EGU2020-3205, updated on 17 Oct 2021
https://doi.org/10.5194/egusphere-egu2020-3205
EGU General Assembly 2020

On the Robustness of Conceptual Rainfall-Runoff Models to Calibration and Evaluation Dataset Splits Selection: A Large Sample Investigation

Feifei Zheng[1], **Danlu Guo**[1,2], Hoshin Gupta[3], and Holger Maier[1,4]

[1]College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, Zhejiang, China (feifeizheng@zju.edu.cn)
[2]Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC Australia (danlu.guo@unimelb.edu.au)
[3]Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA (hoshin@email.arizona.edu)
[4]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA Australia (holger.maier@adelaide.edu.au)

Conceptual Rainfall-Runoff (CRR) models are widely used for runoff simulation, and for prediction under a changing climate. The models are often calibrated with only a portion of all available data at a location, and then evaluated independently with another part of the data for reliability assessment. Previous studies report a persistent decrease in CRR model performance when applying the calibrated model to the evaluation data. However, there remains a lack of comprehensive understanding about the nature of this '*low transferability*' problem and why it occurs. In this study we employ a large sample approach to investigate the robustness of CRR models across calibration/validation data splits. Specially, we investigate: 1) how robust is CRR model performance across calibration/evaluation data splits, at catchments with a wide range of hydro-climatic conditions; and 2) is the robustness of model performance somehow related to the hydro-geo-climatic characteristics of a catchment? We apply three widely used CRR models, GR4J, AWBM and IHACRE_CMD, to 163 Australian catchments having long-term historical data. Each model was calibrated and evaluated at each catchment, using a large number of data splits, resulting in a total of 929,160 calibrated models. Results show that: 1) model performance generally exhibits poor robustness across calibration/evaluation data splits; 2) lower model robustness is correlated with specific catchment characteristics, such as a higher runoff skewness, lower aridity and runoff coefficient. These results provide a valuable benchmark for future model robustness assessments, and useful guidance for model calibration and evaluation.