# Automatic input selection for hydrological modelling: a comparative analysis

Greer Humphrey (1), Stefano Galelli (2), Andrea Castelletti (3), Holger Maier (1), Graham Dandy (1), and Matthew Gibbs (1)

(1) University of Adelaide, Dept. of Civil and Environmental Engineering, Adelaide, Australia, (2) Singapore University of Technology and Design, Pillar of Engineering Systems and Design, Singapore, (3) Politecnico di Milano, Dept. of Electronics, Information, and Bioengineering, Milano, Italy

Input variable selection is an essential step in the development of statistical models and is particularly relevant in hydrological modelling, where potential model inputs often consist of time lagged values of each different potential input variable. While new methods for identifying important model inputs continue to emerge, each has its own advantages and limitations and no method is best suited to all datasets and purposes. Nevertheless, rigorous evaluation of new and existing input variable selection methods is largely neglected due to the lack of guidelines or precedent to facilitate consistent and standardised assessment. This rigorous evaluation would allow the effectiveness of these algorithms to be properly identified in various circumstances. In this paper, we propose a new framework for the evaluation of input variable selection methods which takes into account a wide range of dataset properties that are relevant to real world data and assessment criteria selected to highlight algorithm suitability in different situations of interest. The framework is supported by a repository of data sets to enable standardised and statistically significant testing. This framework is supposed to help promoting the appropriate application and comparison of input variable selection algorithms and eventually serves to provide guidance as to which algorithm is most suitable in a given situation.