Deep dive into regression-based model averaging methods for hydrological signatures: recommendations for methods, pre-processing and ensemble sizes

**Abstract**

It has become more and more common in hydrology to consider multiple estimates of hydrological variables-of-interest (i.e., ensembles) over single model runs. Ensemble members represent different realisations of various model structures, input data, and/or parametrisations. Improved predictions can be made using weighted ensembles with various model averaging methods. Here I perform a large sample comparison of 482 catchments in the continental US with ensemble sizes ranging from 30 to 980 members by running HBV simulations with global, regionalised parameter sets. I compared common linear regression-based model averaging methods with various combinations of constraints on the coefficients; inclusion of an intercept, positivity of coefficients, and coefficients sum to unity. I evaluated their influence on selected hydrological signatures and with different pre-processing options and used the ensemble mean as the benchmark to beat.

I found that, over the full timeseries, none of the model averaging methods are substantially better (by the Kling-Gupta Efficiency) than the ensemble mean, when the ensemble is unbiased. The methods which allow unconstrained coefficients or an intercept have a high chance of producing negative streamflow predictions. Log- or square root transformations lead to all model averaging methods to beat the ensemble mean in different flow quantiles, but the ensemble mean generally provides better autocorrelation than optimised combinations. Overall, the best performing method is the unconstrained Ordinary Least Squares regression. Increasing ensemble sizes above 10-15 members yields constantly diminishing improvements on the performance indicators. Based on the experiments, the ensemble mean gives relatively good performance when one can assume an unbiased ensemble.

For optimising an ensemble, I recommend using a regression method with strictly positive coefficient constraints without an intercept, because it gives the most consistent improvement over the ensemble mean and does not risk negative streamflow values, unlike the unconstrained OLS.