



Data-Driven Modelling: on building nonsensical applications

ROBERT J. ABRAHART (1), NGAHZAIFA AB GHANI (1), and CHRISTIAN W. DAWSON (2)

(1) School of Geography, University of Nottingham, Nottingham, United Kingdom (E-mail: bob.abrahart@nottingham.ac.uk),

(2) Department of Computer Science, Loughborough University, Loughborough, United Kingdom

The search for an optimum set of input drivers is of interest to hydrological modellers in their quest to develop more accurate or more parsimonious data-driven applications. The larger set of potential inputs, from which some preferred sub-set is selected, is seldom if ever questioned. This paper revisits the inadvertent use of virtual or spurious relationships and thereafter applies logic and reasoning to assess recent trends in the selection of input drivers for data-driven modelling of suspended sediment. Numerous models are developed on 'rational inputs' for the Tongue River. Their outputs are compared to published findings derived from data-driven models developed using 'irrational inputs'. The analysis suggests that recent approaches to modelling suspended sediment load are (i) illogical and (ii) deliver misleading expectations with regard to what can or cannot be achieved.