Geophysical Research Abstracts,
Vol. 11, EGU2009-4282, 2009
EGU General Assembly 2009

Data-driven techniques scrutinized: is there one better than the rest?

A. Elshorbagy (2), G. Corzo (1), D. Solomatine (1), and S. Srinivasulu (2)

(1) Unesco-IHE Institute for Water Education, Hydroinformarmatics & Knowedlge Management, Delft DA, Netherlands (d.solomatine@unesco-ihe.org, +31-(0)15-2122921), (2) University of Saskatchewan, 57 Campus Drive Saskatoon, SK S7N 5A9 CANADA

Data-driven (soft computing) modelling techniques have been used for hydrologic prediction for decades. Even though such use has been mostly limited to research, the use of some of them is still on the rise. Researchers have adopted in their work one or more of the techniques. Usually the performance of the adopted technique(s) has been evaluated in contrast to observed data, to another traditional technique, or some other data-driven method. Most of the times one or two techniques have been applied to one or two datasets. Therefore, the results with regard to the performance and reliability of the various techniques have been conflicting and non-conclusive.

In this study, a comprehensive modelling experiment has been conducted to evaluate various data-driven techniques with regard to their prediction accuracy, uncertainty, and computational time. Artificial neural networks (ANNs), genetic programming (GP), evolutionary polynomial regression (EPR), support vector machines (SVM) and instance-based learning, in particular, k-nearest neighbor (k-NN) have been applied to five different hydrologic datasets to model runoff, soil moisture, and actual evaportranspiration. The datasets are considered to be representative of real-life problems in terms of their complexity and non-linearity. Bootstrapping techniques was used to create 12 versions of datasets from each original dataset for uncertainty estimation. Results indicate that none of the techniques can be considered superior in all cases \with regard to various performance measures. However some light can be shed on possible reasons for the performance variability in different cases. The ultimate goal is to design pre-screening tests that allow for selecting the best technique for a case study under consideration, and to achieve it, more research would be needed.