We thank Dr. Refsgaard for his objective review and important comments. This reply provides answers to all of his specific comments:

1. A serious weakness in the manuscript is the lack of emphasis on scientific novelty of the findings of the study. The authors must document the scientific novelties through discussions with reference to state-of-the-art.

The discussion section has been improved; please refer to point 15 for the answer.

2. The Introduction has very good references to state-of-the-art literature. However, it has some serious shortcomings that should be improved:
    - There is no identification of science gaps in state-of-the-art. In order for the present manuscript to document scientific novelty it should address science gap(s).
    - The objectives of the paper/study should be stated towards the end of the Introduction.
    - Lines 12-25, p 7445: This paragraph is a summary of approach and comprises duplication with text in Chapter 2. It does not belong in the Introduction and may be deleted.
    - The last paragraph, p 7445 line 26 - p 7446 line 2, is not required as the manuscript follow a standard format for a scientific paper.

The signaled paragraphs have been rewritten in order to consider the suggestions. The following text has been added from p.7445 line 12.

Our abilities to predict the future hydrological effects to the changes in climate are necessarily limited, even if we had perfect hydrological models (Beven, 2001). Jones et al. (2006) suggest that conceptual and physical based models have a different role in impact assessment, so the former can be used to rapidly assess the impact of different climate scenarios, while the latter can assess the joint impacts of land-use and climate change. Nowadays, the most used approach is to calibrate a hydrological model on current day data and then use the calibrated model to predict the response under changed conditions (e.g. Ludwig et al., 2009; Poulin et al, 2011). However, e.g. Mauser and Bach (2009) have pointed out that any calibration of a model on present conditions may become invalid for the evaluation of climate change impacts. On the other hand, Blöschl and Montanari (2010) argue that we cannot hope to reduce uncertainty by including more detail into the models (as in the case of physical, process-based models).

As mentioned before, most studies on climate change impact have found that the largest source of uncertainty comes from GCM forcing (e.g. Kay et al, 2009). However, hydrological modelling is an important part of the evaluation of the impact of change because it allows us to understand how the hydrological process would react to climate change. The aim of the present study is to assess the contribution of hydrological models to uncertainty in the climate change signal for water resources management. To achieve this, four hydrological models with different structure and complexity are fed with regional climate model outputs for a reference (1971-2000) and a future (2041-2070) periods. The impact on the hydrological regime is estimated through hydrological indicators selected by water managers. In our analysis, the uncertainty from the hydrological model is compared to uncertainty originating from the internal variability of the climate system. This internal variability induced an uncertainty that is inherent to the climate system and that is the lowest level of uncertainty achievable in climate change studies (Braun et al., 2012). It is therefore used as a threshold to define the significance of the hydrological models induced uncertainty. However, the evaluation of the uncertainty associated with the calibration method or model parameters is out of the scope of this study and is covered in many articles (e.g. Poulin et al., 2011; Teutschbein et al., 2011; Kay et al., 2009).

3. Two different emission scenarios are used for the Canadian (A2) and the German case study (A1B). I am concerned that this may influence the results of the study. This is not even reflected upon in the manuscript.

The choice of emission scenarios is limited by the extent of RCM model domains, which is often the case in other studies. For instance, Graham et al. (2007) used different emission scenarios, GCMs, RCMs (and resolutions), and hydrological models. Moreover, the emission scenarios are different for the considered basins, (i.e., the scenarios A2 and B2 for the Baltic Sea Basin, while only the scenario A2 for the Rhine River Basin). Results show that the most important source of uncertainty comes from GCM forcing, which has a larger impact on projected hydrological change than the selection of the emission scenario or the RCM used for downscaling.

In order to mention the influence of emission scenarios on the uncertainty on climate change impact studies, the text has changed on page p.7443 from line11, as described in the response to the comment 4 from the Anonymous Referee.

4. Please provide just a little more information about the basis for the precipitation and climate input in the two catchments – e.g. the number of stations and whether daily or hourly values.

The following text has been added to p.7447 line 4:

The meteorological observation data sets used to calibrate and validate hydrological models and to correct climate simulations are gridded data sets already available for both regions. For Southern Bavaria this has been generated from sub-daily data of 277 climate stations on a 1 km grid with the PROMET model (Mauser and Bach, 2009), while the project partner CEHQ provided its reference data set of daily precipitation and minimum and maximum air temperatures with a resolution of 0.1° for Southern Québec.

5. It appears that three methods are used for bias-correction/downscaling: (i) monthly correction factors for temperature; (ii) LOCI for precipitation; and (iii) SCALMET for the remaining meteorological variables. The two last methods are not explained and the reader has to study literature. It would be useful if the authors provide just a brief description with key characteristics of the methods, so that an otherwise informed reader would be able to assess them.

Additional information has been added in order to clarify the LOCI and SCALMET methods, please refer to the answer to referee #1 comment 2. Note that correction and scaling are two distinct steps during the post-processing of RCM data. We are sorry for the confusion, but hopefully it is clearer now how those correction and scaling steps work.

6. I am concerned that three different calibration techniques are used for the models including automatic SCE for two models, manual trial-and-error for one model and no calibration for the third model. This is far from ideal and may affect the results significantly.

The principal aim of the paper is to assess the contribution of hydrological model uncertainty in the climate change signal by using an ensemble of hydrological models presenting a diversity of structural complexity, yet we do not focus on the uncertainty related to parameters. Poulin et al. (2011) have compared these two sources of uncertainty and have demonstrated that the impact of hydrological model structure uncertainty is more significant than the effect of parameter uncertainty. In addition, in our study the comparison is based on climate change signals, not on absolute values of streamflow, so in our opinion the proposed methodology doesn't affect the main conclusions.

7. The objective functions for HSAMI and HYDROTEL are the sum of squared errors and the root mean square error. Is that not essentially two functions which give the same result, with one being the square root multiplied by a factor of the other?

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.

8. The models have been subject to split-sample tests. This is fine, but information on the results are lacking.

The following text is added at p.7450 line19:

The Nash-Sutcliffe (1970) efficiency coefficient (N.S.) is computed in order to evaluate the performance of the HyMs (Table 3). For the validation period in the *au Saumon* catchment, the daily N.S. has values of about 0.6 for all models, with the exception of PRO, which achieves a value of 0.2. In the *Schlehdorf* catchment, the daily N.S has values of 0.75 for HSA and HYD, but for PRO it is only 0.12. However, the monthly N.S. values are the better for this model (0.57 and 0.61 for *au Saumon* and *Schlehdorf* respectively). Despite the low performance of PRO for daily N.S., it has a comparable performance in the evaluation of hydrological indicators on the reference period (see section 3.1). Calibration and validation processes are more widely described in Ludwig et al. (2012, submitted).

Table 3. Daily and monthly Nash-Sutcliffe efficiency coefficient (N.S.) for the calibration (1990-1999) and validation (1975-1989) periods.

		HSA		**HYD**		**WAS**		**PRO**	
		Cal.	Val.	Cal.	Val.	Cal.	Val.	Cal.	Val.
au Saumon	Daily N.S.	0.74	0.67	0.60	0.64	0.48	0.60	0.37	0.20
	Monthly N.S.	0.79	0.66	0.76	0.72	0.79	0.75	0.59	0.57
Schlehdorf	Daily N.S.	0.83	0.75	0.80	0.76	0.87	0.82	0.34	0.12
	Monthly N.S.	0.92	0.76	0.88	0.78	0.95	0.88	0.76	0.61

9. P 7452 lines 8-14: The test procedure is not well explained and hence not transparent. It is not clear to me how the Wilcoxon rank sum test was done. Which variables were tested? Is it the series of annual values of runoff with the hypothesis that the median values are identical? Or is it the series of five or three values of average runoff originating from the five/three GCMs?

The Wilcoxon test is used to compare samples of climate change signals, obtained with the different HyM.

The text has been changed as following:

The Wilcoxon rank sum test (Wilcoxon, 1945) is used to compare the climate change signals obtained with two different hydrological models. It performs a two-sided rank sum test of the null hypothesis (H_0) that two series of data are independent samples from identical continuous distributions with equal medians, against the alternative that they do not have equal medians (Wilks, 2006). For instance, for a given hydrological indicator (e.g., OMF), we have four climate change signal samples, which have been obtained with the four different HyM. The Wilcoxon rank-sum test tells us, if two samples, obtained from two distinct HyMs (e.g. HSAMI and HYDROTEL), are independent or not (see Sect. 3.3). It should be noted that the climate change signals from the same model are considered as independent, as they come from independent climate simulations.

10. Table 2. The information in this table is not transparent. I recommend that the authors provide some basic statistics (e.g. sample size, mean, standard deviation for the series being compared and p-values to evaluate the level of significance instead of just crosses.

The sample size of the relative change series is 25 values for *au Saumon*, and 9 values for *Schlehdorf*, as established in p. 7452 line 4-7 and p.7455 line 4-6.

The tables are changed as followed:

Table 2 Results of Wilcoxon tests comparing pairs of hydrological models for a) *au Saumon*, and b) *Schlehdorf.* The p-value is shown and the shaded area indicates a rejection of the null hypothesis at significance level of 5%.

a) au Saumon	**HSA-HYD**	**HSA-PRO**	**HSA-WAS**	**HYD-PRO**	**HYD-WAS**	**PRO-WAS**
OMF	0.140	0.001	0.003	0	0	0.816
JDSF	0.421	0.362	0.003	0.641	0.008	0.064
7LF2 SUMMER	0	0	0	0	0.001	0.001
7LF2 WINTER	0	0	0	0.107	0.641	0.237
HF2 SUMMER	0.001	0.449	0.020	0.002	0.222	0.024
HF2 WINTER	0.130	0.923	0.954	0.200	0.107	0.938

b) Schlehdorf	**HSA-HYD**	**HSA-PRO**	**HSA-WAS**	**HYD-PRO**	**HYD-WAS**	**PRO-WAS**
OMF	0.297	0.063	0.001	0.436	0.006	0.094
JDSF	0.241	0.372	0.248	0.422	0.879	0.423
7LF2 SUMMER	0.258	0.730	0.931	0.258	0.077	0.730
7LF2 WINTER	0.077	0	0	0	0	0
HF2 SUMMER	0.340	0.436	0.863	0.863	0.094	0.113
HF2 WINTER	0.0503	0.063	0.0503	0.666	0.730	1

Table 4 is added in order to show the basic statistics of the relative change series presented in Fig.7 and 8.

Table 4 Mean and standard deviation (std) from the relative change series presented in Figs. 7 and 8.

		au Saumon				**Schlehdorf**			
		HSA	**HYD**	**WAS**	**PRO**	**HSA**	**HYD**	**WAS**	**PRO**
OMF	mean (%)	6.7	4.1	11.7	11.8	-6.4	-4.9	-1.7	-3.8
	std	5.2	5.6	5.7	4.8	2.4	2.2	2.0	2.6
JDSF	mean (days)	-13.9	-13.3	-10.7	-12.6	-3.3	-4.7	-4.7	-4.3
	std	2.6	2.4	3.8	4.0	3.18	2.65	3.00	1.79
7LF2 SUMMER	mean (%)	-6.6	-39.1	-31.2	-23.6	-5.4	-8.1	-4.6	-5.7
	std	7.5	5.4	8.7	4.8	6.4	2.9	4.3	4.0
7LF2 WINTER	mean (%)	75.7	38.3	40.1	47.3	-12.5	-5.3	2.8	13.8
	std	17.8	9.9	14.0	20.4	7.4	2.1	1.9	4.0
HF2 SUMMER	mean (%)	13.5	1.3	5.6	21.1	0.4	-5.1	3.5	-5.2
	std	12.8	9.3	10.5	24.5	14.3	7.6	12.4	8.5
HF2 WINTER	mean (%)	4.3	7.9	3.7	3.6	0.4	-5.1	3.5	-5.2
	std	7.7	8.1	8.4	11.2	7.2	8.2	10.2	14.0

11. Figs 7-8: I wonder if there is any need to include the absolute errors. I do not think they provide interesting results here, and they complicate the figures.

Figs. 7 and 8 have changed in order to present only relative changes.

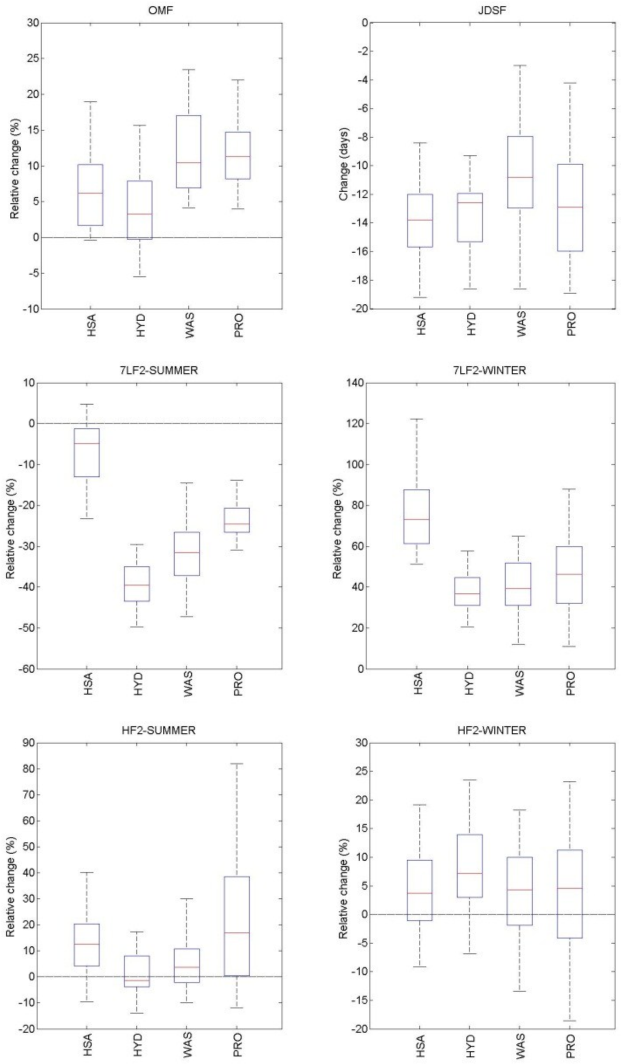

Fig. 7 Changes of hydrological indicators from reference to future period at *au Saumon* (*Haut St-François*, Québec) of overall mean flow (OMF), the Julian day of spring-flood half volume (JDSF), the 2-year return period 7-day low flow (7LF2) in summer and winter, and the 2-year return period high flow (HF2) in summer and winter. For each hydrological indicator, the relative change (as calculated with Eq.2) is presented. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme value

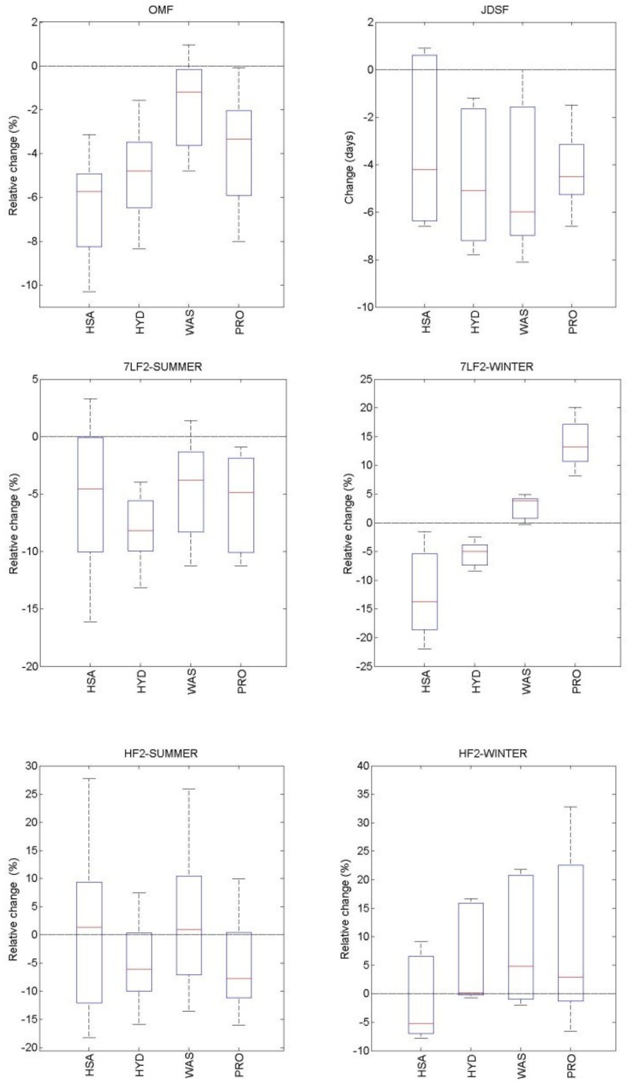

Fig. 8 Same as Figure 7 but for *Schlehdorf*

12. P 7458, lines19-20. The authors recommend using a hydrological model ensemble to fully assess the uncertainty in the climate change signal. This is unclearly phrased, because the hydrological models cannot assess the uncertainty in the climate change signal, but rather the uncertainty on hydrology due to climate change.

The text has changed as followed:

The use of a hydrological model ensemble would thus be recommended in order to fully assess the uncertainty on hydrological indicators due to climate change.

13. P 7458, lines22-23: For the indicators where null hypotheses have not been rejected for any pairs of models the authors recommend that a single conceptual model can be used with some certainty. I think that use of single hydrological models instead of an ensemble of models will, also in these cases, result in underestimation of the uncertainty.

Results show that for some hydrological indicators, the uncertainty due to the natural variability (that is, the irreducible baseline uncertainty) is larger than the uncertainty associated with the hydrological model. Therefore, in a context with limited computationally resources, the less demanding models could be used with a certain level of confidence in order to evaluate the climate change signal on these hydrological indicators. However, we are aware that the generalisation of the conclusions would require more than two sites and should include other sources of uncertainty, as established in the conclusion of the manuscript.

14. The authors recommend (p7460 lines 4-10) that uncertainty in projections added by hydrological models should be included in climate change impact studies. This may well be correct, but it is not documented in the manuscript. The present study only includes natural (GCM initial conditions) climate uncertainty, but not the uncertainty on GCMs nor the uncertainty on downscaling (RCMs + statistical downscaling/bias correction). To draw such conclusion you need to compare the uncertainty generated by hydrological models to the other sources of uncertainty.

Few studies in literature focused on the uncertainty associated with the choice of hydrological model, which is in general considered as small compared with other sources of uncertainty (e.g., Prudhomme and Davies 2008, Wilby and Harris, 2006). Results of this study show that the choice of hydrological model can lead to different climate change signals on hydrological indicators. (The context of this sentence has changed. Please refer to the answer to comment 15.)

15. Chapter 4 conclusions: This chapter should emphasize the novel findings of the study compared to state-of-the-art. This is done to a too limited extend, and should be improved substantially.

The discussion in section 4 has been improved as suggested, from p.7459, line 12:

The principal objective of the paper is to assess the contribution of hydrological models' uncertainty in the climate change signal for water resources management. The results of our study suggest that the added value depends on the hydrological indicator considered and on the region of interest.

Regarding hydrological indicators, Blöschl and Montanari (2010) suggest that that we can have reasonable confidence in predicting hydrological changes that are mainly driven by air temperature (e.g. snowmelt and low flows through evapotranspiration) as opposed to rainfall-driven events like floods. Similarly, Boé et al. (2009) have more confidence to projected changes of low and mean flows. Our results suggest that not only the forcing climate variables but also the hydrological model plays a key role in the uncertainty of projected climate change signal of hydrological indicators.

In the case of high flows and peak time discharge, most of the hydrological models lead to comparable results; therefore, both lumped and calibrated models can be used.

The evaluation of the overall mean flow is more sensitive on the type of model in the Québec catchment than in Bavaria. Therefore, an ensemble of hydrological models should be employed in order to evaluate the range of climate change impacts due to the differences in the process description in different hydrological models. However, the differences in catchment properties (e.g. soil type and topography) can also influence the uncertainty arising from the hydrological model structure (e.g. Kay et al., 2009). As suggested by Blöschl and Montanari (2010), the dependence of local conditions is a distinguishing feature of hydrology that can make the effect of climate change less predictable and more diversified.

The largest relative difference between hydrological model outputs is seen in changes in low flow. Nevertheless, it is important to remember that the HyMs used in this study were not specifically calibrated for low flows, which is reflected in the results for the reference period. Furthermore, this supports the idea that simulation of low flows is an important challenge and need to be improved for low flow management, both in present-day climate and in a perturbed climate (Pushpalatha et al., 2009). Therefore, one must be cautious in the evaluation of climate change impacts on low-flows conditions from a single model. This issue should be re-evaluated with models calibrated over both wet and dry conditions (Seiller et al, 2012).

The GCM is reported to be the most important source of uncertainty in hydrologic climate change impact studies (e.g. Graham et al., 2007, Wilby and Harris, 2006). However, we should still quantify and estimate the uncertainties generated by hydrological modelling. Translation of uncertainty into future risks can provide a valuable contribution to the decision-making process (Beven, 2001). Furthermore, it is necessary to better understand hydrological processes in present climate (e.g., the surface-groundwater interactions to simulate low flows) in order to understand how the perturbed climate will affect future water resources availability.

All in all, we suggest 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. In the absence of an acceptance/rejection criteria (Beven, 2007), all HyMs should be considered equally accurate and therefore should equally contribute to the uncertainty range. The generalisation of this conclusion would require more than two sites and should include the other sources of uncertainty (e.g. internal calibration of HyMs or different GCMs and RCMs).

Another interesting approach is the use of multimodel ensemble to assess structural uncertainties, which has been done by Seiller et al. (2012) to evaluate the relevance of twenty lumped conceptual hydrological models in a climate change context. Results show that using a single model may provide hazardous results, when the model is to be applied in contrasted conditions and generally the twenty-model ensemble gives a better performance.

16. Table 2, caption: It is not the test that is rejected but the hypothesis.

Changed.

17. Fig. 1: The quality of this figure could be improved. The site location map showing Quebec and Bavaria could be reduced (maybe to inserts) and more pace given to the catchment maps. Then it would be interesting to see the location of the precipitation stations on the catchment maps.

The Fig. has changed as in order to include the meteorological stations location.

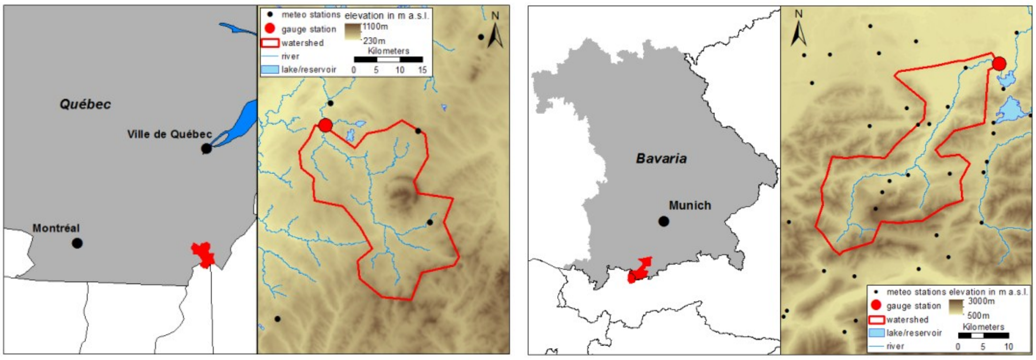

Fig. 1 Location of *au Saumon* and *Schlehdorf* catchments

18. Fig 1 caption uses the term watershed, while the term catchment is used in the text otherwise. Please be consistent.

Fig.1 caption uses “catchments”. However, the manuscript has been revised to be consistent with the terminology.

References:

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

Beven, K.: Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process, Hydrol. Earth Syst. Sci., 11, 460-467, doi:10.5194/hess-11-460-2007, 2007.

Blöschl, G., and Montanari, A.: Climate change impacts—throwing the dice? Hydrol. Process. 24, 374–381, 2010.

Boé, J., Terray, L., Martin, E., and Habets, F.: Projected changes in components of the hydrological cycle in French river basins during the 21st century. Water Resour. Res. 45, W08426, doi:10.1029/2008WR007437, 2009.

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

Jones, R.N., Chiew, F.H.S., Boughton, W.C., Zhang, L.: Estimating the sensitivity of mean annual runoff to climate change using selected hydrological models. Adv. Water Resour., 29 (10), 1419–1429, 2006.

Kay, A.L., Davies, H.N., Bell, V.A., and Jones, R.G.: Comparison of uncertainty sources for climate change impacts: flood frequency in England, Climatic Change, 92, 41–63, 2009.

Ludwig, R., Turcotte, R., Chaumont, D., and Caya, D.: Adapting regional watershed management under climate change conditions - a perspective from the QBIC^3 project. Adv. Geosciences (submitted), 2012.

Mauser, W. and Bach, H.: PROMET - Large scale distributed hydrological modelling to study the impact of climate change on the water flows of mountain watersheds, J. Hydrol., 376, 362-377, doi: 10.1016/j.jhydrol.2009.07.046, 2009.

Nash, J.E., and Sutcliffe, J.V. :River flow forecasting through conceptual models, Part 1 –a discussion of principles. J. Hydrology, 10(3), 282-290. (1970).

Poulin, A., Brissette F., Leconte R., Arsenault R., Maloet J-S.: Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin, J. Hydrol., 409(3-4), 626-636, 2011.

Prudhomme, C. and Davies, H.: Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: future climate, Climatic Change, 93, 197–222, 2008.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, J. Hydrol., 411, 66-76, 2011.

Seiller G., Anctil F., and Perrin C.: Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrol. Earth Syst. Sci., 16, 1171–1189, 2012.

Teutschbein C., Wetterhall F., and Seibert J: Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale. Clim. Dyn. 37, 2087–2105, doi 10.1007/s00382-010-0979-8, 2011.

Wilby, R. L., and Harris, I.: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, Water Resour. Res., 42, doi: W02419, 10.1029/2005wr004065, 2006.