Response to „Review of Staudinger et al. – Quantifying sensitivity to droughts – an experimental modeling approach – by A.F. Van Loon"

In this paper the authors simulated the effects of two precipitation scenarios on streamflow, soil moisture, groundwater and drought indices for a number of catchments in Switzerland and analysed the sensitivity of the catchments to drought. They also investigated the main controls on this sensitivity and found that size, elevation and slope were important.

General comments:

My view is that this paper discusses an interesting topic, namely catchment sensitivity to drought and how catchment properties alter the drought signal. However, the research presented in this paper has some serious issues, both related to the methodology and results and the way these are presented. Here I will list the most important ones.

We thank the reviewer for her efforts reviewing our manuscript. We agree on many of the comments and we are confident that addressing those will improve the manuscript. However, we also have different views on some of the points as discussed below.

- The scenarios of 35 year of progressive drying are not realistic. The authors mention that in their discussion but do not make clear why these scenarios are still useful.

We did not mean the drying scenarios to be realistic in the sense of expected climate change, but we argue that the progressive drying scenarios are nevertheless useful as they show the sensitivity of catchments to extreme drying conditions, in particular in relation to the initial conditions (a dry year follows another dry year). With the scenarios that we chose it is possible to include weather conditions that actually occurred in the studied catchments and combine them with drier than ever initial conditions, which are, however, still based on actually observed precipitation data.

- The authors did 100 model calibrations and averaged the model results. They do this to avoid choosing the "best" parameter set, but introduce other issues. For example when the timing of peaks is different in the model runs, peaks are smoothed in the ensemble mean. Since this is a study on extremes (drought) this might strongly influence the results. I would suggest using the full ensemble of model results predicting the range of possible scenario effects, instead of taking the ensemble mean or choosing the "best" parameter set.

For ensembles that come from different forcing data we would agree. In this study, however, the ensemble comes from different parameterizations using the same forcing. Choosing the ensemble mean of the simulations derived from different parameterizations has less of a smoothing effect as averaging over different driving data series would have. The advantage of using an ensemble of 100 parameter sets is that this gives a more robust picture.

- The entire observation period is used for calibration, no validation of the model results is performed. This is not unusual, but for extrapolation outside the calibrated range (as is the case in this research with the two dry scenarios) validation of the model is needed to estimate its robustness in predicting values outside the calibrated range.

While we certainly agree on the need for validation outside calibration conditions when a model is used for prediction purposes (see also Seibert, 2003), we argue that in this study it is preferable to use as much data as possible to best capture the catchment functioning in the calibrated parameter values of the entire period ranging from dry to wet conditions.

- Many of the results are not surprising and not new. For example the "SoYe scenario did not always result in lower streamflow values compared to the long-term mean, but had rather seasons with pronounced lower flows" and the "timing of the pronounced lower flows appeared to occur simultaneously" (p.7669). That a dry year has lower summer values than the long-term mean is to be expected. And the driest year from the reference run and the driest year from scenario SoYe are actually the same year, only with different initial conditions. It is logical that the low flows occur at the same time in the year and that differences between the scenario and the reference run are small because in Swiss catchments the multi-year memory of the hydrological system is low.

It is true that this result is not surprising, but it shows the general feasibility of the simulations. In a revised version we will shorten this part.

- On the other hand, the results that are surprising were not further investigated. The conclusion that small, high and steep catchments are less sensitive to drought than large, low-lying and flat catchments is counterintuitive, but it is not based on a thorough analysis. I found serious lacks in the statistical analysis of the results: i) only linear relationships have been tested while from the figure non-linear relationships are apparent, ii) the relationship between the tested variables is not investigated (how much is the added effect of the different variables size, elevation and slope, while high catchments are probably also small and steep?), iii) hydrogeology and land use are only investigated with one variable each although there are many possible factors related to geology and land use that might influence the sensitivity of catchments to drought. Some items really need further investigation, for example the negative effect of size and the influence of storage in snow. In the discussion the authors mention that size was also important in the study by Kroll et al. (2004), but it was not discussed whether size was positively or negatively related to low flows. In Kroll et al. (2004) the inclusion of hydrogeological indices was even more important emphasising the importance of hydrogeology in low streamflow prediction. Which parameters can represent the effect of hydrogeology on drought and low flow is still an important question in drought research, so the authors should focus on this analysis to present some novel results.

Regarding the concerns on the statistical analysis of the results raised by the reviewer we would like to emphasize the following:

- ***i) actually we did not use only linear relationships but we also used the spearman rank correlation,***
- ***ii) this is a valid point, which has to be addressed in the revised version,***

iii) we related several primary measures to investigate possible relationships, but agree that measures for hydrogeology and land use are only investigated with one variable each, which we will extend in the revised manuscript.

Also we will further investigate the negative effect of size and the influence of storage in snow.

- The example of 2003 is not representative for other drought events as this event had a very wet winter and spring as initial conditions for a short intense summer drought (see Stoelzle et al. (2014). Therefore, more than one drought event should be investigated in this type of research.

The purpose of this study was not to investigate several drought types, but a more general sensitivity. By choosing the drought 2003 we wanted to picture the effect of drier initial conditions. As mentioned by the reviewer the summer of 2003 was preceded by a wet winter and spring, which even more underlines the different outcome if it would have been a dry preceding winter and spring.

- The paper could benefit from some rephrasing. Sometimes the author's reasoning is hard to follow and at some points I thought the opposite was meant (see some examples below).

Due to these issues I would advise to reject this paper, with possibility for resubmission after major reanalysis and rewriting.

Specific comments:

- P.7661, l.19-20: "additionally different occurrences" > what do you mean?

We mean that they have additionally different occurrences in time and space, which we will clarify in a revised version

- P.7662, l.4: Thanks for citing me, but the correct reference is Van Loon & Van Lanen (2012), full reference: Van Loon, A. F., and H. A. J. Van Lanen (2012), A process-based typology of hydrological drought, Hydrol. Earth Syst. Sci., 16(7), 1915–1946, doi:10.5194/hess-16-1915-2012.

We will correct this in a revised version.

- P.7665, l.1: Since this study focusses on low flow and drought, why not use the Nash-Sutcliffe efficiency based on the logarithm of the discharge?

This is, because we think that the water balance is an important point to be addressed for the general reliability of a hydrological model. We are aware that there are objective functions targeting low flow specifically, but here the choice of a combination of Nash-Sutcliffe and volume error seemed more robust.

- P.7667, l.14-15: Why would you use the days exceeding Q90 if your study is on low flows?

This is done for comparison with the other measures. The information is only the inverse, which is not hard to transform mentally, the resulting indices however are easier to compare.

- P.7667, l.15: “respective”?

“Respective” because each of the ensemble member (different parameterization) has its own reference simulation.

- P.7667, l.17: Was the Q90 calculated from the reference simulation and then fixed for the scenario runs? If Q90 was recalculated the calculation of change is not correct.

The Q90 was reference simulation and kept for the scenarios.

- P.7668, l.14-16: Shouldn’t this be opposite? I see the yellow lines closer to 1, so lower ΔQ in wet years.

What is meant is that the higher elevation catchments (yellow lines) compared to the lower elevation catchments (green lines) show a larger change for the wet years at the beginning of the drying scenario. However, after they crossed the 1-line (long term mean) they show a smaller change than the low elevation catchments. We will clarify the wording in the revised version.

- P.7668, l.19-20: This is only true for scenario SoMo. In SoYe the variability in ΔGW is similar to that in ΔQ.

Correct, but this paragraph talks only about scenario SoMo. We will clarify this fact in the revised version.

- P.7668, l.22: “driest year of reference simulation” > driest in terms of P or Q? And which year is it? Is it the same year in the four example catchments?

It is not the same year for each catchment, which is why it was not mentioned specifically.

- P.7668, l.22-23: Comparison with the long-term mean is not relevant as the driest year from the reference situation is also below the long-term mean.

The scenarios were compared with the long term mean of each DOY (day of year). Depending on the season also the driest year from the reference contains days with higher streamflow than the long term mean.

- P.7668, l.24-26: Not surprising.

As we replied already above, it is not surprising but still worth mentioning.

- P.7669, l.6: Not surprising.

As we replied already above, it is not surprising but still worth mentioning.

- P.7669, l.6-8: I don't think they are so different, probably relative differences are minor. This does not prove differences in sensitivity to drying.

We agree that the differences are small; however, we would not call them minor, at least not for all catchments. It is interesting that the initial conditions can have noticeable impacts even when looking at a whole year. The differences due to initial conditions varied between about 50 and 80%, please note that this is in the same order of magnitude as what might be expected due to climate change (ignoring changing initial conditions). We will add discussion on this in the revised version.

- P.7669, l.9-11: Not surprising.

As we replied already above, it is not surprising but still worth mentioning.

- P.7669, l.10: Why preceding summer? The initial conditions in this scenario are determined by the preceding years.

Yes, but in the humid climate of Switzerland each real year has anyhow the cycle of snow in winter and some rain in summer. So ultimately the last period before the start of the simulations (summer, as of the hydrological year that was used) determines the difference in the initial conditions of the SoMo scenario.

- P.7669, l.14: "diminished"?

Bad formulation indeed, we will rephrase in the revised version.

- P.7669, l.17-18: Again the relative differences are not so large.

See above

- P.7669, l.24: "constant days"?

We will rephrase and use "fixed" instead.

- P.7669, l.24 & p.7670, l.1: Lower elevation catchments are more vulnerable to drought? This requires more investigation.

We agree that while our results in this study indicate that lower catchments are more sensitive to droughts, further investigations are required. We intend to assess the role of snow on runoff vulnerability of lower/higher elevation catchments in an additional sensitivity study. Work on this is ongoing and we will present results in the revised version.

- P.7670, l.15-17: Is that so? Or is it just hard to see because of the same scale of the y-axis?

This is based on the comparison of the storages for each catchment individually before comparing this relation with the others. So, yes that is so.

- P.7670, l.19-21: Rephrase!

We will rephrase this in the revised version.

- P.7670, l.27-28: “IQR” > “Irel”?

Thank you, we will change that.

- P.7671, l.1: Mention prod.no and %forest.

Thank you, we will add that.

- P.7671, l.23-27: What do you mean to say here?

We mean that the weather difference between the different years of a catchment should not overprint the catchment properties. We will rephrase in the revised version.

- P.7672, l.25-27: ?

It is meant that the higher catchments have some snow component to be considered in late spring and early fall, which influence the streamflow in late summer. We will rephrase for clarity.

- P.7673, l.20: Here you mention snow as a possible, but the effect of snow should be investigated in detail to make this study valuable. And what are the “greater storages”?

We agree that while our results in this study indicate the effect of snow, more detailed investigations are required. We plan to investigate the effect of snow in a sensitivity study, by systematically modifying the snow influence in each catchment.

Greater storages == potential subsurface storage features

- Figure 2: Shouldn’t the legend be the other way around (black = SoMo and grey = SoYe)? SoMo was more extreme, wasn’t it?

That is correct, thank you.

- Figure 4: Put the grey line in front of the other lines or make the lines semi-transparent. Otherwise it is not visible when the blue line is higher than the grey line. Furthermore, the x-axis is not very clear. Indicate the years that were plotted.

We will consider the transparency option. The x-axis is indicating DOY day of year, and hence to us the axis seems very clear.

- Figure 6 and 8: The correlation between some variables is significant, but the relationship is not always clear and linear. Refer to the txt for explanation of prod.no.

Regarding the linearity see response above (rank correlation). We will refer to the explanation of the prod.no., thank you.

Reference:

Stoelzle, M., Stahl, K., Morhard, A., & Weiler, M. (2014). Streamflow sensitivity to drought scenarios in catchments with different geology. Geophysical Research Letters.

Seibert, J., 2003. Reliability of model predictions outside calibration conditions, Nordic Hydrology, 34: 477-492.