We thank Referee #1 for his objective review and important comments. This reply provides answers to all of his comments:

Main comments:

1) p.7447 lines 17-22: The climate-simulation ensembles considered by the authors are very limited. For both the Quebec and the Bavaria case, only one GCM, one RCM and one emission scenario are considered. This is very limited to draw general conclusions. State-of-the-art climate change impact investigations nowadays make use of a range of GCMs, RCMs and emission scenarios. The climate simulation ensembles in this paper are limited to changes in the initial conditions (five runs for the Quebec case and three runs for the Bavaria case).

We agree that most of the studies consider several sources of uncertainty, as established on the literature review. However, few studies focus on uncertainty associated with the choice of hydrological model, which is in general considered as small compared with other sources of uncertainty. The climate simulation ensembles used in this study are, in fact, used in order to compare the hydrological models uncertainty with the irreducible uncertainty from natural variability which can only be assessed with multi-member ensemble with the same GCM.

We propose to add the following text in order to clarify the use of the climate simulation ensembles at p. 7488 line 2:

Although the natural variability is just a fraction of the total climate simulations uncertainty, it is irreducible even if perfect models would be available. Therefore, natural variability is used in this study to compare the significance of the uncertainty induced by the hydrological models compared to the irreducible baseline uncertainty.

2) p.7448 lines 12-13: it is unclear at which time scale the LOCI scaling method was applied. At daily scale? The models appear to run with daily or hourly time steps (p.7449 line 14).

The text was changed as follow in p. 7448 line 12:

Similarly, precipitation is corrected with the local intensity scaling method (LOCI, Schmidli et al. 2006), which adjusts 30-years average monthly wet-day frequency and intensity, with a wet-day precipitation threshold of 1 mm (e.g. Chen et al., 2011). Since the LOCI method was developed for daily data, the resulting daily precipitation is redistributed to the sub-daily timescale proportionally to the original RCM precipitation for each day in order to accommodate for a finer temporal resolution of the model data (Muerth et al., 2012). The SCALMET (Marke, 2008) model output statistics (MOS) algorithm then scales all meteorological variables (including also the following uncorrected variables: humidity, wind speed, radiation and cloud cover) from the RCM grid scale to the HyM grid scale using topography as the main predictor for small-scale patterns. SCALMET conserves energy and mass within each RCM grid cell once downscaled on the HyM fine scale grid (Further details on the post-processing of climate simulations can be found in Muerth et al., 2012).

3) p.7452 lines 8-14: The Wilcoxon rank sum test assumes independency between the hydrological model results. I assume this independency condition is not met for results obtained by the same model. Can the authors comment on this?

The hydrological models were driven by the same global and regional climate model ensemble. The members of the ensemble are considered as an independent realisation of climate, both in the reference and the future periods, so the resulted hydrological simulations are also considered independent (see page 7541 line 22-24) , and no statistical test was applied to compare the results from the same model.

The follow text is added to p. 7452 line 15:

It should be noted that the climate change signals from the same model are considered as independent in the sense of estimating the natural variability around it, for they come from independent climate simulations.

4) p.7443 lines 11-20: It surprises me that the authors do not mention the influence of the climate forcing (e.g. emission scenarios). They limit the review to the uncertainty in the climate projections due to different GCMs and RCMs.

As indicated in the modification made above, the objective in this study is to investigate the uncertainty associated to the choice of the hydrological model and to compare it to the smallest possible amount of uncertainty that is associated to the natural variability of the climate system.

The text has changed as follows:

> In the analysis of the impacts on future simulated runoff, Graham et al. (2007) found that the most important source of uncertainty comes from GCM forcing, which has a larger impact on projected hydrological change than the selected emission scenario or RCM used for downscaling. Horton et al. (2006) stress the fact that using different RCMs forced with the same global data set induces a similar variability in projected runoff as using different GCMs, and also that the range of hydrological regimes associated with two considered emission scenarios are overlapping.
>
> Regarding the uncertainty related to the emission scenario, the study of Hawkins and Sutton (2009) for decadal air surface temperature reveals that, in regional climate predictions, this kind of uncertainty makes a small contribution to the total uncertainty for the next few decades.

5) p.7450 lines 6-20: The hydrological models were not explicitly validated for their performance in describing high and low flows, and their performance in terms of flood or low flow frequency distribution. This is very surprising to me given that the models were applied for studying the impact on high and low flows (indicators HF2 and 7LF2) and

because the most interesting/important conclusions drawn from this study are related to these extreme flow conditions.

In this study, we focus in the intercomparaison of four hydrological models in climate change context. These models present a diversity of structural complexity (i.e. lumped, semi distributed and distributed models). One of this models is not calibrated (PROMET) while the rest of models need a calibration-validation procedure. We also took into account the natural variability uncertainty in order to compare the hydrological model uncertainty.

For high flows, results suggest that the uncertainty related to this indicator is more related to the natural variability simulated by climate models than the choice of the hydrological model. The main conclusion of this study found that the uncertainty in projections added by the hydrological models should be included in climate change impact studies, especially for the analysis of mean and low flows. This should be considered as the first step of the uncertainty study. Once this conclusion was established, next steps should include the reduction of this uncertainty trough low and high flow oriented calibration considering a larger number of sites.

If the authors revise their manuscript, better highlighting these and other weak parts of their work, and after meeting the other comments, this paper can in my opinion be considered for publication in HESS.

Other comments:

6) p.7450 lines 11-12: for the HSAMI model the “sum of squares error” is considered whereas for the HYDROTEL model the “root mean squares error” is considered as objective function for the optimization. Does that make a difference? I assume optimization of the two objective functions leads to the same results given that there is only a scaling factor difference.

The calibration procedure of these models was made by two different teams, which make operational use of them, so for this reason we decided to follow their operational procedure. However, we agree with the referee that these two objective functions lead to similar results.

7) p.7450 line 13: That these objective functions favour high flows to the detriment of low flows: I am not convinced of that, because another factor that plays an important role is the autocorrelation: dry spell periods typically have much longer durations than high flow periods; as such the low flows will receive more weight in the objective function.

It is not clear whether the “square” in the objective function equation has a stronger effect than the autocorrelation effect.

Beven (2001) identifies three problems in the use of the sum of squared errors as measure of goodness of fit for rainfall-runoff modelling. The first is that the largest residuals will tend to be found near the hydrograph peaks. Since the errors are squared this can result in the prediction of peak discharge being given greater weight than the prediction of low flows. Secondly, even if the peak magnitudes were to be predicted perfectly, this measure may be sensitive to timing errors in the predictions, and finally, it could be also possible that the residuals at successive time steps may not be independent but autocorrelated in time.

However, we have not established that the "square" in the objective function equation has a stronger effect than the autocorrelation effect. In order to avoid confusions, the sentence of page 7450 line 13, "These objective functions favour a good representation of high flows to the detriment of low flows" will be removed.

8) p.7451 lines 1 & 7: which calibration method is used for the distributions? Does the DVWK approach considers a log Pearson III probability density function for the annual maximum and minimum flows? This is not fully clear from the text.

No post-processing (or calibration) method is applied to the time series of seasonal (summer and winter) maximum daily runoff before the statistical analysis of frequency.

Both the DVWK recommendations for high flows (DVWK, 1979) and low flows (DVWK, 1983) consider the log Pearson III probability function.

The two references are added in p. 7451 line 6:

To calculate 7LF2 and HF2, it is assumed that the time series follow the log Pearson III probability density function, from the German Association of Water (DVWK 1979 and DVWK 1983);

Minor comments:

p.7442 line 13: change "a reference.. and a future …periods" to "a reference…

and a future …period" or to "reference …and future …periods". Same comment for

p.7445 line 17.

The sentences are changed to "a reference…and a future …period"

p.7442 lines 24-26: add "rainfall" to the list of variables affected by the uncertainty

associated to climate scenarios

"rainfall" has been added to the list.

p.7443 line 20 change "GCM" to "GCMs"

Changed.

p.7443 line 28: I suggest to replace "GCMs" by "GCM runs" or "GCM simulations" because the differences are not only due to different GCMs but also different initializations and climate forcing (emission scenarios)

The term has changed to GCM simulations.

p.7444 line 5: replace "model" by "models"

Changed.

p.7444 line 11: replace "China catchment" by "Chinese catchment"

Changed

p.7444 lines 18-19: I suggest to replace "the Canadian Regional Climate Model following the IPCC SRES-A2 scenario" by "one RCM run" because also for the other cited references no details were provided on the RCM model or the emission scenario. Why making an exception for this reference?

Changed.

p.7445 lines 11-12: there is a sudden jump from the cited references to the introduction of the research work by the authors. Please provide a more smooth transition.

The introduction has changed. Please refer to the answer of referee #2 comment 1.

p.7446 line 7-8: change "water systems" to "river basins" p.7446 line 12: change "a managed river systems" to "managed river systems" p.7446 line 23: change "sits"

Changed

p.7449 line 19: "empirical formulation developed by Hydro-Quebec" & "Thornthwaite formulation": please add references

The references have been added as follows:

For the *au Saumon* catchment, HSAMI and HYDROTEL use the empirical formulation developed by Hydro-Québec (Fortin, 2000). For Bavaria, HSAMI still uses the Hydro-Québec formulation while the Thornthwaite formulation (Thornthwaite, 1948) is used in HYDROTEL.

p.7452 line 16: change "model's" to "models" p.7452 line 22: change "climate models" to "climate model projections"

Changed.

p.7452 line 24: change "climate simulations" to "climate change signals obtained"

The hydrological models were forced with the climate simulations, not with the "climate change signals obtained", so we have kept the original sentence.

p.7457 line 27: change "model's" to "model"

Changed.

**References:**

Beven, K.: Rainfall-Runoff modelling. The primer. John Wiley & Sons Ltd., West Sussex, England, 2001.

Chen, J., Brissette, F. P., and Leconte, R.: Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, Journal of Hydrology, 401, 190-202, 10.1016/j.jhydrol.2011.02.020, 2011.

Christensen, J., and Christensen, O.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic Change, 81, 7-30, doi: 10.1007/s10584-006-9210-7, 2007.

DVWK: Empfehlung zur Berechnung der Hochwasserwahrscheinlichkeit. DVWK-Regeln zur Wasserwirtschaft, Verlag Paul Parey, Hamburg, Berlin, 1979.

DVWK: Niedrigwasseranalyse Teil I: Statistische Untersuchung des Niedrigwasser-Abflusses, Verlag Paul Parey, Hamburg und Berlin, 1983.

Fortin, V.: Le modèle météo-apport HSAMI: historique, théorie et application, Institut de recherche d'Hydro-Québec, Varennes, 68, 2000.

Graham, L. P., Hagemann, S., Jaun, S., and Beniston, M.: On interpreting hydrological change from regional climate models, Climatic Change, 81, 97-122, 2007.

Hawkins, E., and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bulletin of the American Meteorological Society, 90, 1095-1107, 2009.

Horton, P., Schaefli, B., Mezghani, A., Hingray, B., and Musy, A.: Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty, Hydrol. Process., 20, pp 2091-2109, doi: 10.1002/hyp.6197, 2006.

Marke, T.: Development and application of a model interface to couple land surface models with regional climate models for climate change risk assessment in the Upper Danube watershed, Fakultät für Geowissenschaften, Ludwig-Maximilians-Universität, München, 2008.

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R.: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, Hydrol. Earth Syst. Sci. Discuss., 9, 10205-10243, doi:10.5194/hessd-9-10205-2012, 2012.

Schmidli, J., Frei, C., and Vidale, P. L.: Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods, Int. J. Climatol., 26, 679-689, doi: 10.1002/joc.1287, 2006.

Thornthwaite, C.W.: An approach toward a rational classification of climate, Geog. Rev., 38, 55-94, 1948.