Hydrol. Earth Syst. Sci. Discuss., 10, C892–C894, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C892/2013/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies" *by* J. L. Salinas et al.

J. L. Salinas et al.

salinas@hydro.tuwien.ac.at

Received and published: 17 April 2013

We would like to thank the reviewer for her/his very thorough and helpful comments on the manuscript. We have addressed the comments as follows:

1. Indeed, the reviewer is pointing out a critical point, which already appeared afterwards to the vast literature review that had to be performed in order to find out which of the studies were to be meta-analyzed. A trade-off needed to be found between the amount of studies we wanted to include in the analysis and how really comparable they are. For example, in flood prediction studies, if in two different papers, two different flood quantiles were estimated, even if the performance measure to evaluate the model is the same, can we really compare the RMSE calculated on a 25-year return period peak discharge and on a 200-year one? For the case of floods, the cross-

validation performance was given, in the great majority of the papers considered, as the root mean squared normalised error (RMSNE) calculated on the 100-year specific flood (discharge per unit area, q100) quantile. This is a very common error measure, combining the bias component and the dispersion error of an estimator (see eg. Kottegoda and Rosso, 1997, pp 243-244(1)). Being the error terms normalised (relative to the value calculated with local data) allows us to compare the error measures for the different studies. The fact that the error measures were calculated on the 100-year quantile is not surprising, as it is a very common designt return period, and is usually a reference value in flood frequency analysis. For the case of low flows, the performance were given, again in the great majority of the papers considered, in terms of the coefficient of determination (R2) in cross-validation for the 95% specific discharge percentile (q95). This is mainly due to the fact that in the field of low flow estimation in ungauged basins, one of the preferred methods is regression (global and/or regional), as one can see from Fig.3, and the R2 is a traditional and well studied performance measure of regression analyses. In this case, the coefficient of determination tells us the amount of explained variance by the model, and is also affected by both bias and dispersion of the estimators. Summarizing, for the two cases (floods and low flows), well-studied, easy to interpret, and commonly reported in the literature performance measures were used. To address the question if the results will be similar if other performances or low flow/flood indexes were used, most probably they will, if one compares them in a consistent way as we tried to do in this study. This actually the rationale behind the level 2 assessment, by comparing the errors catchment by catchment we get consistent results with the level 1, meaning that is independent on the overall/aggregated performance measure considered. All these explanations on the choice of the performance measures and low flow/flood indexes will be included in the corrected manuscript.

(1)Kottegoda and Rosso (1997) Statistics, Probability, and Reliability for Civil and Environmental Engineers. McGraw-Hill, 735 pp.

2. This issue is partially answered in the comments to 1. For clarity, we will include

a couple of sentences indicating that the performance and error measures analyzed were directly taken from the original studies, in the way the authors reported them. To address the question of what we can learn from the median, minimum and maximum values of the criteria functions, we will explicitly say that this actually the motivation for the level 2 assessment, in the sense that the performances reported in level 1 are aggregated, regional measures, and that we could learn much more from the catchment-by-catchment errors as in level 2.

3. The authors completely agree on the uncertainty of the flood and low flow values. And even one step further, when we show the formulas for the error/performance measures and define the "observed" discharges (Qobs), in all the studies considered, the observed values correspond to low flow percentiles or flood quantiles estimated from local data, which have a huge uncertainty themselves as well. In other words, the performances are calculated on uncertain values, on "moving targets" The authors will make a comment on these issues in sections 3.1 and 3.2

4. Number of studies in each subclass will be included in the main text. Also an explicit reference to the high and low performances, and the reason for them in terms of climate or model will be made.

Other comments:

CV was changed to coefficient of variation

Tables:The different low flow and flood indexes will be defined in the main text. The reason for ranges or various values for R2 or RMSNE come from the original papers, where slight variations of the methods were applied, or the same method on different subsamples of the same region. These explanation will be included in sections 3.1 and 3.2 as well.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 411, 2013.