Response to comments of Referee 2

We thank the anonymous referee for the constructive comments which will help to clarify and improve the paper. The comments of the referee are answered below. (See also the general comment).

Main comment 1 and 3

See the general comment to Prof. Sivapalan and the Referees.

Main comment 2

The same comment as one of the detailed comments, see response below.

Detailed comments:

p. 9474 l. 16-19. I do not quite understand why inverse-distance was chosen. If spatial correlation of rainfall is low (proven by low correlation between time series of nearby stations), is it not more sensible to maintain as much variability in the time series? This would plead for a more conservative interpolation approach such as nearest-neighbour.

This method was chosen instead of geostatistical approaches because of the low correlations (Westerberg et al, 2010 as cited and references therein). We also considered comparing the results of IDW with Thiessen polygons initially but as the spatial configuration of the stations changed frequently over time (as a result of the varying availability of data at the stations), the polygons and therefore the weighting of the individual stations would then have changed more substantially over time than with the inverse-distance weighting. In the end, the effect of the interpolation method on the catchment mean areal precipitation used as input to the model would likely have been small.

p. 9478 l. 14-16. By summarizing all information into FDCs, the temporal autocorrelation of the hydrograph is lost. I would at least like to see this issue and its impact on parameter identifiability discussed in the last section. Again note that other authors have considered the use of auto-correlation (e.g. Montanari and Toth, 2007; Winsemius et al., 2009)

(Please see also the general comment and the response to Referee 1 on similar comments.) In the catchments that we studied here, this was not found to be a problem. The parameters in the two models which mainly control the recession periods were well constrained and showed the largest difference in parameter identifiability compared to the Nash-Sutcliffe efficiency, as noted on p 9485 . The falling limbs and throughs of the hydrograph were in both cases more accurately modelled using the FDC-V calibration (fig. 12 and 14), and the simulated hydrographs did not show any problems with this aspect (fig 11 and 16). We will update the discussion of this aspect as noted in the general comment.

p. 9478, l. 16-18. The triangular evaluation function: why was this function selected? Given the uncertainty, is it not more plausible to simply accept all sets within the evaluation points as equally likely?

We do not agree here. In the same way as a statistical distribution will reduce in likelihood away from some central tendency, we would also expect to have more belief in the "best estimate" than at the edges of the range considered.

p. 9479. – p. 9480. Eventually, I understood the selection of EP methods, but it would help to describe these in an equation. The second method: it seems to me that you can expect an unreasonably high density of EP points in the low regions of the flow regime. It seems to me that it is more objective to determine the EPs (and thus the weight of the evaluation on different parts of the flow regime) by the amount of samples rather than the amount of volume. Can you discuss this?

We will revise the wording of this section to clarify (as also mentioned in the response to Referee 1). Other approaches to the EP-selection can of course be considered (as also discussed in the paper). We are not completely sure how the Referee means that the weighting by the amount of samples should be done. If the EPs are chosen by a binning by frequency then that will result in more EPs in the low-flow region, not less, and the highest flows which only represent a small fraction of the flows will be missed. Some type of inverse weighting by the amount of samples could of course be considered and might be tested in future studies (although one would not like to put too much weight on the extremes of the FDC as these are more uncertain, as discussed in the paper). We compared discharge-interval EPs with EPs chosen to reflect the contribution to the overall flow volumes (which might be more directly important in water resources management than flow itself) and showed that the latter gives generally better results. Since small discharges provided small contributions to the flow volume this gives less EPs in the lower part of the flow regime, not more.

p. 9479, l. 22-23. I recommend removing references to a commercial package, unless it provides unique functionality.

OK, done

p. 9489., l. 21. The approach cannot be presented as being fully new.

We partly agree and will change the text accordingly, see general comment posted separately.

p. 9489., l. 4. How are input errors accounted for in this study? The uncertainty of precipitation has not been accounted for.

Line 4 on p. 9489 does not seem to be the right reference to this comment?? We suspect that the comment refers to "(3) influence of input/output errors of an epistemic nature" on line 24 on the same page. We did not mean that we consider input errors explicitly but as discussed in other parts of the paper we think the method is more robust to disinformation in the type of occasional mismatching events in the input and output series, e.g. a peak in discharge but no measured precipitation input for that event.