Hydrol. Earth Syst. Sci. Discuss., 8, C2380–C2383, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C2380/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Comparison of catchment grouping methods for flow duration curve estimation at ungauged sites in France" *by* E. Sauquet and C. Catalogne

E. Sauquet and C. Catalogne

eric.sauquet@cemagref.fr

Received and published: 23 June 2011

1) p. 3242, lines 4-6: If a model is able to generate negative values, this would represent a serious limitation of the models presented to estimate streamflow at an ungauged location. Is there a more formal way that these methods could be constrained rather than by simply replacing negative values to 0.001?

==> Negative values could be found when applying Eq. (2) and Eq. (4) when p is close to 1. One possible option should have been to consider additional constraints (to use for instance Qp(i) = $(b(i))^2$ + a(i)ln(p) instead of Eq.(2)). Several supplementary tests

have been performed and results show that these more sophisticated models do not fit better than the analytical models applied in this paper. The other models provide positive values and are enable to provide zero flow. We may conclude that none of the model is adapted to represent flow intermittence. We decided to replace negative values by 0.001 since it is the level of precision considered in the HYDRO database.

2) I think the use of the concavity index is a clever and interesting way to think about categorizing regions of similarity between flow duration curves.

==> Thank you for this comment.

3) Please add more description to figure 9 in the caption and axis text and add the correlation values to the figure. It is very difficult to see how figure 9 support the statements made on p. 3250, lines 1-25 without more description on the figure itself.

==> We have modified the caption and the figure as well.

4) Because different numbers of catchments were members of the groups, was some weighting applied to the regressions to handle the unequal sample sizes? Or do the authors feel the results are not sensitive to unequal sample sizes used to adjust the regression models?

==> It is difficult to asses the effect of the number of basins involved in the regression and we have not investigated this issue. Our belief is that results are mainly sensitive to the degree of homogeneity within each region. When the number of sites is high, there is certainly an augmented risk to create heterogeneous regions. We do not include weights since no study on the effect of the cluster size was performed in this application.

5) In the discussion about canonical correlation analysis (CCA) on p. 3251-3252, it is interesting that CCA resulted in such poor results for this paper. I appreciate the authors looking into this issue but wonder if more could be said about why CCA performs so poorly. Is there something about flood data that makes CCA well-suited for

application to that problem but not for flow duration curve estimation?

==> The text has been modified.

6) It is interesting that from figures 11 and 12, the visual grouping does not appear to provide a dramatic improvement in the regional regression model. Is this apparent marginal improvement due to the fact the top left graphs in figures 11 and 12 do not have the same scale for the y-axis? An alternative approach to evaluating the grouping would be to compare the results to a random grouping where groups were determined at random and regressions developed from those random groups.

7) p. 3251, lines 24-26: I think that the statement that regression tree performed the best is hard to conclude from figure 11 and 12. Add supporting evidence for this statement. A mention is made of the errors in visual grouping being heteroscedastic but no supporting evidence is provided and, ultimately, I feel this argument is not strong enough to completely discredit the method if the cross-validation results show otherwise.

==> We did not consider the same y-axis for the top graphs in the three last figures since our final conclusion is based on the relative performance of each method measured by the absolute errors calculated for each quantile (the bottom graphs have the same y-axis to facilitate visual comparisons). There is a slight improvement when RT is considered instead of VG (see the whiskers of box plots especially for extreme quantiles). In addition the two criteria RMSE and R^2 demonstrate more objectively the little benefit from using RT.

We test a random grouping. The dataset of 1080 gauged sites was randomly divided into 20 regions. Multiple regressions for each cluster were derived. The cross validation results indicates poor predictive performance of the randomly selected clusters. It confirms the advantage for using VG and RT (see Figure).

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 3233, 2011.

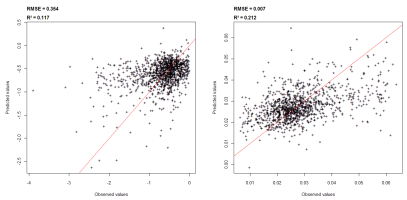

Fig. 1. Cross validation with random grouping