Comment on the novelty of the approach

We thank the two Referees and Prof. Sivapalan for their insightful comments. Here we respond to some general comments, responses to the detailed comments by the referees are given separately.

All the referees and Prof. Sivapalan suggest that the proposed methodology should not be presented as entirely new and mention different related previous studies on the use of different hydrological signatures in model calibration and evaluation as well as other papers that have taken account of discharge uncertainties in model calibration. We believe that our paper is novel in combining FDCs and discharge uncertainty in calibration (previously used by Blazkova and Beven, WRR 2009, as cited) and demonstrating the advantage of using the FDC with volume-weighted evaluation points. On the other hand, we agree that some references to previous work were missing and we will revise the text in accordance.

Comment on lack of timing constraint

The referees are also concerned about the loss of timing information in the FDC relative to using a measure such as the Nash-Sutcliffe efficiency (NSE). This is a valid concern and we agree that it can be surprising how well the FDC calibration worked despite of the lack of explicit timing constraint. We agree that the paper will benefit from discussing this more extensively than before in the discussion. The main objective of this work was to obtain good simulations for water-resource evaluation purposes and not exact timing of peak flows. On the other hand, our results demonstrated that the FDC calibration can produce good reproduction of hydrograph shapes and timing. In the case of the daily model there was no difference in timing between the NSE- and FDC-based calibration. In the case of the hourly model the duration of the peak flows was more uncertain for many events for the FDC-V measure compared to the NSE-derived simulations, although the difference was not very large. The FDC-based simulations in this case included individual simulated hydrographs of similar magnitude as the peak but with a time lag, whereas individual NSE-derived behavioural simulations that had poorer timing consistently underestimated the peak flow. In cases where a larger uncertainty in the timing of the peak flows is not acceptable, additional limits of acceptability on the timing of the peak flows could easily be imposed. It can also be mentioned that an FDC-based evaluation is less sensitive to biases in the timing of precipitation data due to observation practices. We would not, and have not, argued that the advantages demonstrated are universal to all catchment characteristics or model applications, though we think that the approach would be a very useful constraint for most applications. As discussed in the paper, the approach did not provide good results when applied to catchments with snow accumulation and melt. Future applications will reveal for what other types of catchments and models this criterion is not sufficient on its own and where additional criteria are necessary. We will update the discussion on the timing and where the FDC, as a single criterion, might (not) work as well as the objectives. The part in the discussion on the need for other measures in some cases (e.g. like those used in the Blazkova and Beven paper) will also be updated. (See also further discussion of the timing constraint in the response to Referee 1).