Geophysical Research Abstracts
Vol. 19, EGU2017-5722, 2017
EGU General Assembly 2017


# Hydrological data assimilation using Extreme Learning Machines

Marie-Amélie Boucher (1), John Quilty (2), and Jan Adamowski (2)

(1) Department of Applied Sciences, Université du Québec à Chicoutimi, Chicoutimi, Canada (marie-amelie_boucher@uqac.ca), (2) Department of Bioresources Engineering, McGill University, Ste-Anne de Bellevue, Canada

Data assimilation refers to any process that allows for updating state variables in a model to represent reality more accurately than the initial (open loop) simulation. In hydrology, data assimilation is often a pre-requisite for forecasting. In practice, many operational agencies rely on "manual" data assimilation: perturbations are added manually to meteorological inputs or directly to state variables based on "expert knowledge" until the simulated streamflow matches the observed streamflow closely. The corrected state variables are then considered as representative of the "true", unknown, state of the watershed just before the forecasting period. However, manual data assimilation raises concerns, mainly regarding reproducibility and high reliance on "expert knowledge". For those reasons, automatic data assimilation methods have been proposed in the literature. Automatic data assimilation also allows for the assessment and reduction of state variable uncertainty, which is predominant for short-term streamflow forecasts (e.g. Thiboult et al. 2016).

The goal of this project is to explore the potential of Extreme Learning Machines (ELM, Zang and Liu 2015) for data assimilation. ELMs are an emerging type of neural network that does not require iterative optimisation of their weights and biases and therefore are much faster to calibrate than typical feed-forward backpropagation neural networks. We explore ELM for updating state variables of the lumped conceptual hydrological model GR4J. The GR4J model has two state variables: the level of water in the production and routing reservoirs. Although these two variables are sufficient to describe the state of a snow-free watershed, they are modelling artifices that are not measurable. Consequently, their "true" values can only be verified indirectly through a comparison of simulated and observed streamflow and their values are highly uncertain. GR4J can also be coupled with the snow model CemaNeige, which adds two other state variables (snow calorific deficit and the percentage of the watershed covered by snow).

In order to verify the general applicability of the proposed method, it is applied to five watersheds in contrasting hydro-climatic contexts (e.g., nordic, semi-arid, humid, and temperate). We show that the ELMs can be used to successfully update GR4J's state variables, in all cases. Considering the ensemble mean as a deterministic forecast, the Mean Absolute Error between simulated and observed streamflow can be reduced by up to 30 %. We also show that the success of the method is highly dependent on adequate selection of input variables for the ELMs. Overall, this new data assimilation method offers good performance with low computational costs.

**References**

Thiboult, A., Anctil, F. and Boucher, M.-A. (2016) Accounting for three sources of uncertainty in ensemble hydrological forecasting, *Hydrology and Earth System Sciences*, **20**(5), 1809-1825.

Zang, K. and Liu, M. (2015) Outlier-robust extreme learning machine for regression problems, *Neurocomputing*, **151**, part 3, 1519-1527.