Geophysical Research Abstracts
Vol. 21, EGU2019-1208, 2019
EGU General Assembly 2019

Model structure differences influence performance more than number of parameters: findings from a 36-model, 559-catchment comparison study

Wouter Knoben (1), Jim Freer (2), and Ross Woods (1)

(1) University of Bristol, Civil Engineering, Bristol, United Kingdom (w.j.m.knoben@bris.ac.uk), (2) University of Bristol, School of Geographical Sciences, Bristol, United Kingdom

We present results from the first application of the new MARRMoT modelling toolbox in comparative analyses of 36 unique model structures across 559 USA catchments. The research aim is to quantify and understand the ability of these 36 models to simulate streamflow in a wide range of catchments. We answer several questions: is the range of observed model performance the same across all catchments and for different objective functions? Are certain model structures more universally successful than others? Are there unique features that enable some models to succeed?

The Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) implements model structures in a unified framework, ensuring consistency in model setup. Catchments are chosen from the CAMELS data set. Observations for each catchment are separated into a 10-year calibration period (1989-1998) and a 10-year evaluation period (1999-2009). The CMA-ES algorithm optimises each model for each catchment and three objective functions: the Kling-Gupta Efficiency is used in different formulations to assess high, low and combined flow model performance. We quantify model structure uncertainty for each catchment as the difference between the best performing model and the 10th percentile model. If this difference is small, model structure equifinality and uncertainty are high, because many models perform equally well.

For most catchments, at least one model can achieve "acceptable" efficiency values (KGE > 0.6 for 81.8% [high flows], 77.6% [low flows] and 66.4% [combined flows] of catchments). The extent of model structure uncertainty changes along clear climatic gradients. Uncertainty decreases as a larger fraction of annual precipitation occurs as snow, because some models lack a snow module. On the high flow metric, structural equifinality is largest in catchments with an average annual water surplus and a strongly seasonal aridity cycle (most models do well in these catchments). Uncertainty decreases (fewer models do well) in drier catchments and in places with a less pronounced seasonal aridity cycle. With the low flow metric, this pattern is – surprisingly – reversed, and structural equifinality is larger in arid catchments than in seasonally wet places.

The number of free parameters is expected to explain some inter-model differences, but for none of the three metrics, neither for calibration nor evaluation efficiency values, nor in the performance change between periods, does number of parameters correlate with model performance. However, certain model structure elements might explain inter-model differences. For example, some structures include Unit Hydrograph routing as part of the soil moisture routine. These models rank consistently in the bottom third during high flow evaluation and consistently in the top half during low flow evaluation.

Summarising, model structure uncertainty evolves along climatic gradients but with different patterns depending on the objective function. The performance of any model seems less related to its number of parameters and more to its structure. Large-sample model evaluation such as this can highlight models' strengths and assist a modeller in choosing the right one for a specific purpose. This work also indicates an urgent need for catchment-specific benchmarks, because comparing achieved efficiency values between catchments is fraught with difficulty.