Hydrol. Earth Syst. Sci. Discuss., 6, C1404–C1409, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C1404/2009/

Hydrology and Earth System Sciences Discussions

HESSD

6, C1404–C1409, 2009

Interactive Comment

Interactive comment on "*HESS Opinions* "Crash tests for a standardized evaluation of hydrological models"" *by* V. Andréassian et al.

A. Efstratiadis

andreas@itia.ntua.gr

Received and published: 1 July 2009

Evaluating models or evaluating modelling practices?

by A. Efstratiadis and N. Mamassis

The authors investigate the perspectives of a powerful procedure for evaluating hydrological models, inspired from the four-step framework proposed by Klemes (1986). The philosophical foundation of their rationale is indisputable; since falsifiability is an essential element of science (Popper, 1959), nobody could argue against the argument for "tough" evaluation tests to assess the model performance in multiple study basins

– a quest that is indeed far from the common practice, described by the authors as "catchment monograph". In contrast to "monographs", the authors support the use of large data sets (both in time and space), as the ideal working framework for model evaluation. Unfortunately, there are serious limitations in this approach - some are already articulated and commented in Sect. 2.1, while some require a more thorough investigation. We wish to discuss these issues, also taking into account the crash-test pattern.

In our opinion, the model development is a continuous process that oscillates between induction and deduction (Koutsoyiannis et al., 2009). Through induction, observations are used to establish a theory, where the part (sample) is used to explain the whole (population). Through deduction, a theory (formulated either as physical laws or conceptualizations, i.e. models) is used to predict or explain observed phenomena. The (expected) differences between predictions and observed data demand again the use of induction, in order to improve the theory; the process is continued ceaselessly, since more observations are obtained and since better understanding of the processes is gained. Fig. 1 illustrates the induction-deduction cycle using the metaphor of a car, where the increased knowledge (direct via deduction, indirect via induction) is the driver for improved products. Crash tests are part of this cycle; in fact, they are components of the induction phase. Is it the same with hydrological models? Can the proposed crash test significantly contribute to a faithful evaluation thus leading, through induction, to improved models?

To answer these questions, we should define the role of hydrological modelling in the actual technological scene. For, we make the key hypothesis that models are (or should be) designed to support engineering and management decisions, usually in conjunction with other water resource simulation tools (e.g. hydrosystem operation models). In contrast to "bulimic" scientists, who wish to develop theories (and models) of the most general applicability, engineers are typically interested on ensuring the optimal model fitting to the specific study area, even if this requires manual interventions to its math-

HESSD

6, C1404–C1409, 2009

ematical structure (the “monograph” approach). From a possibly naïve but certainly practical point-of-view, a model working satisfactory well in the specific hydrosystem would be by far preferable from a model providing moderate performance, although in a large spectrum of catchments. Thus, the honest response of a practical engineer against the validity of the crash-test would be negative. Those who attempt to be both scientists and engineers remain very sceptic against models that are qualified as general-purpose tools – in contrast to cars, whose “generality” is an obvious requirement for their commercial success (except if models should be promoted in the same manner to industrial products).

Our experience with highly complex and, simultaneously, heavily modified hydrosystems (Rozos et al., 2004; Efstratiadis et al., 2008) suggests the necessity for integrated tools that also ensure flexible schematization and parameterization, for which significant effort is required to provide consistent calibrations. In this context, the implementation of a large-scale crash evaluation test seems to be utopian (due to time and data limitations – a topic already discussed by the authors) and probably misleading. For instance, the box-plot diagram for the proxy basin test, showing the model performance over the set of 600 catchments, indicates that the model has poor predictive capacity (efficiency < 0.50) for about half of the examined catchments, while the mean efficiency is negative (i.e. the model predictions are worse than the predictions based on the historical average). Is this a surprising result or an evident statement related to well-known model limitations, which would be easily recognized by an experienced hydrologist? Was the same model structure (parameterization) adequate for all catchments? Was the same calibration methodology followed? Was the unique criterion (efficiency) representative for evaluating the overall model performance? What about uncertainty and equifinality?

We think that model building and calibration is not and will never be an automatic computer exercise, although powerful optimization algorithms are useful to speed up the parameter estimation procedure. In contrast, ensuring realistic and reliable represen-

HESSD

6, C1404–C1409, 2009

Interactive Comment

tation of all modelling aspects is an exhaustive task – particularly when moving from simple lumped approaches to more detailed schemes, in order to simulate multiple fluxes at multiple sites. In this context, a balanced scientific and engineering approach is essential in order to better conceptualize the specific processes (kind of deduction) and take advantage of all available specific information (kind of induction); the last issue also includes soft data, which is a relatively new dimension in hydrological modelling (Seibert and McDonnell, 2002; Efstratiadis et al., 2008). This requires deep understanding of the model advantages and weaknesses, the data limitations, and, over all, the peculiarities of the system under study – overview that is impossible to have when dealing with large data sets.

Undoubtedly, the crash tests are not useless. But we should investigate very carefully the results for each specific catchment, to get safe conclusions. Going back to the car analogy, the crash-test results depend on the test rigidity and the car use. Always there might be a test that is so rigid that even an excellent car will fail. On the other hand, it is not realistic to specify a crash test for a car that will be used in frozen tundra, a rainforest, a desert or inside an Arabic Casbah. In hydrological modelling the situation is much more complicated – the test results depend not only on the data and site characteristics but also (and maybe more) on the modelling practice. Hence, a state-of-the-art tool, implementing the most sophisticated theory, would easily fail due to misuse issues, such as inefficient representation of process interactions, over-parameterization, lack of parsimony, blind use of optimization in calibration, etc. (Nalbantis et al., 2007).

To summarize, we think that the major quest is how ensuring the best possible adaptation of a specific model to a specific area, while the model transposability in space has limited practical interest. The proposed crash test framework could be improved by also incorporating engineering judgment, which in its current formulation (resembling a numerical exercise) seems to be totally missing. In fact, hydrological systems are too complex and too unique to be treated as “faceless” components of such a black-box procedure. So, instead of evaluating models themselves, wouldn’t be much more useful to evaluate modelling practices?

References

Efstratiadis, A., Nalbantis, I., Koukouvinos, A., Rozos, E., and Koutsoyiannis, D.: HYDROGEIOS: a semi-distributed GIS-based hydrological model for modified river basins, Hydrol. Earth Syst. Sci., 12, 989-1006, 2008.

Klemes, V.: Operational testing of hydrologic simulation models, Hydrolog. Sci. J., 31, 13-24, 1986.

Koutsoyiannis, D., Makropoulos, C., Langousis, A., Baki, S., Efstratiadis, A., Christofides, A., Karavokiros, G., and Mamassis, N.: HESS Opinions: "Climate, hydrology, energy, water: recognizing uncertainty and seeking sustainability", Hydrol. Earth Syst. Sci., 13, 247-257, 2009.

Nalbantis, I., Efstratiadis, A., and Koutsoyiannis, D.: On the use and misuse of semi-distributed rainfall-runoff models, XXIV General Assembly of the International Union of Geodesy and Geophysics, Perugia, International Union of Geodesy and Geophysics, International Association of Hydrological Sciences, 2007 (http://www.itia.ntua.gr/en/docinfo/744/).

Popper, K.: The logic of scientific discovery, Routldege, London, UK, 513 pp., 1959.

Rozos, E., Efstratiadis, A., Nalbantis, I., and Koutsoyiannis, D.: Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows, Hydrol. Sci. J., 49(5), 819-842, 2004.

Seibert, J., and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration, Water Resour. Res., 38(11), 1241, doi: 10.1029/2001WR000978, 2002.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 3669, 2009.

HESSD

6, C1404–C1409, 2009

Interactive Comment

HESSD

6, C1404–C1409, 2009

Interactive Comment

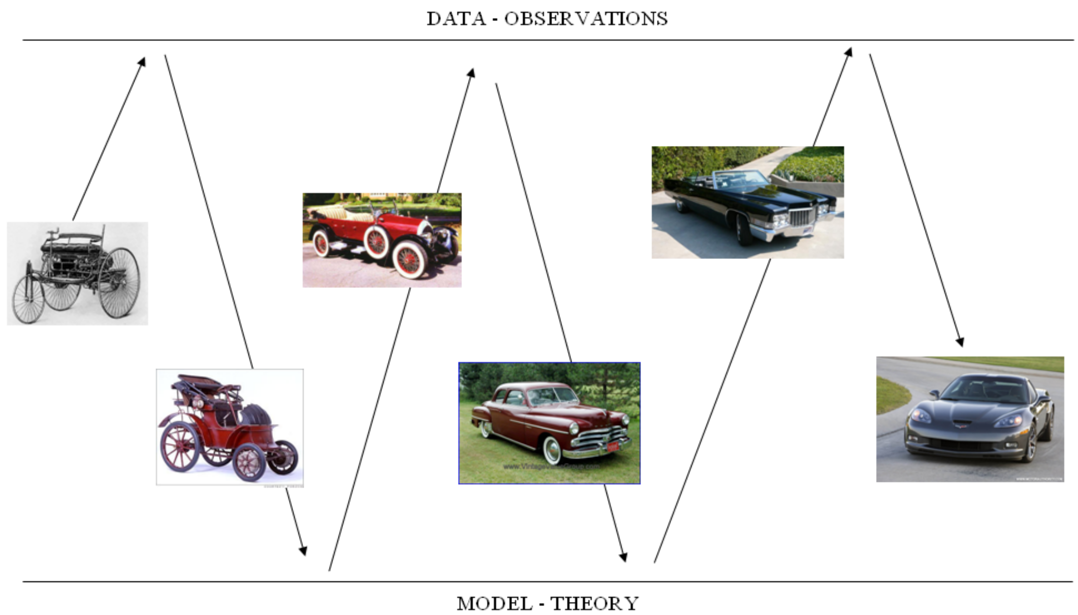

Fig. 1. Schematic layout of the induction-deduction cycle for the car industry.