

## ***Interactive comment on “Are streamflow recession characteristics really characteristic?” by M. Stoelzle et al.***

**M. Stoelzle et al.**

michael.stoelzle@hydrology.uni-freiburg.de

Received and published: 20 December 2012

Referee: This manuscript compares the results obtained from recession curve fitting for different (a) selection criteria for recession periods and (b) curve-fitting methods for the selected recessions. The findings are interesting, but the presentation could be improved in several ways.

Primary concerns: 1) Say more about how the differences in recession timescale and inferred storage are logical consequences of the methods selected. For example, fitting a lower envelope (Brutsaert) will by design give longer recession timescale than fitting to an average recession rate (Kirchner). As another example, because the relationship between  $Q$  and  $-dQ/dt$  is in fact not linear on a log-log scale, excluding high flows (the

C5917

first few days after peaks) would be expected to affect the retrieved linear regression parameters  $a$ ,  $b$  in a systematic way. Thus, the discussion should as much as possible explain the differences seen in terms of the intrinsic properties of the methods employed.

Authors Reply: We agree that estimations of recession time and storage depletion depend on the intrinsic properties of the different methods, which focus differently on different parts of the recession. In fact, this is one of the aspects we want to highlight with our comparison in order to debunk the notion that one could just choose any method as the relative ranking of catchments would remain the same. Interestingly the catchment ranking among the original recession analysis methods is not at all consistent (cf. p. 10576, lines 3–6 and Table 1), thus the question arise how reliable the estimation of one recession characteristic (e.g. storage depletion) with a certain RAM could be (regardless which fitting model is used). Further on, we agree that parameters  $a$  and  $b$  are affected in a systematic way by different extraction procedures (e.g. higher slopes will more or less lead to lower intercepts), but we also found an inconsistent catchments' ranking (low to moderate Spearman's  $\rho$  in Table 1 of the discussion paper) and a wide range of slopes (also with moderate adj.  $R^2$ ) in the scatterplots of calculated storage depletion from each combination of RAMs (Fig. 4 in the discussion paper). These results suggest that it is hard to reveal a direct link between properties of RAMs and corresponding model parameterization. We will improve the description of the link between intrinsic methodological differences and our results.

2) Along the same lines, the different methods were devised under different theoretical assumptions and largely for different purposes. It probably is not fully correct to say that the differences found between methods "elucidate a considerable uncertainty", since these differences are not some quasi-random scatter; rather, depending on the application, one of the methods may well be the most appropriate one to use to estimate a recession curve, while others are inappropriate.

Authors Reply: We agree that the word "uncertainty" is not correct and as a conse-

C5918

quence the mentioned paragraph in the conclusion could be rephrased (page 10583, line 26 – page 10584, line 2): "We suggest that the limited comparability of recession characteristics derived with different RAMs elucidate the distinctiveness of individual analysis methods. . ." However, the attached distributions of slope  $b$  (Fig.1) indicate that neither the extraction methods nor the fitting models lead to distinguishable estimations of parameter  $b$ . As all methods have strength and weakness and as the real recession parameterization is unknown it seems to be problematic to find one, single adequate RAM for a certain application. For example, a lower envelope (LE) provides a much wider distribution of estimated slopes  $b$  compared to the linear regression fitting model (REG). This suggests that, for example, a estimation of spontaneous groundwater outflow with LE-model might lead to a much higher slope  $b$  than the other fitting models REG and BIN (for the same catchments), thus to faster receding streamflow recessions, even though the lower envelope should represent the lowest decline in a  $dQ/dt$ - $Q$ -plot for a certain value of  $Q$ . This hampers the choice of the best method for a certain application as the bias of one method seems to be too large to distinguish the different parameter estimations.

Secondary concerns: 1) Give more characteristics of the catchments used (area, climate, land cover) and ideally a citation for the streamflow data.

Authors Reply: We will add more catchment characteristics, information about climate and land cover and a reference for streamflow data to the revised paper (Paragraph 3, page 10574, lines 4-11): We used daily streamflow data (1971-2009) of 20 meso-scale catchments (between 26 and 954 km<sup>2</sup>) in the state of Baden-Württemberg in Germany (Fig. 2). Apart from catchment areas varying two orders of magnitude they also represent a wide range of physiographic and hydrogeological characteristics such as different geology (e.g. metamorphic, limestone, sandstone), drainage density (0.2 - 1.8 km<sup>-2</sup>), mean slopes (5-38%) or mean altitudes (226 – 850 m a.s.l.). The three common land-covers are forests (20-92%), greenlands (2-36%) and crops (0-66%), urban areas within the catchments are negligible. All catchments can be classified as

C5919

humid with annual precipitation ranging from 770 up to 1710 mm, some of these are partly snow influenced. The runoff regimes are mostly snow-rainfall dominated with a peak in spring and typically summer low flows.

2) Is the  $\rho$  at the beginning of the last paragraph on p. 10575 different from the  $\rho$ -bar that appears afterward? Also, should be "Spearman's" in the Table 1 caption.

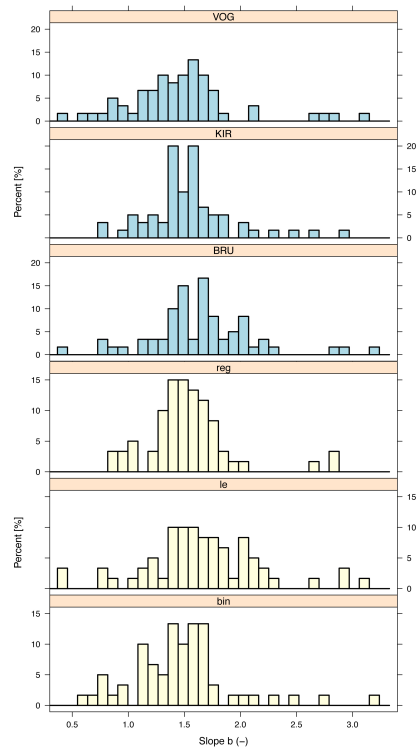
Authors Reply: Yes, these  $\rho$ 's have different meanings. The  $\rho$ -bar is the calculated mean out of 3 values within one extraction procedure. We will rewrite the sentence in the revised manuscript: "Spearman's  $\rho$  is highest within KIR (with a mean  $\rho$ -bar of 0.92 for all three model fittings), but ranking is also relatively good for BRU ( $\rho$ -bar = 0.88) and VOG ( $\rho$ -bar= 0.82)." We will also correct the caption of Table 1.

3) Fig. 3: If there are only 20 watersheds, how can there be multiple outliers from the "whiskers extending to upper and lower 5% percentiles"?

Authors Reply: As the estimation of intercepts depends on the derived slopes we will remove the intercept-boxplot from this figure. The colored boxplots (Fig. 3 in discussion paper) span interquartile range, the whiskers extend according to the R-manual for the boxplot()-function to the most extreme data point which is no more  $n$  times ( $n=3$  in our study) the interquartile range from the box. Points that lie outside this range are marked as outlier (crosses). Perhaps the term 'outlier' is misleading. We will change the caption accordingly to simply state what crosses here mean.

4) Fig. 4: Give the same number of significant digits (e.g. 2) for the regression coefficients in all cases.

Authors Reply: We will provide a revised version of Figure 4 with slightly differing values for the Kirchner extraction procedure due to a calculation correction of the weighted linear regression (KIR.le, KIR.reg, KIR.bin) and the same number of significant digits for all regression coefficients.



**Fig. 1.** Relative frequency of slope  $b$  for different extraction procedures (VOG, KIR, BRU) and different model fittings (REG, LE, BIN).