Response to the comment of C. Luce (Referee)

We would like to thank the reviewer for his frank assessment of the manuscript. Below is our response to the issues raised in the review. The original comment is printed in plain font, our response is printed in italics.

This was a challenging paper to review. It leaps firmly into the midst of a swirling field of debate about how to use trends, projections, and sensitivities to inform estimates of potential futures, a valuable and necessary discussion for the community. It seems to do so, though, with little sensitivity to some of the tensions in that field of work, perhaps intentionally (?). Given the potential value in engendering further discussion on this debate and more openly explaining and exploring the logic embedded in alternative methods, I will bite on the offered bait. Readers find in this manuscript, on the one hand, a very interesting, even engaging, introduction written by some of the luminaries in hydrology about one of the principle challenges in the field. On the other hand, part way through the manuscript, the narrative becomes enmeshed in speculation. While some of the speculative leanings were hinted at in the introduction, they were overt in the synthesis and following sections. Specifically, the authors postulate that concordance and discordance among the three approaches can directly inform decisions on which are correct or incorrect. They do so without support of evidence from this analysis or citation of previous evidence that conclusions about projections derived from concordance are correct. Although these issues make the current manuscript difficult to follow, a reframing of the argument may be able to use much of the same information in a more constructive context. That context would be asking whether they can do what they did. There is greater value in discussing myriad reasons why there might be disagreement among these methods rather than attempting to resolve those disagreements through, as yet, unvetted assumptions.

The reviewer states that "the authors postulate that concordance and discordance among the three approaches can directly inform decisions on which are correct or incorrect." We would state this slightly differently in saying that we postulate that concordance and discordance among the three approaches are indicators of the confidence one can have in the projection.

The Good:

There was much to appreciate about this paper. It offers a discussion of the challenges facing us in estimating effects and consequences of climate change and the importance of correct estimates for water resources management. They open with a general discussion of how trend information has been applied in contrast to more strictly mechanistic reasoning. I appreciate the opportunity in that for learning about other work in this area, as well. There are also some good lessons and warnings about different reasoning approaches, for example a concise description of concerns about the "upward" approach based on uncertain precipitation. I particularly appreciated several examples wherein logic, deductive, and inductive reasoning were noted as useful tools for interpretation, and then summarized in the first paragraph at the top of page 13072.

The paper also works with a large dataset condensed to a few representative examples. This assisted in taking in the information from a humanly-comprehensible set of time series while providing a sense of both the spatial diversity (and spatial correlation) and temporal diversity to ensure that patterns are not emergent from a few preselected sites or times. In short, it was rich in both spatial and temporal diversity without overwhelming. In this it was aided by well-constructed graphics. A few questions remain, but on the net, substantial information was made readily available to the readers to evaluate claims.

The Concerns:

Ultimately, the paper raises many questions about alternative methods for projecting the future, which is of great value. In this case they do so by applying those alternative methods and comparing results. In doing so they ride roughshod over a number of potential objections related to each method (though enumerating a few as they did). If the intended purpose were to explore where the various objections or errors in logic lead each method potentially astray, so as to offer a reference or catalog on how we can, and do, go wrong in our projections, I could see much value. Instead, the authors venture in the introduction that the three different methods can be reconciled by expert judgment, and reveal in the synthesis section that they evaluate differences primarily (or maybe just initially) on agreement between alternative methods, stating, “The confidence one has in the projection will depend on how strongly the pillars agree, and on their individual uncertainties,” and “The confidence bounds of the individual projections are a starting point for assessing the credibility of each pillar,” and (in the conclusion) “In all cases, the confidence in the combined projection will depend on how closely the pillars agree, and on the individual uncertainties.” I am aware of no studies (and they cite none) demonstrating the truth of these statements, and they do not test them in this manuscript.

The main concern of the reviewer seems to be the premise of the paper that agreement between results of alternative methods is an indicator of the credibility while variation between the results is an indicator of the uncertainty of the projections. We apologise for not explicitly providing supporting evidence for this statement which we are doing now. The IPPC Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections (Knutti et al., 2010, p. 2), for example, has: “Ensemble: A group of comparable model simulations. The ensemble can be used to gain a more accurate estimate of a model property through the provision of a larger sample size, e.g., of a climatological mean of the frequency of some rare event. Variation of the results across the ensemble members gives an estimate of uncertainty.” This is exactly what we are doing in this paper. The premise underlying this paper is exactly the one underlying all IPCC (and most other) ensemble projections. The Good Practice Guidance paper further has “Ensembles made with the same model but different initial conditions only characterise the uncertainty associated with internal climate variability, whereas multi-model ensembles including simulations by several models also include the impact of model differences. Nevertheless, the multi-model ensemble is not designed to sample uncertainties in a systematic way and can be considered an ensemble of opportunity.” We are doing multi-model ensembles which are not a systematic sampling but do provide insight into uncertainty and credibility, at least according to the IPCC point of view. We agree that this premise involves assumptions but it certainly is good practice. In the revised manuscript we make the basis of the premise more explicit and give full justification.

Knutti, R., G. Abramowitz, M. Collins, V. Eyring, P.J. Gleckler, B. Hewitson, and L. Mearns, 2010: Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections. In: Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, and P.M. Midgley (eds.)]. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland.

I acknowledge their sentence saying, “here, the analysis aims at understanding the reasons for the disagreement, by checking the credibility of each projections based on the data used and the assumptions made.” This is a wonderful sentiment. I also acknowledge examples of physical reasoning provided in the following section (7.2). However, the examples provided were brief and simplified in their analysis and subject to alternative physical reasoning to that offered by the authors. There were also no systematic rules or principles beyond “consistency” offered for evaluating the alternatives, no generalization beyond each case study analyzed by the experts. Rather than highlight the complexity and potentially the equivocal nature of the comparisons, they indicate that the correct answer is most likely where there is consensus among multiple potentially untenable lines of logic.

Probably at the heart of my questions is that the first and third approaches use trend extrapolation in a fairly direct way, either of the phenomenon of interest directly (low flow) or the precipitation and temperature driving that behavior. These are offered as nominally equivalent replacements for climate projections from GCMs without reasonable (or any) consideration of the various low-frequency climate contributions to those trends. I've certainly heard the name Hurst brought up any time I even present an historical trend, and I know this group has previously published on the subject. I don't know of any circumstance where historically derived trends are accepted unquestioningly as an expectation for an ongoing rate of change. It would seem that I would need to accept raw extrapolation of a 30-year trend as a reasonable estimate in order to accept the reasoning of this paper. In essence, there are multiple layers of assumption – linearity in trend and process, causality by time or temperature alone as a basis for extrapolation – necessary to allow us to hold all pillars in equal stead, itself a seeming assumption for the proposed reconciliation process.

Again, we were probably not clear on the role of the trend extrapolation methods. We fully agree that historically derived trends should not be accepted unquestioningly as an expectation for an ongoing rate of change and already say so a number of times in the paper. More importantly, we intend to paint on a broader canvas. The trend extrapolation methods are examples of projection approaches that differ from the usual GCM based scenarios. The aim of the paper is not to promote the extrapolation of trends but to illustrate the value of using different methods based on different data. Another model type that could be equally well used within the same framework would be "trading space for time" (see, e.g. Perdigão and Blöschl, 2014). Yes, there are multiple layers of assumptions but the paper does not hinge on them. Rather the paper hinges (as pointed out by the reviewer) on the premise that consistency/inconsistency between different methods is an indicator of certainty/uncertainty. In the revised manuscript we highlight the broader perspective and explicitly state that the trend extrapolation is an example rather than a recommended method.

Perdigão, R. A. P., and G. Blöschl (2014) Spatiotemporal flood sensitivity to annual precipitation: Evidence for landscape-climate coevolution, Water Resour. Res., 50, 5492-5509, doi:10.1002/ 2014WR015365.

We can shorthand the "three pillars" in concise terms as: 1. Direct extrapolation of a trend in flow 2. Calculation of flow from GCM-projected climates using a model 3. Calculation of flow from trend-extrapolated climates using the same model (P.S. A table – perhaps not quite this perfunctory – might be a useful way to summarize and contrast the pillars.) "Flow" need not be the variable of interest, and we can conceptually generalize to other hydrologic outcomes, some of which have nonlinear relationships with climate forcings at varying time scales. On the basis of this alone, why might we expect the 1st and 3rd "pillars" to match in all but the trivial 0-trend case? We know that the mean of a non-linear process is not the same as the non-linear process operating on the means of the inputs. The presentation of the third alternative also seems to offer eerily stationary variance in projections (perhaps I misinterpret the red-lines in the plots?) that contradicts some well recognized expectations (e.g. Field et al, 2012). These points are entirely aside from the fact that the trends in climate for the third is based on 1948-2010, while that for the first is 1976-2008. If the first and third pillars are not really rigorously framed, they come across as "strawmen" proposals in contrast to the more conventional GCM-based approach. At the same time, generous criticism is offered for GCM precipitation projections in the introduction (probably well deserved), which lends a certain frailty to that pillar as well. Are the authors trying to warn us that the three pillars of hydrologic projection are made of straw; that we should be watching for the big bad wolf? It does not seem to be their intent, but it is a difficult feeling to escape.

As noted above, the aim of the paper is not to promote the extrapolation of trends but to illustrate the value of using different methods based on different data. We are now making this clearer in the revised paper.

Perhaps the disconnect for me in reading this paper is related to my own slow work about reconciling GCM projections against trends (See Luce and Holden, 2009 and Luce et al., 2013 for instance). It seems that there should be utility in contrasting trends in climate and flow with GCM and hydrologic model retrospectives. It is important to question and hone our precipitation expectations, which seem so deeply uncertain from GCMs. But challenging the GCM projections with raw extrapolations of flow or climate seems like a weak challenge, particularly given that we know there are other periodical trends potentially superimposed. I fear that without demonstrated rigor in the trend analysis, the kind of effort the authors offer will be dismissed by our partners in the climate and atmospheric sciences community.

The reviewer seems to imply here that the trend analysis in the paper lacks rigor, while the methods used in Luce and Holden (2009) and Luce et al. (2013) do provide the necessary rigor. May be we are missing the point here, but it seems to us that Luce and Holden (2009) and the present manuscript are very similar with respect to the trend estimation and its interpretation. Luce and Holden (2009) estimate trends in the distribution of annual runoff at 43 gages and interpret the detected trends in the context of snow melt and climate indices, not unlike the interpretations of this paper. They also make the implicit assumption that the trend will continue into the future when they make management recommendations (which is obviously about the future), e.g. "Water allocation will become increasingly difficult with increasingly low annual streamflows" (p. 4). We therefore cannot see why the Luce and Holden (2009) approach would have advantages over the one used in this paper. Luce et al. (2013) provide more process detail on the comparison between GCM results and trend analyses. We do take the point that more quantitative process detail would strengthen the paper. We have therefore added, where appropriate, quantitative support of the process interpretations in the spirit of Luce et al. (2013).

On a more technical level, their method did not assume a Gaussian distribution of residuals around the trend line while the method used in this paper does. To adopt more rigor, we therefore compared the trend estimates with those using a nonparametric approach based on bootstrapping to estimate distribution-free confidence intervals. The results are given in supplement A of this response. The bootstrap distributions of predicted values turn out to be very close to Gaussian so the results change very little. The expected changes never differ by more than 4% from those of the method used in this paper, and their 95% confidence bounds never differ by more than 21% (period 2021-2050) and 33% (period 2051-2080) from those of this paper. However, we do see the value of the nonparametric approach and have adopted it therefore in this paper, replacing the Gaussian approach in the original manuscript.

I perceive the scientific community already taking on permutations of these three "pillars" through a range of scientific methods examining the sensitivity and consistency aspects through careful dissecting of trends of different time scales and variability from a range of climate processes. I acknowledge that these examinations are commonly of limited spatial scope and perhaps tediously meticulous, but do we have to abandon our sense of caution to effectively make a challenge? Have the various local efforts at incremental progress become too diffuse in their effect? Do we need to consider alternatives that have a touch of the outrageous? Perhaps so, and I'm open to the manuscript doing so; it just seems like a position that requires some justification given the other excellent ongoing work in the community, only a small portion of which is cited.

As mentioned above we now give more detailed justification of the approach adopted.

A Suggestion:

It seems the paper would most benefit from a more questioning stance; asking whether they can do what they would like to do – unless they are able to cite someone else who has it successfully. It would be wonderful and useful if they (or presumably in the future, "we") could apply their approach of comparing among the three pillars. If section 7.1 were framed more in the context of developing a hypothesis about how the three approaches (perhaps with slight refinements for 1 and 3 to acknowledge the potential need for anthropogenic attribution) could frame a genuinely systematic approach to reconciliation, the manuscript would come across more constructively. Then section 7.2 would presumably demonstrate that, in fact, the projections in agreement are more likely to occur. At the very least I would expect it would generate an excellent discussion on potential futuring practices that is informed by some thorough analysis of a large data set.

We take the reviewer's point of adopting a more questioning stance. We have condensed the manuscript by 30% and changed the perspective throughout the paper to better highlight the causes of the differences between the methods.

A Perspective?

This final question is not intended to require modification of the manuscript or response by the authors. It is just here as a point of consideration or perspective relative to the overall framing offered by the current manuscript, which may or may not be helpful to briefly ponder. An underlying conceptualization of all three pillars is in determining the rate of change. One lesson from the various climate modeling exercises is a monotonic trend in temperature. If we do not societally change our fossil energy consumption practices, it is not a question of "if" we will reach 3, 4, or 6 C increases, just "when". If we resolve our temperature uncertainty to instead be a temporal uncertainty, we can recast our questions to be about the sensitivity to temperature and a plausible range of precipitation. Is the timing question so important that we should prioritize that as our fundamental question in hydrology over assuring that we can adequately describe the hydrological system response to a generalized "warming" of 2 to 6 C? Should our three pillars have a heavy weight on timing, or by accepting the eventuality, focus on hydrologic process or sensitivity?

Sincerely, Charles Luce

References:

Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M.: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX). A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp., 2012.

Luce, C. H., and Holden, Z. A.: Declining annual streamflow distributions in the Pacific Northwest United States, 1948-2006, Geophys. Res. Lett., 36, L16401, doi:10.1029/2009GL039407, 2009.

Luce, C. H., Abatzoglou, J. T., and Holden, Z. A.: The Missing Mountain Water: Slower Westerlies Decrease Orographic Enhancement in the Pacific Northwest USA, Science, 342, 1360-1364, DOI: 10.1126/science.1242335, 2013.

SUPPLEMENT A

Original CI

Table #2 Trend projections FOR MID OF PROJECTION PERIOD 2035 for (2021-2050) and 2065 for (2051-2080)

	Hoalp	Muhlv	Gurk	Buwe
Predicted discharge 2050 (m^3/s)	0.28 m^3/s (0.19, 0.38) m^3/s	0.67 m^3/s (0.36, 0.97) m^3/s	1.17 m^3/s (0.48, 1.87) m^3/s	0.02 m^3/s (-0.10, 0.14) m^3/s
Change 2050 (%)	+42% (-5, +88)	-10% (-51, +32)	-36% (-74, +1)	-89% (-156, -21)
Predicted discharge 2080 (m^3/s)	0.35 m^3/s (0.20, 0.51) m^3/s	0.58 m^3/s (0.07, 1.09) m^3/s	0.74 m^3/s (-0.42, 1.90) m^3/s	-0.08 m^3/s (-0.29, 0.12) m^3/s
Change 2080 (%)	+78% (1, 156)	-21% (-91, +48)	-60% (-123, +3)	-145% (-258, -33)

BOOTSTRAPED CI (5000 replications)

Table A.2 Trend projections FOR MID OF PROJECTION PERIOD 2035 for (2021-2050) and 2065 for (2051-2080)

Table 2

	Hoalp	Muhlv	Gurk	Buwe
Predicted discharge 2050 (m^3/s)	0.28 m^3/s (0.19, 0.37) m^3/s	0.68 m^3/s (0.45, 1.02) m^3/s	1.19 m^3/s (0.58, 2.00) m^3/s	0.02 m^3/s (-0.14, 0.14) m^3/s
Change 2050 (%)	+39% (-7, +71)	-8% (-41, +34)	-36% (-72, -1)	-90% (-177, -22)
Predicted discharge 2080 (m^3/s)	0.35 m^3/s (0.22, 0.45) m^3/s	0.60 m^3/s (0.15, 1.14) m^3/s	0.74 m^3/s (-0.23, 2.01) m^3/s	-0.08 m^3/s (-0.33, 0.12) m^3/s
Change 2080 (%)	+74% (0, 123)	-21% (-79, +51)	-59% (-113, +9)	-148% (-282, -36)

Figure A.1. Bootstrap distribution of trend projection for Hoalp, period 2065 for (2051-2080)

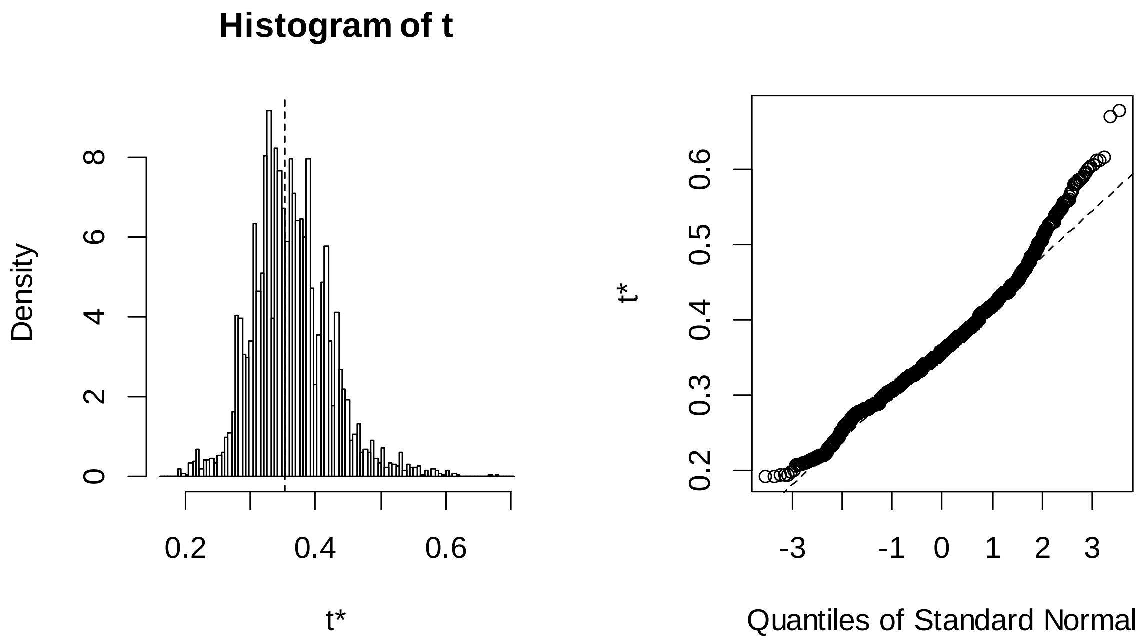

Figure A.2. Bootstrap distribution of trend projection for Muhlv, period 2065 for (2051-2080)

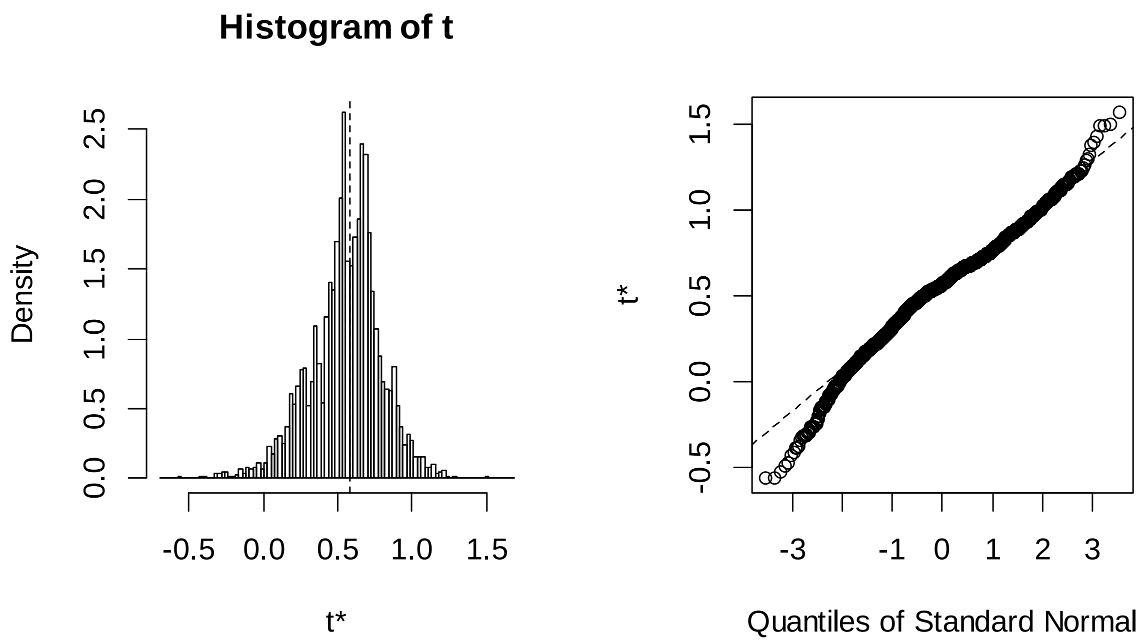

Figure A.1. Bootstrap distribution of trend projection for Gurk, period 2065 for (2051-2080)

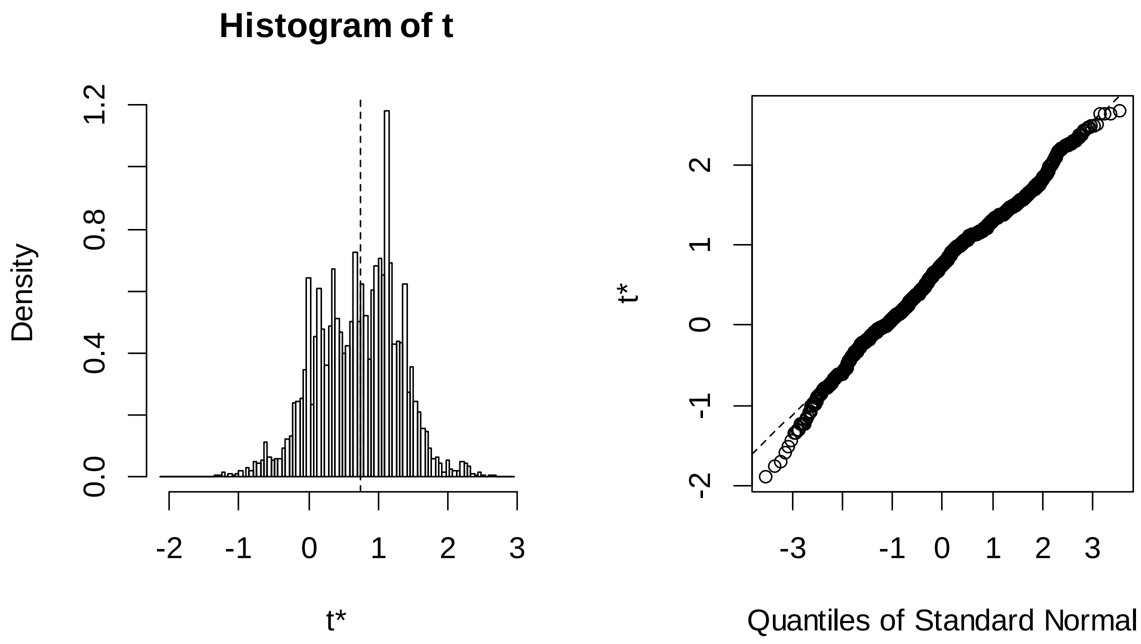

Figure A.2. Bootstrap distribution of trend projection for Buwe, period 2065 for (2051-2080)