How to distinguish between hydrological models appropriate for impact studies and unsuitable models? Crash-test necessity

Alexander Gelfan (1,2)

(1) Water Problems Institute, Russian Academy of Sciences, Moscow, Russia, (2) Moscow State University, Hydrological Department, Moscow, Russia

During the last decades there is a large gap between progress in improving hydrological models, methods of their calibration, their ability to assimilate different types of data, etc. on the one hand, and outdated methods of model evaluation and testing on the other hand. We still use easy-to-pass tests and very "soft" statistical criteria of model efficiency. As a result, there are a lot of "good" models that pretend to be suitable for impact studies, and the number of such models grows like snowball. However though performance of all models is mainly good, but, most likely, many of them are useless. In order not to be overwhelmed by pseudo-good models, we have to be able to distinguish between models appropriate for impact studies and unsuitable ones, to understand the grounds for credibility of a given hydrological model.

It seems reasonable to search an answer on the question posed in the session's title within the framework of the following pragmatic argumentation, which is based on three well-established statements (Klemeš, 1986; Coron et al., 2011; Andreassian et al., 2013; Refsgaard et al., 2014; Thirel et al., 2015). First, predictive hydrological model can never be universally validated but its performance can be evaluated for situations that imitate the "target" conditions of the model application. Second, if the model does not perform well, it implies that the model is most likely to be inadequate in the "target" conditions. Third, the opposite is not true: a lack of disagreement does not necessarily result in the model applicability for these conditions, however appropriate evaluation design increase credibility for and decrease uncertainty in the model results.

Because the model's predictive ability usually cannot be tested directly from data (the latter, most probably, never be available), specific test ("crash-test") is necessary allowing one to reinforce the model's applicability. In the presentation, two issues related to crash-test developing are discussed: test's design and test's performance measures. Examples are provided demonstrating that models successfully passed ordinary test (e.g. split-sample test with Nash-Sutcliffe performance measure) are rejected by the crash-test, i.e. these models can be assumed as a poor hypothesis for the behavior of basin and unsuitable for impact study.