



Validation of uncertainty estimates in hydrologic modelling

M. Thyer (1), K. Engeland (2), B. Renard (3), G. Kuczera (1), and S. Franks (1)

(1) chool of Engineering, University of Newcastle, Callaghan, NSW, Australia (mark.thyer@newcastle.edu.au), (2) SINTEF Energy Research, N-7465 Trondheim, Norway, (3) Cemagref, UR HHLY, Lyon, France

Meaningful characterization of uncertainties affecting conceptual rainfall-runoff (CRR) models remains a challenging research area in the hydrological community. Numerous methods aimed at quantifying the uncertainty in hydrologic predictions have been proposed over the last decades. In most cases, the outcome of such methods takes the form of a predictive interval, computed from a predictive distribution. Regardless of the method used to derive it, it is important to notice that the predictive distribution results from the assumptions made during the inference. Consequently, unsupported assumptions may lead to inadequate predictive distributions, i.e. under- or over-estimated uncertainties. It follows that the estimated predictive distribution must be thoroughly scrutinized (“validated”); as discussed by Hall et al. [2007] “Without validation, calibration is worthless, and so is uncertainty estimation”.

The aim of this communication is to study diagnostic tools aimed at assessing the reliability of uncertainty estimates. From a methodological point of view, this requires diagnostic approaches that compare a time-varying distribution (the predictive distribution at all times t) to a time series of observations. This is a much more stringent test than validation methods currently used in hydrology, which simply compare two time series (observations and “optimal” simulations). Indeed, standard goodness-of-fit assessments (e.g. using the Nash-Sutcliffe statistic) can not check if the predictive distribution is consistent with the observed data.

The usefulness of the proposed diagnostic tools will be illustrated with a case study comparing the performance of several uncertainty quantification frameworks. In particular, it will be shown that standard validation approaches (e.g. based on the Nash-Sutcliffe statistic or verifying that about $p\%$ of the observations lie within the $p\%$ predictive interval) are not able to discriminate competing frameworks whose performance (in terms of uncertainty quantification) is evidently different.