We thank the AE and the referees for their interest and constructive comments on our paper that helped to improve its quality. The response to each referee and the comments of the AE is given in bold text below, with references to the updated manuscript version. A version of the updated manuscript text with tracked changes is attached at the end of this document, which shows how the text has been updated (but no tracked changes for how paragraphs have been moved in the discussion part, for clarity).

Response to Referee 1: D.A. Hughes

This paper represents the results of a very comprehensive study of data and modelling uncertainties in a relatively data scarse region and therefore makes a valuable contribution to hydrological modelling theory and (potentially) practice. In general terms the paper is also well written, but I found some of the explanations of the methods a bit confusing. However, I am not sure that they can be simplified and perhaps they would become clearer if the previous papers are consulted (something I admit that I did not do). I did, however, find that the discussion section seemed a bit long and somewhat repetitive. I would therefore encourage the authors to look at making the finala section more concise trying not to repeat too much of what is already in other parts of the text.

Reply: We thank D.A. Hughes for his positive words about our paper and the constructive comments that helped to improve and clarify the paper. We agree that the discussion section needs rewriting and have therefore shortened it and added subheadings to give it a better structure and clearer presentation (Section 6 in the revised paper).

A final comment relates to the high degree of uncertainty in the simulations (and some of the observed data). I would like to have seen some comments about this in terms of the practical use of water-balance results. Mention is made of robust predictions under different circumstances and the possible need for more regionalised information. What does this really mean in terms of the use of modelling results for '..effective management of these resources' and can such uncertain results be of any value for water resources management? I realise this is not the main topic of the paper, but I do think that some concluding (possibly even speculative) remarks could be made about this issue.

Reply: This is a very interesting point, especially when it comes to predictions in ungauged basins in a region where data inconsistencies can be expected. We accounted for many different types of uncertainties when making our predictions, and in basins where the data were found to be reliable this resulted in generally reliable simulations where the water balance was constrained according to the regionalised FDCs (where FDCs have a long history of use for different types of water management, e.g. Vogel and Fennessey, 1995). The width of the predicted uncertainty was therefore dependent on the uncertainty in the regionalised FDCs and was in the best cases almost equal to that from the local calibration and in the less accurate cases much wider.

The uncertainty estimates from our method give much more information for water management than deterministic model simulations would have had. Having a prediction with high uncertainty is also much more valuable than having no information at all for an ungauged catchment, but when using that prediction for water-resources management the quality of the information that went into making that prediction needs to be taken into account. In using this method for a completely

ungauged basin in this region it would thus be advisable to carefully scrutinise the quality of the precipitation input data to assess potential effects on the predictions.

In the cases where the data were inconsistent, our analyses showed the need for additional information and improved data, which is important knowledge for water-resources management. Since our method would be used for predictions for ungauged catchments in a region with other nearby gauged catchments, much information about the dataset consistency would be found by making the types of analyses for the gauged catchments as we made here and by testing the method in cross-evaluation for the gauged catchments first to learn about the different types of uncertainties that are affecting the simulations. In this region it was found that for many basins the predictions should often not be expected to be accurate for each individual day because of input data errors, which should be kept in mind when the information is used for water management.

We have added some remarks about this to section "6.4 Concluding remarks" Line 710–718 in the revised manuscript.

Other minor comments: The reference to 1000-2500mm lower estimates of precipitation (end of section 4.1) is very important but not perhaps emphasised enough as a major source of uncertainty.

Reply: We agree that this is an important problem for the two basins where it occurred, however this problem of largely overestimated precipitation in the CRN073 dataset only occurred for two Panamanian basins that were clear outliers on the Budyko curve, and was not found to be a general problem. For these two basins no behavioural simulations were found in the local calibration. Since these data inconsistencies were identifiable from the data screening we made, this highlights the value of making such analyses in this type of regional modelling. However, in completely ungauged basins the discharge-dependent data screening methods we used would not be able to identify such data problems, and in the paper we therefore stressed the need to develop data screening methods that do not rely on observed discharge data. We have added a short note to emphasize the precipitation uncertainty in the end of section 4.1, line 297–298.

Line 22 of section 4.5: The sentence 'Simulations with correlation in deviations across successive EPs then obtain a lower weight..' is not very clear to me and perhaps can be better explained.

Reply: This means that a simulation with a systematically over- or underestimated FDC for (part of) the flow range will get a lower weight, but that such simulations are still acceptable. We found it important for allow for such (non-stationary) biases since the data analyses showed that they were frequent in the discharge and model input data. The rating-curve analysis of the Honduran stations showed several stations with under- or overestimated discharge and residuals that varied systematically with flow, and there were also temporally non-stationary rating curves. The screening for dataset inconsistencies and visual analyses of the data series also showed that several stations had likely non-stationary errors in the precipitation data. We have added "…, i.e. a systematically under- or overestimated FDC for (part of) the flow range can still be behavioural but get a lower weight." to the end of this sentence, line 443–445.

Minor errors: Last line of section 5.3 'constraint' should be 'constraints' or 'provide an additional constraint'. Similarly line 4 at the top of the 2nd paragraph of scetion 6 (constraints).

Reply. Thanks, we will change this.

References

Vogel, R. M., and Fennessey, N. M.: Flow Duration Curves .2. A Review of Applications in Water-Resources Planning, Water Resour Bull, 31, 1029-1039, 1995.

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 have used 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 have clarified this in the revised manuscript, line 244–245. 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. We have added some discussion about this on line 684–686.

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 have added this recommendation to the revised manuscript, line 616. We have also restructured the discussion section (Section 6) so that the part about data screening ("6.1.3 Detection and impact of dataset inconsistencies") follows immediately after the section about precipitation data uncertainty ("6.1.2 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 in section 6.3, line 699–701 in the revised manuscript. While general guidelines on this question would be valuable, we do not think these can be derived from our study alone, and discuss the need of further studies on line 701–703. 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 have shortened and restructured it into several subsections accordingly. We have added 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 have included a more explicit motivation of the uncertainty estimation method chosen in the end of section 6.1.1, line 586–591 and the end of Section 6.1.3, line 635–641 in the revised manuscript, and also referred to the previous debate about this issue there.

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.

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 have included a discussion about this in the revised version in section 6.1.1 and 6.1.3 (see also response to A. E. Sikorska above).

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, has been included at the end of the introduction section in the revised manuscript, line 132–133.

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 have included 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, line 426–430.

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 our 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, line 434, in addition to the Westerberg et al. (2011) reference already there, and refer the reader to these previous papers. We have also added a short definition of the behavioural simulations in the abstract, line 37.

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 have restructured and shortened the discussion section in the revised manuscript (see also response to the first two reviewers).

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 have added a reference to Table 1 in Westerberg et al. (2011) for the model equations to the text in Section 3, line 244.

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.

Response to the AE, Patricia Saco

Editor Initial Decision: Reconsider after major revisions (14 Apr 2014) by Dr. Patricia Saco
Comments to the Author:
This manuscript presents an interesting methodology that uses regionalized flow duration curves to constrain prediction uncertainty in water-balance modelling of ungauged basins.

We have received three positive reviews of the manuscript. The reviewers have provided very valuable comments that, if properly addressed, will help enhance the manuscript. They all agree that the paper's contribution is substantial and useful, but also agree that some portions of the manuscript need to be improved. The authors, in the response letters have shown and they have carefully looked at these comments and have stated that the paper will be revised to account for the reviewers suggestions.

In my own opinion the paper is well written and its content is novel and valuable. After looking at the reviewer's comments and my own assessment of the work I consider that the paper will be suitable for publication if the authors adequately address the reviewers' comments, particularly those that will help improve the clarity of the methodology. Some of the major points raised by the reviewers that need to be carefully addressed in the revised manuscript are:

We thank the AE for the positive comments about our paper.

1) Reviewer 1 has pointed out some lack of clarity in the methodology section and has made some suggestions for improvement. D.A. Hughes also suggests including a discussion on the use of "uncertain results" for water resources management that will be extremely useful for the readership of HESS.

We have clarified several methodological issues in response to all the reviewers' comments and revised and restructured the discussion part in section 6 as detailed above. We have also included discussion about the use of the uncertain results in section 6.4.

2) Reviewer 2, has listed a number of specific points that will help clarify the methodological approach (including selection of parameters ranges, a better explanation on the effect of precipitation uncertainty on simulation results, and clear explanation on the transferability of the methodology to a different study area).

We revised the parameter bounds and re-ran the model with the new wider bounds such that all the results in the updated manuscript version, and the related figures, are based on the updated simulations. This resulted in some minor changes to the results but did not change any conclusions about the method. We have also discussed the role of precipitation uncertainty and made some

recommendations regarding the use of the method in other areas, however, this is a first study and in our opinion more studies are needed before more conclusive recommendations can be made.

3) Reviewers 2 and 3 have both pointed out the need to discuss the existence of alternative approaches for uncertainty estimation, and a justification for the selection of the methodology used in this study. It will indeed be beneficial to briefly discuss these alternative approaches, the rationale for the selection of this particular approach, and to include the appropriate references in the revised manuscript.

In summary, though the revisions needed are only moderate, they should address not only the main reviewer's concerns (listed above) but also their minor comments, as this certainly will help to improve the revised manuscript.

We have included more discussion and justification about the selection of the uncertainty estimation methodology in the paper, with reference to relevant previous papers and the more extensive discussion in these previous papers.

Regional water-balance modelling using flow-duration curves with observational uncertainties

I. K. Westerberg[1,2,3], L. Gong[2], K. J. Beven[2,4], J. Seibert[2,5], A. Semedo[2,6], C.-Y. Xu[7,2], S. Halldin[2]

[1] {Department of Civil Engineering, University of Bristol, Queen's Building, University Walk, Clifton BS8 1TR, UK}

[2] {Department of Earth Sciences, Uppsala University, Villavägen 16, 75236, Uppsala, Sweden}

[3] {IVL Swedish Environmental Research Institute, P.O. Box 210 60, 10031, Stockholm, Sweden}

[4] {Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK}

[5] {Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland}

[6] {CINAV – Escola Naval, Base Naval de Lisboa, Alfeite, 2810-001 Almada, Portugal}

[7] {Department of Geosciences, University of Oslo, Postboks 1047 Blindern, 0316, Oslo, Norway}

Correspondence to: I. K. Westerberg (ida.westerberg@bristol.ac.uk)

Abstract

Robust and reliable water-resources mapping in ungauged basins requires estimation of the uncertainties in the hydrologic model, the regionalisation method, and the observational data. In this study we investigated the use of regionalised flow-duration curves (FDCs) for constraining model predictive uncertainty, while accounting for all these uncertainty sources. A water-balance model was applied to 36 basins in Central America using regionally and globally available precipitation, climate and discharge data that were screened for inconsistencies. A rating-curve analysis for 35 Honduran discharge stations was used to estimate discharge uncertainty for the region, and the consistency of the model forcing and evaluation data was analysed using two different screening methods. FDCs with uncertainty bounds were calculated for each basin, accounting for both discharge uncertainty and, in many cases, uncertainty stemming from the use of short time series, potentially not representative for the modelling period. These uncertain FDCs were then used to regionalise a FDC for each basin, treating it as ungauged in a cross-evaluation, and this regionalised FDC was used to constrain the uncertainty in the model predictions for the basin.

There was a clear relationship between the performance of the local model calibration and the degree of dataset consistency – with many basins with inconsistent data lacking behavioural simulations (i.e., simulations within predefined limits around the observed FDC) and the basins with the highest dataset consistency also having the highest simulation reliability. For the basins where the regionalisation of the FDCs worked best, the uncertainty bounds for the regionalised simulations were only slightly wider than those for a local model calibration. The predicted uncertainty was greater for basins where the result of the FDC-regionalisation was more uncertain, but the regionalised simulations still had a high reliability compared to

the locally-calibrated simulations and often encompassed them. The regionalised FDCs were found to be useful on their own as a basic signature constraint; however, additional regionalised signatures could further constrain the uncertainty in the predictions and may increase the robustness to severe data inconsistencies, which are difficult to detect ~~in~~ for ungauged basins.

1 Introduction

Knowledge about the temporal and spatial variability of water resources is essential for effective management of these resources, for preventing water-related disasters, and for fostering cooperation and avoiding conflict over trans-boundary waters. Mapping of this variability requires hydrologic models in situations where: 1) discharge data are of insufficient quality, 2) predictions are required for time periods with no monitored discharge, or 3) predictions are required for basins without discharge monitoring stations. Model-parameter values and their uncertainty ranges can be estimated by calibration to measured data in the first two cases whereas the last case requires a regionalisation procedure. Discharge data are non-existent, intermittent or non-available for many basins, which make Predictions in Ungauged Basins (PUB) an important prerequisite for comprehensive water-resources mapping (Bloeschl et al., 2013). However, estimating the response of an ungauged basin always involves some uncertainty, and one of the features of the PUB science plan was the development of methods to constrain that uncertainty (Hrachowitz et al., 2013; Sivapalan et al., 2003). In this study we addressed uncertainties in the observational data, the hydrological model parameterisation and the regionalisation method (based on regionalised flow-duration curves, FDCs).

Conceptual water-balance models have traditionally been regionalised by transferring parameter values from gauged to ungauged basins using some measure of hydrologic similarity or a regression with model parameter values as dependent variables and physical characteristics of the basins as independent variables (Seibert, 1999; Jakeman et al., 1992; Parajka et al., 2005; Xu, 2003). Such procedures are often limited by their assumption of model-parameter independence and incomplete assessment of predictive uncertainty for gauged and ungauged basins (McIntyre et al., 2005; Bardossy, 2007; Buytaert and Beven, 2009).

Wagener and Montanari (2011) discuss a convergence of approaches for PUB in recent years where regionalisation is based on the expected functional behaviour of the ungauged watershed rather than the model and its parameters. Watershed behaviour has been quantified in the form of information or “signatures” derived from discharge or other types of data for model calibration in recent studies (Winsemius et al., 2009; Son and Sivapalan, 2007; Yu and Yang, 2000; Castiglioni et al., 2010; Westerberg et al., 2011b; Blazkova and Beven, 2009; Yadav et al., 2007). Many of these studies have been made within a set-theoretic approach for uncertainty estimation (e.g. Blazkova and Beven, 2009; Yadav et al., 2007; Winsemius et al., 2009), but Bayesian statistical approaches have also been used (e.g. Bulygina et al., 2009). The types of information that have been used include recession curves (Winsemius et al., 2009), slope of the FDC (Yilmaz et al., 2008; Yadav et al., 2007), base-flow index (Bulygina et al., 2009), spectral properties (Montanari and Toth, 2007), and flood discharge and snow-water equivalent frequency quantiles (Blazkova and Beven, 2009). Calibration approaches focused on matching hydrological signatures thus allow regionalisation to be performed directly on a wide range of hydrologic information, which is then used to constrain model parameters at ungauged sites. Yadav et al. (2007) regionalise constraints on expected watershed response behaviour in the UK and account for uncertainty in the regionalisation method. Kapangaziwiri et al. (2012) use regionalised signature constraints for runoff ratio (long-term ratio of runoff over precipitation) and slope of the FDC in combination with prior parameter estimation. Yu and Yang (2000) regionalise FDCs and calibrate their model against a performance measure based on specific exceedance percentages of the FDC using an optimisation algorithm.

Uncertainties in observational data affect the information content of data and derived signatures and it is therefore important to estimate and account for these uncertainties also in rainfall-runoff model regionalisation (Hrachowitz et al., 2013). However, as noted in the recent review by McMillan et al. (2012) no studies have so far explicitly investigated the role of observational uncertainties in this context. Discharge-data uncertainty can often be estimated based on rating-curve analyses and has received increasing attention in recent years. Relative errors of around 10–20% for medium to high flows, with higher ranges for low flows (50–100%) and out-of-bank flows (40%) are typically reported (McMillan et al., 2012). The main uncertainties relate to the approximation of the true stage-discharge relation by the rating curve. Discharge data are therefore especially uncertain in alluvial rivers with non-stationary stage-discharge relationships (Jalbert et al., 2011; Guerrero et al., 2012) and for flow conditions outside those used for constructing the rating curve. Model input data, especially precipitation, are also affected by sometimes substantial uncertainties that are more difficult to estimate and may have non-stationary characteristics, e.g. because of temporal changes in the number and quality of precipitation gauges (Westerberg et al., 2010; Brath et al., 2004). In some cases the observational uncertainties can be so large that the model forcing and evaluation data are physically inconsistent (Beven and Westerberg, 2011), e.g. because of inferred actual evaporation greater than potential evaporation (Kauffeldt et al., 2013) or runoff ratios greater than one (Beven et al., 2011). Such data inconsistencies will be “disinformative” in calibration of a model built on such assumptions. Datasets can be screened for inconsistencies prior to modelling (Kauffeldt et al., 2013; Beven et al., 2011), however, identification of inconsistent data might prove difficult in cases where auxiliary information is not available or where disinformation is not easily identified.

The aim of this study was to investigate if regionalised FDCs could be used to reliably constrain water-balance prediction uncertainty in ungauged basins, while estimating and analysing uncertainties in the observational data and regionalisation method as well as the model parameterisation. We used the FDC-calibration method of Westerberg et al. (2011b) together with regionalised FDCs, therefore also testing this method for a wider range of basins than in the previous study. A variety of approaches have been used for regionalisation of FDCs (reviewed by Bloeschl et al., 2013), including the fitting of a frequency distribution (Castellarin et al., 2004) or a parametric equation (Yu et al., 2002) to the FDCs where the parameters are regionalised through regression with basin characteristics as independent variables. Holmes et al. (2002), building on the work of Burn (1990a, b), use a region-of-influence (ROI) approach to predict FDCs for the UK, with a dynamic definition of a ROI based on hydro-geologic similarity. While some studies explore uncertainty in the regionalised FDCs (e.g. Yu et al., 2002)~~,~~ and data uncertainties in snow-model regionalisation (He et al., 2011) and rainfall and parameter uncertainties in modelling a poorly gauged urban basin (Sikorska et al., 2012), none has, to our knowledge, accounted for discharge and input-output data uncertainties in FDC or rainfall-runoff model regionalisation.

2 Study area and data

2.1 Study area

Central America is a region with a highly variable climate in both space and time despite its small extent (around 520,000 km^2). This has resulted in many water-related disasters; flooding with severe consequences such as inundations and destruction of important crops, promulgation of landslides, and loss of lives (Waylen and Laporte, 1999); and sustained droughts with severe consequences for hydro-power generation, water supply, irrigation and tourism (George et al., 1998). The characteristics of the complex regional climate have been well studied (e.g. Alfaro, 2002; Amador et al., 2006; Magaña et al., 1999; Enfield and Alfaro, 1999), but there are relatively few published hydrological modelling studies (but see e.g. Birkel et al., 2012; Westerberg et al., 2011b; Hidalgo et al., 2013). One reason for the scarcity of peer-reviewed literature is the difficulty to access comprehensive and good-quality hydro-meteorological data, and several studies point to the need for data quality control in this region (Aguilar et al., 2005; Westerberg et al., 2010; Flambard, 2003). The regional precipitation regime has a less marked seasonal variability on the Atlantic Coast compared to the Pacific Coast, where around 80% of the precipitation falls in the rainy season from May to October–November (Portig, 1976). There is also a rainfall minimum, the so-called midsummer drought or *veranillo* in July–August on the Pacific Coast, resulting in a bimodal regime with two peaks in June and September–October (Magaña et al., 1999). The spatiotemporal variability of precipitation is high, since precipitation is often convective, and associated with different mechanisms such as hurricanes, tropical storms, and easterly waves in the atmosphere (Peña and Douglas, 2002). Temperature variability is low, with a greater diurnal than ~~annual range~~seasonal variation that is characteristic of the tropics. Climate variability on an inter-annual time scale is pronounced with large differences between wet and dry years; this variability is modulated by ENSO (El Niño/Southern Oscillation) and Atlantic sea-surface temperatures (Diaz et al., 2001; Enfield and Alfaro, 1999).

2.2 Model forcing data

The water-balance model we used was driven with daily precipitation and daily potential evaporation data and calibrated and evaluated using daily discharge. Comprehensive local climate and discharge datasets covering the whole of Central America are difficult to obtain as observation data are either non-existing or cannot be made available with a reasonable effort. We therefore used globally or regionally available gridded meteorological data in this study. In early attempts with the regional model, potential evaporation calculated from ERA-Interim (Dee et al., 2011) climate variables at a 0.75° resolution and TRMM precipitation data (Huffman et al., 2007) with a spatial resolution of 0.25° were used for the period 1998–2009. However, this resulted in inconsistently simulated hydrographs in a few test basins since the TRMM precipitation did not compare well to local precipitation data. We therefore used daily precipitation data from the CRN073 dataset (Magaña et al., 1999; Magaña et al., 2003) at a spatial resolution of 0.5° that covers Central America, Mexico and the Caribbean region for the period 1958–2000. It is based on station data from the national weather services blended with satellite precipitation estimates for the oceans. The station data cover different time periods resulting in time-varying errors and some obvious in-homogeneities could be seen for many stations in the late 1990s, which may result from inclusion of malfunctioning automatic rain gauges. Since the temporal coverage of this dataset did not overlap sufficiently with the potential evaporation calculated from the ERA-Interim data, we used the WATCH Forcing Data (WFD; Weedon et al., 2010) for the period 1958–2000 at a

0.5˚ spatial resolution. The WFD provide bias-corrected variables based on the ERA-40 reanalysis (Uppala et al., 2005) and we used specific humidity, atmospheric pressure, 2-metre air temperature, 10-metre wind speed, net shortwave radiation and net long-wave radiation to calculate potential evaporation using the Penman-Monteith FAO-56 equation (Allen et al., 1998). Specific humidity was first converted to relative humidity using a mixing-ratio method and 10-metre wind speed was converted to 2-metre wind speed using a logarithmic relationship (Allen et al., 1998). Prior to the calculation of potential evaporation, the quality of the WFD data was evaluated using~~by~~ daily weather data (Global Surface Summary of the Day, or GSOD) from the National Climatic Data Center (NCDC, 2011). The comparison was made for 18 half-degree cells spread over the study area, each of which contained at least one GSOD station with at least five years of daily data. The evaluation showed that WFD air temperature and the WFD-derived relative humidity were reasonably correlated with GSOD data although with average biases of -1.7˚C and +6 % respectively. No significant correlation was found between WFD and GSOD wind-speed data, which is often the least sensitive variable for the estimation of potential evaporation on the daily scale. The WFD radiation components showed good agreement when compared with radiation components derived from sunshine hours recorded at the airport in Tegucigalpa, Honduras.

2.3 Discharge data and basin delineation

The discharge data were obtained from the Global Runoff Data Centre (GRDC, 2010), which includes data from 91 discharge stations from all Central-American countries except Belize. Daily data were only available for 77 stations of which none were located in Guatemala or El Salvador. In addition to these 77 stations we included two Honduran stations (Paso La Ceiba on the Choluteca River and La Chinda on the Ulúa River) for which daily discharge and its uncertainty had been calculated using a time-variable rating curve in a fuzzy regression based on estimated uncertainties in the stage and discharge measurements (Westerberg et al., 2011a describe the calculation for the Paso La Ceiba basin). The total period for which there were data for at least one station was 1952–2009, with most of the data available 1965–1994. We used official rating curves and stage-discharge measurements for another 35 stations in Honduras (see section 4.2) to estimate discharge-data uncertainty for all GRDC stations in this study. Paso La Ceiba and La Chinda were included in this dataset together with three of the GRDC stations; but discharge time series were not available for the remainder and they could therefore not be included in the rest of the study.

The GRDC discharge data and the station locations were analysed to select stations with: 1) a sufficient number of years with data (≥5 years), 2) discharge that appeared to have sufficient quality from a visual inspection of the time series, 3) no detected influence from major dams in the basin during 1965–1994, and 4) a location that was not in the basin of another of the stations. Obvious outliers in the series (values orders of magnitudes too large) were removed. This procedure resulted in a set of 36 basins that could potentially be used for regionalisation. These basins (Fig. 1) were delineated from the HydroSHEDS elevation data (Lehner et al., 2008), a gridded global hydrography dataset with the highest resolution (3") publicly available at present. Upstream areas for HydroSHEDS pixels were derived by Gong et al. (2011). The basins were registered in the HydroSHEDS flow network overlaid with 0.25˚x0.25˚ cells. Only the ~~active~~ parts of the boundary cells that were in the catchment, as delineated by the HydroSHEDS pixels, contributed discharge to the downstream gauging station. The GRDC station coordinates sometimes had a low precision and were adjusted to obtain basins with the right basin area using visual inspection of river locations from satellite images and/or coordinates of higher quality from local sources. We used a tolerance of 10%

difference between the area reported in the GRDC database and that obtained from the delineation together with a visual inspection of basin boundaries. Since a large part of Central America is mountainous, the greatest source of uncertainty in basin areas is likely the exact location of the stations and not the precision of the delineation algorithm. While all calculations were made on a depth per unit area basis, uncertainty in catchment area has a direct effect on the water balance calculation. Many discharge series had frequent gaps and the temporal availability of data at the stations varied substantially in the region, with most data available for Panama and the least for Costa Rica (Fig. 2).

3 Regional water-balance model

We tested a simple lumped version of the water-balance model WASMOD (Xu, 2002) that was previously used with good results in Honduras (Westerberg et al., 2011b), and we used the same model equations as in this earlier study. The model has four parameters (sampling ranges for uncertainty estimation given in parenthesis); for actual evaporation ([0, 1] -), routing of fast flow ([0, 1] d~~ay~~$^{-1}$), fast flow ([e$^{-11\cancel{7}}$, 1~~e~~$^{-4}$] mm^{-1}) and slow flow ([e$^{-12\cancel{9}}$, 1] mm$^{0.5}$ d~~ay~~$^{-1}$), see model equations in Table 1 in Westerberg et al. (2011b). These parameter intervals where used for all catchments since no information on parameter regionalisation was available. The 0.25˚ spatial resolution used with the TRMM and ERA-Interim data in the early model version was retained for the CRN073 and WFD data at a 0.5˚ scale since the centre locations of the CRN073 and WFD cells differed by 0.25˚. The precipitation and evaporation data were interpolated to the higher resolution using nearest-neighbour interpolation. Monte Carlo simulations with 15~~0~~0,000 model runs were performed for each basin using uniformly sampled parameter values and a four-year model warm-up period.

4 Method

This study was carried out in five steps (Fig. 3): 1) observational uncertainties were first analysed and estimated through: a) a screening for ~~a)~~ dataset inconsistencies, b) estimation of discharge uncertainty using a rating-curve analysis, and c) estimation of the temporal uncertainty in FDCs stemming from short time series; 2) regionalisation of FDCs; 3) local calibration of the water-balance model using all available data (for comparison to the regionalised results); 4) regional modelling by constraining the uncertainty in basins treated as ungauged with the regionalised FDCs; and 5) posterior performance analysis of the results. We used the period 1965–1994 because of a comparably large availability of discharge data, and since the CRN073 precipitation data did not show the same occurrence of in-homogeneities as in the later period.

4.1 Screening for data inconsistencies

The consistency of the model input and evaluation data for each basin was evaluated for both long-term averages and the daily time-series scale. The long-term analysis used a Budyko curve (Budyko, 1974), which shows the relationship between the aridity index (long-term ratio of potential evaporation over precipitation) and the runoff ratio ~~(long-term ratio of runoff over precipitation)~~. The Budyko relation was plotted to identify stations with inconsistent data; either a runoff ratio greater than one or inferred actual evaporation greater than potential evaporation (grey areas in Fig 4). The second quality check was the calculation of the correlation between the Current Precipitation Index (CPI; Smakhtin and Masse, 2000) and discharge for intermediate and high flows. The CPI is essentially the sum of the Antecedent Precipitation Index (API, Kohler and Linsley, 1951) and the precipitation on the current day and was calculated using a decay coefficient of K = 0.85 (the lowest value in the range quoted by Smakhtin and Masse) so that for day t the index is,

$$I_t = I_{t-1}K + R_t \qquad (1)$$

where R_t was precipitation at day t. All basins with a correlation between CPI and discharge lower than 0.3 were identified in red on the Budyko curve (Fig 4). It could be seen that these basins were mostly located in the inconsistent, grey areas in Fig. 4 (except for one station that had a correlation greater than 0.3 despite an unrealistic runoff ratio, which in this case might result from an uncertain basin area). The long- and short-term analyses thus gave similar results, which increased our confidence in the screening methods.

There were four basins with unrealistic runoff ratios (>>1) and these were excluded leaving a final 32 basins for the regionalisation. The four excluded basins were all small basins in the mountainous parts of Costa Rica (maximum elevations between 1800–3000 m.a.s.l.) and the precipitation data at a scale of 0.5° were likely not sufficiently representative for these basins. There were three basins with runoff ratios close to one as well as low correlations between discharge and CPI, which indicated that the data may be inconsistent, but these were kept for further study since such runoff-ratio values may be a result of discharge-data uncertainty. Two additional basins (Laja Blanca and Boca de Cupe) had combinations of aridity-index and runoff-ratio values that were far from the theoretical line but were not excluded (Fig. 4 and Table 1 in Appendix A). Both basins were located in the easternmost part of Panama and had seemingly too high mean annual precipitation values, which might be a result of poor coverage of local precipitation stations in the CRN073 dataset in that area. Mean annual precipitation 1971–2002 presented by the Panamanian hydroelectric company show around 1000–2500 mm year^{-1} lower values (ETESA, 2007), which indicates a major source of uncertainty.

4.2 Estimation of discharge uncertainty

Stage-discharge measurements for the 35 discharge stations in Honduras (basin areas 110–21400 km^2, see also Section 2.3) were used to estimate the uncertainty in the discharge data as an upper and lower uncertainty bound. These 35 stations had rating curves that had been classified as having an acceptable or good quality in a previous Honduran water-resources project and the rating-curve equations reported in that project (Flambard, 2003) were used here. Rating-curve data from other countries were not available and it was assumed that the errors of the reported discharge data were similar to those in Honduras, i.e., that the Honduran stations were representative for measurement practices and conditions in the region. The discharge uncertainty could therefore be underestimated in cases where discharge data from the other countries include stations with poorer rating curves. Site-dependent uncertainties, e.g. related to a poor choice of measurement location, could not be quantified. For many stations there was considerable temporal variability in the rating measurements. For these stations a rating curve for a period with many measurements covering a large part of the flow range was selected. The residuals along each rating curve were then calculated as a percentage of the rating-curve-calculated discharge corresponding to the same stage measurement. To facilitate comparison between the residuals at different stations for different flow ranges, the discharge data were normalised by the mean discharge for each basin, using mean discharges reported in the Honduran national water-balance study (Balairón Pérez et al., 2004) as we had no discharge time-series data. The normalised discharge was grouped in frequency intervals limited by the percentiles 1, 5, 10,... , 95, 100; the 1 percentile was used instead of zero to exclude the very lowest flows that resulted in large relative residuals because of division by values close to zero. The 2.5 and 97.5 percentile values for the residuals belonging to each group of normalised discharges were calculated and used together with the median normalised discharge in each group to calculate the rating-curve uncertainty as a function of the normalised discharge. Exponential and power-~~type~~ law functions were fitted to the positive and negative residual percentiles respectively, and these functions were then used to estimate discharge uncertainty for all the GRDC stations in the regionalisation.

When mean daily discharge is calculated, it is important to realise that the actual observations might have been collected with different temporal resolutions. If stages are not registered continuously this can result in a commensurability error in daily discharge data especially if measurements are taken in-between flow peaks. In Honduras, three measurements were taken during the day and in some cases more around flow peaks (Westerberg et al., 2011a; Flambard, 2003). The size of this error depends on the size and response time of the basin, with larger values for small basins and those that have a quick response. We used a value of 17%, previously estimated using 15-minute-resolution stage data for the 1766 km^2 Paso La Ceiba basin in Honduras which responds quickly to rainfall and is comparably small (Westerberg et al., 2011a). The estimate can therefore be considered conservative for most of the stations in the regionalisation. In Costa Rica, stage was recorded continuously using limnigraphs; this error source was therefore excluded for these stations. For the other countries we had no information on the stage-recording method and the Honduran practice was assumed. An estimated error in the actual stage reading of 5% was also added to the uncertainty bounds, as previously used in the fuzzy rating-curve method by (Westerberg et al., 2011a). The different uncertainties were assumed to be additive when calculating the daily discharge uncertainty. This is a simplification that may have resulted in overestimated uncertainty bounds.

4.3 Calculation of FDCs and temporal uncertainty from short time series

The discharge uncertainty estimates were used in the calculation and regionalisation of FDCs for all basins. The FDC, traditionally calculated for a period of record, describes the time duration that a certain flow is equalled or exceeded, and is a compact signature of runoff variability that has often been regionalised to ungauged basins (Bloeschl et al., 2013). Our regionalisation was based on data for the period 1965–1994 and in all the following analyses only years with at least 80% complete data (either calendar year or hydrological year depending on reported format) were used to avoid biases in the FDCs. First, evaluation points (EPs) were defined as specific exceedance percentages on the FDCs (using the same method as Westerberg et al., 2011b). The choice of EPs emphasises different aspects of the hydrograph; some previous studies have only used low-flow EPs for FDC regionalisation (e.g. 30–99% exceedance by Castellarin et al., 2004), while others have used EPs covering almost the entire flow range from 0.1 to 99% exceedance (Mohamoud, 2008). We did not include the very lowest or highest flows since these would likely be associated with the largest uncertainty, but used a volume-weighting method for calculating EPs (Westerberg et al., 2011b), which resulted in simulations with a good match to the whole flow range in this previous study. This means that EPs for each basin (*local EPs*) were determined so they were evenly spaced according to the area under the FDC (that equals the volume of water contributed by flows in a certain magnitude range) with increments of 5%. This resulted in 19 EPs when excluding the maximum and minimum flows. The same EPs had to be used for all basins in the regionalisation and we chose these as the median EP values of all the different sites for each of the 19 EPs (*regional EPs*) The calibration using the at-site data for each basin was assessed using both the local and regional EPs to evaluate the effect of this difference. Uncertain FDCs consisting of the best-estimate specific discharge with uncertainty limits were calculated using the observed discharge data and their estimated uncertainty bounds. This calculation of the uncertainty in the FDC implied an assumption that the uncertainty may consist of non-stationary bias rather than a random error (see also Westerberg et al., 2011b).

Varying temporal data availability (stations that do not have data covering the whole 30-year period used for the regionalisation, Fig. 2) results in added uncertainty to the calculated FDCs because the FDC based on the available data might differ from that for the entire period. We estimated this *temporal uncertainty* in the upper and lower uncertainty bounds as a function of the number of years with data using the nine stations that had long-term data (at least 80% complete daily data in total in 1965–1994). Seven of these were located in Panama, one station in Honduras and one in Nicaragua. In terms of the variability of the FDCs, these stations covered most of the observed range of the normalised FDC discharge values. There were between 5–29 years of data at all the stations in the modelling period 1965–1994 and the uncertainty was estimated using all possible consecutive 5, 6, ..., 29-year periods and 1000 randomly generated series of non-consecutive years. For the latter the order of the years was not maintained and individual years could not be selected more than once per realisation when the 5–29-year series were generated. The uncertainty was calculated from the realisations as the 2.5 and 97.5 percentiles of the percentage uncertainty in the specific discharge values at the upper/lower uncertainty bounds for the FDC EPs. The largest uncertainty from the two sampling schemes (random and consecutive) for each number of years with data was used. This temporal uncertainty was finally added to the FDC uncertainty bounds as a function of the number of years of discharge data at each station in 1965–1994.

4.4 Regionalisation of FDCs with uncertainty

These uncertain FDCs were regionalised using a weighted linear combination of the N most similar basins. We defined similarity based on a number of climate and basin characteristics which all had been found to be related to the FDC discharge values in a correlation analysis (Table 1). These characteristics were standardised by subtracting the mean and dividing by the standard deviation for all basins. The similarity was then calculated using the similarity measure defined by Burn (1990a, b) as the Euclidean distance in the space spanned by the standardised characteristics (Eq. 2):

$$d_{it} = \sqrt{\sum_{m=1}^{M} \left(X_{mi} - X_{mt} \right)^2} \quad (2)$$

d_{it} is the Euclidean distance between the target basin t, and basin i in the data pool; X_{mi}, is the standardised characteristic m for basin i. While geographic distance was not included explicitly, differences in the characteristic QLONG essentially agree with geographic distance because of the spatial distribution of the basins. The weights for each basin in the regionalisation were, similar to Holmes et al. (2002), calculated based on the relative inverse distances (Eq. 3):

$$w_{it} = \frac{\frac{1}{d_{it}}}{\sum_{i=1}^{N} \frac{1}{d_{it}}} \quad (3)$$

w_{it} was the weight of basin i in prediction of target basin t and N was the number of basins in the data pool. For calculating the predicted FDCs using these weights the uncertain discharge at each EP was defined as a fuzzy number with a triangular membership function defined by the lower, crisp (best-estimate) and upper uncertainty limits. The uncertainty in the regionalisation was accounted for through a weighted aggregation of the fuzzy discharge at each EP using the N most similar basins. The general weighted mean operator for fuzzy numbers ~~by~~ of Dubois and Prade (1980) was used to aggregate these membership functions to a new membership function; the individual membership functions were rescaled so that the area under the curves equalled the weights w_{it} and then summed over the range of the support (Fig. 5). The 2.5, 50 and 97.5 percentiles of the cumulative distribution of the aggregated membership function were finally used as lower, crisp and upper uncertainty bounds for the regionalised FDC.

The FDC regionalisation was evaluated in a jack-knife cross-evaluation by excluding one basin at a time because the low number of stations did not allow for separate calibration and validation sets. The correspondence between the predicted and observed FDC-discharge uncertainty bounds at the EPs was evaluated by two measures. The *reliability* of the predicted uncertainty bounds was calculated as the overlapping range between the observed and simulated uncertainty bounds as percentage of the observed range. The *precision* of the predicted uncertainty bounds was calculated as the overlapping range as percentage of the simulated range. These measures were previously used by Westerberg et al. (2011b) and Guerrero et al. (2013). They are similar to the ones used by Yadav et al. (2007) and Breinholt et al. (2012), but differ in that they incorporate an estimate of the uncertainty in the observed discharge data, where that estimate consists of an upper and lower bound that allows for non-stationary biases in-between the bounds.

4.5 Local and regional water-balance modelling

The simulated uncertainty from the Monte Carlo runs was first constrained (in a local calibration) using limits of acceptability in the extended Generalised Likelihood Uncertainty Estimation (GLUE) method (Beven, 2006) for the locally calculated FDCs (Westerberg et al., 2011b). This was done both for the local EPs and the regional (median) EPs used in the regionalisation, using the discharge data for each station in 1965–1994 (Section 4.3). Behavioural simulations were required to be within the limits of acceptability defined from the discharge-data uncertainty at each of the 19 EPs. Then the simulations were constrained with the regionalised FDCs. In both cases an informal likelihood was calculated in the same way as Westerberg et al. (2011b), using the sum of a triangular weighting at each EP of the simulated value relative to the observed data and its limits of acceptability. Simulations with correlation in deviations across successive EPs then obtain a lower weight but can still be behavioural if they are inside all limits of acceptability, i.e. a systematically under- or overestimated FDC for (part of) the flow range can still be behavioural but get a lower weight. The simulated uncertainty bounds were 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.

4.6 Posterior performance analysis

The resulting simulated uncertainty bounds were analysed, as with the FDC regionalisation, by calculating two different model diagnostics that assess the similarity between the uncertainty bounds for the simulated and observed discharge. *Reliability* was in this case defined as the percentage of time that the simulated and observed uncertain intervals overlapped, and *precision* was in the same way as for the FDC regionalisation the overlapping range expressed as a percentage of the simulated range, but here calculated as the average value for the number of days with observations. All the model diagnostics were calculated for low, intermediate and high flows separately. Low flows were defined as flows smaller than the median flow, high flows as flows that were exceeded 1% of the time, and intermediate flows were all flows in between these limits.

5 Results

5.1 Estimation of discharge uncertainty

The analysis of discharge uncertainty for the 35 Honduran stations showed that five stations had most medium to high-flow residuals in the range ±10% of the discharge calculated from the official curves. The remainder had larger deviations and the 2.5 and 97.5 percentiles of the distributions were around ±25%, with larger percentage uncertainties for low flows (Fig. 6). Underestimation was larger than overestimation and there were sometimes poor rating-curve fits to the lowest measurements. For some stations the average residual values varied with flow as a result of poorly fitted rating curves. The exponential and power-~~type~~ law functions fitted to the positive and negative residual percentiles respectively fitted well to the data with adjusted R2-values of 0.80 and 0.98 (Fig. 6). Uncertainty values for normalised discharges smaller/larger than the smallest/largest point used in the fitting were set to the smallest/largest value when these functions were used to calculate the discharge uncertainties for the GRDC stations. The final calculated uncertainty in discharge after the stage and temporal commensurability error had been added varied between -266% and +64% of the crisp discharge for the low-flow range and between -52% and +45% for the high-flow range, where negative (positive) values denote underestimation (overestimation) as in Fig. 6. The uncertainty ranges for the lowest flows were larger than the previously calculated discharge-uncertainty limits at Paso La Ceiba (Westerberg et al., 2011a) and La Chinda as an effect of larger uncertainty in the fitting of some official rating curves. The medium to high flow range was almost identical to that for Paso La Ceiba but around 5% larger in this calculation than that for La Chinda where the non-stationarity in the stage-discharge relationship was less pronounced compared to at Paso La Ceiba.

5.2 Calculation and regionalisation of FDCs with uncertainty

The added uncertainty to the FDC discharge as a result of time series shorter than the 30-year modelling period varied in the range of 3–45% (4–33%) for the upper (lower) uncertainty bound (for time series with 5–29 years of data). This temporal uncertainty was added to the uncertainty bounds for the FDC discharge values for the stations with incomplete time series data before the regionalisation. The FDCs showed great variability in the region; normalised discharge (by mean discharge) varied in the range 3.8–27 (0.05–0.59) for the lowest (highest) regional EP at an exceedance percentage of 0.52% (75%). The number of surrounding basins to be included in the FDC regionalisation was chosen as eight as a trade-off between increase in reliability and decrease in precision (Fig. 7). In 12 of the 32 basins the regionalised FDCs encompassed the observed FDCs (reliability = 100% for all EPs). At some of these basins (e.g. no. 5, 12, 18, 22, and 24, Fig. 7) there were also high precision values. There were six stations where the minimum reliability was less than 50% (Fig. 7). Observations from these stations plotted in the upper and lower extremes of the Budyko curve and included the most extreme FDCs in the region in terms of shape and magnitude of specific discharge, two of these stations had been identified as having likely disinformative data. The poorer performance for the most extreme FDCs was not surprising given that the linear weighted combination method used for regionalisation makes it difficult to predict the most extreme FDC shapes. There was a clear relation between runoff ratio and precision (not shown), with higher precision in humid basins (except for Guatuso, no. 1, which had an inconsistent runoff ratio of 1.05 and a greatly underestimated regionalised FDC at all EPs). Examples of regionalised FDCs for four stations, including one of the best (San Francisco, no. 24) and one of the worst (Tamarindo, no. 16), are given in Fig. 8.

5.3 Water-balance modelling using local calibration

Local calibration of the model parameters to the observed FDCs resulted in behavioural simulations in 26~~4~~ of the 32 basins using the regional EPs, of which basin no. 17 had no behavioural simulations when using the local EPs (Fig. 9). The basins with no behavioural simulations included ~~four~~ three basins in northern Costa Rica (no. ~~1~~2–4) that had runoff ratios of different magnitudes but approximately the same mean annual precipitation (Table 1 in Appendix A), as well as the two Panamanian stations (no. 27 and 28) that deviated substantially from the Budyko curve (Fig. 4). The differences in the reliability and precision between the simulations calibrated using local and regional EPs were small (Fig. 9). There were 13 basins for the regional EP calibration with reliability ≥~~greater than~~ 50% for low, intermediate and high flows. Unrepresentative precipitation data likely had an important contribution to the poorer performance in the other basins since a visual inspection showed obvious differences between basins with lower and higher high-flow reliability (Fig. 10). To further test this hypothesis, the correlation between the observed discharge for intermediate and high flows and CPI was plotted against the high-flow reliability for the local calibration with regional EPs (Fig. 11), and it could be seen that the basins with poor performance also had a poor agreement between CPI and observed discharge. For some basins (Fig. 10, bottom) there appeared to be a frequent timing difference of one day for the flow peaks, which may be related to commensurability uncertainty between precipitation and discharge stemming from precipitation measurements taken in the morning but discharge representing daily averages (Westerberg et al., 2011b). This may have had an impact on the values of the reliability and precision measures (it would lead to lower values, especially for high flows), but would have had little impact on the FDC-calibration.

5.4 Regional water-balance modelling

The reliability of the regionalised simulations was comparable to that of the local calibration, with generally higher values for ~~intermediate and high flows and sometimes lower values for low flows for~~ the regionalisation with some exceptions for intermediate (Guatuso, basin no 1, see below) and high flows (Fig. 12a–c). The precision values were often lower, in particular for low and intermediate flows; this was in general related to the wider uncertainty bounds for the regionalised simulations (as a consequence of the greater uncertainty in the regionalised FDCs).

The predicted uncertainty bounds for the regionalised simulations always overlapped with the locally-calibrated simulation bounds (except for Guatuso, basin no. 1, which had an inconsistent runoff ratio of 1.05 and a regionalised FDC that was greatly underestimated), and also encompassed them for a large part of the time for most basins (Fig. 12d, 100% overlap as percentage of the locally-calibrated bounds means that they are encompassed). The overlap in percentage of the regional bounds (with a low value indicating relatively wide regional bounds) showed a similar pattern to the precision of the FDC regionalisation. There was also a clear relation for the aridity index with relatively wider regionalised bounds in more arid basins (Fig. 12e), which appears to be a result of relatively greater uncertainty for regionalised FDCs in arid basins in combination with narrow locally-calibrated bounds as a result of few behavioural simulations in the most arid basins. Similar results with greater uncertainty in regionalisation in arid basins were also found by Bloeschl et al. (2013).

There was almost no difference between the locally and regionally simulated hydrographs where the regionalisation of the FDCs worked best (e.g. Camaron, basin no. 22, Fig. 12 and Fig. 13). Where the regionalised FDCs had wider uncertainty bounds, the predicted simulation uncertainty was greater than that from the local calibration (e.g. Balsa, no 6 and Agua Caliente, no. 12, Fig. 13). In such cases additional regionalised information, e.g. recession behaviour (Winsemius et al., 2009), might provide additional constraints. For basins where the regionalisation worked less well, such as at Guanas (no. 14, that, except for Guatuso, had the poorest regionalisation results of the stations with behavioural local simulations) there was, apart from wide uncertainty bounds, also a systematic shift to the uncertainty bound for the less well regionalised part of the flow range (here high flows) but still a high degree of overlap with the locally-calibrated uncertainty bounds (Fig. 12 and Fig. 13). There were six basins with behavioural simulations when the wider regionalised FDCs were used to constrain the simulations but not when using the local data (e.g. Guardia, no. 2). In all these cases the data seemed inconsistent when inspecting the time series of discharge and precipitation.

6 Discussion and concluding remarks

This study has explored a method for predictions in ungauged basins based on FDCs that accounts for uncertainty in the observed data, the FDC-regionalisation method and the model parameterisation. This method is novel in for the first time explicitly incorporating observational uncertainties in rainfall-runoff model regionalisation; uncertainty in discharge from rating-curve analyses, uncertainties stemming from the use of short discharge time series, and analyses of uncertainties stemming from disinformative data. It also addresses the need for reliable predictions in ungauged basins in developing regions, where data limitations are often important, as highlighted by Hrachowitz et al. (2013).

6.1 Estimation and impact of observational uncertainties

6.1.1 Discharge data uncertainty

Discharge-data uncertainty can often be an important source of error (McMillan et al., 2010), which to our knowledge has not previously been accounted for in regionalisation~~, and only a few previous studies have considered uncertainties in the regionalisation method (Yadav et al., 2007; Seibert, 1999; Kapangaziwiri et al., 2012; McIntyre et al., 2005)~~. We estimated the uncertainty in the GRDC discharge data using 35 rating stations in Honduras, with the assumption that measurement practices and rating-curve derivation were similar in the rest of the region. The different uncertainties in the discharge-uncertainty estimation were assumed to be additive which may have resulted in overestimated uncertainties. It was, however, likely a conservative estimate that reflected the lack of information about site-specific conditions. The estimated discharge uncertainty was similar but somewhat higher to that reported in the review by McMillan et al. (2012), with the largest uncertainties for low flows for many stations as a result of poor rating-curve fits in combination with higher natural variability and relative measurement uncertainties for low flows (Pelletier, 1988). Patterns could be seen for some of the Honduran discharge stations in the variation of the residuals as a function of normalised flow as a result of poor rating-curve fits. An assumption of errors with a simple structure within the bounds was therefore not appropriate when the estimated uncertainty bounds were used for the GRDC discharge station data in model evaluation, but the limits-of-acceptability approach we used allowed for non-stationary biases within the observed uncertainty bounds.

6.1.2 Precipitation data uncertainty

Overall, precipitation ~~-~~data quality was probably the most limiting factor. The WFD variables used to calculate potential evaporation differed somewhat to local station data, but precipitation ~~-~~data quality is more important than evaporation ~~-~~data quality in many cases (Paturel et al., 1995). Because of lack of information about the magnitude of the precipitation errors, we only treated this uncertainty source implicitly through data-screening analyses and visual inspections of the time series. The CRN073 precipitation data were the best available gridded data for the Central-American region. However, because of the high spatial variability of precipitation (Alfaro, 2002; Magaña et al., 1999), the resolution of the CRN073 data was not sufficient for many basins – in particular those located in mountainous regions where runoff ratios greater than one were found likely because of underestimated precipitation. In such circumstances no hydrological model that assumes mass balance can be expected to give good predictions (Beven et al., 2011). There were also noticeable time-variable errors in the precipitation dataset as a result of changes in station density and/or measurement equipment.

6.1.3 Detection and impact of dataset inconsistencies

The two methods that were used to screen the dataset for inconsistencies between the runoff and climate data~~, and they~~ gave mostly similar results. The disinformative outliers on the Budyko curve resulted from runoff ~~coefficients~~ ratios much greater than one (Section 6.1.2~~in small mountainous basins and therefore likely because of underestimated precipitation~~) and from some basins with overestimated precipitation compared to higher-quality local information. Most basins with low discharge-CPI correlation were outliers on the Budyko curve, with often obvious mismatches between the precipitation and discharge data time series, and there was a strong relation between the discharge-CPI correlation and high-flow reliability in the local calibration. This suggests that this method was useful for identifying inconsistent data in this region, and we recommend the use of data-screening methods in future regional studies. It should be remembered, however, that there may be shorter informative periods even if long-term averages are inconsistent, and matching peaks in precipitation and discharge should not be expected under all circumstances. Event-based runoff ratios may be useful to identify data with inconsistent events in basins with low baseflow but require sub-daily data in most basins (Beven et al., 2011).

Identification of disinformative data prior to modelling may not always be possible, and another method for dealing with such data inconsistencies is therefore to use model-evaluation criteria that are robust to moderate disinformation (Beven and Westerberg, 2011).

Calibration focused on hydrological signatures, such as FDCs, could be expected to be more robust to moderate disinformation, such as the presence of a few events with inconsistent inputs and outputs (Westerberg et al., 2011b). Our study combined these two methods for addressing ~~disinformative~~ the significant data uncertainties in studies of this type, and both were necessary considering that all disinformation could not be identified in the data screening and that the calibration method in some cases resulted in behavioural simulations even with highly disinformative input data. The latter cases can be detected ~~in posterior performance analyses and data screening~~ in gauged catchments, but calls for discharge-data independent data screening methods and/or the use of multiple signature constraints in ungauged catchments. Further research is needed to investigate the effects of disinformation on signature calibration and how best to estimate the effect of observational uncertainties on the values of different types of signatures. ~~Similarly to Kauffeldt et al. (2013) we found many disinformative data in the large-scale datasets we used and our analyses highlighted the importance of addressing such inconsistencies prior to and during modelling, especially considering the generally poor availability of information about the original data used to construct the datasets and their errors.~~

The choice of an appropriate likelihood in the face of the errors that affect hydrological inference has been discussed in great detail (Beven et al., 2012; Clark et al., 2012). In this study we found a high presence of non-stationary errors in the model input and evaluation data with little information about the magnitudes. This made the informal likelihood function we used a suitable choice since it allowed implicitly for some of these errors without requiring an error model to statistically represent the error characteristics.

6.2 The use of FDCs for regional water-balance modelling

The regionalised simulations were generally reliable compared to local simulations in the basins where behavioural simulations were found in local calibration. In the basins where the regionalisation of the FDCs worked best there was little difference between the regionalised and local simulations. Where it worked less well the predicted uncertainty was sometimes much wider than the local uncertainty bounds and the most extreme FDC shapes were less well predicted, leading to some systematic shifts to the uncertainty bounds compared to the local calibrations in those cases. Greater uncertainty in the regionalised compared to the local FDCs reduced their information content for constraining model predictive uncertainty in ungauged basins. This was especially important in the presence of disinformative input data, where simulations within the regionalised FDC uncertainty bounds were found in some basins but not within the locally~~-~~-estimated FDC bounds that were narrower.

In local model calibration, posterior~~-~~-performance analyses are useful to check whether the chosen signatures (e.g. the FDC) provide sufficient constraints for the particular modelling application (type of model structure, basin, climate, etc.) or whether additional information is needed to constrain the simulations (Westerberg et al., 2011b). However, in regionalisation such analyses cannot be made for the ungauged catchments and it would be advisable to always apply several different regionalised signatures (Yadav et al., 2007; Castiglioni et al., 2010) to ensure greater robustness of the predictions – especially in the presence of completely disinformative input data. ~~When using this regionalisation method it~~It would, however, still be ~~is~~ important to perform data screening and posterior performance analyses in the nearby gauged basins ~~to learn about the different errors sources that affect the simulations there, as well as the different types of constraints that are needed to constrain the simulations~~ since similar behaviour, uncertainties and conditions might be expected ~~in nearby~~

~~ungauged basins~~. ~~There is therefore a need to~~The use of other signatures requires further investigation of~~e~~ how observational uncertainties affect the uncertainty in different types of signatures and their regionalisation~~, as well as to develop discharge-data independent screening tools for model-input data for ungauged basins~~.

The method for FDC calibration developed by Westerberg et al. (2011b) was here tested for a wider range of basins and resulted in a high reliability in the local calibration in basins where the data screening indicated that the data had good quality. An assessment of the performance for different ~~parts of the~~ hydrograph~~ (base flow, troughs, peaks, rising and falling limbs)~~ aspects ~~as in the previous study~~ and of different ways of choosing the EPs on the FDCs, as in the previous study, was not made here but would be useful to assess the performance of the FDC calibration for the wider range of hydrological conditions in this study. It could be seen that in arid basins the discharge was generally more constrained in recession periods compared to in humid basins (likely as a result of the more non-linear FDC shape), indicating that recession information (e.g. Winsemius et al., 2009; McMillan et al., 2013) might be useful to further constrain the uncertainty bounds in the latter case. Further conclusions on the strengths and weaknesses of the FDC calibration for this wider range of basins could also be drawn through the use of different model structures, e.g. different conceptualisations of groundwater storage and runoff generation in groundwater-dominated basins. The parsimonious model structure used here might be overly simple in many cases even if it showed good results previously at Paso La Ceiba (Westerberg et al., 2011b). Compared to those results, the ~~total~~ average reliability was ~~much~~ lower here (8~~7~~6%, compared to 95% previously), with the main difference between the simulations being the precipitation data. The CRN073 precipitation used here had a correlation of only 0.77 with the locally-interpolated precipitation in that study. It might also be possible to estimate the prior parameter ranges based on catchment and climate characteristics, however such an analysis was outside the scope of this paper and would also be affected by disinformation in the regionalisation data.

6.3 Regionalisation of FDCs with uncertainty

The FDC-regionalisation method was based on a fuzzy aggregation of the FDCs from the hydrologically most similar basins, which accounted for uncertainty in the data as well as the regionalisation relation. It resulted in generally reliable results except for the most extreme FDC shapes. This was because of the weighted combination of the FDCs in combination with relatively few gauged stations for a quite heterogeneous region. We found it important to include climate as well as basin characteristics in the definition of hydrologic similarity since rainfall is a dominating factor in shaping the hydrological regime in Central America (George et al., 1998; Waylen and Laporte, 1999). The representativeness of the climate data likely affected the calculation of hydrologic similarity and therefore the FDC regionalisation. The different lengths of the discharge series resulted in a temporal uncertainty that we estimated as a function of the number of years with data. The FDC-regionalisation approach we used was similar to that of Holmes et al. (2002) who used a much larger set of basins. The effect of the chosen number of hydrologically similar catchments was evaluated in a cross-evaluation, and we recommend performing this type of analysis to inform the choice. ~~Like them, we also found better results by using a normalisation with mean discharge instead of basin area. This left the residual problem of estimating mean discharge for the ungauged basins which was problematic and led to the use of area instead. A better method for predicting mean discharge could likely improve the regionalisation results~~Further conclusions about the advantages and disadvantages of the regionalisation method could be drawn by t~~.~~

~~T~~esting it~~this FDC regionalisation method~~ in other regions with better-quality ~~precipitation~~ data~~ and long-term discharge series would enable further conclusions about its advantages and disadvantages~~.

6.4 Concluding remarks

The FDC contains important information about hydrological behaviour that is needed for most water-balance investigations in ungauged basins, and it is therefore of interest on its own as well as a basic regionalised model constraint in many cases. Further research will be required to reveal what additional regionalised information is needed to ensure robust predictions under different circumstances and how uncertainties in such additional regionalised information can be reliably estimated. This study provides a strong demonstration of the need to assess the quality of the data used to inform the estimation of ungauged basin responses in a regionalisation study. The potential for non-stationary epistemic errors and hydrological inconsistencies means that the regionalisation might be subject to significant uncertainties that are difficult to estimate by standard statistical methods. This implies that deterministic predictions might be misleading, and that explicit recognition of uncertainty should be used in decision making. Where the estimates of uncertainty are particularly high, further data collection might be valuable in making decisions for water-resources management.

Appendix A: Discharge stations and basin characteristics

Table 1. Discharge stations and basin characteristics, indices calculated for 1965–1994 except for RR and E_{POT}/P that were calculated for the period of discharge record (i.e. the same as in the Budyko plot, Fig. 4)

No	River@Station	Lat. (˚)	Long. (˚)	Area (km²)	RElev[1] (m)	E_{POT}/P[2] (-)	RR[3] (-)	MAP[4] (mm)	RL5[5] (days)	NYr[6]
1	Rio Frio@Guatuso	10.67	-84.82	287	1787	0.47	1.05	2869	129	7.0
2	Tempisque@Guardia	10.55	-85.58	972	1877	0.72	0.26	2213	186	5.0
3	Tenorio@Rancho Rey	10.47	-85.16	236	1742	0.46	0.38	2869	129	11.0
4	Rio Canas@Libano	10.43	-85.02	132	1346	0.49	0.29	2869	129	7.0
5	Rio La Barranca @Guapinol	10.03	-84.58	197	1920	0.58	0.55	2452	208	5.0
6	Grande de Tarcoles @Balsa	9.93	-84.38	1660	2688	0.53	0.50	2438	215	9.0
7	Grande de Candelaria@El Rey	9.67	-84.30	667	2393	0.55	0.55	2490	209	7.0
8	Rio Terraba@Palmar	8.97	-83.47	4825	3798	0.41	0.67	2952	197	11.0
9	Estrella@Pandora	9.73	-82.95	634	2190	0.48	0.77	2653	205	7.0
10	Sixaola@Bratsi	9.55	-82.88	2131	3759	0.47	0.97	2562	210	7.0
11	Humuya@Guacamaya	14.74	-87.64	2621	2081	0.75	0.27	1525	251	13.0
12	Agua Caliente @Agua Caliente	14.67	-87.32	1578	1865	0.72	0.34	1493	265	13.0
13	Guayape@Guayabilas	14.59	-86.29	2229	1757	0.60	0.21	1770	244	15.0
14	Coco@Guanas	13.50	-85.95	5527	1739	0.98	0.17	1304	291	17.7
15	Rio Villa Nueva @Puente	12.93	-86.83	1044	1568	1.04	0.26	1458	283	13.0

16 El Tamarindo @Tamarindo	12.25	-86.71	217	310	1.29	0.17	1410	273	25.7
17 Brito@Miramar	11.38	-85.95	235	385	0.98	0.21	1645	244	21.7
18 Grande de Matagalpa @Paiwas	12.78	-85.12	6498	1514	0.71	0.35	1782	238	10.0
19 Mico @Muelle de los Bueyes	12.07	-84.53	1673	938	0.51	0.37	2587	197	13.0
20 Chiriqui Viejo @Paso Canoa	8.53	-82.83	805	3350	0.34	0.72	3394	164	30.0
21 Chiriqui@Interamericana	8.42	-82.35	1331	3267	0.32	0.89	3850	155	28.0
22 Tabasara@Camaron	8.07	-81.63	1172	2206	0.37	0.72	3346	210	29.7
23 San Pablo @Interamericana	8.20	-81.25	756	1820	0.36	0.65	3213	211	27.4
24 Santa Maria@San Francisco	8.22	-80.97	1379	1812	0.41	0.62	2911	202	29.7
25 La Villa@Atalayita	7.87	-80.53	1019	917	0.64	0.46	1929	247	30.0
26 Rio Grande @Rio Grande	8.43	-80.50	505	1654	0.52	0.46	2471	197	29.7
27 Chucunaque @Laja Blanca	8.40	-77.83	2963	1031	0.32	0.26	4088	62	10.7
28 Tuira@Boca de Cupe	8.05	-77.57	2409	1803	0.17	0.21	5378	24	20.4
29 Chagres@Chico	9.26	-79.51	409	904	0.46	0.76	3167	186	9.0
30 Changuinola @Valle del Risco	9.28	-82.53	1692	3276	0.39	0.96	3124	189	23.0
31 Rio Ulua@Chinda	15.12	-88.20	8579	2757	0.73	0.47	1511	256	29.0
32 Rio Choluteca @Paso La Ceiba	14.29	-87.06	1805	1664	0.88	0.17	1268	287	13.7
33 Rio Toro@Veracruz	10.5	-84.22	196	2611	0.42	1.29	3016	131	7.0
34 Sarapiqui@Puerto Viejo	10.46	-84.00	825	2833	0.38	1.36	3261	141	11.0
35 Naranjo@Londres	9.46	-84.07	224	2932	0.50	1.61	2578	210	7.0
36 Pejibaye@Oriente	9.82	-83.68	231	2051	0.51	1.77	2371	222	7.0

[1] RElev is the elevation range in metres

[2] E_{POT}/P is the aridity index, where E_{POT} is potential evaporation and P is precipitation, here calculated for the period with discharge data at each station

[3] RR is the runoff ratio, total runoff divided by total precipitation calculated for the period with discharge data at each station

[4] MAP is the mean annual precipitation

[5] RL5 is the average number of days per year with precipitation below 5 mm

[6] NYr is the number of years with 80% complete data in a year or hydrological year in 1965–1994

Acknowledgements

The authors gratefully acknowledge the European Union (FP6) funded Integrated Project WATCH (Contract No. 036946) for the meteorological data. This work was funded by the Swedish International Development Cooperation Agency grant number 75007349, and used the high-performance computing resources at Uppsala Multidisciplinary Center for Advanced

Computational Science (UPPMAX). We thank the staff at SERNA, Honduras for providing data for the Honduran basins. We also thank the staff at CIGEFI, University of Costa Rica, in particular Ms. Beatriz Quesada, for their kind assistance with the CRN073 dataset. The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° 329762. This is also a contribution to the CREDIBLE consortium funded by the UK Natural Environment Research Council (Grant NE/J017299/1). We thank Denis Hughes, Anna Sikorska, and the anonymous referee for constructive comments that helped to improve the manuscript.

References

Aguilar, E., Peterson, T. C., Obando, P. R., Frutos, R., Retana, J. A., Solera, M., Soley, J., Garcia, I. G., Araujo, R. M., Santos, A. R., Valle, V. E., Brunet, M., Aguilar, L., Alvarez, L., Bautista, M., Castanon, C., Herrera, L., Ruano, E., Sinay, J. J., Sanchez, E., Oviedo, G. I. H., Obed, F., Salgado, J. E., Vazquez, J. L., Baca, M., Gutierrez, M., Centella, C., Espinosa, J., Martinez, D., Olmedo, B., Espinoza, C. E. O., Nunez, R., Haylock, M., Benavides, H., and Mayorga, R.: Changes in precipitation and temperature extremes in Central America and northern South America, 1961-2003, J Geophys Res-Atmos, 110, D23107, doi:10.1029/2005JD006119, 2005.

Alfaro, E. J.: Some characteristics of the precipitation annual cycle in Central America and their relationships with its surrounding tropical oceans, Tópicos Meteorológicos y Oceanográficos, 9, 88-103, 2002.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration – guidelines for computing crop water requirements, FAO, Rome, 300 p., 1998.

Amador, J. A., Alfaro, E. J., Lizano, O. G., and Magana, V. O.: Atmospheric forcing of the eastern tropical Pacific: A review, Progress in Oceanography, 69, 101-142, 2006.

Balairón Pérez, L., Álvarez Rodríguez, J., Borrell Brito, E., and Delgado Sánchez Sánchez, M.: Balance hídrico de Honduras - documento principal, CEDEX, Madrid, 2004.

Bardossy, A.: Calibration of hydrological model parameters for ungauged catchments, Hydrol Earth Syst Sc, 11, 703-710, 2007.

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

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., Smith, P. J., and Wood, A.: On the colour and spin of epistemic error (and what we might do about it), Hydrol Earth Syst Sc, 15, 3123–3133, 2011.

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.

Birkel, C., Soulsby, C., and Tetzlaff, D.: Modelling the impacts of land-cover change on streamflow dynamics of a tropical rainforest headwater catchment, Hydrolog Sci J, 57, 1-19, 2012.

Blazkova, S., and Beven, K. J.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, Water Resour Res, 45, W00B16, doi:10.1029/2007WR006726, 2009.

Bloeschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H. H. G. eds. Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales, Cambridge University Press, Cambridge, 2013.

Brath, A., Montanari, A., and Toth, E.: Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model, J Hydrol, 291, 232-253, 2004.

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.

Budyko, M. I.: Climate and life, Academic press, London, 1974.

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.

Burn, D. H.: Evaluation of Regional Flood Frequency-Analysis with a Region of Influence Approach, Water Resour Res, 26, 2257-2265, 1990a.

Burn, D. H.: An Appraisal of the Region of Influence Approach to Flood Frequency-Analysis, Hydrolog Sci J, 35, 149-165, 1990b.

Buytaert, W., and Beven, K.: Regionalization as a learning process, Water Resour Res, 45, W11419, doi:10.1029/2008wr007359, 2009.

Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., and Brath, A.: Regional flow-duration curves: reliability for ungauged basins, Adv Water Resour, 27, 953-965, 2004.

Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., and Montanari, A.: Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach, Adv Water Resour, 33, 1235-1242, 2010.

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.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q J Roy Meteor Soc, 137, 553-597, 2011.

Diaz, H. F., Hoerling, M. P., and Eischeid, J. K.: ENSO variability, teleconnections and climate change, Int J Climatol, 21, 1845-1862, 2001.

Dubois, D., and Prade, H.: Fuzzy Sets and Systems: Theory and Applications, Academic Press, San Diego, CA, 1980.

Enfield, D. B., and Alfaro, E. J.: The dependence of Caribbean rainfall on the interaction of the tropical Atlantic and Pacific oceans, J Climate, 12, 2093-2103, 1999.

ETESA. Mapa de isoyetas anuales (1971-2002): http://www.hidromet.com.pa/Mapas/Mapa_de_Isoyetas.pdf, access: 2011-03-01, 2007.

Flambard, O.: Rapport Actividad 1 - Inventario, análisis y diagnostico de la red hydrometeorologica, ref D4168/RAP/2003-00059-B (limited acc.). EDF, Toulouse, 60, 2003.

George, R. K., Waylen, P., and Laporte, S.: Interannual variability of annual streamflow and the Southern Oscillation in Costa Rica, Hydrolog Sci J, 43, 409-424, 1998.

Global Runoff Data Centre: http://grdc.bafg.de, access: 2010-02-23, 2010.

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, doi:6700-6715, 10.1002/wrcr.20533, 2013.

Guerrero, J. L., Westerberg, I. K., Halldin, S., Xu, C. Y., and Lundin, L. C.: Temporal variability in stage-discharge relationships, J Hydrol, 446, 90-102, 2012.

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, 2011.

Hidalgo, H. G., Amador, J. A., Alfaro, E. J., and Quesada, B.: Hydrological climate change projections for Central America, J Hydrol, 495, 94-112, 2013.

Holmes, M. G. R., Young, A. R., Gustard, A., and Grew, R.: A region of influence approach to predicting flow duration curves within ungauged catchments, Hydrol Earth Syst Sc, 6, 721-731, 2002.

Hrachowitz, M., Savenije, H. H. G., Bloschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) a review, Hydrolog Sci J, 58, 1198-1255, doi:10.1080/02626667.2013.803183, 2013.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G. J., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F., and Wolff, D. B.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J Hydrometeorol, 8, 38-55, 2007.

Jakeman, A. J., Hornberger, G. M., Littlewood, I. G., Whitehead, P., Harvey, J. W., and Bencala, K. E.: A systematic approach to modelling the dynamic linkage of climate, physical catchment descriptors and hydrological response components, Math. Comp. Sim., 33, 359-366, 1992.

Jalbert, J., Mathevet, T., and Favre, A. C.: Temporal uncertainty estimation of discharges from rating curves using a variographic analysis, J Hydrol, 397, 83-92, 2011.

Kapangaziwiri, E., Hughes, D. A., and Wagener, T.: Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa, Hydrolog Sci J, 57, 1000-1019, doi:10.1080/02626667.2012.690881, 2012.

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C. Y., and Westerberg, I. K.: Disinformative data in large-scale hydrological modelling, Hydrol Earth Syst Sc, 17, 2845-2857, 2013.

Kohler, M. A., and Linsley, R. K.: Predicting the runoff from storm rainfall, US Weather Bureau Research Paper 34, Washington DC, US, 1951.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, Eos, Transactions, AGU, 89, 93-94, 2008.

Magaña, V., Amador, J. A., and Medina, S.: The midsummer drought over Mexico and Central America, J Climate, 12, 1577-1588, 1999.

Magaña, V. O., Vásquez, J. L., Perez, J. L., and Perez, J. B.: Impact of El Niño on precipitation in México, Geofísica Internacional, 42, 313-330, 2003.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, Water Resour Res, 41, W12434, doi:10.1029/2005wr004289 2005.

McMillan, H., Freer, J., Pappenberger, F., Krueger, T., and Clark, M.: Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions, Hydrol Process, 24, 1270-1284, doi:10.1002/Hyp.7587, 2010.

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

McMillan, H., Gueguen, M., Grimon, E., Woods, R., Clark, M., and Rupp, D. E.: Spatial variability of hydrological processes and model structure diagnostics in a 50 km^2 catchment, Hydrol Process, doi:10.1002/hyp.9988, 10.1002/hyp.9988, 2013.

Mohamoud, Y. M.: Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves, Hydrolog Sci J, 53, 706-724, 2008.

Montanari, A., and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resour Res, 43, W05434, doi:10.1029/2006wr005184, 2007.

Parajka, J., Merz, R., and Bloschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrol Earth Syst Sc, 9, 157-171, 2005.

Paturel, J. E., Servat, E., and Vassiliadis, A.: Sensitivity of Conceptual Rainfall-Runoff Algorithms to Errors in Input Data - Case of the Gr2m Model, J Hydrol, 168, 111-125, 1995.

Pelletier, P.: Uncertainties in the single determination of river discharge: a literature review, Canadian Journal of Civil Engineering, 15, 834-850, 1988.

Peña, M., and Douglas, M. W.: Characteristics of wet and dry spells over the Pacific side of Central America during the rainy season, Monthly Weather Review, 130, 3054-3073, 2002.

Portig, W. H.: The climate of Central America, in: World Survey of Climatology, edited by: Schwerdtfeger, W., Elsevier, New York, 405-464, 1976.

Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, Agr Forest Meteorol, 98-9, 279-293, 1999.

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.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences, Hydrolog Sci J, 48, 857-880, 2003.

Smakhtin, V. Y., and Masse, B.: Continuous daily hydrograph simulation using duration curves of a precipitation index, Hydrol Process, 14, 1083-1100, 2000.

Son, K., and Sivapalan, M.: Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, Water Resour Res, 43, W01415, doi:10.1029/2006wr005032, 2007.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q J Roy Meteor Soc, 131, 2961-3012, 2005.

Wagener, T., and Montanari, A.: Convergence of Approaches towards Reducing Uncertainty in Predictions in Ungauged Basins (PUB), Water Resour Res, W06301, doi:10.1029/2010WR009469, 2011.

Waylen, P., and Laporte, M. S.: Flooding and the El Nino-Southern Oscillation phenomenon along the Pacific coast of Costa Rica, Hydrol Process, 13, 2623-2638, 1999.

Weedon, G., Gomes, S., Viterbo, P., Österle, H., Adam, J., Bellouin, N., Boucher, O., and Best, M.: The WATCH forcing data 1958-2001: a meteorological forcing dataset for land surface- and hydrological-models, 2010.

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.

Westerberg, I. K., Guerrero, J.-L., Seibert, J., Beven, K. J., and Halldin, S.: Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras, Hydrol Process, 25, 603–613, doi:10.1002/hyp.7848, 2011a.

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, 2011b.

Winsemius, H. C., Schaefli, B., Montanari, A., and Savenije, H. H. G.: On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, Water Resour Res, 45, W12422, doi:10.1029/2009wr007706, 2009.

Xu, C.-Y.: WASMOD - The water and snow balance modeling system, in: Mathematical Models of Small Watershed Hydrology and Applications, edited by: Singh, V. J., and Frevert, D. K., Water Resources Publications LLC, Highlands Ranch, Colorado, U.S, 555-590, 2002.

Xu, C. Y.: Testing the transferability of regression equations derived from small sub-catchments to a large area in central Sweden, Hydrol Earth Syst Sc, 7, 317-324, 2003.

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, 2007.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour Res, 44, W09417, doi:10.1029/2007wr006716, 2008.

Yu, P. S., and Yang, T. C.: Using synthetic flow duration curves for rainfall-runoff model calibration at ungauged sites, Hydrol Process, 14, 117-133, 2000.

Yu, P. S., Yang, T. C., and Wang, Y. C.: Uncertainty analysis of regional flow duration curves, J Water Res Pl-Asce, 128, 424-430, 2002.

Table 1 Basin and climate characteristics. Climate indices calculated for 1965–1994

Characteristic type	Characteristic name	Unit	Description
Climate	PSTD	mm	Standard deviation of daily precipitation.
Climate	RL5	days	Number of days per year with $P < 5$ mm. Used to characterise the length of the region's highly variable dry season.
Climate	P/E_{POT}	[-]	Ratio of average annual precipitation and average annual potential evaporation, a wetness index previously used for regionalisation by Yadav et al (2007).
Topography	DPSBAR	m/km	Index of watershed steepness from the UK Flood Estimation Handbook, the average of the steepest drainage path slope for each cell in the basin (Bayliss, 1999)
Topography	RELEV	m	Elevation range, calculated as maximum minus minimum elevation
Location	QLONG	decimal degrees	Longitude of discharge station

Fig. 1 The Central-American region, elevation distribution and the location of the studied basins and the Honduran rating stations.

Fig. 2 Temporal availability of data for each discharge station, countries in parenthesis (CR = Costa Rica, HN = Honduras, NI = Nicaragua, and PA = Panama)

Fig. 3 Schematic description of the method used in this study

Fig. 4 Budyko curve showing the relationship between the aridity index and the runoff ratio for periods with discharge data at each station in 1965–1994 (Fig. 2). Areas outside the theoretical limits of the Budyko curve (indicating inconsistent data) are marked in grey. Basins with a correlation between CPI (Eq. 1) and discharge for intermediate and high flows of less than 0.3, also indicating data inconsistencies, are plotted in red.

Fig. 5 Regionalisation of uncertain FDCs using the general weighted mean operator for fuzzy numbers by Dubois and Prade (1980) for each EP. The individual membership functions for the fuzzy FDC discharge for each of the N surrounding stations were rescaled so that the area under the curves equalled the weights and then summed over the range of the support to a new membership function for the regionalised FDC (top). The 2.5, 50 and 97.5 percentiles of the cumulative distribution of the aggregated membership function were then used as lower, crisp and upper uncertainty bounds for the regionalised FDC (red circles).

Fig. 6 Rating-curve residuals for 35 Honduran stations (one colour per station) and 2.5 and 97.5 percentiles of the residuals in each group (the groups were differentiated by frequencies of 1, 5, 10... 95, 100%) plotted against the median normalised (by mean discharge) discharge in each group. Functions were fitted to the 2.5 and 97.5 percentiles against the median normalised discharge in each group respectively to calculate rating-curve uncertainty as a function of the normalised discharge. The residuals were calculated as rating-curve discharge minus observed discharge as a percentage of the rating-curve discharge and the plot excludes a few smaller and larger residuals to improve the visibility for the main flow range.

Fig. 7. Reliability and precision of the FDC regionalisation, with different numbers of hydrologically similar basins included in the regionalisation (top) and the minimum and maximum values for each station for the chosen number of basins (N=8, bottom).

Fig. 8 Examples of regionalised and observed uncertain FDCs. Both discharge and EP exceedance percentage values are shown in log space. The thin/dashed lines represent the best-estimate discharge data and the thick lines the upper and lower uncertainty bounds.

Fig. 9 Number of behavioural simulations using local calibration to FDCs with local and regional EPs, and using regionalised FDCs (top), reliability (middle) and precision (bottom) measures for low, intermediate and high flows, for local and regional EPs respectively in local calibration.

Fig. 10 Precipitation, observed and simulated discharge (mm day^{-1}) at Bratsi, station no. 10 (top), one of the stations that had a poor correlation between observed discharge and CPI

(0.12), and at Paiwas, station no. 18 (bottom) that had a high correlation between observed discharge and CPI (0.60). The simulated discharge was calibrated using FDCs calculated from local observed discharge and using the regional EPs.

Fig. 11 High-flow reliability for the local calibration with regional EPs plotted against the correlation coefficient between the Current Precipitation Index (CPI, Eq. 1) and observed discharge for intermediate and high flows. Basins without behavioural simulations were assigned a reliability of zero.

Fig. 12 Comparison of observed and simulated uncertainty bounds for simulations constrained with local and regionalised FDCs for a) low, b) intermediate and c) high flows for the 24 basins that had behavioural local simulations; d) comparison of regionally constrained and locally-calibrated uncertainty bounds, the overlapping range between these bounds is expressed as a percentage of the width of the locally-calibrated and the regionalised bounds respectively and the 10th percentile and median values of the distribution for each time series are shown; e) width of the regionalised bounds as a percentage of the width of the overlapping area between the regionalised and the locally-calibrated uncertainty bounds, then taken as the average value for the whole time series, plotted against the aridity index.

Fig. 13 Precipitation (dark blue), comparison of simulated uncertainty bounds from regionalisation (red) and local calibration (black) with observed discharge (light blue) at Camaron (no. 22 with the best FDC-regionalisation), Guanas (no. 14 ~~with~~that, except for Guatuso, had the poorest FDC-regionalisation when there were behavioural local simulations), Balsa (no. 6 with high FDC-regionalisation uncertainty), Agua Caliente (no. 12 with a good FDC-regionalisation but poorer data consistency and local calibration), and Guardia (no. 2 with inconsistent data and no local behavioural simulations). The regionalised (red) and observed uncertain (blue) FDCs are shown in log-log space (right in each plot) together with the correlation between discharge and CPI for intermediate and high flows. The observed FDCs are plotted as used in the local calibration, i.e. without added temporal uncertainty.