Response to Referee #3

We thank Referee #3 for the profound evaluation of the paper and the helpful comments. Accordingly, the following revisions were made and added to the reviewer comments (our response is highlighted in blue and italic type):

Schneider et al presented an interesting study with the WaterGap Model. The results of projected changes in streamflow and analyses for different climate zones in Europe are of interest. This is a well-written paper and I recommend it for publication, pending further explanations from the authors and the following revisions.

Remarks: - Introduction: please provide an overview of previous studies that focussed on climate change impact on streamflow and flow extremes in Europe. References to relevant studies are missing (e.g. Dankers and Feyen, 2009; Feyen and Dankers, 2009)
Response: We agree on this. The introduction was extended by an additional paragraph covering related studies and how our study can be distinguished.

- While the Watergap model simulations were performed on a high spatial resolution 5' x 5' (which is about 6 x 9 km2 for Central Europe as mentioned in 9199 line 12) the authors decided to use output from GCMs rather than RCMs to force the hydrological model. I am really wondering why the authors used GCM data and not RCMs data. I think that use of RCM output (as available for example from the ENSEMBLES project) would be more logical choice, so the authors need to explain why they used GCM rather than RCM data.
Response: The analysis was conducted in the context of project work (i.e. within the WATCH project) which provided climate projections for the three applied GCMs. Apart from the spatial resolution (0.5° in WATCH vs. 25km in ENSEMBLES), we think that the appliance of these GCM projections are a reasonable choice for our impact analysis for the following reasons:

- *GCM output was available on a daily time step and covered the whole of Europe*
- *a statistical bias correction procedure was conducted on daily precipitation as well as on minimum, maximum and mean surface air temperature based on transfer functions. In comparison to the delta-change approach, which was often used before, this procedure provided more suitable climate forcings for the use in hydrological impact studies as more reliable data of changes in hydrological variability and extremes were obtained.*
- *In addition, the WATCH project provided daily observed time series for the climate variables (WATCH forcing dataset, WFD) which were used as the reference data for the statistical bias correction and could be used for the calibration of WaterGAP3.*

- On page 9200, line 17 they mention that they simply disaggregated the climate forcing to 5', but does this mean that 0.5â´LŸ grid with climate forcing were simply split up to 5' climate forcing? Please explain this, as it does not seem the most elegant approach for producing high resolution streamflow projections for Europe.
Response: The climate forcings were simply split up to the 5 arc minute grid cell raster and we agree that more elegant approaches for the downscaling to higher spatial resolution are available. We used this simple approach as the WATCH-forcing data were applied for the bias correction on 0.5° x 0.5° resolution. By applying a more sophisticated approach on downscaling, the link between bias-correction and WATCH forcing data could get lost.

- The authors use output of three GCM experiments to account for some uncertainty. They explain why they focussed on experiments for SRES A2. However, it would be good to explain the choice of the three GCMs in their study and why they didn't select a larger number of climate models.
Response: Done. The choice of the three GCMs was included in the text. Please see also response two questions prior to this.

- Please explain why the baseline was performed with the WATCH–forcing data (pag 9200 line 13).
Response: We agree that the use of the WATCH forcing data introduces additional uncertainty in the analysis. Hence, we have conducted new WaterGAP runs for the baseline forced with the GCM data of CNCM3, ECHAM5 and CNCM3. The comparison against the WATCH forcing data was replaced in the paper by the new results.

- While the authors mention that the WaterGap model is calibrated for Europe using 221 GRDC gauging stations (page 9200 line 1-3) I think it would be good to explain briefly how this calibration was performed (which data set used, which period, which parameters). It is also necessary to include results (an extra figure or table) to summarize the quality of the streamflow simulations by the WaterGap model for the historical (baseline) period or otherwise to refer to a paper/report that shows an evaluation of the WaterGap modelling performance on 5' with this reference dataset.
Response: A reference to the calibration process (Döll et al. 2003) was included in the paper which describes the complex calibration process in detail. Additionally, further information about the calibrated parameter was added to the text. Since Döll et al. 2003, the calibration process has not changed. We agree that for the 5' version of WaterGAP no complete validation results are published in the scientific literature so far. Only in Schneider et al. 2011, a validation was done for ecological important flood flow events. Hence, validation results are provided for the 5' version of WaterGAP as supplement to the paper.

- Overall, I think the selected hydrologic parameters of the IHA are a good choice. However, I was wondering why a 1-day minimum flow and Julian data in 1-day minimum flow are selected and how meaningful these parameters are for this assessment. I would expect that a minimum flow for certain duration would be more relevant for analysis of changes in low flow than 1-day minimum flow.
Response: Olden and Poff provided an analysis on redundancy that stated that the IHA cover almost the total range of available indicators, but some of them are redundant. Based on redundancy matters, we selected 12 out of the 32 indicators, whereas we took the mean and the variability of change for each indicator. Certainly, if two indicators are redundant it can be discussed which indicator is more reasonable. Impacts on low flows are well represented in our study by the 1-day minimum magnitude, the timing of the 1-day min flow and the number of low pulses. We decided to use the low flow magnitude (instead of the mean duration of low pulses) due to the following reasons:
- *the low flow magnitude has a strong impact on the concentration of pollutants and hence on water quality related issues*
- *it defines other habitat attributes such as flow velocity and water temperature*
- *it determines the wetted area and the habitat volume and hence the habitat suitability (e.g. it concentrates fish and other organisms benefiting predators)*

- Döll et al. (2009) and Döll and Zhang (2010) also used indicators of ecologically relevant flow using the WaterGap model on a global scale. Please include these studies as references.
Response: Done. Döll and Zhang (2010) were implemented in a new paragraph on related studies in the introduction and also mentioned under Results.

Minor remarks:
- Page 9194, line 26: what do you mean with "validation on robustness"? Please rephrase.
Response: Changed. The goal of our study was to show how homogenous the results are for each climate zone rather than a test on robustness. Hence, the phrase 'robustness' is inappropriate and was removed from the paper.

- Page 9195, line 17: increase in evap is not only due to higher temperatures.
Response: We agree as evaporation also depends on factors such as radiation, wind speed, air humidity and soil water content. Hence, the sentence was rephrased as follows: "Higher

temperatures could potentially increase evaporation rates at surfaces and transpiration by plants which leads to a reduction in runoff (Frederick and Major, 1997)"

- Page 9196, line 15: Do you mean "thermoelectric cooling"?
Response: Done. Sentenced rephrased. The cooling of thermal power plants was meant.

- Page 9201, line 13 and header in Table 3: explain what you mean with ensemble median results. (you used only three GCMs).
Response: In our analysis for the paper, we calculated the ensemble median and the ensemble mean of the three climate projections. The impact on the results was not crucial and did not affect the general findings of our study. It can certainly be discussed if the ensemble median or the ensemble mean is more appropriate. As we make use of only three GCMs in our study, we agree and included the results of the ensemble mean in the paper.

- Page 9203, line 2: "threshold exceedance" is a difficult term to understand as reader and I think this term needs more explanation in this sentence.
Response: We agree on this. The sentence was removed and a more detailed explanation is given at the end of Chapter 2.2: "As natural flow regimes can be modified in various ways, those indicators showing a threshold exceedance (i.e. the alteration to natural extends ±30% or 30 days respectively) were summed presuming that each indicator has the same weight."

- Pag 9204, line 22: for consistency I would suggest to use "snow cover", which is presented in Fig 2.
Response: Done. Sentence rephrased.

- Pag 9207, line 22: "In the following section,"
Response: Done. Sentence rephrased: "In this chapter, …"

- Pag 9208, line 5: check capital "central location"
Response: Done.

- Pag 9208, line 20: I would recommend to use the word "increases " rather than "elevates"
Response: The word 'increase' is used quite often in our paper. To have some alternation in such terms, we implemented the term 'elevated' at this part of the text.

- Pag 9210, line 17: replace "they" by "these events"
Response: Done.

- Pag 9212, line 11: decreasing by 8.0% in the ensemble median (is this an average for the whole region)?
Response: Sentenced rephrased to make it clearer: "While mean winter precipitation increases only slightly, mean summer precipitation is decreasing by 8.0% over the whole region in the ensemble mean."

- Pag 9214, line 4-6: "Our results show that besides other anthropogenic factors, climate change may severely alter natural patterns of flow. I miss the reference to Döll and Zhang (2010) who focussed on impacts of both climate change and anthropogenic impacts on streamflow (see reference below).
Response: Done. Reference included in the introduction as well as in the results & discussion chapter.

- Fig 5: This figure is not very well readable. Perhaps you could increase the contrast in colours (e.g. dark green or white background for European continent) to increase readability of the figure.
Response: Done. Contrast increased.

- Fig 8-13: Would be interesting to include also observed annual flow to show how simulated flow corresponds with observed flow.

Response: In our analysis we are addressing natural flow regimes and the impact of climate change on those in the 2050s. That is, human impacts such as water abstractions and dam regulation are not considered and were disabled in all WaterGAP runs for this study. We are afraid that adding the existing (observed flow) to the charts will lead to misconceptions as then existing flow regimes of the present are compared to natural flow regimes that would have existed in the same period, but without anthropogenic water abstractions and dam regulation. Hence, we decided not to included observed flows in Fig. 8-13. However, a validation of simulated existing flow against observed data (including figures and efficiency criteria) at 238 stations was added as supplement to this paper.

References

Dankers, R., Feyen, L., 2009. Flood hazard in Europe in an ensemble of regional climate scenarios. Journal of Geophysical Research-Atmospheres, 114.

Döll, P., Fiedler, K., Zhang, J., 2009. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. Hydrology and Earth System Sciences, 13(12): 2413-2432.

Döll, P., Zhang, J., 2010. Impact of climate change on freshwater ecosystems: a global-scale analysis of ecologically relevant river flow alterations. Hydrology and Earth System Sciences, 14(5): 783-799.

Feyen, L., Dankers, R., 2009. Impact of global warming on streamflow drought in Europe. Journal of Geophysical Research-Atmospheres, 114.