Reply to the comments of Dr. Stefan Hagemann (in blue):

Major remarks

Within the last few years, the bias correction of climate model output has become a hot topic within climate impact research. Here, the authors investigate the impact of bias correction on simulated runoff regimes and the relative change in selected runoff indicators over two mountainous catchments. They used bias-corrected RCM output to force four different hydrology models of various complexities. The paper is a valuable contribution to the bias correction topic and associated uncertainties attached to the GCM-RCM-Hydrology model modelling (HM) chain. Generally, the paper is well structured and concisely written. But the conclusions section needs some rewriting. Especially the following issues should be clarified and discussed more thoroughly.

1. Even though the paper generally writes about bias correction, it actually uses a specific bias correction method: “Local intensity scaling” for precipitation, additive correction for temperature, both on a monthly basis. Thus, in some aspects the results might be specific for this combination of methods.

That's true, so we changed line 10-11 of page 10228 to "The focus of this work is on the impact of the bias correction methods used in our study on simulated runoff characteristics and their climate change signals." Furthermore, we mention this limitation by adding that we should validate on "a larger set of both, climate simulations and bias correction methods to improve the robustness of this conclusion." (page 10229 line 21). Just to be safe, we also highlighted "bias correction of regional climate simulations" in line 24 of page 10228, as we did not investigate correction of GCM data as an input to hydrological models.

2. Results and main conclusions may be specific for the climate regime covered by the two catchments considered, i.e. mountainous regions with snowfall/snowmelt dominated hydrological regimes.

It's true that our results only represent specific types of catchments. Therefore, we added in line 17 (page 10229) "Based on our results for two humid, snowmelt affected catchments, we assume..." and in line 20 "However, this methodology should be validated upon a wider set of catchments representing different runoff regimes (spatial validation) and...".

3. It is written (P. 10229 – line 9-10) that “Bias correction can be seen as mostly unnecessary to obtain the climate change signal.” In this respect, you have to specify what climate change signal means. Clearly, if you want to obtain the climate change signal from the corresponding climate model, you don’t need bias correction. Presumably you want to obtain the climate change signal in the simulated runoff. If only one or a few GCM-RCM-HM ensembles are used, this is certainly not correct if the climate change signal of the climate model output is changed by the bias correction. Thus, your conclusion requires the use of a large ensemble where these changes in the climate change signal of the climate model output cancel each other.

True, we have to clarify the conclusion by adding "of the investigated hydrological indicators.", as you presume! Yet, regarding you comment on the need for a large ensemble, this is the assumption we

state in line 7 ("based on our results and in this large ensemble context..."). Of course this ensemble is not very large , but the effect is already visible. Hope you are fine with that?

It is written that "In particular, more strongly biased climate simulations are more likely to have their climate change signal affected by bias correction." This, in fact, is a mathematical property of the bias correction method and depends on the structure of the bias. A systematic constant (independent of the value of the specific variable) bias is easy to correct, e.g. by subtracting the bias, and this would not change the associated climate change signal, even if the bias would be very large. Using a quantile mapping based transfer function approach (statistical bias correction) is used, the impact of the bias correction method of the climate change signal has been visualized graphically by Haerter et al. (2011). Dependent on the specific method the mean signal changes or remains unchanged.

We did not formulate clearly here. It better should read: "In particular, the effect of bias correction on the change signal of hydrological indicators is larger for more strongly biased climate simulations, while the average signal of a large ensemble including those simulations is hardly affected."

4. In the method section (p. 10214-10215), research questions are described. This is usually part of the paper's outline/purpose in the introduction section and should consequently be moved to the introduction.

We agree that the research question should be moved at the very end of the introduction. Because of this, the lines 12-16 of page 10211 have been integrated in the research questions moved here. Of course, the definition of the change signal etc. were left in 2.2, but this had to be reformulated a bit.

Why do you mention that the HYDROTEL model uses two potential ET formulations? Do you use both? If yes, do you count this as an extra hydrological model (I don't think so)? Is this important for the paper's outcome? If not please restrict yourself to one formulation.

We are sorry for this confusion! HYDROTEL was set up using the approach of Fortin (2000) in Quebec only. Because this approach was only tested in Quebec before plus there is more data available in Bavaria, HYDROTEL was set up in Bavaria using the Thornthwaite approach. This is not ideal, but the consequences for our study are small. The HYDROTEL team also tested the Fortin approach in Bavaria and found no big differences during the reference period. To clarify, we rewrote 10219/27 - 10220/5:

"c) In HYDROTEL, potential ET is computed by an empirical formulation (Fortin, 2000) for Quebec or by the Thornthwaite approach for the Bavarian region. Potential ET is then reduced to an actual value based on soil water availability. d) "HSAMI also estimates evapotranspiration with the empirical formulation of Fortin (2000) using minimum and maximum air temperature only."

5. The description of the hydrological model ensemble (Sect. 2.4) lacks some clarity in the presentation. I suggest making a table with the different model characteristics that allows an easier comparison between the models. Please provide also the spatial resolution (in ° or km) for each of the models.

Good idea, Table 2 (Characteristics of the hydrological model ensemble.) has been added.

	HSAMI	HYDROTEL	WASIM-ETH	PROMET
Model type	conceptual	mixed	mixed	process-based
Resolution (temporal, spatial)	24h, lumped	24h, HRUs (hydrological response units)	24h, fully distributed (1x1 km^2)	1h, fully distributed (1x1 km^2)
Meteorological inputs	Temperature, precipitation		Temperature, precipitation, humidity, wind speed, radiation	
Evapotranspiration (ET)	Potential ET, empirical (Fortin, 2000)	Potential ET, Fortin or Thornthwaite	Potential ET, Penman-Monteith	Actual ET, Penman-Monteith
Soil water model	Saturated & unsaturated zone reservoirs	3 soil layers, infiltration approach	Multiple layers, Richards' equation	4 soil layers, Philip equation
Snow pack model	Temperature-index approach	Temperature-index incl. energy balance	Temperature-index approach	Snow pack energy balance

Minor Comments

In the following suggestions for editorial corrections are marked in *Italic*.

p. 10206 – line 7 … layer of *uncertainty*.

OK, that's clearer.

Introduction – p. 10210 In the discussion of bias correction, I suggest referring also to Themeßl et al. (2010) who compared different bias correction methods and found that quantile mapping based approaches show the best performance in reducing biases, particularly at high precipitation quantiles. Also noteworthy is the study of Hagemann et al. (2011) who showed that bias correction may have an impact on the climate change signal for specific locations and months.

These are noteworthy papers, so we included the following on page 10210 line 26 before "Hence...":

"Hagemann et al. (2011) report that bias correction of GCM data may affect the hydrological climate change signal in specific locations and seasons."

And on page 10211 line 8 before "River runoff...":

"The methods were selected for their simplicity and have some inherent flaws. The monthly correction may create jumps in the corrected datasets between months, and following Themeßl et al. (2011) LOCI performance is slightly inferior to the quantile mapping approach, especially at high precipitation intensities."

p. 10211 – line 10-12 Sentence is difficult to read. Please rewrite!

Hopefully replaced with something better: "From the simulated daily runoff for the reference period, hydrological indicators characterizing mean, high and low flows as well as the timing of the spring flood are computed and compared to observations."

p. 10212 – line 10 … *winter, only* ….

Changed.

p. 10212 – line 15 … *Alps, therefore, the* …. It seems you are using too much ";" in places where they are not used in English. Please check manuscript appropriately.

Yes, thank you for that hint. We tried to eliminate all of those...

p. 10216 – line 23 to p.10217 – line 2 Lengthy sentence is difficult to read. Please rewrite!

OK. Now split up in four sentences: "...takes advantage of three relations: a) Elevation dependencies... b) Physical relationships... c) Empirical monthly..."

p. 10217 – line 4 … couple *a* RCM …

OK.

p. 10217 – line 10 It is written: “This cold bias is also present at a much larger scale in the corresponding CGCM simulations (not shown), suggesting that large temperature biases in the driving data propagate through the modeling chain.”

This is one possible explanation. If CGCM and CRCM use similar model formulations and parameterizations, it might also point to a model problem that this specific model family might have over Haut Saint-Francois. As CGCM has not been used to force different RCMs over this area, no specific conclusion can be made.

True, in this paper no conclusions can be made regarding this. And it is also irrelevant, so we deleted the second half of the sentence "... suggesting that large temperature biases in the driving data propagate through the modeling chain."

p. 10217 – line 29 It is written: … a monthly correction is performed at the RCM grid point scale on air temperature by subtracting the 30-yr mean monthly biases.

This leads to unrealistic jumps in daily temperature time series at the end of the month to the beginning of the next month , which may cause some problems. Did you take this into account? Please add a note on that!

Added the following note: "Since the biases in temperature vary only weakly between months (Fig. 3), the discontinuity introduced in the corrected data is very small."

p. 10218 – line 20-23 Complicated sentence is difficult to read. Please rewrite!

OK, now it reads: "In terms of climate change signal from the chosen RCMs for the 2050 horizon, the *Haut Saint-Francois* region is projected to see its temperature increase by about 3 °C with up to 4 °C in Winter. Precipitation is projected to increase by up to 30% in Winter, about 20% in Spring and Fall and to decrease slightly during Summer months."

p. 10226 – line 2-3 … climate *models* for winter … … contributes to *runoff*.

As this sentence also included a ";" we now changed it to: "...projected by the climate model simulations for winter (Fig. 4), leading to faster snow melt and hence to a shift in peak runoff."

p. 10226 – line 4 … compared *to* the …

Of course!

p. 10226 – line 14 It is written: “… the importance of the ensemble is front and center here …”
I don’t understand! Please rewrite!

Hopefully this is better: "However, this highlights the importance of an ensemble projection, as the other two RCMs show a wide range of positive and negative signals both with and without bias correction."

p. 10226 – line 16-17 It is written: “ … give a pessimistic outlook on the possibility of reaching a conclusion …” I don’t understand! Please rewrite!

Yes, a little complicated. Now it reads: "At last, the large amount of uncertainty in the ensemble results demonstrates that it is hard to reach a conclusion for the high flow indicator."

p. 10227 – line 3 … impact *on* the …

Sure.

p. 10227 – line 15 It is written: "… is clearly shown in the results for all scenarios …" I don't understand. You do not use different emission scenarios! Please rewrite!

True, this sentence can be misunderstood. We rewrote this to: "The effect of using an ensemble of multiple climate models is clearly shown by the few rejections of the null hypothesis, when indicators are analyzed based on all available RCM runs." Accordingly, figure 9 has be changed.

p. 10227 – line 17 … all *for* the …

Sure.

Fig. 3, 4, 7, 8 The legend text and axis descriptions are too small! Please increase size!

Fig. 3 You should indicate temperature and precipitation panels in the figure caption as well as the catchments, such as you are doing for Fig. 4.

Fig. 4 … temperature (*upper panels*) and precipitation (*lower panels*) over…

Fig. 7,8 The dots are small and hard to separate Figs. need to be improved.

OK, the figures have been redone (see below). We hope that they are well done now.

All in all, thank you very much for your constructive comments! We think with your help some parts of the paper could be clarified and rounded out.

Yours,

Markus Muerth and co-authors

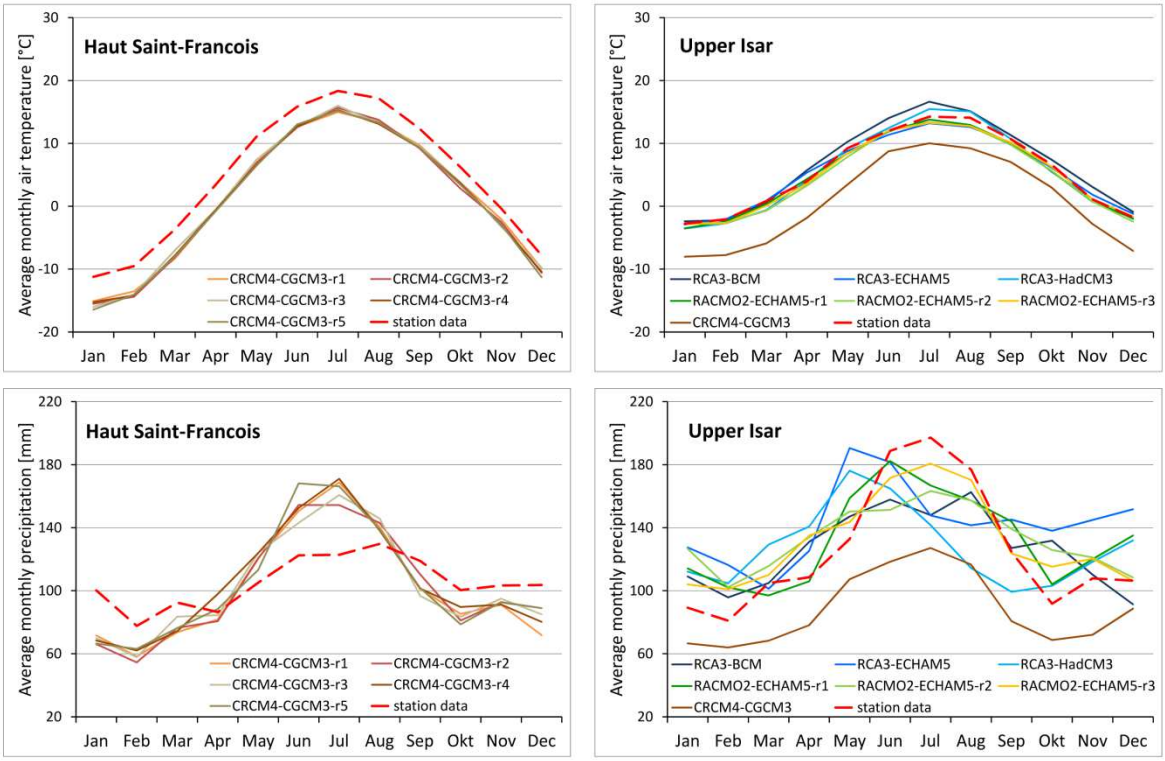

Fig. 3: Climatology of air temperature (upper panels) and precipitation (lower panels) over the main catchments Haut Saint-Francois (left panels) & Upper Isar (right panels) derived from climate models and observation. (Acronyms refer to RCM-GCM combinations and runs in a GCM member ensemble.)

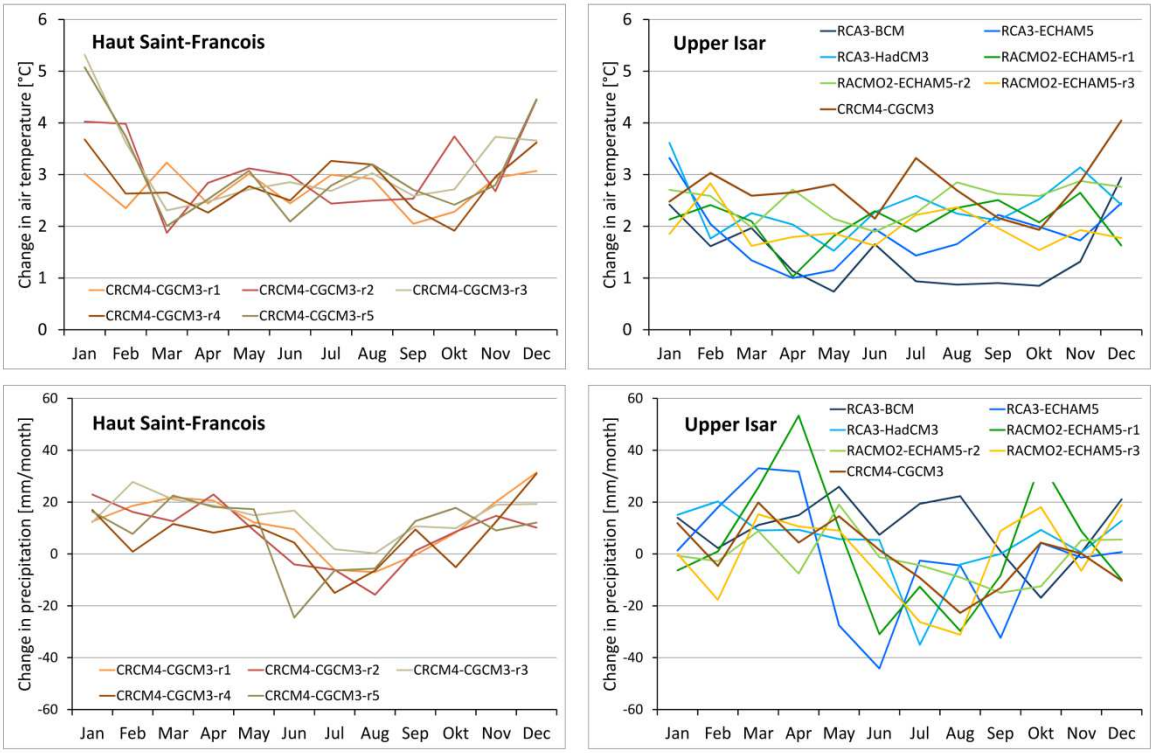

Fig. 4: Climate change signal of air temperature (upper panels) and precipitation (lower panels) over the two main catchments Haut Saint-Francois (left panels) & Upper Isar (right panels) between the

reference (1971-2000) and the future (2041-2070) period. (Acronyms refer to RCM-GCM combinations and runs in a GCM member ensemble.)

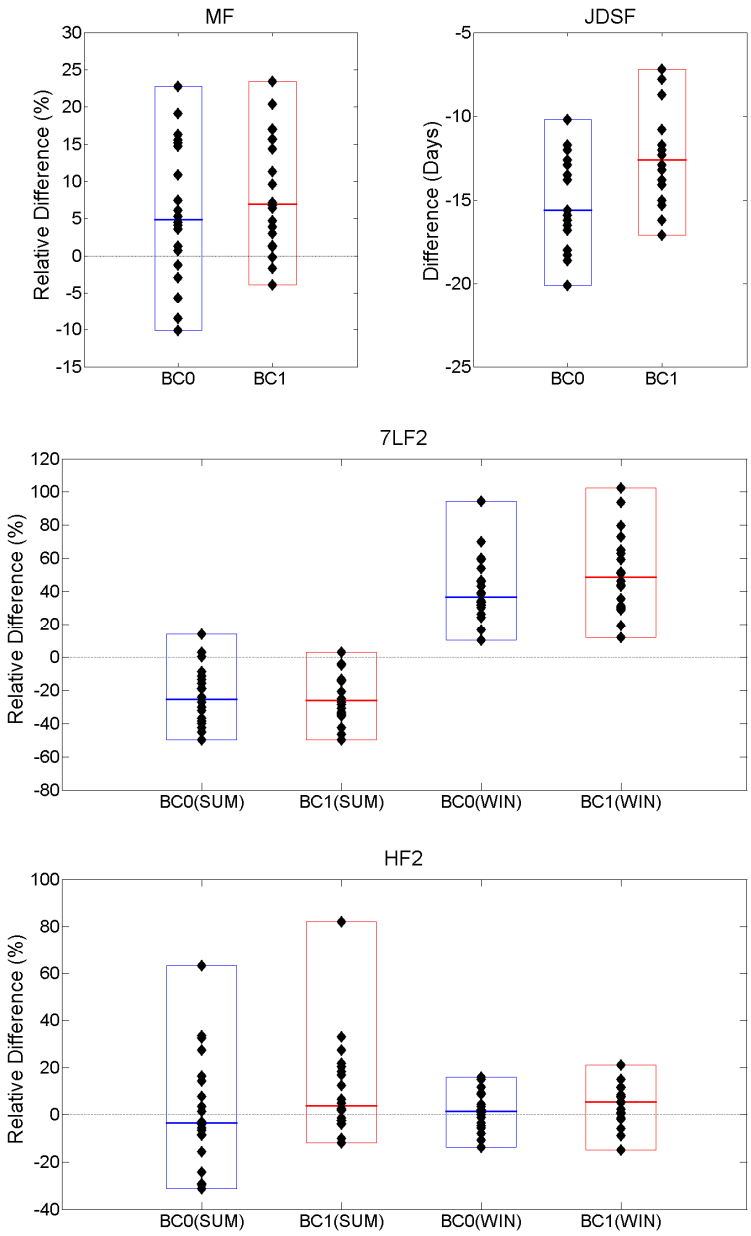

Fig. 7: Relative change of the investigated indicators between reference and future period at Saumon based on five members of the CRCM-CGCM ensemble over Quebec with horizontal bars indicating the median value.

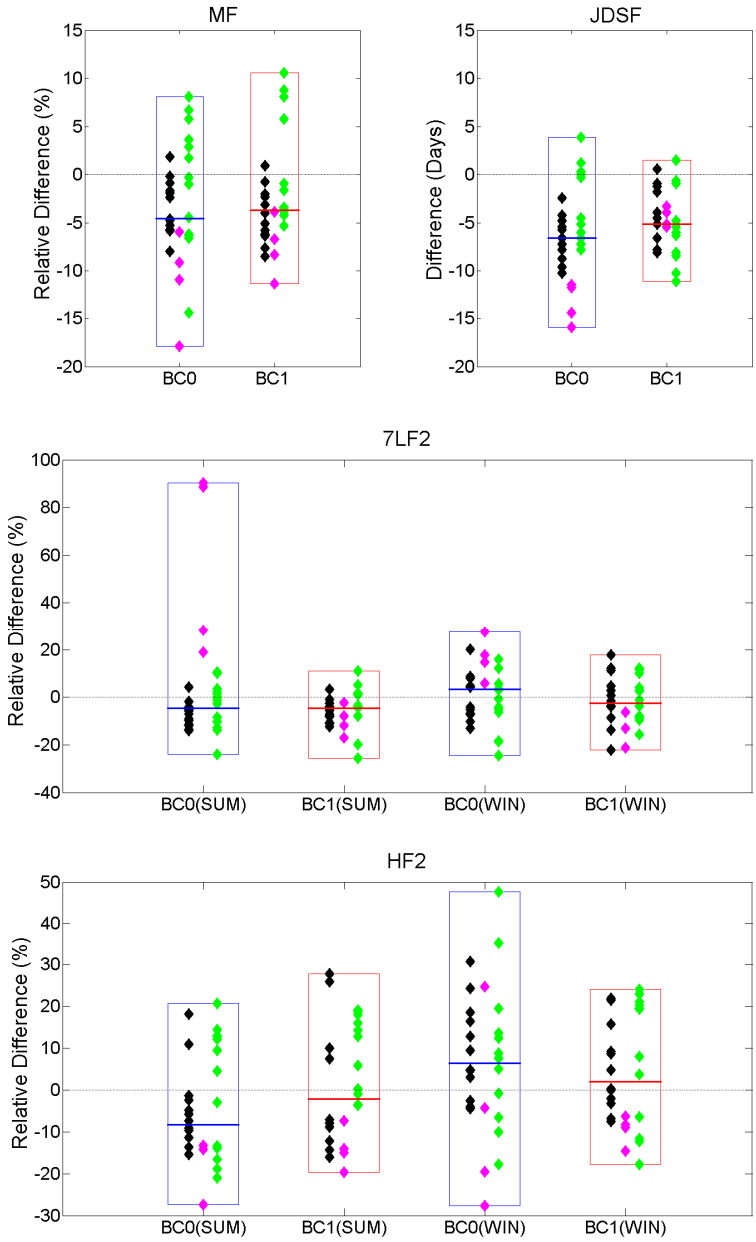

Fig. 8: Relative change of the indicators between reference and future period at Schlehdorf. The black dots indicate the RACMO simulations driven by ECHAM. Green dots specify RCA simulations driven by different pilots (BCM, ECHAM & HadCM); pink dots indicate the CRCM-CGCM simulations; horizontal bars indicate the median of all dots.