Review of Forzieri et al. – Ensemble projections of future streamflow droughts in Europe – by A.F. Van Loon

In this paper the authors simulated future changes in streamflow drought deficit and low flow indices across Europe by using forcing data from a variety of models (GCMs and RCMs) with a climate change scenario as input for a hydrological model. They also included the effect of a scenario of changes in water use by coupling the hydrological model to a water abstraction model. They conclude that rivers in large parts of Europe will be negatively affected by climate change and that water use will aggravate drought conditions especially in Southern, Western and Central Europe.

This research is interesting and relevant and the topic deserves publication in HESS. Although the paper is quite long and could at some points be more concise, I think that it is very well written and the results are presented in a clear manner. As reviewer I really appreciate the effort that the authors took on writing a good manuscript.

My main concern with the research presented is related to the fact that the authors neglect the effect of using multiple climate change scenarios, multiple hydrological models, and multiple water use scenarios. For example, Hagemann et al. (2013) found that the "spread resulting from the choice of the hydrology model is larger than the spread originating from the climate models over many areas." Here, the authors mention in the conclusions that "hydrological uncertainty – here not accounted for – may further increase the variability in the low flow projections as suggested by the considerable discrepancy between large-scale hydrological models in the evaluation of drought propagation", but they do not mention the uncertainty related to using different climate change scenarios and water use scenarios. The whole idea of scenarios is that all of them should be used to get a clear picture of the range of possible futures as scenarios are all possible realisations of the future. I understand that doing the analysis for a multitude of climate scenarios, a multitude of climate models, a multitude of hydrological models, and a multitude of water use scenarios is very demanding and might not be feasible for one research group in a limited time span. However, I think it is very important to clearly discuss this issue in the current paper. Now the reader cannot compare results of this study with other studies, e.g. that of Feyen and Dankers (2009), because both a different climate change scenario and a different hydrological model were used. These issues cannot be neglected. The least the authors should do is to underpin the choice of climate scenario and hydrological model in the introduction or methods section and discuss the effect of these choices in the discussion or conclusion section. Furthermore, the authors could test the effect of the four different water use scenarios, because I estimate that that is most easy to implement in their modelling scheme.

Another general point is that the authors state in the introduction that drought is a natural phenomenon, but in the results and discussion section use the word drought also for the situation influenced by water use. The authors should refer to the discussion of the definitions of drought and water scarcity, as it is summarised by the European Expert Group on Water Scarcity and Drought at the following website: http://www.globalwaterforum.org/2013/08/26/how-to-distinguish-water-scarcity-and-drought-in-eu-water-policy/.

Specific comments:

p.10721, l.1-2: Please provide references.

p.10722, l.4-5: "Undisturbed catchments": as this does not relate to human influence anymore, use "however" to show the contract with the previous sentence.

p.10722, l.6-8: "dryness": vague term > leave it out.

"agricultural": better to use the term "soil moisture drought" as other sectors might be impacted by low soil moisture levels.

"water supply": related to water distribution issues > change to "water availability" p.10723, l.27-28: "multiple driving climate scenarios": this is not done in this study. Please clarify.

p.10725, l.5-6 & p.10726, l.26-27: By bias-correcting precipitation and temperature, but not bias-correcting the other meteorological variables that are needed for the calculation of potential evapotranspiration, like vapor pressure, wind speed etc., inconsistencies will arise between precipitation, temperature and PET data. Please discuss the implications.

p.10727, l.11: Please provide some basic information on the calibration of the LISFLOOD model. This is needed because the calibration of the model is used as an argument in the discussion on p.10747, l.16.

p.10728, l.12: "Maximum Likelihood method" > introduce the abbreviation (ML), because it is used later on in the paper.

p.10729, l.1: "Q80": I guess you used a fixed threshold, so equal threshold throughout the year (see l.28 "annual analysis")? Please add this. I also guess you recalculated the threshold for observed and simulated time series (see p.10737, l.12)? This is important because it implies that you are only considering relative differences between droughts in observations and simulations in Fig. 3.

p.10731, l.15-17: "the multi-model average or median can be expected to outperform individual ensemble members": this is also shown for low flow and drought, see Gudmundsson et al. 2012 and Stahl et al. 2011. You might want to include these (or comparable) references.

p.10732, l.6-9: You might want to consider including a formula to express this statement more clearly.

p.10733, l.15-17: Where do the data come from? From the EWA database?

p.10733, l.21: Is the minimum contributing upstream area of 1000 km2 the result of your selection on p.10730?

p.10734: Be very careful with the use of the r2 as statistical measure in validation as a high offset or a negative correlation also gives a high r2, so either disregard r2 in validation or give it much less attention than EF, which is a much better measure for this purpose.

p.10735, l.6-7: "7 day average minimum flows": is that all annual 7-day minimum flows averaged per station?

p.10735, l.24-28: Could this underestimation also be related to the omission of reservoirs in the simulation?

p.10736, l.18-20: Or the simulations are more peaky than the observations because the model response to precipitation is too fast. This is shown for many large-scale hydrological models in many studies. Please consider showing some example time series of simulations and observations and their thresholds and deficit volumes.

p.10736, l.28: "capturing extreme streamflow droughts" > "capturing the statistics of extreme streamflow droughts"

p.10737, l.6-7: "(2) an incorrect parameterization of the groundwater storage due to bias in the observed winter precipitation": how is the groundwater storage in the frost season influenced by the winter precipitation which falls as snow and does not result in recharge to the groundwater system?

p.10737, l.16-22: Maybe you should mention here already that you recalculate the frost and nonfrost season for the future.

p.10738, l.1-2: "at the location itself": what do you mean with location? Grid cell? Or gauging station (fig.1)? If grid cell, then what is the contributing area? If gauging station. then only include the station locations in Fig.4 (like in Fig.7).

section 3.3: the headings of the subchapters are unequal. You could add "in the nonfrost season" to the heading of subsection 3.3.2, 3.3.3 and 3.3.4.

p.10740, l.5: “Langness, Isohaara and Dau Gavpil, Neuhausen” > “Langness, Isohaara, Dau Gavpil, and Neuhausen”.

p.10740, l.13-15: Vague sentence, please rephrase.

p.10741, l.17-19: The changes in actual ET can be quantified from the model output.

p.10745, l.14-17: Vague sentence, please rephrase.

p.10746: The uncertainty in climate change scenario and water use scenario does increase significantly over time. Please mention this.

p.10750, l.2: “increased competition for water”: this is not a result of this research. Be very careful with these broad conclusions.

p.10752, l.1: “aus der Beek” > “Aus der Beek”

Table 1: Move footnote “b” to next line.

Fig. 5 & 6: The text in the figures is not legible. Please increase figure, increase font size or find another way to provide the information.

Fig. 9: “20yr minimum flow” > “20yr return level minimum flow”

Fig. 11: “Welch’s test” > “Welch’s t test”

References:

Feyen, L., and Dankers, R.: Impact of global warming on streamflow drought in Europe, Journal of Geophysical Research, 114, D17 116, doi: 10.1029/2008jd011438, 2009.

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., and Koirala, S.: Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe, Journal of Hydrometeorology, 13, 604–620, doi: 10.1175/JHM-D-11-083.1, 2012.

Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voss, F., and Wiltshire, A. J.: Climate change impact on available water resources obtained using multiple global climate and hydrology models, Earth Syst. Dynam., 4, 129-144, doi:10.5194/esd-4-129-2013, 2013.

Stahl, K., Tallaksen, L. M., Gudmundsson, L., and Christensen, J. H.: Streamflow data from small basins: a challenging test to high resolution regional climate modeling, Journal of Hydrometeorology, 12, 900–912, doi: 10.1175/2011JHM1356.1, 2011.