



## **Global evaluation of runoff from ten state-of-the-art hydrological models**

Hylke Beck (1), Ad de Roo (1), Albert van Dijk (2), Jaap Schellekens (3), Emanuel Dutra (4), Gabriel Fink (5), and Rene Orth (6)

(1) European Commission, Joint Research Centre (JRC), Ispra, Italy (hylke.beck@jrc.ec.europa.eu), (2) Fenner School of Environment & Society, Australian National University, Canberra, Australia, (3) Inland Water Systems Unit, Deltares, Delft, The Netherlands, (4) European Centre for Medium-Range Weather Forecasts, Redding, UK, (5) Center for Environmental Systems Research (CESR), University of Kassel, Germany, (6) Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland

Observed streamflow data from 966 medium sized catchments (1000 to 5000 km<sup>2</sup>) around the globe were used to comprehensively evaluate the daily runoff estimates (1979–2012) of six global hydrological models (GHMs) and four land surface models (LSMs) produced as part of Tier-1 of the earthH2Observe project. The models were all driven by the WATCH Forcing Data ERA-Interim (WFDEI) meteorological dataset, but used different datasets for non-meteorologic inputs and were run at various spatial and temporal resolutions, although all data were re-sampled to a common 0.5° spatial and daily temporal resolution. For the evaluation, we used a broad range of performance metrics related to important aspects of the hydrograph. We found pronounced inter-model performance differences, underscoring the importance of hydrological model uncertainty in addition to climate input uncertainty, for example in studies assessing the hydrological impacts of climate change. The (uncalibrated) GHMs were found to perform better than the LSMs in snow-dominated regions, and the ensemble mean was found to perform only slightly worse than the best (calibrated) model. The models generally showed an early bias in the spring snowmelt peak. We further found that, despite adjustments using gauge observations, the WFDEI precipitation data still contain substantial biases which propagate in the simulated runoff. Overall, more effort should be devoted to calibrating and regionalizing the parameters of macro-scale models.