Geophysical Research Abstracts, Vol. 9, 06974, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-06974

Neural networks and clustering in estimation of the total model uncertainty of hydrologic models

D.L. Shrestha, D.P. Solomatine

UNESCO-IHE Institute for Water Education ({d.shrestha, d.solomatine}@unesco-ihe.org, sol@ihe.nl)

Hydrological models which are generally used to determine the runoff hydrographs are often faced with different sources of uncertainty including natural randomness, input data, model parameters, model structure etc. Due to interaction between these sources of uncertainty it is not always possible to separate the individual contributions to the total model uncertainty. For the decision making process especially in the flood forecasting and warning systems total model uncertainty (aggregate affect of all sources of uncertainty) is more important than the individual sources of uncertainty.

In this study total model uncertainty is considered. Uncertainty can be expressed as the probability distribution of the output model errors, without making a distinction between various sources of uncertainty related to parameters, data or model structure. The essence of the method for estimating model prediction uncertainty using data-driven models was presented by Shrestha and Solomatine earlier (published in 2006 in the Neural Networks Journal, at the Int. Conference on Hydroinformatics, and at EGU-2006). This presentation develops the method further and addresses the issues unresolved earlier. The method is termed UNcertainty Estimation based on local Errors and Clustering (UNEEC).

The direct analytical estimation of the probability distribution of the model errors is often impossible; but it can be expressed it in the form of its quantiles, or prediction limits (interval) of the model prediction. The model input space is partitioned into different zones or clusters having similar values of model errors. To achieve that, first a data matrix (training set) is built by combining historical model inputs and corresponding model error. This data is partitioned using, e.g., fuzzy c-means clustering. The prediction interval is constructed for each cluster and propagated from each cluster to the examples according to their (fuzzy) membership in each cluster.

The same data is used to construct a regression numerical prediction model (neural network). This model is applied to estimate the prediction intervals (limits) for new (or test) data.

The method was applied to estimate uncertainty of hydrologic models. Various approaches to constructing the data set adequately describing the uncertainty structure were tested. The results show that the method generates interpretable uncertainty estimates, and it can be a valuable tool for assessing the uncertainty of the model outputs.