Response to Referee 2: A. E. Sikorska

This paper presents an approach to constrain prediction uncertainty in water-balance modelling for ungauged catchments by means of regionalized flow duration curves. Specifically, the authors investigated parametric uncertainty of a simple hydrological model, uncertainty in observational data and in the regionalization method. The analysis is based on the comprehensive dataset of 36 basins in Central America with the area ranging from 132 to 8579 km2 and with long term discharge records from 1965-1994 years.

Generally, the paper is well organized and constitutes a significant contribution to hydrological studies because across the world a significant portion of catchments remains ungauged. However, I have a few specific comments to the authors that, I believe, will help improving the manuscript.

Reply: We thank A. E. Sikorska for her positive comments about the manuscript and the specific comments that helped to improve the manuscript.

1) The approach is tested with a water-balance model, WASMOD. The parametric uncertainty of this model was estimated by sampling randomly parameter values from the defined ranges (Sect. 3). The choice of sampling ranges, however, is not well justified neither in this paper nor in the previous one (Westerberg et al., 2011). The selection of sampling ranges can play an important role in the estimation of prediction uncertainty. Furthermore, model parameters for all catchments are always sampled from the same ranges. Should you include any weighting factor for model parameter priors depending on some catchment characteristics such as a catchment area?

Reply: We agree that the selection of the parameter ranges can play an important role. In the previous paper there was a well-defined peak in the response surface for all parameters, but for some of the parameters we agree that this choice of ranges was not necessarily the best for all catchments in this study, where we are also using a different time period and lower-quality regional datasets. We therefore re-ran the model for all catchments with wider intervals for the fast-flow parameter ([e^{-11} 1]) and slightly wider bounds for the slow flow parameter ([e^{-12} 1]) (the routing and the evaporation parameters were already set to their maximum intervals). We also increased the number of Monte Carlo runs to 150,000 simulations for each basin when using the wider bounds. This did not change any conclusions from the analyses or the main patterns in the result analyses (fig. 9, 11 and 12), but resulted in smaller changes to the uncertainty bounds for most catchments in the local (fig 10) and regional simulations (fig 13), with sometimes wider bounds and a few behavioural simulations were found in two basins with inconsistent data (Guatuso, and Guayabilas) that had none previously. In the revised version of the paper we will use the updated simulations with the wider parameter intervals.

We agree with the reviewer that it would be interesting to have prior parameter ranges that depend on catchment characteristics; however this would require a regionalisation analysis that is outside the scope of our paper. We will clarify this in the revised manuscript. Climate characteristics such as aridity might be an important characteristic for such a regionalisation; however, an added complication that needs to be considered in such an analysis is that the setting of the parameter ranges would then also be affected by disinformation in the datasets.

2) In the discussion (line 11 p. 15704) the authors state that the precipitation-data quality was probably the most limiting factor in uncertainty estimation. This is an important statement because

most of catchments suffer from the lack of sufficient rainfall information. Recent studies have showed that the uncertainty in precipitation data strongly influences simulation results (e.g. McMillan et al., 2011). Although, the authors are aware of that, this needs some more emphasis and some recommendations in this respect could be given.

Reply: We stated that precipitation-data quality was probably the most limiting factor based on the results from the data-screening analyses in which we identified many datasets with inconsistent data. In many of the catchments with low correlations between CPI and discharge there were obvious mismatches between peaks in precipitation and discharge (e.g. Fig. 10). There is a high spatial and temporal variability of precipitation in Central America, resulting from the interaction of many different precipitation-generating mechanisms with the high mountain range that stretches through the region (see section 2.1 and references therein). In addition, quality control of data at the local scale has been identified as important, with as much as 22% of the daily precipitation dataset in a previous study using 60 gauges for a catchment in Honduras being rejected because of poor quality (Westerberg et al., 2010). When making analyses for a long time period for a larger region such as here, one should also expect non-stationary errors in the data as a result of different number and types of gauges being used for different time periods, as well as fewer gauges being available for the regional scale compared to a detailed local dataset. We found that our methods for analysing data information content through the screening procedures were important to use, and we recommend using such analyses also in other studies. We will add this recommendation to the revised manuscript. We will also restructure the discussion section so that the part about data screening follows immediately after the section about precipitation data uncertainty, thus giving more emphasis to this problem.

3) Based on the results and Fig. 7, using information from more catchments in the regionalization method leads to the increase in prediction reliability and to the decrease in prediction precision. In this regards, a choice and a number of selected catchments and cross sections may be of the essential relevance. This is an important issue when translating the method into another study and should be discussed.

Reply: In using the method for other basins we recommend performing the same cross-evaluation of the effect of the number of hydrologically similar catchments used in the FDC-regionalisation as shown in Fig. 7, to justify this choice. We have added a sentence about this in the discussion section. While general guidelines on this question would be valuable, we do not think these can be derived from our study alone. More similar studies are needed to relate the optimal number to station density and variability (incl. climate, geology, land use, etc…)

4) Although, generally the paper is well written, I share the first Reviewer's concern that the Sect. 6, i.e. Discussion and concluding remarks, is too long and slightly repetitive. This makes it difficult to follow and decreases the overall strength of the take home message. I would recommend to rewrite this section by splitting it into two separate subsections. I would also expect summarising recommendations for using the method and its usefulness for other studies.

Reply: We agree that this section needs rewriting and will shorten and restructure it into several subsections accordingly. We will add a recommendation about the evaluation of the FDC-regionalisation (see reply to the previous comment) to the existing discussion about the need to try it in a region with better-quality data to be able to draw further conclusions. As stated in the

previous reply we think further studies are needed before conclusive recommendations can be made.

5) My last comment relates to the chosen method of uncertainty estimation, namely the Generalized Likelihood Uncertainty Estimation (GLUE). Although, the methodology of uncertainty estimation is not the focus of this paper, more promising and rigours methods would be more adequate such as Bayesian methods with a realistic likelihood function (e.g. Mantovan and Todini, 2006; Reichert and Mieleitner, 2009; Del Giudice et al., 2013; Evin et al., 2013). I would like the authors to elaborate on that especially when discussing the limitations of their study.

Reply: We agree that the methodology of uncertainty estimation is not the main focus of this paper, but still an important issue. We found a high presence of non-stationary epistemic errors in the input and evaluation data for which there was little information about their absolute magnitudes or character (including rating-curve residuals that vary with flow range and non-stationary rating curves but lack of site-specific information, and substantial non-stationary precipitation errors and inconsistencies in input-output data combinations). We do not believe that the assumptions behind formal Bayesian likelihood measures that rely on an explicit model of the structure of the errors would be suitable in the presence of these errors, as have been extensively discussed by some of the authors of this study previously (e.g. Beven et al., 2012; Beven and Westerberg, 2011; Beven et al., 2008). We will include a more explicit motivation of the uncertainty estimation method chosen in the revised manuscript, and also refer to the previous debate about this issue in the discussion section in the revised version.

References:

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., and Rieckermann, J.: Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, Hydrol. Earth Syst. Sci., 17, 4209-4225, doi:10.5194/hess-17-4209-2013, 2013.

Evin, G., Kavetski, D., Thyer, M., and Kuczera, G.: Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration, Water Resour. Res., 49, 4518–4524, doi:10.1002/wrcr.20284, 2013.

Mantovan, P. and Todini, E.: Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, J. Hydrol., 330, 368–381, doi:10.1016/j.jhydrol.2006.04.046, 2006.

McMillan, H., Jackson, B., Kavetski, D., and Woods, R.: Rainfall Uncertainty in Hydrological Modelling: An Evaluation of Multiplicative Error Models, J. Hydrol., 400, 83–94, doi:10.1016/j.jhydrol.2011.01.026, 2011.

Reichert, P. and Mieleitner, J.: Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, Water Resour. Res., 45, W10402, doi:10.1029/2009WR007814, 2009.

Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C.-Y.: Calibration of hydrological models using flow-duration curves. Hydrol. Earth Syst. Sci., 15, 2205–2227, doi:10.5194/hess-15-2205-2011, 2011.

References

Beven, K. J., Smith, P. J., and Freer, J. E.: So just why would a modeller choose to be incoherent?, J Hydrol, 354, 15-32, DOI 10.1016/j.jhydrol.2008.02.007, 2008.

Beven, K. J., Smith, P., Westerberg, I., and Freer, J.: Comment on "Pursuing the method of multiple working hypotheses for hydrological modeling" by M. P. Clark et al., Water Resour Res, 48, W11801, Doi 10.1029/2012wr012282, 2012.

Beven, K. J., and Westerberg, I. K.: On red herrings and real herrings: disinformation and information in hydrological inference, Hydrol Process, 25, 1676-1680, 2011.

Westerberg, I. K., Walther, A., Guerrero, J.-L., Coello, Z., Halldin, S., Xu, C. Y., Chen, D., and Lundin, L.-C.: Precipitation data in a mountainous catchment in Honduras: quality assessment and spatiotemporal characteristics, Journal of Theoretical and Applied Climatology, 101, 381-396, DOI: 10.1007/s00704-009-0222-x, 2010.