Hydrol. Earth Syst. Sci. Discuss., 10, C301–C306, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C301/2013/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Comparative assessment of predictions in ungauged basins – Part 1: Runoff hydrograph studies" *by* J. Parajka et al.

Anonymous Referee #3

Received and published: 11 March 2013

General comments (P = Page; L = line):

This synthesis paper aims to compare studies predicting runoff hydrographs in ungauged basins. The methodology consists first on analyzing the median Nash-Sutcliffe Efficiency criteria (NSE) from 34 studies reported in the literature involving 3874 catchments, and second on a more detailed analysis of individual basins. The results discuss the performance NSE of various regionalization methods depending on the climate, the type of the method, the data availability, the model complexity, etc. The topic is novel and I appreciated the large efforts undertaken by the authors to synthesize the majority of the international literature on the topic. The paper is clearly structured and I enjoyed reading it. While the title announces "runoff hydrograph studies", the whole paper is based on the interpretation of only the NSE. My main comments concern the

justification of the choice of the NSE criteria in comparison to other criteria, the original data analysis, the impact of data uncertainty on the NSE values and the consequences on paper results such as the significance of the classification of methods, and some secondary comments.

Specific comments:

1. The choice of the Nash-Sutcliffe criteria (NSE): Since 1970, the NSE is an international well-known hydrological standard, and I totally agree to choose it as a criterion because probably this is the only available information on model performance. The NSE is useful to compare different methods or models on the same catchment or using the same set of data, but comparing NSE among various basins is not so evident, and other criteria can be analyzed : i) The NSE is one among other performance criteria such as the error on the total volume, the error on runoff coefficient, the NSE calculated on the root square of the discharge or on the Log of the discharge etc. (please cite the large literature on the criteria functions used in hydrology; e.g. a synthesis in Dawson et al., 2007). Conventional objective functions such as the root mean square error, the NSE, or the index of agreement were largely discussed in the literature because they tend to emphasize the high flows, and consequently, are oversensitive to extreme values and outliers (Legates and McCabe,1999). On the opposite, the mean absolute percent error tends to emphasize the low flows. It is not evident that the paper obtains similar results when using other criteria. The choice of the NSE must be discussed and the results discussed if other criteria were chosen. The NSE is also very sensitive to high discharge data (especially values and frequencies of peak-flows in comparison to the mean discharge value). If there is a high heterogeneity in time discharge series, a low NSE value may result from bad simulations on a very few number of data corresponding to high peak-flows. Hence the comparison of NSE among basins is not trivial. For all these reasons, I suggest that the authors discuss the large international literature on the significance of the NSE criteria especially when used in different basins (see for example Schaefli and Gupta, 2007), and justify the choice and

the use of only the NSE and why not other complementary criteria? If other efficiency criteria than NSE are available in some of the literature used in Levels 1 or 2, it will be very interesting to see if we obtain similar (or different) results than those obtained with the NSE.

2. Data analysis: A very important data based was analyzed in this paper. However, it is not clear how the NSE values were identified in Table 1. It will be pedagogic to explain on one study case (one line from Table 1), how the NSE values were extracted from literature and then used in this study; a short explanation can be added in an appendix. What can we learn from the whols set of original papers, and from the median, minimum and maximum values of NSE? Do the authors of the original paper use other criteria functions? This explanation will be helpful to discuss the significance and the uncertainty on the value considered of the NSE.

3. Uncertainty on data: All discharge data, especially during peak-flows, are measured with high uncertainties. Please discuss the impact of data uncertainty on the NSE: What will be the impact on the NSE of uncertainties especially on high discharges, and does the uncertainty on data can impact the main results and classifications presented in this paper? In order to reduce the impact on the NSE value of uncertainty on peak-flows especially during floods and inundation, does the use of the NSE applied on the root square of the discharge or the Log of the discharge will modify the results?

4. It will be also interesting in the discussions to comment the highest and lowest values of the NSE for each category analyzed (Figures 2 to 6): please indicate the reference and if possible comment why some studies gave very good values of the NSE (close to 0.9) while others gave low values of NSE (0.4 – 0.5).

5. The number of data used to compare climate regions (x-axis in Fig 2), regionalization methods (Fig 3 and 5), number of catchments (Fig 4), number of model parameters (Fig 6) can vary drastically among regions, methods, etc. When comparing methods, the conclusions depend on the number of available data. The authors must indicate clearly

the number of points used for each interpretation, and some details can be added in order to improve the clarity of the paper: i) for example P 384, L 17-26 and P 385, L 1-2: The paper indicates the number of results for “spatial proximity” (33 results), “parameter regression” (17 results), “model averaging” (11 results) and “regional calibration” (4 results) but didn’t give the number for the “similarity group”. I guess that the total number of results will give 75, but it is not evident. Moreover, the number of data used per class is not given for the other applications (Figures 2 to 6)! I suggest that the authors add on the x-axis of Figures 2, 3 and 4 and for each type of climate (on Fig 2), regionalization method (on Fig 3) and number of catchments (on Fig 4) the number of results (or the number of points) used in each column. Ii) I count 73 points (and not 75) on both Figures 5 and 6; please check. Please indicate also the number of points used for each regionalization method (Fig 5) and for each class of models (Fig 6). Iii) For the same reasons, please also indicate on Figures 7, 8 and 9 the number of points used or each class of the x-axis.

6. Data characteristics: It will be very helpful for the reader if additional characteristics of the data used are given in complementary to Table 1. For example and if available, for each study in Table 1: i) the total number of basins per study; ii) the range of variation of area, rainfall, discharge, runoff coefficients, aridity index, elevation, etc. (probably available for Level 2); iii) the range of variation of the NSE and other error criteria if available.

Other comments:

P 380, L 24-25: The paper states that 4 characteristics are analyzed, while only three are given.

P 380, L 18: It is stated that there is 34 studies in Table 1 which results in a total of 75 assessments. However it is not clear how many results are derived from each study. Please indicate on Table 1 the number of results from each study.

Table 1: I didn’t understand the significance of the various values of the runoff model

efficiency for a given study: i) only one value vs a list of values separated by a comma; ii) a range of values (e.g. 0.62-0.71). I find approximately 70 values (unique value, or range of values) in the column “Runoff model efficiency”; does these values related to the 75 assessments cited above?

Table 2: i) the first line of the Table indicates that there is 33 studies for Level 1 while it is indicated in the abstract (P 376, L 6) and in section 3 (P 380, L 15) that there is 34 studies; please clarify. ii) Please indicate the number of results (in brackets) for the three options of Level 2 even if the number of results is equal to the number of studies.

Figure 2: i) It is not evident to check that the total number of points is 75; please indicate on the x-axis the number of points used (same remark for Fig 3, 4, 5 and 6). Ii) Figure 2 shows only one line, while the legend indicates “Lines” and not “Line”. Figure 6: On the x-axis, please put “Number of model parameters” instead of “No of model. . .” Figures 7 and 9: Please indicate the number of studies (and/or results) used for each class (represented on the x-axis).

Figure 9: In the title of the third figure, please replace (aridity index 3 1) by (aridity index > 1).

References:

Dawson, C. W., Abrahart, R. J. and See, L. M. (2007) HydroTest: A web-based toolbox of evaluation metrics for the standardized assessment of hydrological forecasts. Environmental Modeling Software 22, 1034–1052.

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

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

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 375, 2013.