Reply to review #1 (T. Bosshard)

We would like to thank T. Bosshard for his thoughtful comments on our work and for the recommendations he stated. We think the paper has benefited substantially from these comments.

General comments

The study is an important contribution to the field of climate-impact modeling in Alpine catchments, particularly addressing the issue which model chain element the uncertainties arise from. The literature review includes the most relevant publications, the paper is well-structured and the methods are mostly well explained. My main criticism is the way the uncertainties are quantified (see 9. comment below). Along with some other minor changes, I recommend acceptance for publication.

Detailed comments

1. Chapter 3.1, page 3179, line 8: It would be good to indicate that the HadCM3Q3 is the low sensitivity version of HadMC3Q0.

Thank you for this comment. We added this information in the revised manuscript.

2. Chapter 3.1, page 8180, lines 15-17: You derive the lapse rate based on two temperature stations, only. Why did you not include other temperature stations? I guess within the search radius of 40 km, there should be other stations as well. I'm just a bit worried to derive a linear regression relationship based on two data points, only.

Thanks for this valuable comment. We agree with you that deriving near-surface lapse rates from two stations only may be subject to uncertainty. However, long-term temperature records are relatively scarce in mountain regions, especially at high elevation zones.
In the revised manuscript, we used a number of additional temperature stations to verify the calculated lapse rates. For the period from 1985 to 2000, the lapse rates derived from the two stations are very similar to the lapse rates based on seven stations. This justifies that the chosen values are representative. We added a new figure (Fig. 2), in which we compare the different lapse rate calculations and added a small discussion in the revised manuscript.

3. Chapter 3.4: HQsim only requires temperature and precipitation as input variable. So, how is the evapotranspiration derived? I might be described in the referenced papers, however, I couldn't access these papers. Therefore, it would be good to include a short statement about the way evapotranspiration is modelled.

We feel sorry for this. Evaporation is modelled according to the concept of Hamon`s potential evaporation in dependency of the water availability. We added this information in the revised manuscript.

4. Chapter 3.4, page 8184, lines 1-2: There should be more information about the validation of the hydrological model. Fig. 3 is not very helpful as it only shows one year. Why did you chose this particular year? I think it is also necessary to include a validation of the target variables in the impact study, i.e. the mean annual cycle (Fig. 7) and the exceedance probability distribution (Fig. 8).
Regarding the issue of the transferability of the model parameters to future climates: How does the performance change in the period 1971-2005? As we have already experienced some temperature increase during that period, it might be very interesting to see how the performance changes. I would like to state that I know that such a presentation is not standard practice yet. However, in my view, it would help the community to get a hand on the issue of the model transferability (e.g. Merz et al. (2011))

Thank you for these valuable comments. We agree with you that more information about the validation of the hydrological model would be helpful. In the revised manuscript, we added a validation of the target variables, i.e. the mean annual cycle (Fig. 4b) and the exceedance probability distribution (Fig. 4c), as suggested by the reviewer.
The original figure shows runoff simulations for the year 1975 (Fig. 4a). This is a fairly typical hydrological year for this catchment characterised by snow melt induced spring floods as well as floods during the summer season, which were caused by heavy precipitation events. Therefore, the year can be considered as being broadly representative. We added this information in the revised manuscript.

Concerning your second comment (the transferability of the model parameters), we found no significant changes in the model performance during the period from 1971-2005. We added this information in the revised manuscript.
There are some years which are simulated very well by the model, while others are simulated satisfactory. However, 'good' and 'bad' simulations can be found during the whole investigation period.
We think that a more general testing procedure is necessary to get robust conclusions on parameter transferability (e.g. performing split-sample tests as done for example by Merz et al. (2011)). Although some temperature increase is already observed during the investigation period, the variability in the performance among different years is very large, making it difficult to obtain clear trends.

5. Chapter 4.1, page 8185, lines13-22: From Fig. 4.b), I have the impression that there is a pronounced tendency towards underestimated seasonal mean runoff (i. e. out of the 6x4 seasonal data points, most of them are lower than the 4 reference data

points). Since the local scaling corrects for the mean, I would expect the water balance to be equal over the long run. Deviations should only be due to natural variability, thus, there should be a balanced set of over- and underestimated seasonal mean runoff set. It seems though that the statistical characteristics of the bias-corrected scenarios lead to a systematic underestimation. How come? Do you know something about it? Could it be related to a bias in the wet-day frequency, causing higher evapotranspiration? It is only a very small effect and most probably does not have any impact on the validity of the study, but it would be very interesting to know more about it. I leave it to the authors to decide how much they want to discuss this issue.

We agree with the referee that most simulations underestimate seasonal mean runoff values. However, the relative deviations of around half of the simulations (12 out of 24 simulations) vary between +5% and - 5% and thus, are relatively small.
When having a look on the winter season, it can be seen that all control simulations underestimate the values of the reference simulation. This could be due to adjusting both variables, i.e. temperature and precipitation, independently, which may have destroyed the physical consistency between the two variables. This aspect is already discussed in chapter 5.
When having a look on the remaining seasons, it can be seen that the simulation based on the QQ-mapping technique provides better estimates than the simulations based on the local scaling technique. This could be the result of possible errors in the wet-day frequency, which are not accounted for in the local scaling approach (as suggested by the reviewer). The biases in the wet-day frequency are relatively large, e.g. while around half of the observed days are dry, only around one quarter of the days in the bias-corrected control simulations are dry (by using the local scaling technique). Thus, the bias-corrected control simulations contain too many low precipitation (‘drizzle’) days, which may significantly affect evapotranspiration. Similar results were also reported by Stoll et al. (2011) for a catchment in Switzerland. We discussed this aspect in the revised manuscript.

6. Chapter 4.2: The title sounds a bit misleading to me. Also in chapter 4.3, uncertainties related to climate models and downscaling are discussed.

We changed it to “Uncertainty in climate projections” in the revised manuscript.

7. Chapter 4.2, page 8186-8187, lines 25-5: This needs some more discussion of other studies. I would have expected that the RCMs have more freedom to develop there own atmospheric circulation during summer due to flat pressure fields. Thus, the uncertainty due to the RCMs should be higher in summer. Other studies (e.g. Hingray et al., 2007) have shown that regional scaling relationships are more important in summer than in winter.

We discussed this aspect in the revised manuscript:

"These results are in partial disagreement with previous studies. Results from the PRUDENCE project (10 RCMs forced by 1 GCM; Christensen and Christensen 2007) have shown that the largest uncertainty over central European areas (Jacob et al. 2007) and catchments (Rhine, Danube; Hagemann and Jacob 2007) occurs during the summer. Here, the regional climate is less constrained by the boundary forcing due the importance of local scale processes, such as convection and land - atmosphere interactions. For precipitation, our results agree with those mentioned above except for July where the limited sample size of 3 RCMs likely leads to an underestimation of RCM uncertainty. For temperature, the largest RCM uncertainty occurs from March to June while during the summer months of July and August the RCM uncertainty is rather low. This is likely caused by the mountainous location of the watershed where snow-related processes, especially the snow-albedo feedback (see, e.g. Hall and Qu 2006), have a dominant impact on the warming signal during the snowmelt period in spring. Model differences in the representation of these processes lead to different strengths in the snow-albedo feedback and, thus, to larger uncertainties in the projected warming signal."

8. Chapter 4.2, page 8187, lines 11-12: According to chapter 3.3.3, the QQ mapping is done using a 31 day moving window. This is very much similar to a monthly calibration. Please clarify this inconsistency.

We do not agree with the referee. Especially at the boundaries of each month, differing results are obtained when using either a monthly calibration or a moving window approach. We added a clarifying sentence.

9. Chapters 4.2 and 4.3: The uncertainty measure is not defined. One can deduce that the range from minimum to the maximum ensemble member is used as a measure for the uncertainty, but this needs to be defined explicitly. Furthermore, the choice of the minimum-maximum range as a measure of the uncertainty needs to be discussed. While it is certainly an easy measure to interpret, there are some statistical disadvantages. The minimum-maximum range does not make use of the data from the ensemble members in between. For example, in the temperature panel of Fig. 5a), the months 10 and 11 have a similar uncertainty according to your definition, but in month 11, the third members is situated more in the middle of the uncertainty range than in month 10. If one estimated the uncertainty in terms of variances, this would lead to a different result. Another disadvantage is that the range is not normalized by the number of samples. While you use 3 samples for GCM, RCM and bias-correction, you have 20 samples for the hydrological model parameter uncertainty. This might distort the importance of the different uncertainty sources. And as a last point to discuss, the measure does not allow for an additive partitioning of the uncertainty into contributions of the 4 uncertainty sources, i.e. the sum of the 4

uncertainty ranges does not equal the range of the full ensemble. So far, a proper discussion of these limitations is missing. Maybe even a new sub-section in section 3 would be appropriate to cover this methodological aspect of the study.

Thank you very much for this valuable comment. We totally agree with the referee that the uncertainty measure was not defined clearly. In this study the range from minimum to the maximum was used as a measure for the uncertainty. We added a new sub-section in section 3, as suggested by the reviewer, where we clarified this aspect and discussed the limitations of this uncertainty measure.

10. Chapter 4.3.2, page 8188, lines 15-18: Shouldn't it be the other way around: Small precentage changes in summer translate in large absolute changes since discharge is higher in summer than in winter (Fig. 3)?

Thank you for spotting this mistake. We modified the corresponding sentence.

11. Chapter 4.3.2, page 8189, line 11: I would say, the largest uncertainty range due to hydrological model parameters amount to about 35

We found an error in Fig. 4 and corrected this in the revised manuscript. We feel sorry for this mistake. Now, the comment as stated in the original manuscript is correct.

12. Chapter 4.3.2, page 8189, line 14: What do you mean by "certain condition"? Large biases of the hydrological model that potentially transfer in large projection uncertainty or rather large climate changes that force the hydrological model to simulate runoff regimes that it has not been calibrated to? Or anything else? I think the issue of the model parameterisation is an important one and merits some more detailed discussion for the reader. Otherwise, every reader will interpret it differently.

Thank you for this comment. We aimed to say that large uncertainty in hydrological modelling may be transferred in large projection uncertainty. We added a clarifying sentence.

13. Chapter 4.3.3: Mean high flow is not defined. From Fig. 8, I understand that you discuss the whole exceedance probability distribution, thus you cover the whole range from low to high flows.

We agree with the referee that mean high flow was not defined. In the revised manuscript we used the term exceedance probability distribution. We modified the corresponding phrases.

14. Chapter 5, page 8190, lines 17: Using local scaling, also biases in the variability

(e.g. wet day frequency) are inherited from the climate models.

Here seems to be a misunderstanding. In this paragraph we compare the range of the climate change signals obtained by using the two different model chains. The ECHAM5_RACMO_DELTA approach only considers changes in the mean between the control and scenario run. The ECHAM5_RACMO_SCAL simulation, instead, considers both changes in the mean and the variability. As can be seen in the figure, the results differ for both simulations, which is mainly the effect of considering climate variability (e.g. changes in the wet-day frequency) or not.

15. Chapter 5, page 8193, lines 9-15: I do not fully agree with the author's conclusion. Even with a 20 year period, the model cannot be calibrated on conditions in the climate future when the temperatures are considerably above present day conditions (Coron et al., 2012). Other studies with shorter calibration periods have found the same result that hydrological parameter uncertainty is a negligible source of uncertainty (Prudhomme and Davies, 2009; Schäfli et al., 2007). Also, from Fig. 3 I would say that the uncertainty band width of the hydrological simulations is too narrow since the observed runoff is often outside of the uncertainty range. Based on this result, I would rather speculate that the uncertainty due to the hydrological model parameters is underestimated. The issue of the model transferability to future climates is an important topic and in my view, probabilistic parameter sets do not fully solve the problem. At the same time, it is clear that one cannot blame the hydrological model alone as also all the other model chain components suffer from the same problem. Thus, I do not expect the author's to solve this problem. It is clearly out of scope of this paper, but I would like to have a more diverse discussion of the issue.

Thank you for this comment. We added a more comprehensive discussion about parameter transferability in the revised manuscript.

16. Chapter 5, pager 8194, lines 16-21: I like it that you mention the issue of varying one uncertainty source while the others are kept constant. I suggest to inlcude the keyword "interactions" between the uncertainty sources in this text passage.

We modified this paragraph in the revised manuscript.

17. Figure 7: The suffixes DELTA and SCAL seem to be wrong when compared to the other figures.

Thank you for spotting this mistake. We modified the figure in the revised manuscript.

18. Fig. 8a)-d): Please include a line at zero to enhance the readability.

We modified Fig. 8 according to your suggestion.

Reply to review #2 (R. Weingartner)

We would like to thank R. Weingartner for his help to improve our manuscript. We found the comments very helpful.

General comment

The paper of Dobler et al. is an important contribution to the discussion of uncertainty related to climate-impact studies. It is well structured. The design of the study is described in an exemplary manner. The results of the study are mostly clearly elaborated and compared. A clear measure of uncertainty, however, is missing or not sufficiently elaborated (s. detailed comments below). The final discussion in the conclusion chapter provides a differentiated view of the findings and also their limitations. The most relevant references are considered. The title of the paper should include the location of the study because the alpine setting is an important and specific aspect of the study. All-in-all, I recommend to accept this paper for publication after some minor revisions.

We change the title to "Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed".

1. Abstract, p. 8174, line 2. "less attention ..": I do not fully agree with this statement which is repeated in other paragraphs of the paper, too. In contrary, the components of uncertainty are intensively discussed today

We agree with the referee that there are an increasing number of studies that assess uncertainty in climate change impact studies. However, compared to the number of studies that have focused on the impact itself, only relatively few have addressed uncertainty originating from different components. Moreover, while most of these studies have assessed uncertainty resulting from climate models, especially the GCMs, only few attempts have been carried out to study the effects of other components, e.g. the bias-correction approach or the hydrological model parameters. Thus, we think the statement is correct and would like to keep it.

2. Chapter 6.2, study area: The Lech catchment is described as "a typical Alpine valley". Please elaborate more in detail what you mean with typical? Typical for which regions of the Alps? If you discuss this aspect then you could also draw some conclusions on the spatial representativeness of your findings (this point is not addressed in the conclusions).

Thank you for this valuable comment. We deleted the term 'typical' as it was confusing anyway. We added a discussion about the spatial representativeness of our findings in the conclusions.

3. p. 8180, line 2: Describe the grid box (spatial resolution) and the procedure described in this short sentence more in detail.

We added the spatial resolution of the grid box and clarified the procedure in the revised manuscript.

4. p. 8180, line 13 - 22: How can you describe the temporally variable lapse rate with only two stations? I assume there are much more station available within the Lech catchment as well as in the region.

Please see comment no. 2 in the reply to reviewer#1.

5. Performance for present climate conditions, p. 8185, line 7ff: Why did you not include the delta change method in your evaluation of the performance?

In case of the delta change approach the reference simulation is regarded as control simulation. We added a clarifying sentence.

6. Fig. 4 (a), p. 8207: the scale of y-axis is too large ([0, 125]) to provide a realistic impression of the model performance on a daily basis. Perhaps you could introduce relative deviations.

Fig. 4a indicates that all models are able to represent the seasonal cycle. However, we don't think it is appropriate to indicate relative deviations, because the deviations are mainly affected by natural climate variability and thus, may not show the model performance itself.

7. Fig. 5 to Fig. 8: Although these figures depict the uncertainties which originate from different sources for different parameters, they are not directly comparable at first glance due to an inconsistent caption of the single diagrams, e.g., the uncertainty resulting from GCMs is shown in Fig. 5 as "(a)", in Fig. 6 as "(i)", in Fig. 7 as "(a)" and in Fig. 8 as "(a)". This is somehow confusing. A direct caption "GCM uncertainty" (instead of (a) or (i)) may be helpful.

Thank you very much for this comment. We followed your suggestion and modified the figures.

8. Fig. 6 and Fig. 8: Each figure indicates the "size of impact range originating from uncertainty source". This seems to be a measure of uncertainty. The calculation of the measure "percentage points" is not described in the paper. Furthermore, the paper refers not in detail to these specific diagrams which are very important for comparing the uncertainties which result from the different sources.

In the revised manuscript we added a new section (sect. 3.6) in which we describe the uncertainty measure (please see comment 9 in the reply to reviewer #1). In Fig. 6, 7 and 8 we added the information that the differences (percentage points) between the minimum and maximum values are plotted.

9. Mean high flow, p. 8185, line 7ff: The title of this chapter is misleading as floods of different return periods (exceedance probabilities) and not mean high flows are investigated.

Thank you for this comment. We totally agree with the reviewers comment. We modified the title of this chapter.

10. Fig. 8: Please elaborate why the uncertainties are largest for floods of smaller size. i.e. with a large exceedance probability.

The figure shows that for rare events (both high and low flows) the uncertainties are highest. However, as we focus on high flows, we only mention the increasing uncertainty when focusing on larger floods (right part of the figure).

11. p. 8191, line 14: Why is the Lech basin complex? Was does this mean? Is it more complex than other alpine basins? cf. point 2 of the detailed comments.

We agree with you that the word complex is misleading in this context. We modified this sentence in the revised manuscript.

12. p. 8191, line 25 and 26 ! p. 8192, line 2: The sentence on p. 8191 suggests that the GCM structure is the most important source of uncertainty. If this statement is correct, the sentence on p. 8192 "Uncertainty related to the choice of RCMs is found to be on comparable magnitude“ is misleading.

Thank you for this comment. We modified this sentence in the revised manuscript.