Reply to Anne Van Loon

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.

The authors would like to thank Anne Van Loon for her positive comments on the manuscript. Please find below the detailed answers to the general and specific comments.

****

General comments

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.

We modified lines 87-93 to better explain that one of our objective is a temporal extrapolation of the daily drying probability (eg Fig 11) in regions, based on discrete observations (~5/years, raw data in Fig 5).

Figure 11 shows the number of consecutive days with simulated RPoD > 20% to characterize the severity in both time and space of drying. Figure 11 is derived from the reconstructed continuous time series. Figure 5 results from ONDE observations, i.e. statistics are based on five inspections per year between May and September.

As discrete data, ONDE observations cannot provide any information on the persistence of dry conditions between two consecutive dates of observation. In summer, rewetting is possible after convective rainfall episodes or inversely short-lived events of drying may occur between two dates with “Flowing” state. The rewetting-drying events may have significant impacts on communities whose survival is conditioned by the duration/frequency of drying. The duration of drying is of importance for ecologists, as one key driver of the composition and persistence of aquatic species (e.g. Kelso and Entrekin, 2018). During recent decades, hydrologists and ecologists have been working on developing metrics to quantify alterations of the river flow regime and their consequences for the ecosystems (Poff and Ward, 1990; Richter et al., 1996; Snelder et al., 2009; De Girolamo et al., 2017). Most of the metrics (e.g. Poff et al., 1997; Olden and Poff, 2003; D’Ambrosio et al., 2017) are determined on the basis of continuous time series of daily discharge.

In that sense, the objective of this study is to provide information through probability of drying to a daily time step using discontinuous data from the ONDE network.

****

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.

This work is the first step towards a more ambitious project: modelling the dynamic of daily flow-states at the ONDE sites, *i.e.* at the reach scale. The idea was to start by the coarsest spatial scale before developing tools adapted to the local scale. Thus, the effort was mainly put on the temporal aspects but at the regional scale. It has consisted in identifying the robust and significant indices that are related to flow intermittence at the regional scale. The indices allow extrapolating information from the discrete observation data and will be introduced in the set of potential explanatory variables or considered as proxies of the probability of dry conditions in the modelling framework for each ONDE site. Other local potential explanatory drivers such slope, riparian vegetation, presence of pools, rainfall, evapotranspiration, permeability, water abstraction, etc. that are scarcely available at the country scale, will be collected in the next steps.

****

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.

The authors agree with this remark. We modified the text (lines 465-481 and lines 544-551) to better explain the international relevance of our results.

To our knowledge, no study has proposed to reconstruct daily flow states time series of headwater streams at the country scale as France (> 500 000 km²) using discrete observations in time and space. In the literature, studies at national scale remain focused on the detection and the mapping of IRES because these rivers are historically poorly investigated and their proportion in existing hydrographic networks remains inaccurate or misunderstood (Nadeau and Rains, 2007; Snelder et al., 2013). Recently, several studies proposed alternative methodologies in order to estimate metrics in ungauged IRES (Gallart et al., 2016) or to predict daily streamflow in river basin experiencing flow intermittence (De Girolamo et al., 2017) but remain applicable at local scale.

From a methodological point of view, our method relating discrete drying observation to continuous daily gauging data seems robust across the highly diverse (climate and topography) regions of France, and provides good predictions in an independent region excluded from the calibration process (PoC). These two results suggest a potential application of our approach in other countries.

Citizen science has proved to create opportunities to overcome the lack of hydrological data and lead to densify the flow state observation network (Turner and Richter, 2011; Buytaert et al., 2014). Note that the paper demonstrates the value of the ONDE network and thus promotes such a kind of network whose creation is less expensive than the installation of additional gauging stations, to survey flow intermittence.

****

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)?

We modified the section 2.6 to better explain our methodology.

Datasets considered as inputs in the equations on pages 9 and 10 are successively the dataset 2011-2017 then the dataset 1989-2017. The non-exceedance frequency of discharge and groundwater levels F is computed at a daily time step leading to extrapolate daily values of RPoD. Parameters for each HER2-HR combination *a*1; *b*1 and *F*0 for model LLR and *a*2; *b*2 for the LR model are successively determined by regression (Figure 3) using the calibration data between 2012 and 2016 of the two datasets (calibration period when ONDE observations are available over the whole year). The number of piezometers and gauging stations selected in each HER2-HR combination is different according to the datasets used as inputs (see section 2.4 and 2.5), leading to different values of F over the common period 2011-2017.

Table 2 shows the inter-annual NSE of both models with the two datasets as inputs while Figures 7 and 8 show the average NSE over the entire calibration period between 2012 and 2016. The values of NSE, during the year 2017, concern the validation period. The current Table 2 may be confusing and the calibration and validation NSEs have been specified in the revised paper:

		2011-2017 dataset						1989-2017 dataset					
		Calibration					**Valid.**	Calibration					**Valid.**
		2012	2013	2014	2015	2016	**2017**	2012	2013	2014	2015	2016	**2017**
LLR model	May	0.2	0.0	0.5	0.5	0.6	**0.4**	0.2	0.0	0.3	0.0	0.7	**0.2**
	June	0.6	0.3	0.8	0.5	0.8	**0.5**	0.6	0.3	0.5	0.3	0.8	**0.5**
	July	0.7	0.5	0.6	0.6	0.8	**0.7**	0.7	0.5	0.5	0.4	0.8	**0.6**
	August	0.8	0.6	0.7	0.7	0.8	**0.6**	0.7	0.5	0.5	0.5	0.8	**0.6**
	Sept.	0.7	0.8	0.6	0.6	0.7	**0.6**	0.6	0.7	0.5	0.5	0.6	**0.6**
	May - Sept	0.8	0.8	0.7	0.7	0.8	**0.7**	0.8	0.7	0.5	0.6	0.8	**0.7**
LR model	May	0.2	0.0	0.5	0.1	0.6	**0.3**	0.3	0.0	0.3	0.0	0.7	**0.2**
	June	0.6	0.5	0.8	0.5	0.8	**0.4**	0.6	0.4	0.5	0.3	0.7	**0.4**
	July	0.7	0.6	0.5	0.6	0.8	**0.6**	0.7	0.4	0.5	0.4	0.8	**0.6**
	August	0.7	0.6	0.7	0.6	0.7	**0.6**	0.6	0.4	0.5	0.4	0.7	**0.5**
	Sept.	0.6	0.8	0.6	0.7	0.7	**0.6**	0.5	0.6	0.4	0.5	0.6	**0.6**
	May - Sept	0.8	0.8	0.7	0.7	0.8	**0.7**	0.8	0.7	0.5	0.6	0.8	**0.7**

The first sentence of Section 3.2.3 is redundant with Section 3.2.1. We modified this paragraph to focus more on the annual performance of each model in the revised version.

The first paragraph of Section 3.3.2 briefly presents the model performance by graphically comparing simulated RPoDs with $RPoD_{ONDE}$. This paragraph only confirms the conclusions given above and have been shortened.

****

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.

We do agree that mixing natural and human-influenced stations bias the conclusions of the analysis and we modified the discussion (lines 487 to lines 499).

Here, the selection of the gauging stations inherits from previous studies and from the long expertise of the time series available in the HYDRO database. We have excluded stations with heavily modified river flow regime. As this selection is the result of expertise, we cannot be sure that there is absolutely no human action that may impact low flows.

The HYDRO database managers (section 2.4) consider as strongly influenced, gauging stations located on rivers regulated by dams, reservoirs or important water abstractions precisely localized, or on channelized rivers (e.g. diversion channel). As for the HYDRO gauging stations, ONDE sites are located on headwater streams without major human influence.

Regarding alteration issues in our datasets, we do not have access to the exact location and the volumes of water withdrawal for irrigation purposes. However, due to their upstream location, water availability is expected to be low, which may limit potential withdrawals and as consequence flow alteration at ONDE sites. Piezometers have been identified as involved in groundwater/surface water exchanges (section 2.5) and they experience seasonal fluctuations similar to the headwater streams monitored by the HYDRO database. The level of alteration of groundwater levels by water withdrawal is unknown because no information is available. However, in sedimentary plains where agricultural crops dominate the landscape, we are not sure that no human action affects low flows. Hopefully all the basins are not strongly affected by abstraction.

****

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.

Both models are equally suitable from a theoretical perspective and they demonstrate similar performance over the period 2012-2016. However, out of the calibration period (*i.e.* 1989-2011), both models are facing unexperienced climate conditions. As detailed in the last paragraph of the Discussion, the tails of the logarithmic curve and of the logistic curve are different and induce distinct values when the average of the non-exceedance frequencies *F* is close to 0%. As an illustration of the divergence of the models, maps of $RPoD_{LR}$ and $RPoD_{LLR}$ are displayed with *F* fixed to 1% in Figure 14. Predictions from the LLR model are thus larger than those from the LR model during generalized drought. We are not able to identify which model provides the more realistic values out of the conditions experienced over the calibration. Hence, we consider that presenting the results of these two models is of interest and keeping the two models is a way to put into perspective the estimated values - in particular those around the years 1989 to 1991 in response to extremely dry conditions.

****

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.

We added the definition of RPoD in section 2.2. Observed values of RPoD ($RPoD_{ONDE}$) is calculated as follows:

$$RPoD_{ONDE}(d) = \frac{(Ndrying)_{\mathrm{HER2-HR}}}{(Nflowing + Ndrying)_{\mathrm{HER2-HR}}}$$

where *d* denotes the observation date of the ONDE network, Ndrying and Nflowing are the number of drying and of flowing statuses observed at ONDE sites located in a same in a HER2-HR combination at the observation date *d*, respectively.

****

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)?

We provided the details for computing F in the section 2.6.

F is computed for each HER2-HR combination:

Let us consider a day (d) and the gauging stations and piezometers available in the HER2-HR combination.

The non-exceedance frequency of the discharge observed at the day *d*, Fq, is determined for each gauging station using the flow duration curve. In the same way, the non-exceedance frequencies of the groundwater levels Fgw observed the same day is determined for each piezometer.

The average of the non-exceedance frequencies (F) is calculated following the next equation:

$$\mathrm{F(d)} = \frac{\sum_{i=1}^{i=Nq} Fq_i}{Nq} \times \frac{Nq}{(Nq+Ngw)} + \frac{\sum_{j=1}^{j=Ngw} Fgw_j}{Ngw} \times \frac{Ngw}{(Nq+Ngw)} = \frac{\sum_{i=1}^{i=Nq} Fq_i + \sum_{j=1}^{j=Ngw} Fgw_j}{(Nq+Ngw)}$$

with Fq_i: the mean non-exceedance frequency of discharge at the gauging station i calculated between d and d-5; Fgw_j: the mean non-exceedance frequency of groundwater levels at the piezometer j calculated between d and d-5; Nq: the number of gauging stations selected in a HER2-HR combination and Ngw: the number of selected piezometers selected in the HER2-HR combination. The non-exceedance frequency combining discharge and groundwater levels characterize a general hydrological state at a HER2-HR scale.

****

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?

This section (section 3.2.3) has been modified and the revised manuscript presents NSEs for the 2017 validation year. Table 2 has been modified and presents these additional results. Figure 10 has been modified and shows the dispersion between predicted RPoD and drying observed at ONDE sites in the scatter plot during the validation year 2017 (Fig. 10a and 10b) in comparison with the year 2012 which obtains the better NSE during calibration period (Fig. 10c and 10d). The NSE obtain in 2017 are 0.72 with the LLR model and 0.68 with the LR model against respectively 0.83 and 0.81 in 2012. The performance is slightly lower in 2017 but remains acceptable with NSEs close to 0.7 and both models seem able to predict RPoD out of the calibration period.

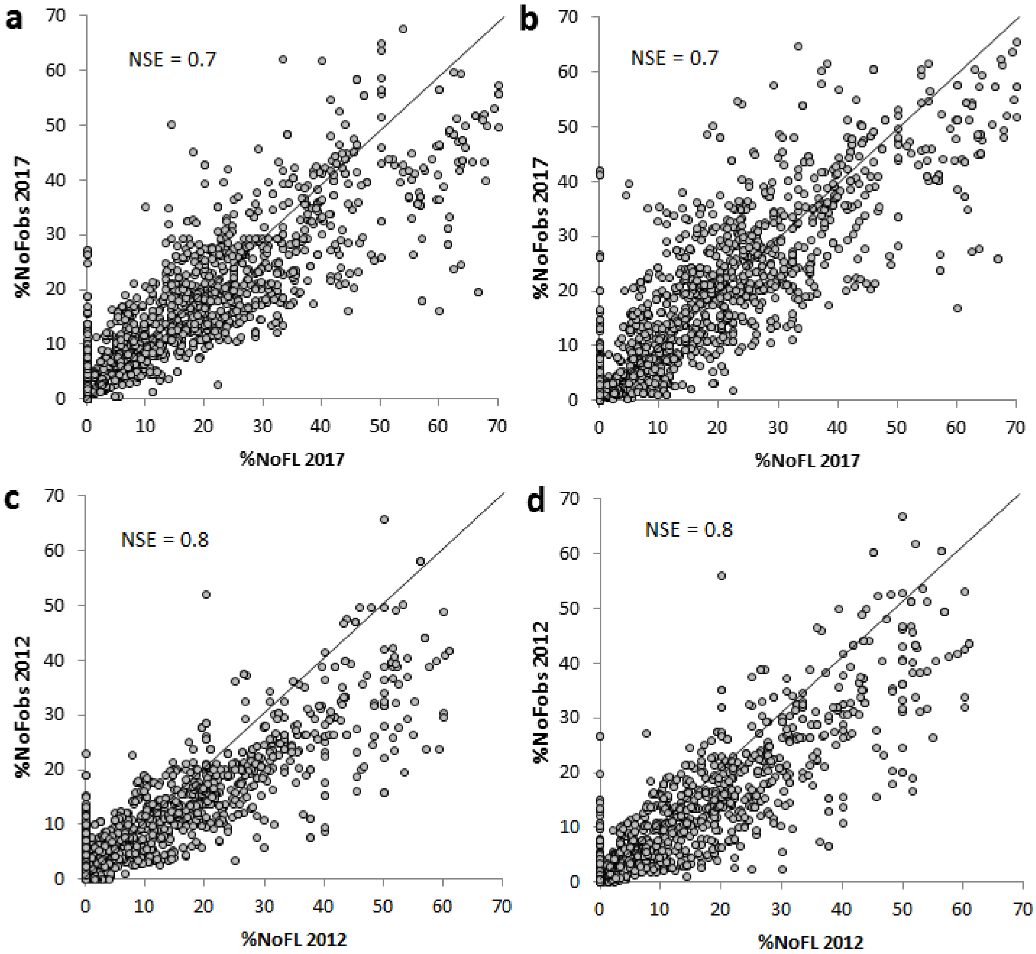

Figure 10. Scatter plot of the predicted RPoD (x axis) and drying observed at ONDE sites (y axis) in 2017 and 2012 simulated with the 2011-2017 dataset by: (a) and (c) the LLR model and (b) and (d) the LR model.

****

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.

The authors wanted to highlight the perspectives to this work and to show the possible ways to predict RPoD at the local scale. The authors have taken this remark into account and shortened this part of the conclusion.

****

Textual comments:

Thank you for your very attentive reading, all your corrections/suggestions have been taken into account. We also took into account the remark about the concept of RPoD which will be better detailed and we will present the equation to compute the values of $RPoD_{ONDE}$ (Eq. 1; Page 6, L140-145). This formula can also be applied to derive the values of $RPoD_{POC}$ (Page 7, L168-170).

References

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., De Bi??vre, B., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D. M., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T., Tilahun, S., Van Hecken, G. and Zhumanova, M.: Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development, Frontiers in Earth Science, 2, doi:10.3389/feart.2014.00026, 2014.

D'Ambrosio, E., De Girolamo, A. M., Barca, E., Ielpo, P. and Rulli, M. C.: Characterising the hydrological regime of an ungauged temporary river system: a case study, Environmental Science and Pollution Research, 24(16), 13950–13966, doi:10.1007/s11356-016-7169-0, 2017.

De Girolamo, A. M., Barca, E., Pappagallo, G. and Lo Porto, A.: Simulating ecologically relevant hydrological indicators in a temporary river system, Agricultural Water Management, 180, 194–204, doi:10.1016/j.agwat.2016.05.034, 2017.

Gallart, F., Llorens, P., Latron, J., Cid, N., Rieradevall, M. and Prat, N.: Validating alternative methodologies to estimate the regime of temporary rivers when flow data are unavailable, Science of The Total Environment, 565, 1001–1010, doi:10.1016/j.scitotenv.2016.05.116, 2016.

Kelso, J. E. and Entrekin, S. A.: Intermittent and perennial macroinvertebrate communities had similar richness but differed in species trait composition depending on flow duration, Hydrobiologia, 807(1), 189–206, doi:10.1007/s10750-017-3393-y, 2018.

Nadeau, T.-L. and Rains, M. C.: Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy1: Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy, JAWRA Journal of the American Water Resources Association, 43(1), 118–133, doi:10.1111/j.1752-1688.2007.00010.x, 2007.

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Research and Applications, 19(2), 101–121, doi:10.1002/rra.700, 2003.

Poff, N. L. and Ward, J. V.: Physical Habitat Template of Lotie Systems: Recovery in the Context of Historical Pattern of Spatiotemporal Heterogeneity, Environmental Management EMNGDC, 14(5), 1990.

Richter, B. D., Baumgartner, J. V., Powell, J. and Braun, D. P.: A method for assessing hydrologic alteration within ecosystems, Conservation biology, 10(4), 1163–1174, 1996.

Snelder, T. H., Lamouroux, N., Leathwick, J. R., Pella, H., Sauquet, E. and Shankar, U.: Predictive mapping of the natural flow regimes of France, Journal of Hydrology, 373(1-2), 57–67, doi:10.1016/j.jhydrol.2009.04.011, 2009.

Snelder, T. H., Datry, T., Lamouroux, N., Larned, S. T., Sauquet, E., Pella, H. and Catalogne, C.: Regionalization of patterns of flow intermittence from gauging station records, Hydrology and Earth System Sciences, 17(7), 2685–2699, doi:10.5194/hess-17-2685-2013, 2013.

Turner, D. S. and Richter, H. E.: Wet/Dry Mapping: Using Citizen Scientists to Monitor the Extent of Perennial Surface Flow in Dryland Regions, Environmental Management, 47(3), 497–505, doi:10.1007/s00267-010-9607-y, 2011.