Hydrol. Earth Syst. Sci. Discuss., 6, C1392–C1395, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C1392/2009/

Hydrology and Earth System Sciences Discussions

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

Dr Andréassian

vazken.andreassian@cemagref.fr

Received and published: 29 June 2009

Comments by Kieran M. O'Connor on "Crash tests for a standardized evaluation of hydrological models" by Andréassian et al., 2009

It is nearly a quarter of a century since Klemeš (1986) introduced his hierarchical scheme for the systematic evaluation of hydrological models which was a generalisation of the split-sampling test then (and indeed still) in vogue. A penetrating, frank, honest and unbiased assessment of progress (or lack thereof) in model hydrological evaluation is long overdue, leading to open discussion and perhaps the evolution of a better way of conducting our modelling studies in the future. This is precisely what this paper delivers and it is to be warmly welcomed for opening up a topic which modellers

have, with some honourable exceptions, largely ignored or swept under the carpet. This is not to suggest that we are consciously being dishonest or irresponsible but all too often we seem to fall into the trap of genuinely believing in the scientific framework or a least the operational efficacy and generality our own models, shunning all criticism (I like the reference to alcoholics!), instead of harbouring a healthy scepticism for the whole business. Our goals are too often restricted to specific case studies and objectives or to the (often selfish) validation of our own model constructs, usually applied in very limited hydrological conditions, using inadequate, unrepresentative and often uncertain data. Based on the "Why write one paper where two will do!" philosophy, a series of papers by an individual or modelling group, where each paper may contain only incremental new model developments, may entirely lack a coherent critical aggregate overview of model evaluation.

The authors return to the Klemeš (1986) approach as their starting point, a very good place to start, advocating its structured and rigorous application, using larger and longer data sets, applied to a wide variety of catchments. They tease out the excuses advanced by modellers for not doing this, virtually none of which are scientifically justifiable but are still understandable considering the physical working environment of most modellers. The authors suggest that we should "always present the results of model-related discussions with distributions of model performance over large and varied data sets". They courageously provide an instructive and informative example of the Klemeš approach, with the inevitable conclusion that there is "considerable room for improvement". That's an honest evaluation!

Introducing a novel concept in hydrological modelling, the authors propose a 'test to destruction' of our models, a time-honoured engineering principle, which they call "crash tests", by applying them in a wide variety of situations, challenging and otherwise, and not only recording the failures but embracing them as a means of identifying model weaknesses and possibly improving the models. Clearly, a problem cannot be solved if it is remains unacknowledged! Sadly, journal editors and reviewers are generally

not keen to acknowledge model failures by publishing such studies and moreover it is galling for the modeller when a model into which much celebrated science has been infused performs significantly worse that a crude black-box one. All this requires a change in the mindset of researchers, authors and editors!

Critical to the advancement of model evaluation practice is the availability of the "large and varied data sets needed to develop and test hydrological models, to ensure their generality, to diagnose their failures and to improve them". It would seem that only one of the international agencies, such as WMO, would be in a position to develop such a set of benchmark data for this purpose, made freely available to all modellers. In addition, an agreed extensive set of model performance indices should be specified, which could be automatically applied in all model development studies, bearing in mind that the modelling objectives may differ and the necessity of 'comparing like with like'. Clearly, a performance index directly related to or derived from the objective function used for model calibration will suggest better model performance than the value of the same index based on the results using a different objective function.

The authors also consider a hybrid approach whereby "single catchment analyses and large data set studies work together" but their logic supporting this concept is, in my opinion, the weakest element of this paper. For me, al least, it is not very coherent or convincing!

Finally, the authors advocate making proper use of their proposed "crash tests", but this assumes that the hydrological modelling community will now turn over a new leaf and mend their ways. I'm a little sceptical about that, but I hope I'm wrong! At least the authors have courageously aired a long-hidden serious deficiency in hydrological modelling, one which sooner or later we must confront! I am hopeful that this paper will generate much discussion and perhaps inspire a new generation of modellers. The authors are to be congratulated on making this contribution.

Professor Kieran M. O'Connor (Retired), Dept. of Engineering Hydrology, National

University of Ireland, Galway (Block R), University Road, Galway. IRELAND.

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