Hydrol. Earth Syst. Sci. Discuss., 9, C1166–C1172, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C1166/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Ideal point error for model assessment" *by* C. W. Dawson et al.

C. W. Dawson et al.

a.shamseldin@auckland.ac.nz

Received and published: 29 April 2012

The authors are grateful to both Reviewers for their efforts in reviewing our manuscript. In these closing comments we set out our proposals for revising and improving the submitted manuscript.

Reviewer #1 suggested only minor editorial changes be made. We interpret the lack of any criticism of the paper content as satisfaction on the part of the Reviewer, and we are, therefore, pleased to have received a supportive review.

Reviewer #2 made a number of outspoken criticisms of the manuscript, and on the basis of these criticisms questioned whether the work is of sufficient merit for publication. It is important to note that these criticisms were not generally technical – and that the numerical basis of our work was not challenged. Instead, these criticisms were con-

textual and related to the relevance and importance of the arguments developed in the manuscript. These criticisms had four key foci:

a) That the paper is dealing with an issue that is of limited hydrological interest;

b) The approach used is insufficiently 'profound' and does not add significant knowledge;

c) The theoretical treatment of IPE would be better replaced with a 'case study' approach;

d) That our application of naive model benchmarks as a basis for standardising ideal point error (IPE) values is incorrect.

In each case, we assert that we have offered clear and justified rebuttals of the Reviewer's criticisms with detailed responses (see Short Comment: 'Response to Reviewer #2'). However, we also accept from the Reviewer's reaction to our manuscript that some improvement is required. Below, we indicate where we believe amendments may/may not be warranted and indicate our intended revisions.

In some cases the Reviewer's statements appear to reflect a general view that the theoretical examination of a new hydrological model performance metric is of insufficient interest to the hydrological community to warrant publication and offers no profound insights to the discipline. This view is strongly contradicted by recent evidence from published studies in which theoretical and empirical investigations of the meaning and usefulness of individual performance metrics has been a clear feature. Indeed, the Nash Sutcliffe Efficiency (NSE) index (Nash and Sutcliffe, 1970) alone has been the basis of at least six papers in the last five years (Schaefli and Gupta, 2007; Criss and Winston, 2008; Jain and Sudheer, 2008; Gupta et al., 2009; Ruesser et al., 2009; Moussa, 2010). This strongly suggests that there is, in fact, both interest and merit in undertaking a detailed examination of the most recent metric that has been proposed: the IPE (Elshorbagy et al., 2010a,b).

In other cases, the Reviewer's comments appear to extend either from an ideological stance that we argue is not representative of the breadth of views within published hydrological studies, or from a misunderstanding about the ideas advocated by the paper. For example, the Reviewer's assertion that model benchmarking represents an unacceptable approach to developing IPE contradicts the ideas presented in numerous, well-cited papers concerned with interpreting model performance metrics (e.g. Nash and Sutcliffe, 1970; Legates and McCabe, 1999; Siebert, 2001; Schaefli and Gupta, 2007; Criss and Winston, 2008; Jain and Sudheer, 2008; Reusser et al., 2009; Moussa, 2010). Indeed, the most widely used NSE index is based on benchmarking model outputs against the mean discharge of the calibration period. Similarly, the Reviewer's interpretation of our work as recommending the transfer of a naive model developed on one set of data to an independent data set as a basis for relative model evaluation is not correct. At no point in the manuscript do we recommend transferring a model from one data series to another.

We do, however, accept that the paper could be improved through the development of arguments that more clearly demonstrate the hydrological interest of the paper, and the knowledge that it delivers. Such arguments should also demonstrate how the approach developed in our paper is consistent with other, well-cited examples from the hydrological literature; including examples from this journal. We are grateful to Reviewer #2 for highlighting areas in which the strengthening of our arguments would be beneficial.

To this end, we propose the following core revisions to the paper, in addition to more minor editorial adjustments:

1. A new introductory section that repositions the work within a more specific context of the challenges surrounding data-driven modelling (DDM) and highlights the knowledge that the paper delivers to this group. Indeed, the common approach for evaluating the performance of different data-driven hydrological models is solely metric based - making limitations in an evaluation metric a profound problem for DDMs. The desire to have a method of combining multiple metrics into a single measure has been of key interest

to data-driven modellers as a way of enhancing the means of model comparison. This is the reason that IPE was originally developed and published (in this journal). However, there are significant numerical and technical limitations in the original version of IPE that make its direct application in DDM studies problematic or impossible – and these limitations will have wider hydrological relevance too. By more clearly specifying these limitations in our revised paper, and highlighting the approaches we offer for overcoming them, the hydrological interest of the paper and the knowledge delivered by the manuscript, will be made clearer.

2. A new paragraph and associated tabulation, that contextualises the basic, theoretically-driven approach adopted in this paper, and justifies it with respect to the approaches adopted in other published papers concerned with evaluating different performance metrics used in hydrological modelling. There are several, well-cited examples of papers from the last decades (including from this journal), that elucidate how a complex metric performs in a theoretically-driven manner; using artificially engineered errors computed on a relatively simple hydrological data series (e.g. Krause et al., 2005; Cloke and Pappenberger, 2008). Such approaches ensure that the results have a degree of transferability that is made impossible if a case-study method (as argued by Reviewer #2) is employed. We advocate that the use of artificially engineered errors helps to interpret real-world cases in which more complex combined errors occur. The new paragraph and tabulation will thus demonstrate how our paper compares to these methods, avoids the problems of specific case-study results, and elucidates the strengths and weaknesses of IPE using a theory-driven approach. The text will also be able to counter the suggestion made by Reviewer #2 that the paper is of little hydrological interest by highlighting the numerous citations attached to the papers to which our work relates.

3. A new section that explains the importance of benchmarking, and the way that it is employed in our work, more clearly. There appears to be some confusion on behalf of Reviewer #2 about our use of benchmarks. This is evidenced by their incorrect in-

terpretation of our conclusions (see 'Response to Reviewer #2). We do not advocate the transference of our t+4 naive model to other data series – this would be nonsense as our data series is independent of other hydrological series. We advocate that performance metrics delivered by a simplified, autoregressive AR(1) naive model should be used as the baseline measure in the IPE equation, to which the performance of other, more complex models is then compared. In this way, model evaluation using IPE becomes a relative assessment tool and the performance benefits of different models developed on the same series can be made relative to a standard model benchmark. This approach is similar to the method used by Moussa (2010) in their study of the NSE index, and responds to the ideas of Seibert (2001). The approach has the benefit of enhancing the degree of transferability of IPE results between models developed on different data series. Without the use of a benchmark, each IPE value is entirely dependent on the data series upon which the model(s) were developed – and there is no accepted standard data series for model comparison available to hydrologists which can avoid the necessary use of case study data. As case study data series are independent, so too are the IPE scores and there is, therefore, no valid basis upon which IPE scores can be compared. By benchmarking IPE to a common, naive model rather than a data series, the IPE score of models developed on independent data series can be compared on the basis of the extent to which each model's performance exceeds that of a baseline. This, therefore, delivers greater transferability for IPE and delivers a profound improvement to model metric comparison. The revised paper will present these arguments in a clearer manner, together with supporting citations, and strengthen the justification for our benchmarking of IPE.

References:

Cloke, H.L., Pappenberger, F. 2008. Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures. Meteorological Applications, 15(1), 181-197.

Criss, R.E, Winston, W.E. 2008. Do Nash values have value? Discussion and alternate

proposals. Hydrological Processes, 22(14), 2723-2725.

Elshorbagy, A., Corzo, G., Srinivasulu, S. and Solomatine, D.P. 2010a. Experimental investigations of the predictive capabilities of data-driven modelling techniques in hydrology – Part 1: Concepts and methods. Hydrology and Earth System Science, 14, 1943-1961.

Elshorbagy, A., Corzo, G., Srinivasulu, S. and Solomatine, D.P. 2010b. Experimental investigations of the predictive capabilities of data-driven modelling techniques in hydrology – Part 2: Application. Hydrology and Earth System Science, 14, 1931-1941.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F. 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. Journal of Hydrology, 377(1-2), 80-91.

Jain, S.K., Sudheer, K.P. 2008. Fitting of hydrologic models: a close look at the Nash-Sutcliffe index. Journal of Hydrologic Engineering, 13(10), 981-986.

Krause, P., Boyle, D.P., Base, F. 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geoscience, 5, 89-97.

Legates, D.R., McCabe, G.J. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35(1), 233-241.

Moussa, R. 2010. When monstrosity can be beautiful while normality can be ugly: assessing the performance of event-based flood models. Hydrological Sciences Journal, 55(6), 1074-1084.

Nash, J.E., Sutcliffe, J.V. 1970. River flow forecasting through conceptual models 1: a discussion of principles. Journal of Hydrology, 10(3), 282-290.

Reusser, D.E., Blume, T., Schaefli, B., Zehe, E. 2009. Amnalysing the temporal dynamics of model performance for hydrological models. Hydrology and Earth Systems

Science, 13, 999-1018.

Schaefli, B., Gupta, H.V. 2007. Do Nash values have value? Hydrological Processes, 21(15), 2075-2080.

Seibert, J. 2001. On the need for benchmarks in hydrological modelling. Hydrological Processes, 15(6), 1063-1064.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 1671, 2012.