Review of BEAUFORT et al. "Extrapolating regional probability of drying of headwater streams using discrete observations and gauging networks"

By Anne Van Loon

General comments:

I would like to congratulate the authors with this interesting paper. In the paper, they use a number of databases (official networks and citizen science data) to predict regional drying of headwaters in France, which gives interesting information on spatial and temporal variability of drying. The data, approach and results are robust. I do have a few fundamental and technical questions (see below), but I hope these can be solved easily by the authors.

Firstly, the authors need to explain why a regional assessment of headwater drying is needed. What is the benefit of Figure 11 over Figure 5? The patterns of drying are the same, so Figure 5 would be sufficient to indicate hotspots of drying within France and temporal variability in drying. In the discussion, the authors point out that for accurate IRES management estimation of "drying at the reach scale is needed" (p.18 l.427) and in the conclusion they mention that the approach does not allow for characterisation of drying in "nearby streams within the regions" (p.21 l.495). So if local scale information is so important and this method cannot be used to extrapolate between streams in one region, then why do we need the regional scale? Why go to coarser resolutions if you have detailed observation data at least for some rivers? In this way you lose spatial information without gaining anything in return.

Secondly, the paper is focused on France. This in itself is not a problem, since the methodology and results are interesting and useful beyond France, but the author fail to put their findings in a broader perspective in the discussion. Literature on IRES research from outside France should be discussed and the authors should clarify what is new and interesting about this work from an international perspective. On p.19 l.443-452, the authors mention how their results are consistent with previous studies, which is great, but they should additionally point out what their study adds. If this is not done, the study would be better placed in a Journal like Journal of Hydrology – Regional Studies.

Thirdly, I would like the authors to help the reader more in understanding the methodology. Figure 3 is helpful, but in the manuscript it is not always clear which data was used for what. Especially when explaining the equations on page 9 and 10, the authors could be clearer on which dataset was used, which time period. Also in the Results section it should be clarified when they are referring to calibration results, validation with POD data, or validation with the year 2017. For example, the first paragraph of Section 3.2.3 is quite confusing, because it discusses the performance of the models in the calibration period, which was already discussed in Section 3.2.1. Table 2 should be explained better; how is it different or similar to the information presented in Figures 7&8? Also, in the first paragraph of Section 3.3.2 the authors state that "the simulated RPoD fit well to RPoDONDE" (l.349), but wasn't that already discussed in Section 3.2.1 (Figure 7&8)?

Fourthly, it is unclear whether natural and/or human-influenced sites are selected in this study. In Section 2.4, the authors mention that the "observed discharges were not or only slightly altered by human actions" (p.7 l.164), but they do not specify whether the other datasets, i.e. groundwater levels, ONDE and POD observations, are near-natural too. This is important, as the authors mention in the discussion, “the basins are subject of intense agriculture with important water withdrawals during summer. Abstractions greatly reduce the water availability in rivers and in aquifers which are no longer able to support the low water levels and lead to increased flow intermittence. The responses of biological communities to artificial flow intermittence is still poorly understood compared to natural IRES.” (p.19 l.435-439) If near-natural and human-influenced data are mixed in the predictions, it will be very difficult to understand the reasons for the regional patterns in drying and the statements about the highest drying occurring in sedimentary plains due to the low elevation gradient and dependence on rainfall might be flawed.

And finally, it is unclear why two statistical models are used throughout the paper. If they are equally suitable from a theoretical perspective, two (or more) models could be used for testing, but then the best model should be used to simulate the final results.

Specific comments:

The regional probability of drying needs to be explained. In Section 2.6 the authors only mention that RPoD is calculated, but they never explain how this variable is calculated exactly.

The weighted average of the non-exceedance frequencies (F) needs to be explained better. According to the Discussion section discharge and groundwater levels are combined (l.411-412), but this is not explained clearly enough in the Methods section (l.202-203). How are these non-exceedance frequencies of groundwater and discharge averaged since they have such different shapes and ranges (see Figure 3). And what do the authors mean with “with respect to the relative proportions of gauging stations and piezometers” (l.203-204)?

The authors conclude that “both models seem able to predict RPoD out of the calibration period” (l. 330-331), but do a NSE of 0.4 and 0.5 warrant such a statement?

A significant part of the Conclusion section discusses future work. Is that relevant for this manuscript? I would suggest leaving those paragraphs out as they distract from the main message of this paper.

Textual comments:

- l.20-21: make the time periods consistent. The abstract mentions 2011-2017 and 1989-2017, whereas the introduction mentions 2012-2016 and 1989-2016 (l.97-98).

- l.155: RPoDpoc not explained yet. The concept of RPoD needs to be explained first, before you can introduce different versions.

- l.194: change “we merged 20 of the 280 regions with a neighboring one located in the same HER1” to “we merged the NER2-HR combination with a neighboring one located in the same HER1. This was done for 20 of the 280 regions.”

- l.307: Fig. 10 > Fig. 9

- l.317: “both models use with the 1989-2017” > what was meant here?

- l.322: change “whatever the dataset” to “independent of dataset”

- l.372-373: this is an unclear sentence. What “information on headwaters” was missing? And why were they unable to predict?

- l.375: what do you mean with “better capture spatial distribution of IRES located at the upstream the hydrographic network”?

- l.460: reformulate “a crushing of extreme values”

- Fig. 8: Nash > NSE

- Fig. 12: explain dates in figure

- Please check the English of the manuscript, for example the consistency between verb and subject, e.g. l.248 “an increase”, l.250 “reaches” & “occurs”, l.369 “observed”, l.465 “requires”