



## **Two-step calibration and proxy-basin validation of ensemble rainfall-runoff predictions in a Swedish mesoscale catchment**

Jean-Francois Exbrayat (1), Neil R Viney (2), Jan Seibert (3,4), Hans-Georg Frede (1), and Lutz Breuer (1)

(1) Institute for Landscape Ecology and Resources Management, Research Centre for BioSystems, Land Use and Nutrition (IFZ), Justus-Liebig-University Giessen, Heinrich-Buff-Ring 26, 35392 Giessen, Germany, (2) CSIRO Land and Water, Canberra, Australia, (3) University of Zurich, Switzerland, (4) Stockholm University, Stockholm, Sweden

**Abstract:** This study proposes to evaluate the effect of ensemble modelling in a two-step calibration and validation approach. Five different rainfall-runoff models, LASCAM, LASCAM-S, INCA, SWAT and HBV-N-D, were applied to the mesoscale River Fyris catchment (Mid-Eastern Sweden). Runoff in this region shows a spring flood peak as well as high flows during fall. The five models are conceptually different in their way to describe the catchment hydrology as illustrated by various spatial calculations units, flux processes, water storages and layers. For two non-nested sub-catchments daily discharge data were available for the five-year long study period. Models were setup using the same climatic data set. Single-objective calibrations were conducted for each discharge record and model. The calibration criterion was the average value of Nash-Sutcliffe Efficiencies calculated for discharge (focusing on high flows) and logarithmic discharge (focusing on low flows). In a second step, multi-model ensembles were compiled using the calibrated runs as ensemble members by the following different merging schemes:

- daily mean of the predictions for each day,
- daily median of all ensemble members,
- daily weighted mean, with weights set according to the respective calibration criterion values,
- multiple linear regression ensembles using the single runs as independent variables and the observations as dependent variables, and
- multiple linear regression ensembles compiled in the same way but with constraining the regression to a zero intercept.

The same calibration criterion than for the single models as well as the bias were computed for each generated ensembles. They were used as goodness-of-fit descriptors and their values were compared between single ensemble members and multi-model ensembles. The uncertainty bounds described by the calibrated runs and the corresponding set of generated ensembles were analysed in terms of range of predictions (or spread) and matching rate (or skill).

A proxy-basin validation approach was then performed to simulate ungauged basin conditions. Calibrated parameter sets of one discharge station were used to generate predictions for the other discharge record. Weights and regression coefficients computed in the above mentioned calibration step for each catchment were utilised for the uncalibrated predictions of the other catchment. The two previously defined criteria were also calculated for the newly created ensemble predictions. They were used to investigate the evolution of the quality of the single predictions between members and compiled ensembles. In the same way the evolution of the described uncertainty bounds between members and full set of generated ensembles was addressed.

Improvement was achieved by merging single runs in ensembles, even with only 2 members, fulfilling the ensemble approach aim. At the same time, uncertainty bounds of the predictions were always reduced for the ensembles compared to single model calibrations and these bounds included most of the measured discharges. We therefore concluded that the application of multi-model ensembles in hydrology was one way to overcome structural model uncertainty issues.