Reply to Anonymous Referee 3

The manuscript by Westerberg et al. (2013) presents a method to estimate predictive uncertainty in conceptual hydrological modeling of ungauged river basins by using flow-duration curves as information source. The idea is to account for output data uncertainty when transferring parameters inferred in gauged watersheds to similar ungauged watersheds. The methodology for uncertainty assessment combines fuzzy regression analysis and informal inference methods.

In my view the paper is well written and its topic is relevant for the HESS audience since it stresses the need to account for different uncertainty types in hydrological modeling. There are however some critical issues that need to be addressed before publication.

Reply: We thank Referee #3 for the review and the positive comments about the manuscript.

I. The scientific method used for uncertainty analysis is not the most appropriate one. Indeed, after having discussed all the flaws of the GLUE methodology (e.g., Mantovan et al. [2007], Stedinger et al. [2008], Clark et al. [2012]) it is astonishing that this "pseudo-Bayesian" approach is used without any explanation of its appropriateness and shortcomings. It seems necessary, at least to properly justify why this approach has been preferred given the availability of new promising statistical approaches for uncertainty analysis (e.g., Renard et al. [2010], Reichert and Schuwirth [2012]). More importantly, the authors should clearly discuss the limitations of the interpretation of the resulting uncertainty bounds. As Clark et al. [2012] pointed out, GLUE uncertainty estimates appear to lack quantitative significance and the use of "new triangular pseudo-likelihoods" do not seem to solve this problem nor other fundamental weaknesses of GLUE. If the uncertainty intervals are not even intended to encompass the relevant fractions of validation data what is the meaning of these predictions and how can we practically use them?

Reply: The different views on what an appropriate likelihood function should be have been discussed in great detail before (see e.g. Clark et al., 2012; and Beven et al., 2012, and references therein), and we do not think this needs to be repeated here in detail. Whether the structure of the errors that affect the modelling can be described statistically in a likelihood function, or whether they have a more complex and non-stationary epistemic character that cannot be represented by a simple statistical description without overestimating the data information content is an important issue. In the present study, the high presence of non-stationary epistemic errors (about which there is little information about their magnitudes) in both the model input and evaluation data make the informal likelihoods we use particularly suitable, since there is no assumption about purely random errors, or biases of a certain stationary/simple structure. We agree that this motivation could be stated more explicitly in the paper and will include a discussion about this in the revised version in section 6.1.1 and 6.1.3.

With regards to the interpretation of the GLUE uncertainty bounds, these have a clear interpretation with respect to uncertainties in the observed data used to set the limits of acceptability. In this paper the uncertainty bounds are calculated at each time step as the 2.5 and 97.5 percentiles of the likelihood weighted distribution of the simulated discharge of all behavioural parameter-value sets as stated in Section 4.5. The behavioural criteria was set based on the estimated uncertainty in the observed FDC, where every simulation that is inside the estimated uncertainty in the observed FDC at each evaluation point is considered behavioural and

given a weight depending on how close to the best-estimate observed value it is. The uncertainty bounds therefore have a clear interpretation relative to the estimated uncertainty in the observed FDC.

II. The citation of other studies dealing with uncertainty analysis in ungauged basins and concerning errors in calibration data, especially those applying formal statistical methods, is quite limited. In order to present a more balanced view I suggest to discuss at least the following papers:

Honti et al. [2013]: uses a recent Bayesian approach to deal with several uncertainty types (included observation uncertainty which is disentangled from the other contributions) to reliably quantify the uncertainty of flow duration curves and discharge.

Sikorska et al. [2012]: shows how to assess runoff predictive uncertainty in ungauged basins by using autroregressive error models.

Renard et al. [2010]: tries to quantify different uncertainty components in a Bayesian framework by also separately accounting for uncertainties in the measured runoff.

Reply: We agree that these all are relevant papers, but find it difficult to include the whole range of relevant papers on uncertainty analyses. We cite the important paper by McMillan et al. (2012), which reviews different approaches for estimating and accounting for calibration-data (discharge) uncertainties, and we specifically mention two papers for rating-curve analyses for alluvial rivers/non-stationarity that is of particular relevance in our case. The focus of our paper is ungauged basins and specifically the use of signatures in model regionalisation, and we have cited important papers in this respect (including the formal Bayesian approaches of Bulygina et al, 2009 and He et al., 2011). Two of the papers suggested by the reviewer are not about ungauged basins. The Sikorska et al. 2012 paper, about rainfall and parameter uncertainties for a poorly gauged urban basin, will be included at the end of the introduction in the revised manuscript.

Minor points:

i. “Reliability” and “precision” should be also defined in relation to the probabilistic performance measures of “reliability” and “sharpness” (see e.g., Breinholt et al. [2012]). How do these concepts relate?

Reply: The reliability and precision measures were previously used by Westerberg et al. (2011), Guerrero et al. (2013), and Coxon et al. (2013). They are similar to the measures used by Yadav et al. (2007) and Breinholt et al. but differ in that they incorporate the estimate of the uncertainty in the observed discharge data, where that estimate consists of an upper and lower bound that allow for non-stationary biases in-between the bounds (e.g. because of the rating-curve errors that in some cases varied strongly with flow range). We will include this explanation with reference to the Yadav et al. and Breinholt et al. papers as well as the references to the previous papers where the measures were first used at the end of section 4.4.

ii. Define “behavioral simulations”: for researchers not familiar with the previous papers of the authors it can be hard to understand this concept without further explanation.

Reply. The extended GLUE uncertainty estimation method using limits of acceptability as we used here was proposed by Beven (2006), and the method for using it with FDCs as in this paper is described by Westerberg et al. (2011). Instead of using a traditional lumped performance measure, models are considered behavioural or acceptable if they produce simulations inside the observed uncertainty in the evaluation data, in this case the observed FDC. In the paper we explicitly defined the behavioural simulations in section 4.5 "Behavioural simulations were required to be within the limits of acceptability defined from the discharge-data uncertainty at each of the 19 EPs". In order to not increase the length of the already long paper with further explanations of the limits of acceptability method, we have included a reference to the original Beven (2006) paper in section 4.5, in addition to the Westerberg et al. (2011) reference already there, and refer the reader to these previous papers. We will also add a short definition of the behavioural simulations in the abstract.

iii. The Discussion is currently a big block of text. It think it would help understanding it better if the authors would structure it into subsections.

Reply. We agree and will restructure and shorten the discussion section in the revised manuscript.

iv. Section 3 (Model) is not optimally structured: first, the model would fit better in the methods; second, the description of the model structure is mixed with the prior definition and the numerical implementation of the uncertainty analysis routine. I think these three concepts should be separately explained and better organized.

Reply: It is true that Section 3 of the paper is presented concisely. However the details are available in the past papers cited and we feel that sufficient detail is given here in that the concentration is on the use of the model for the regionalisation methodology. We will add a reference to Table 1 in Westerberg et al. (2011) for the model equations to the text in Section 3.

References

- Breinholt, A., Møller, J., Madsen, H., and Mikkelsen, P.: A formal statistical approach to representing uncertainty in rainfall-runoff modelling with focus on residual analysis and probabilistic output evaluation-distinguishing simulation and prediction, J. Hydrol., 472–473, 36–52, doi:10.1016/j.jhydrol.2012.09.014, 2012.

- Clark, M. P., D. Kavetski, and F. Fenicia, Reply to comment by K. Beven et al. on "Pursuing the method of multiple working hypotheses for hydrological modeling", Water Resour. Res., 48, W11802, doi:10.1029/2012WR012547, 2012.

- Honti, M., Stamm, C., and Reichert, P.: Integrated uncertainty assessment of discharge predictions with a statistical error model, Water Resour. Res., 49, 4866– 4884,doi:10.1002/wrcr.20374, 2013.

- Mantovan, P., Todini, E., Martina, M., Reply to comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology", Journal of Hydrology, Volume 338, Issues 3–4, Pages 319-324, http://dx.doi.org/10.1016/j.jhydrol.2007.02.029, 2007.

- Reichert, P. and Schuwirth, N.: Linking statistical bias description to multiobjective model calibration, Water Resour. Res., 48, W09543, doi:10.1029/2011WR011391, 2012.

- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46, W05521, doi:10.1029/2009WR008328, 2010.

- Sikorska, A. E., Scheidegger, A., Banasik, K., and Rieckermann, J.: Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models, Hydrol. Earth Syst. Sci., 16, 1221–1236, doi:10.5194/hess-16- 1221-2012, 2012.

- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder , Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, Water Resour. Res., 44, W00B06, doi:10.1029/2008WR006822, 2008.

References

Beven, K. J.: A manifesto for the equifinality thesis, J Hydrol, 320, 18-36, 2006.

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.

Breinholt, A., Moller, J. K., Madsen, H., and Mikkelsen, P. S.: A formal statistical approach to representing uncertainty in rainfall-runoff modelling with focus on residual analysis and probabilistic output evaluation - Distinguishing simulation and prediction, J Hydrol, 472, 36-52, DOI 10.1016/j.jhydrol.2012.09.014, 2012.

Bulygina, N., McIntyre, N., and Wheater, H.: Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis, Hydrol Earth Syst Sc, 13, 893-904, 2009.

Clark, M. P., Kavetski, D., and Fenicia, F.: Reply to comment by K. J. Beven et al. on "Pursuing the method of multiple working hypotheses for hydrological modeling", Water Resour Res, 48, W11802, Doi 10.1029/2012wr012547, 2012.

Coxon, G., Freer, J. E., Wagener, T., Odoni, N., and Clark, M. P.: Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments Hydrol Process, doi:10.1002/hyp.10096, 10.1002/hyp.10096, 2013.

Guerrero, J. L., Westerberg, I. K., Halldin, S., Lundin, L. C., and Xu, C. Y.: Exploring the hydrological robustness of model-parameter values with alpha shapes, Water Resour Res, 49, 6700-6715, 10.1002/wrcr.20533, 2013.

He, M. X., Hogue, T. S., Franz, K. J., Margulis, S. A., and Vrugt, J. A.: Corruption of parameter behavior and regionalization by model and forcing data errors: A Bayesian example using the SNOW17 model, Water Resour Res, 47, W07546, 10.1029/2010wr009753, 2011.

McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol Process, 26, 4078-4111, doi: 10.1002/hyp.9384, 2012.

Sikorska, A. E., Scheidegger, A., Banasik, K., and Rieckermann, J.: Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models, Hydrol Earth Syst Sc, 16, 1221-1236, DOI 10.5194/hess-16-1221-2012, 2012.

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 Sc, 15, 2205-2227, DOI 10.5194/hess-15-2205-2011, 2011.

Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Adv Water Resour, 30, 1756-1774, DOI 10.1016/j.advwatres.2007.01.005, 2007.