Replies to the comments by David Post

We would like to thank David Post for his generally positive comment and for his suggestions for improving our manuscript.

Below, reviewer comments are in italic font and our replies are in normal font.

This is a nice example of a study that attempts to determine exactly what it is about rainfall-runoff models that means they are not capable of predicting well runoff under changed climate conditions. One major thing that would improve the paper would be to quantify for the reader what the change in relevant hydroclimatological characteristics during the verification period actually are. The authors state that the area was subject to significant climate changes, but do not tell us what these actually were. Were the evaluation periods drier/hotter? If so, by how much. What were the relative runoff coefficients?

> Thanks for this advice. Over the study period, precipitation, air temperature and E_{ref} increased, as shown in Fig. 5 (a–b) and Fig. 7. We will add a description of the changes in the hydro-meteorological conditions to Section 2.1.
>
> Over the period 1977–2014 annual precipitation increased by 32 ± 23 mm yr^{-1} or 2.4 ± 1.7 % per decade (based on undercatch corrected SPARTACUS data), air temperature increased by 0.45 ± 0.09 °C per decade and global radiation increased by 5.1 ± 0.9 W m^{-2} per decade on average over the study catchments. In contrast, discharge did not show strong trends and the average trend over the study period was 0.2 ± 3.1% per decade (Duethmann and Blöschl, 2018).
>
> Relative annual runoff coefficients (based on observed discharge data and undercatch corrected SPARTACUS precipitation data) vary in the range of 0.22 and 0.86 between the study catchments. The runoff coefficients significantly decreased ($p \leq 0.05$) in 29 and significantly increased in 4 of the 156 study catchments. (A large number of catchments showed insignificant decreases, probably due to the large interannual variations of the annual runoff coefficients).

Despite these issues, I have just three comments on improving the paper:

1. The title is misleading. Almost every model will predict discharge changes in response to climate change. The questions is why they do not 'accurately' predict discharge changes? The addition of a qualifier like 'accurately' would be useful.

> We will revise the title and add 'correctly' as a qualifier. The title then reads 'Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change?'.

2. Changes in anthropogenic influences are largely ignored as the authors claim that the catchments are largely unregulated and existing diversions were introduced before the beginning of the study period. I would question this. While the diversions may be in place before the beginning of the study period, are there operating rules related to this diversions which may vary from year to year, for *example allowing larger diversions during periods of low flow (or vice-versa). I ask as we have identified catchments in Australia that not only behaved abnormally (gave lower than predicted yields during the Millennium drought), but that have not returned to 'normal' yields post-drought. One hypothesis for this is that farmers sank groundwater bores to access an alternative water supply during the drought when they were unable to pump from surface water. Any lowering of the groundwater table resulting from this activity would obviously lead to lower than expected yields. Once this 'sunk cost' had been incurred, there would be no benefit to farmers in ceasing the pumping of water from these bores, thus they may still be doing so post-drought. Such anthropogenic influences are of course hard to determine (and even harder to quantify), but the authors would do well to keep them in mind.*

Changes in private abstractions by households or farmers are indeed difficult to get hold of. In Austria, water abstractions for irrigation are much less important than in Australia. With respect to anthropogenic impacts on water resources, diversions for hydropower generation are much more relevant than abstractions for irrigation. Irrigation in agriculture is most relevant in small areas east, southeast and northwest of Vienna, where estimated irrigation amounts of agricultural areas exceed 10 mm/year (BMLFUW, 2011). In most parts of Austria, estimated irrigation amounts of agricultural areas are less than 1 mm/year. The fraction of arable land in our study catchments is only small (5% on average over the catchments) and the catchments in our study hardly overlap with those areas where agricultural areas receive a large amount of irrigation. At this stage we therefore assume that changes in irrigation amounts are not a major source for the deviations between the simulated and observed discharge changes.

3. The assessment that problems with the model calibration can be the source of the poor performance during the evaluation period is a good one. In particular, that processes that are relevant in the calibration period are not present (or 'activated' to use the author's terminology) in the calibration period. I am not sure that extending the calibration period from 5 to 25 years will actually evaluate whether this is the case. It may be that these processes will be seen in the 25 year period, but it may not. One thing that could be done is to compare the model that is calibrated on the evaluation period (or perhaps part of it) to the model that is calibrated on the calibration period. If different processes are dominant in the evaluation period, this would be seen in how these models perform on an independent data set.

In the original version of the model, the model is calibrated in a 5 yrs period and then evaluated in 6 other 5 yrs periods. For example, the model calibrated in 1978-1982, is evaluated in 1983-1987, 1988-1992 and so on. If this model performs well in calibration (e.g. in 1978-1982) but performance is worse in evaluation (e.g. 1983-1987), this might be due to a process that was seen in the evaluation but not in the calibration period. If the model is calibrated over a 25 yrs period that includes both 1978-1982 and 1983-1987, the process that was relevant only in 1983-1987 is now included in the calibration period. If there are potentially additional processes that are however not seen in the 25 yrs calibration period, these processes cannot explain the decrease in model performance when e.g. calibrating in the first 5 yrs of this period and evaluating over the other 20 yrs of this period.

References

Duethmann, D., and Blöschl, G.: Why has catchment evaporation increased in the past 40 years? A data-based study in Austria, Hydrol. Earth Syst. Sci., 22, 5143-5158, 10.5194/hess-22-5143-2018, 2018.

BMLFUW (2011): Irrigated areas in Austria – final report (Bewässerte Flächen in Österreich – Endbericht), in German. https://gruenerbericht.at/cm4/jdownload/download/28-studien/470-39-abschaetzung-der-bewaesserungswuerdigen-flaechen (last access 11. March 2020).