Data-driven Modelling in Water-related Problems

Lecturer: Dr. D.P. Solomatine
UNESCO-IHE Institute for Water Education, Delft, The Netherlands

One of the ways to classify models is based on contradicting and comparing behaviour (processes described on the basis of physical laws) – and collected data (that also contains information about the modelled system). The former models are often called process models (or behavioural, physically based), and the latter could be naturally called data-driven models.

This tutorial will discuss the place of data-driven modelling (DDM) in Hydroinformatics, relation to process models, data pre-processing, cover main principles of machine learning and computational intelligence as engines of DDM, methods and techniques used, and present practical examples of DDM and the relevant software. The techniques considered include both widely accepted methods as neural networks and fuzzy rule-based systems, as well as relatively less known methods as M5 model trees, committee machines, boosting and support vector machines.

The tutorial is based on the course taught at the Hydroinformatics section of IHE-Delft, the experience of practical applications of this technology, and the experiences of other institutions active in this area. The practical applications of various techniques in the water-related issues include: flood control, rainfall-runoff modelling, reservoir optimization, real-time control, surge water level prediction, interpretation of aerial photos. With respect to the short course given by the author in Cardiff at Hydroinformatics-2002, this tutorial will include the new material and research results, e.g., in the area of committee machines and user-transparent data-driven models, together with the new practical applications.

The participants will receive fully versions of the two software tools:

- *GLOBE* for global and genetic optimisation;
- *NeuralMachine* for neural network and weighted local regression modelling.

*Software used:* GLOBE, Weka, NeuralMachine, AFUZ.
Syllabus: 4 Lectures, 9:00 – 16:45

Lecture 1  9:00 – 10:30
Quick start: example of a data-driven model (neural network).
Modelling in the framework of hydroinformatics. Data-driven and physically based models. Importance of data. Choice of variables. Model calibration, verification, testing.
Computational intelligence, data mining, machine learning and soft computing. Connections to statistics and optimisation.
Main types of machine learning: classification, association, clustering, numerical prediction. Hypotheses, concepts, instances, attributes. Types of data: nominal, ordinal, interval, ratio.
Decision trees, classification and association rules. Example: real-time control of a water system.

Tea Break

Lecture 2  10:45 – 12:15

Lunch Break

Lecture 3  13:30 – 15:00
Other types of ANNs and their use for prediction and classification.
Practical applications of data-driven methods in hydrological modelling and flood control.
Committee machines. Examples of their use in flow prediction problems.

Tea Break

Lecture 4  15:15 – 16:45
Support vector machines (SVM) in classification and prediction.
Evolutionary methods as a part of computational intelligence, and as a multi-extremum optimisation method.
Introduction to fuzzy rule-based systems (FRBS). Practical applications in rainfall data reconstruction, hydrological modelling and flood control.
Main notions of the non-linear dynamics and chaos theory. Practical applications in prediction of ocean water levels.
Discussion. Course evaluation. Closing.