Response to Editor comment

I would like to thank the three reviewers for their excellent comments and suggestions. I would like to accept the paper, but pending major revisions. In particular a correctly revised paper should
1. Better establish the connection with previous work, particularly on the idea of using Flow Duration Curves for model calibration and model evaluation. The review of Prof. Sivapalan contains important points and suggestions that should be forcefully addressed in a revision. Also, an earlier paper by Winsemius also uses a signature based approach to rainfall runoff modeling, but considers explicitly uncertainty in the original data used in the FDC. Actually, other Bayesian methods explicitly consider data uncertainty as well. Data assimilation papers provide detail about this. Main point, the ideas in this paper cannot be presented as novel and new, and the paper should do a much better job in mentioning and discussing previous contributions. If this is not done, I suspect the paper will not reach its full potential.

We have major revisions to the paper, including adding a wider range of references to previous work, in addition to those that we had already cited concerning the previous use of FDCs in model calibration and Bayesian approaches to model calibration. We have clarified the novelty of the approach presented in the way that it takes uncertainty in the discharge estimates into account. The method has also been placed in the context of using other measures in the rewritten Discussion and Conclusions section of the paper.

2. Establish the limited information content of Flow Duration Curves for inferring rainfall - runoff model parameters that consider timing and onset of the hydrograph. In response to this comment, I have done a few calibrations myself (other model and data) and have noticed that the proposed procedure works relatively well for simple (low parameter) problems, but exhibits difficulty inferring the appropriate parameters for more complex models.

We have demonstrated that the FDC calibration can provide as good results as NSE-based measures for models using daily time steps, but that timing is an issue for sub-daily time steps. We have noted that these conclusions might not necessarily be general to all catchments. The original discussion on timing limitations has been extended substantially, also mentioning that additional criteria could be needed for less parsimonious models, or where peak-flow timing is an important requirement.

3. How would the presented results be if the discharge data are transformed prior to the analysis? Are the findings still the same? Indeed, the NSE has several shortcomings as it comes to low and intermediate flow (NSE emphasizes peak flow), and therefore a transformation of the data might yield completely different conclusions.

We have added results on the use of transformed discharge measures that are now discussed in the Discussion and Conclusions section.

4. The NSE threshold used to differentiate between acceptable and non-acceptable behavior is subjective and strongly influences the outcome of the analysis. Moreover, major improvements have been made in the past decade with respect to model calibration, and very few studies rely on the NSE only to retrieve the appropriate parameters. A study by Yapo et al. (1996) uses multiple measures (BIAS, RMSE, etc) simultaneously. Therefore, NSE might be an insufficient benchmark; and other studies are available that have much better treated discharge error. Non parametric approaches are available that estimate the discharge error from the measurements itself. I also would like to refer to a study by Schaefli and Gupta that have clearly established the limitations of

NS type of criteria. I am looking forward to a manuscript that correctly incorporates the many suggestions of the reviewers and the main points considered herein.

We fully understand the limitations of the NSE as a calibration measure (having already contributed to this debate elsewhere). The limitations of NSE as a measure and the study by Schaefli and Gupta as well as other studies were already cited in the original version of the paper. However, the NSE is only being used as a comparison in this study since it is sensitive to timing errors which is expected to be the main concern with using the FDC. Within the application of GLUE we have used different thresholds to illustrate the effect of this choice. However, the methodology that we are advocating in this paper does not have this limitation since the Limits of Acceptability are set based on the discharge uncertainties independent of any model runs.