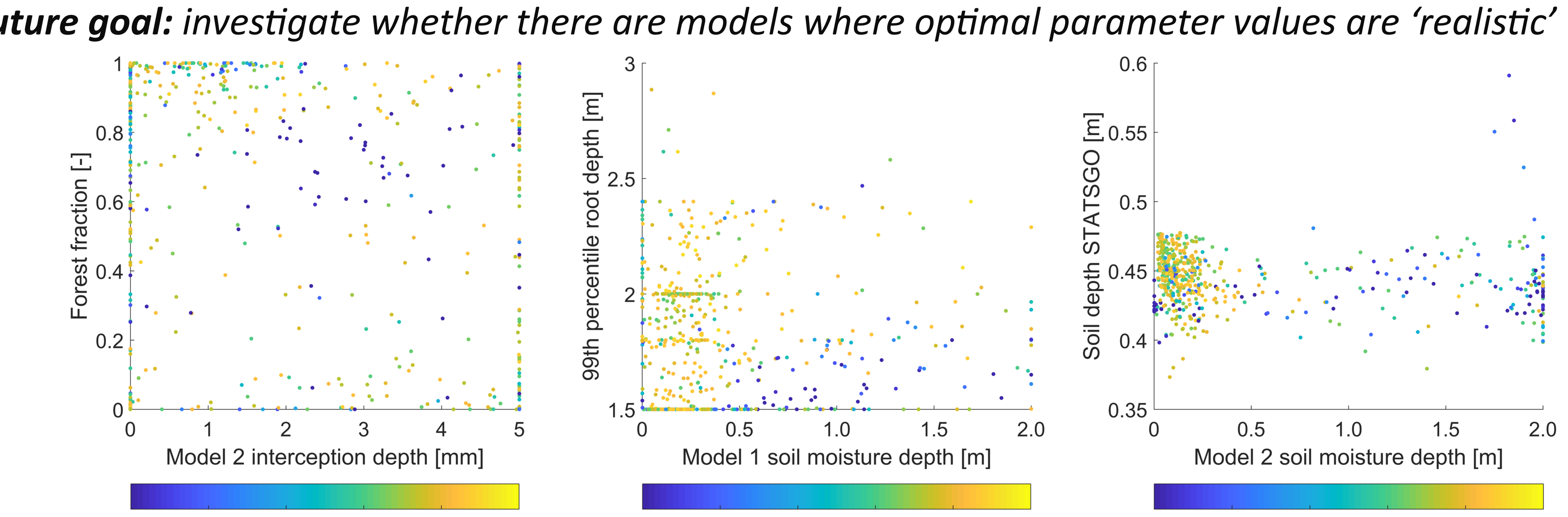


(fig. 3): [Top row] Spatial organization of calibration (a, c) and evaluation (b, d) performance for models 1 (a, b) and 2 (c, d). [Bottom row]: Spearman rank correlation between model performance and catchment descriptors

Calibrated parameter values lack realism. There is no relationship between 'optimal' parameter values and catchment attributes (fig. 4)

This statement holds for other catchment attributes not shown here.

Future goal: *investigate whether there are models where optimal parameter values are 'realistic'*



(fig. 4): Comparison of catchment attributes and parameter values calibrated for high flow simulation. The colour scheme shows objective function value for the parameter set [KGE(Q), high flows]

[1] Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. Hydrology and Earth System Sciences, 18(2), 463–477. https://doi.org/10.5194/hess-18-463-2014

[2] Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrology and Earth System Sciences, 21, 5293–5313. https://doi.org/10.5194/hess-2017-169

[3] Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology, 279(1–4), 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7

[4] Savenije, H. H. G. (2010). "Topography driven conceptual modelling (FLEX-Topo)." Hydrology and Earth System Sciences, 14(12), 2681–2692. https://doi.org/10.5194/hess-14-2681-2010

Acknowledgements: W. Knoben is supported by WISE CDT (EPSRC grant EP/L016214/1).