

[Home](#)

[Online Library HESS](#)

- [Recent Final Revised Papers](#)
- [Volumes and Issues](#)
- [Special Issues](#)
- [Library Search](#)
- [Title and Author Search](#)

[Online Library HESSD](#)

[Alerts & RSS Feeds](#)

[General Information](#)

[Submission](#)

[Review](#)

[Production](#)

[Subscription](#)

[Comment on a Paper](#)

#### Journal Metrics

 **IF 2.462**

 **5-year IF 2.670**

 **SNIP 0.856**

 **SJR 0.099**

[Definitions](#)

ARCHIVED IN

  
PORTICO

[Volumes and Issues](#) [Contents of](#)

Hydrol. Earth Syst. Sci., 14, 1931-1941, 2010  
www.hydrol-earth-syst-sci.net/14/1931/2010/  
doi: 10.5194/hess-14-1931-2010

© Author(s) 2010. This work is distributed  
under the Creative Commons Attribution 3.0 License.

## Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology

A. Elshorbagy<sup>1</sup>, G. Corzo<sup>2</sup>, S. Srinivasulu<sup>1</sup>, and D. P. Solomatine

<sup>1</sup>Centre for Advanced Numerical Simulation (CANSIM), Department of Civil and Geological Engineering, University of Saskatchewan, Saskatoon, SK, S7N 3A5, Canada

<sup>2</sup>Department of Hydroinformatics and Knowledge Management, UNESCO Institute for Water Education, Delft, The Netherlands

<sup>3</sup>Water Resources Section, Delft University of Technology, Delft, The Netherlands

**Abstract.** A comprehensive data driven modeling experiment is presented in a two-part paper. In this first part, an extensive data-driven modeling experiment is proposed. The most important concerns regarding the data driven modeling (DDM) techniques and data were handled, compared and evaluated, and the basis on which findings and conclusions were drawn are discussed. A concise review of key articles that present comparisons among various DDM techniques is presented. Six DDM techniques, namely, neural networks, genetic programming, evolutionary polynomial regression, support vector machines, M5 model trees, and  $k$ -nearest neighbors are proposed and explained. Multiple linear regression and naïve models are also suggested as baseline for comparison with various techniques. Five datasets from Canada and Europe represent evapotranspiration, upper and lower layer soil moisture content, a rainfall-runoff process are described and proposed, in the second part for the modeling experiment. Twelve different realizations (groups) each dataset are created by a procedure involving random sampling. Each group contains three subsets: training, cross-validation, and testing. A modeling technique is proposed to be applied to each of the 12 groups in each dataset. This way, both prediction accuracy and uncertainty of the modeling techniques can be evaluated. The description of the data, the implementation of the modeling techniques, results and analysis of the findings of the modeling experiment are deferred to the second part of this paper.

[Final Revised Paper](#) (PDF, 384 KB) [Discussion Paper](#) (HESSD)

Citation: Elshorbagy, A., Corzo, G., Srinivasulu, S., and Solomatine, D. P.: Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology, Hydrol. Earth Syst. Sci., 14, 1931-1941, doi:10.5194/hess-14-1931-2010. [Bibtex](#) [EndNote](#) [Reference Manager](#) [XML](#)