Dear Editor Bettina Schaefli,

Please find below our reply to your Editorial summary. As you will find, the manuscript has been significantly modified to address yours and the reviewers' criticisms.

Sincerely,

S. Fatichi, S. Rimkus, P. Burlando, R. Bordoy, and P. Molnar

Responses to the Editorial Summary

The three reviewers and the additional comments are all rather critical about the scientific significance and the quality of the research in the form it is presented in the submitted manuscript. The discussion of the paper shows however that the submitted manuscript essentially failed to convincingly report the novelties of the research and the relevance of the results and conclusions. I am convinced that the research underlying the manuscript is very interesting for the readers of HESS but the manuscript requires some substantial reworking. As far as I can see, this will not involve many more new simulations but a better presentation of what has been done

Given that the paper deals with a topic and a case study which I know very well myself (see namely Hingray et al., 2010;Tobin et al., 2012), I will provide an additional Editor review in a separate comment. Hereafter, I would like to summarize the discussions and the main comments of the reviewers.

Main comments

1)*What are the novel contributions of the paper / new findings?. The authors summarize the objectives of their work and new results very well in their response to reviewer 2. This should be reflected in the new paper, namely in the abstract, the introduction and the conclusion.*

Reply: We modified and significantly re-wrote abstract, introduction, discussion and conclusion to emphasize the novelties of the paper.

2) *Relevance of the results given that only one GCM and 2 RCMs are used for one greenhouse gaz emmission scenario. This important critic has been convincingly discussed by the authors arguing that they account for the stochastic variability of climate and that their projections do not go beyond 2050. This point should become very clear in the new manuscript.*

Reply: We reinforce our argument by modifying the discussion of this topic in the introduction (Section 1) and by introducing a new Figure (Figure 1) prominently at the beginning of the manuscript. We think this should cover exhaustively all the reviewers' concerns.

3) *Future scenario without modified water management / hydropower production rules. The authors argue that keeping the management equal to the present-day situation is the best they can do. Furthermore, they argue that the future simulations with the present-day management still convey interesting findings whereas, in response to reviewer 3, they argue that a future scenario simulation without hydropower infrastructure is not interesting. This is not entirely convincing. The authors report elevation dependent climate change impact effects which are processed through the hydraulic infrastructure with irrealistic management rules (present-day for future scenario). What is the relevance of such results? Would it not be more interesting to also report the "natural" response? Reporting this natural response would be particularly important for comparison with subsequent work in other Alpine regions. Providing material for comparative studies is indeed of prime* *importance. I am furthermore not convinced that keeping the management to present-day situation is the best and only option at this stage (please refer to my editor review).*

Reply: Please see our detailed multiple responses to the Editorial review. We respectfully disagree that the management rules (present-day target-level for future scenario) are unrealistic at the annual scale. Flexibility in management will be rather limited in conditions of a decreasing water resource. Market and demand induced short-term management changes which are conversely very likely are difficult to define as they interact with the energy policy of the Swiss confederation, which is presently under investigation and consolidation. Performing simulations of future scenarios without infrastructure was numerically not feasible at this point, beside the fact that it also represents a purely hypothetical condition.

4) *Too simplistic glacier model (uniform ice thickness). The line of argumentation of the authors is essentially "that's the best we can do" and they discuss the fact that related errors would essentially appear as timing errors (in terms of when the new regime will appear).*

Reply: The simplistic assumption on the initialization of the ice thickness and its possible consequences are now widely discussed in the manuscript in Section 4.0, where the reader is warned about the possible shortcoming of our methodology.

5) *The paper reports climate change impact on extreme events without showing evidence that the modeling system does a good job in simulating extreme events. This point is to my view very critical. Most climate change impact studies do not discuss extreme events because the researchers think that neither their meteo scenario production scheme nor their hydrological model does a good job in simulating extremes. Both points should be extensively discussed. Personally, I do not believe that a model, which does a good overall job (water balance, seasonality, spatial distribution of components), does necessarily a good job for extreme events. And this problem is certainly not "smoothed out" by analyzing differences in extreme events rather than absolute values. I would even argue that a model that is calibrated (with whatever method) with observed meteo data (station data, gridded data) has a very low probability to produce reliable results for extreme events if it is run with *generated* data (spatially or temporally downscaled precipitation and temperature).*

Reply: We now included a Figure (Figure 4) to show the performance of the model in reproducing extreme events (annual maxima and minima) during the observational period. We fully agree that the results obtained for annual maxima and minima are very uncertain and should be evaluated critically (we wrote this explicitly in Section 4). However, we still think it is worth to present them for at least one reason. We acknowledge that there are limitations in the capability of the stochastic downscaling to reproduce extreme meteorological events and of hydrological models in translating them into precise flood estimation. However, accounting for differences obtained in multiple realizations and multiple years, in a full stochastic framework, partially alleviate these limitations and may tell something about changes in annual maximum and minimum flow. The uncertainty, which makes conclusions based on the analysis of deterministic simulations very weak or impossible, as stated by the Editor, is strongly reduced when many simulations are compared and changes between two periods are analyzed (we did not present any change in the absolute values of maximum flows or for high return periods, just annual maxima). Analyzing differences in annual maxima rather than absolute values is definitely more robust and can be done even if the hydrological model skills are less than perfect, as we saw in our simulations for the historical period where annual maxima for some station are overestimated of 1.5 times, but the interannual variability in annual maxima is almost perfectly captured. Furthermore, maximum flow generated in different years and in different stochastic simulations are likely triggered with contributions from different parts of the catchment and also for different underlying hydrological reasons (e.g., if a annual maximum is obtained in October or in June, the snowmelt component might have a different weight). Furthermore, if the hydrological model is not totally wrong the generation of a flood of bigger or smaller magnitude in a relative sense is mostly dictated by the climate forcing. Therefore, we see the only plausible explanation that an increase (in stochastic sense, which in our case does not exclude a no-change scenario) is driven, at least to a certain extent, by the underlying climate change signal and not by artifacts of the hydrological model or of the stochastic methodology. While, we repeatedly acknowledge and we are aware that there is a huge uncertainty in this type of analysis, we still think there is some merit in their presentation. We now clarified these points in the discussion (Section 4).

6) *Used methods are not well presented. I have to agree with the reviewers that I do not think that the level of presented details for the climate change scenarios is sufficient. The corresponding subsection just gives a suit of references without sufficient details. Even after reading the paper several times, I do not have a complete picture of all the steps. I would like to have supplementary material (online only) discussing in sufficient detail the used methods. This should include a table with the change factors (I guess they are monthly factors) which might be re-used by later studies. It would also be nice to have a sketch of the entire scenario production procedure, including the step of producing input to the gridded hydrological model (the interplay of gridded RhiresD data and station data is not very clear). Supplementary material might also be of use for other methodological details.*

Reply: We substantially re-wrote the Section 2.4 “Generation of current and future climate forcing” to introduce or better explain methodological details. Furthermore we added additional explanations and a more detailed description of the methodology in the Supplementary Material (Text S2 and Figure S1 and S2). Regarding the table of factors of change, this can only be done with a cumbersome representation and useless for most readers and so we decided not to include it. Estimated factors of change are different for all of the stations for all of the statistics of precipitation and air temperature, for the different months and for the different future decades (thousands of factors of change). However, we are more than open to share this information with any interested reader who may get in contact with us.

7) *Use of a manual rather than an automatic calibration procedure, absence of uncertainty analysis. The paper does not present any details about the parameters, neither on their values, nor the spatial variability. Are all the parameters distributed in space? How many of them have no direct physical meaning? How were the soil properties related to model parameters? Were the manually adjusted parameters distributed in space? At the moment, the reader is left with the impression that the results just fall out of the model and that we have to believe them without any further insights. And we do not get new insights into how to set up a spatially distributed model for similar case studies. There should also be a better justification for the absence of an uncertainty analysis, especially for extreme events. And finally: there is a huge literature on selecting hydrologically meaningful model performance criteria; from my point of view, computational time is the only limiting factor to automatic calibration and uncertainty analysis.*

Reply: We substantially modified the Section 2.3 which describes the calibration of the model, addressing the most important of the Editor comments. However, we think that this manuscript is not the appropriate venue for a long discussion about model parameters, calibration strategy, etc., which will deviate very much from its focus. We can provide model parameters to any interested reader, but we do not find necessary to include them in the paper.
Strategies for “manual calibration” of distributed hydrological models are difficult to codify, since they are mostly related to the experience and hydrological knowledge of the person performing the calibration. We did some experiments within our group in using Topkapi-ETH to compare different calibration strategies, e.g., manual vs. various automatic and manual from different “users” too (sociologic experiment). We are still elaborating the results and we aim to publish this research in a very different context. Results seems to show that the added value of automatic calibration for this model is marginal. Moreover, we have to say that we respectfully disagree with the fact that

computational time is the only limiting factor to automatic calibration. This would be the case in a "ideal world" where all the variables of interest are measured and can be included within the calibration strategy. However, this is unlikely the case in many real applications for which the only available calibration and validation variable is the discharge. An automatic calibration in a context of a model constrained only to river discharge and of a strongly engineered streamflow regulation is likely to lead to parameter changes, which compensate the model limitation in reproducing that part of variability that is not due to natural processes but to anthropogenic controls. This would essentially lead to better diagnostic metrics but likely also to value of the parameters which compensate for the lack of knowledge of artificial controls.
Furthermore, we would also need to evaluate the actual advantage gained from automatic calibration. Is it really worth investing a large time effort (all the simulations of this study took about 1268 days of single core computations) for just (most of time marginally) improved validation metrics. We think that for lumped conceptual models, this is likely to be the case (because the improvement is typically dramatic), as it would be important for specific purposes, such as a targeted calibration for flood forecasting (where small improvements are already significant). However, if the scope is to use a process-based model to mimic complex hydrological process, is this really the case? Since we believe that we have a different in comparison to the Editor but we are open to a scientific debate, we would be glad to continue this exchange with the Editor in a different context (see also discussion in Fatichi et al., 2012b, Section 4).