Response to comments of Referee 1

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

Identifiability of routing parameters

(Please also see the discussion in the general comment). In the two catchments we investigated in the paper, the information in the FDCs was sufficient to constrain the model parameters to obtain realistically simulated hydrographs. At least for the types of model structures and catchments where we tested this approach, an implicit constrain on the timing was made when the model was forced to reproduce the distribution of discharge within the limits of acceptability. This is likely because the distribution of discharge is conditional on the distribution of the precipitation input in magnitude and time. If the routing parameters had been completely wrong this would not have resulted in an acceptable distribution of discharge, so in many cases there would be some constraint on the timing already in the distribution of discharge combined with the distribution of the inputs in time and magnitude. If there is a component in the model and catchment that delays the runoff response to the input (e.g. snow storage or perhaps significant deep groundwater contributions) the FDC calibration would not work as well on its own (we have tested this for snow storage). If there is a more uniform distribution of precipitation and discharge this might also be more problematic for the FDC-calibration (this remains to be tested) – future studies should be made to clarify in which cases the FDC alone is sufficient as a constraint. As suggested in the general comment there is no reason why it should not be combined with other constraints if needed, which could easily be applied within the framework of limits of acceptability used here. We do not suggest using the FDC-calibration blindly without checking the correspondence with the observed discharge, e.g. with the type of posterior analysis we made, to assess whether the simulated hydrographs are acceptable or whether additional constraints are needed depending on the requirements of the study.

We will update and expand the section in the discussion that discusses the limitations of the FDC-calibration and clarify the objectives.

Comparison to traditional Nash-Sutcliffe efficiency (NSE)

We used the traditional NSE as this is a metric that is very sensitive to timing errors and as it is familiar to most hydrologists. The aim here was to show that the FDC calibration on its own could result in simulated hydrographs that had a good match to the observed data. The main concern of the Referee with the FDC-calibration is that the timing is not constrained, and therefore we think it is more relevant to compare with simulations by the untransformed NSE that are known to be very sensitive to timing errors. Better time-series based hydrograph criteria than the NSE approach could of course be used, such as a transformed NSE or a multi-criteria calibration but this still does not address the problems of how to handle uncertain discharge data, to define a threshold for what criteria values are acceptable, or non-overlapping model input and evaluation data. Furthermore we don't agree with the Referee that the comparison to NSE was insufficient as we used two different overlap measures, different behavioural thresholds values, comparison of the scaled scores for five

different aspects of the hydrograph and an analysis of mean scaled scores and the simulated hydrographs in addition to the FDCs in Figures 8 and 9.

Comparison with time-series based approaches

We already cited a number of papers (within different modelling philosophies) that explicitly account for discharge uncertainty in hydrograph-based approaches in the introduction of the paper. As noted in the general comment we will add more citations to other relevant work and we will expand the discussion on the benefits and limitations of the FDC-calibration. The reasons for using the NSE-metric are given above.

Final comment on rescaled FDC

The final comment appears to be a misunderstanding as this was simply a weighting by volume to identify the evaluation points on the normal FDC. It can also be seen as a weighting by the area under the FDC which represents volume of water contributed by flows smaller than a certain magnitude when considering the time step. This will be clarified in the paper.