Geophysical Research Abstracts
Vol. 18, EGU2016-5032-1, 2016
EGU General Assembly 2016

Looking beyond general metrics for model evaluation - lessons from an international model intercomparison study

Laurène Bouaziz (1), Tanja de Boer-Euser (2), Claudia Brauer (3), Gilles Drogue (4), Fabrizio Fenicia (5), Benjamin Grelier (4), Jan de Niel (6), Jiri Nossent (7,8), Fernando Pereira (7), Hubert Savenije (2), Guillaume Thirel (9), and Patrick Willems (6)

(1) Deltares, Hydrology, Delft, Netherlands (Laurene.Bouaziz@deltares.nl), (2) Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, P.O. Box 5048, NL-2600 GA Delft, The Netherlands, (3) Hydrology and Quantitative Water Management Group, Wageningen University, The Netherlands, (4) Laboratoire LOTERR, Université de Lorraine, Metz, France, (5) Eawag, Dübendorf, Switzerland, (6) Hydraulics division, Department of Civil Engineering, KU Leuven, Kasteelpark Arenberg 40, BE-3001 Leuven, Belgium, (7) Flanders Hydraulics Research, Antwerp, Belgium, (8) Vrije Universiteit Brussel (VUB), Department of Hydrology and Hydraulic Engineering, Brussel, Belgium, (9) Irstea, Hydrology Research Group, Antony, France

International collaboration between institutes and universities is a promising way to reach consensus on hydrological model development. Education, experience and expert knowledge of the hydrological community have resulted in the development of a great variety of model concepts, calibration methods and analysis techniques. Although comparison studies are very valuable for international cooperation, they do often not lead to very clear new insights regarding the relevance of the modelled processes. We hypothesise that this is partly caused by model complexity and the used comparison methods, which focus on a good overall performance instead of focusing on specific events. We propose an approach that focuses on the evaluation of specific events. Eight international research groups calibrated their model for the Ourthe catchment in Belgium (1607 km2) and carried out a validation in time for the Ourthe (i.e. on two different periods, one of them on a blind mode for the modellers) and a validation in space for nested and neighbouring catchments of the Meuse in a completely blind mode. For each model, the same protocol was followed and an ensemble of best performing parameter sets was selected. Signatures were first used to assess model performances in the different catchments during validation. Comparison of the models was then followed by evaluation of selected events, which include: low flows, high flows and the transition from low to high flows. While the models show rather similar performances based on general metrics (i.e. Nash-Sutcliffe Efficiency), clear differences can be observed for specific events. While most models are able to simulate high flows well, large differences are observed during low flows and in the ability to capture the first peaks after drier months. The transferability of model parameters to neighbouring and nested catchments is assessed as an additional measure in the model evaluation. This suggested approach helps to select, among competing model alternatives, the most suitable model for a specific purpose.