Synthesis of my review :

Even if this paper appear to be rather long and sometimes "dense", I really appreciated reviewing this paper. I am very happy to congatulate authors for such an amount of work and very useful information and analyses on the drought history over France, since 140 years. Having an experience on data-rescue and long-term historical reconstructions, I consider that this work could have many applications, both in terms of reseach activities or operational hydrology. This work could also help hydrologists to communicate with water managers, decision-makers or stakeholders, in order to show them exemples of long-term hydrological variability.

The authors would like to thank Referee 2 for his positive comments on the manuscript and the specific and technical comments (in italic below) that will lead to improve the manuscript. The detailed answers to the specific comments are presented below.

I really hope that SCOPE hydro time-series would be available soon ?

SCOPE Climate and SCOPE Hydro will be made available as soon as possible in forthcoming data papers. In the meantime, preliminary packed datasets are available upon request to the authors.

I would rate the scientific significance and quality as Excellent. However, I rate the presentation quality as Fair to Good, because some paragraphs appear to be difficult to understand, even with carefull attention. I would like to invite authors to improve the explanation in a more pedagogical way of § 2.2.2 (Bias correction and Schaake Shuffle) and 3.2.2 (spatial matching procedure, also used for the ensemble case). This could undermine our appreciation of the quality of the paper, even if § 4 and § 5 are very Interesting.

Following our responses to comment from referee #1, we will only present the SCOPE Climate dataset in section 2.2.2 and describe the entire SCOPE method (with more details) in the appendix. Efforts will be made to improve the understanding of section 3.2.2. The issue may however come from the low resolution of Figure 4 which can lead to a confusion of event matching.

It might not be the objective of the authors, but a paper in two parts could be easier to read, with a first part considering the methodology (basicaly from 20CR-SANDHY-SUB datasets to SCOPE climate) and a second part considering hydrological analyses and the discussion (basicaly, SCOPE Hydro and hydrological analyses).

Writing up a two-part paper with this material would be quite difficult. See the responses to comments above on restructuring the climate part.

Major comments :

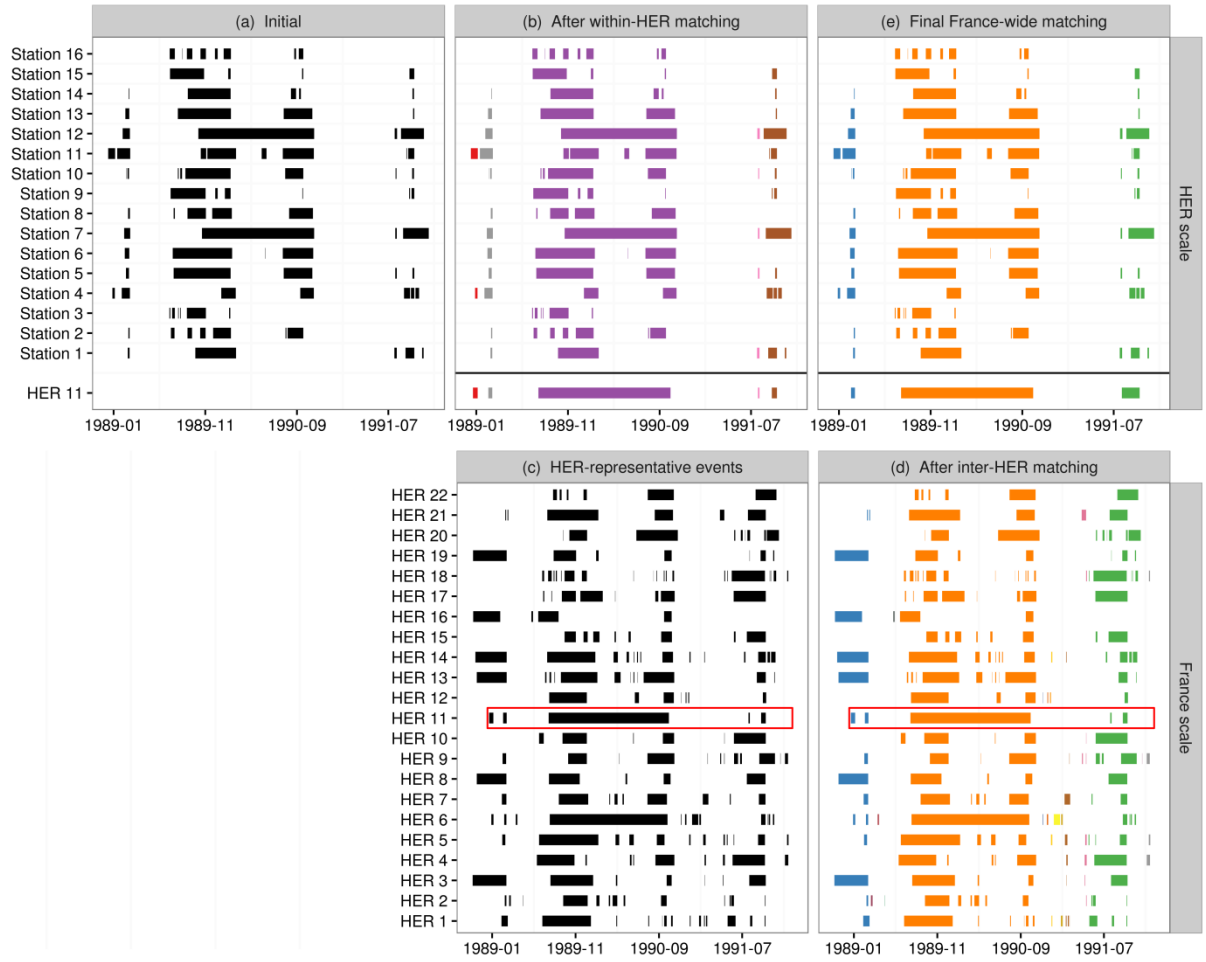
§2.2.2 SCOPE Climate : this paragraph presenting the bias correction via a resampling-based correction approach and improvement of spatial coherence via Schaake Shuffle should be improved in order to be easily understood ;

Indeed, SCOPE will be detailed in the appendix (see previous answer).

§3.2.2 spatial matching procedure : the overlapping process is not clear. This paragraph should be improved in order to be easily understood

** the step from Fig 4a to Fig 4b is not clear on this example : I don't understand why two independent events are considered for red and grey colors, while there is only one event considered with the purple color ? Station 11 event definition should be continous during period covered by red and grey colors ?*

This come from the lack of resolution of Figure 4. The correct figure is the following:



Grey and red events are indeed two different events in 4(a), which was not clear in the version included in the manuscript.

** the step from Fig 4d to Fig 4e is not clear on this example : again, I don't understand why an event could be discontinuous, for the two blue and two green events ?*

4d to 4e is only a local report of the inter-HER matching to the stations of the HER. After the inter-HER matching, the first two events of HER 11 are matched together (two bars with the same blue color). If you go back to 4b, the local events corresponding to these two blue bars are the red and grey events. So after the inter-HER matching, these two independent events are matched together, giving the blue color instead of the grey and red. But this misunderstanding may also come from the same problem than previously (it was not possible to distinguish the two independent event for the grey and red bars).

Fig 7 : again, I don't understand why there is only two spatio-temporal events and not four ?

This figure only illustrates the report of the spatial matching from average events to SCOPE Hydro events. The colors you can observe in 7a are already the result of the spatial matching on average events (spatial matching which is not shown before, as Figure 4 corresponds to Safran Hydro and not average events). So the spatial matching on average events gives two different events: one orange and one green. 7b shows the report of these two events to the SCOPE Hydro events. At the top of the figure, Safran Hydro events are only shown for information as it is used in the discussion (this is the only line which refers to a result shown in Figure 4). Maybe it would be clearer to choose two different colors for average events and events of the 25 members (than orange and green) as it is completely independent of Figure 4.

§3.1 hydrological modeling : since the aim of this study is to represent particularly well drought events and that it is well-know that hydrological models are performing poorly on drought, why authors didn't consider an objective function based on hydrological signitures specific for drought, such as distribution of drought duration, severity, etc (VCN 10, VCN30, ...) ?

SCOPE Hydro has been created in order to be used in any type of hydrological studies. For this reason, the objective function has been chosen to give equal weights to high and low flows.

Minor comments :

p4, l26 : problem with the length of the line ;

This will be corrected with the final version.

p6, l11 : it might be out of the scope of this paper, but have you tryed to analyse the 20CR-SANDHY-SUB bias using a weather type classification (the seasonal classification is interesting but, beyond seasons weather type proportion might change from a season to another) ? ;

We didn't try but indeed, it is a nice suggestion to better understand the origin of biases.

p7, l21 : KGE is expressed as KGE = 1-SQRT (...) ;

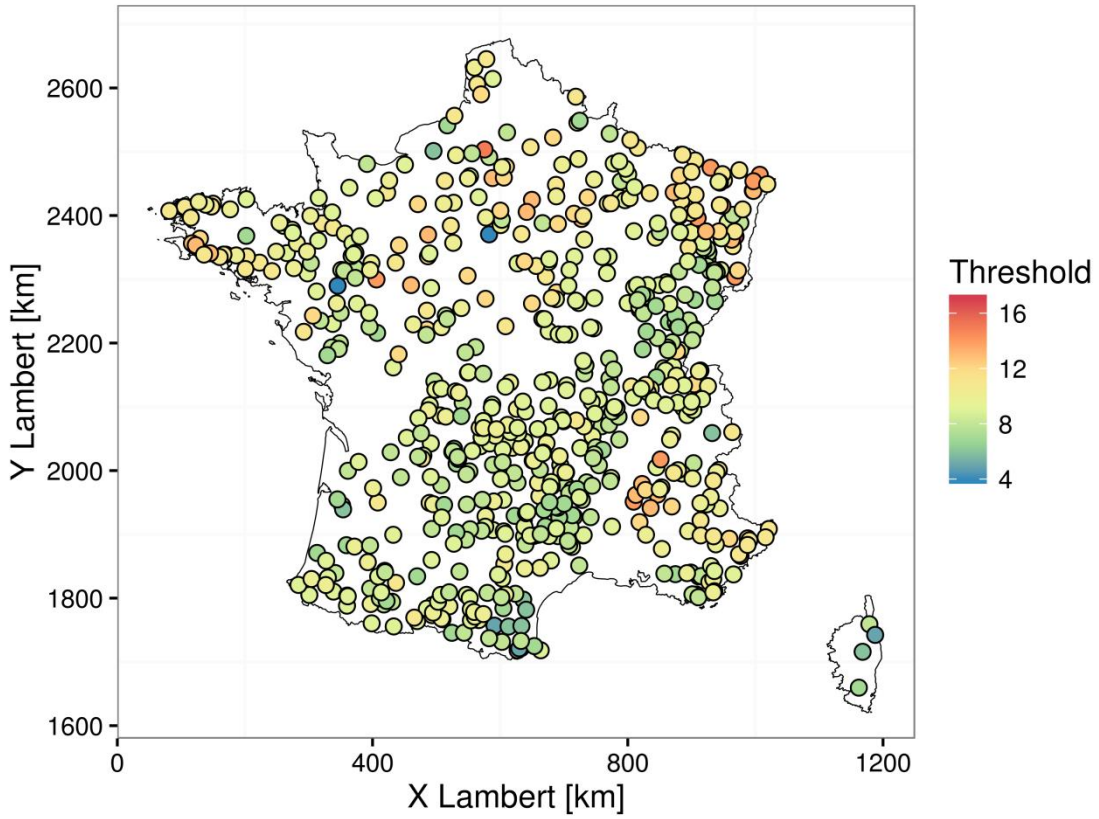
Indeed, this will be corrected.

p7, §3.1 : a table with quantiles of catchments caracteristics and summary of performances (KGE, r, alpha, beta) might be interesting (as Table 2, in Pushpalatha et al., 2012);

It might indeed be interesting and we will include it in the forthcoming paper describing SCOPE Hydro in detail. Please note that KGE values are available in the manuscript in Fig. 2.

p14, l12 : is the number of members to consider an event (10 on Fig 6 example) adapted from one station to another or roughly selected for the 662 stations ? If it's different from one station to another : give some quantile to precise the variability of this threshold ? Have you tested an unique value for the whole station sample ? ;

This number is different from one station to another. Below is the map of the final values:



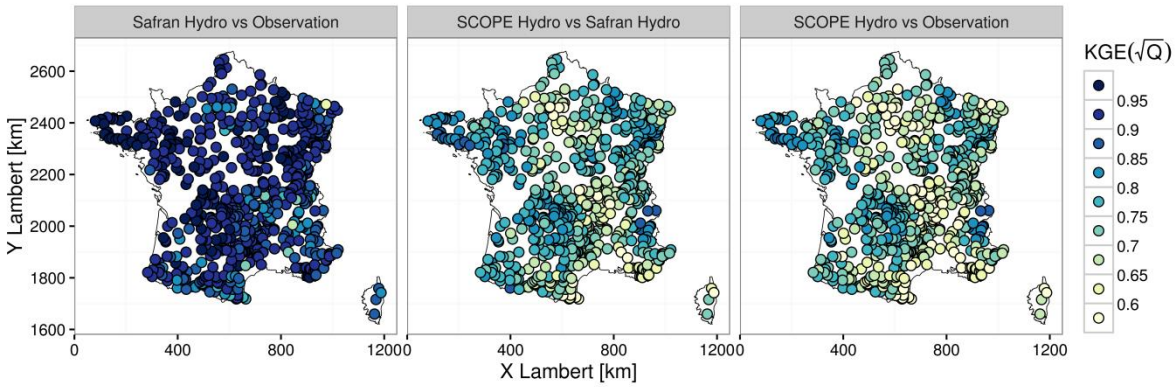
A unique value has not been tested as having a different value for each station allows a re-calibration against Safran to improve low-flow event identification. But except for specific stations, values are kept between 8 and 12.

p17, l4, Fig 2 : I would appreciate to see distributions or boxplots of r, alpha, beta and KGE criteria ;

See our response to a comment above. The manuscript is already very long and dense and we would rather not add more figures not directly related to extreme low-flow events.

p17, §4 : again, it might be out of the scope of this paper, but it could be interesting to caracterise SCOPE hydro performances for drought simulation using hydrological signatures and/or probabilistic criteria, such as CRPSS, etc. ? ;

Cf. Response to comment above, and response to referee #1 for a figure showing the median performance of SCOPE Hydro and Safran Hydro in terms of KGE and KGE decomposition (alpha for variance, beta for bias and r for linear correlation on the calibration period).



p18, figure 10 & p19, figure 11 : for the ones not used to duration values and severity values, it could be interesting to put a panel on these figures with the distributions of event durations and severity obtained with the Observation or Safran Hydro. Another option would be to add a second y-axis with the quantiles corresponding to the duration/severity values ? ;

As the information brought by figures is already very dense, we would be prefer not to add such a panel. However, the idea of translating values in mm /days into long-term quantiles would be an interesting way of presenting the results, and we will keep it in mind for further analyses.

p22, fig 14 : what is the total spatial extent of the 622 hydrological stations ? what is the proportion of gauged surface over the France surface ? ;

The gauges surface corresponds to around 41% of the France surface.

p22 : It would be interesting to distinguish snow-dominated catchments and raindominated catchments and show a figures with the spatial extent of drought, given these two main processes (snow/rain)? ;

This would be a very nice extension of this paper. As this paper is mainly a methodological paper, we do not which to extend the results.

p25 & p26 l14-22 : given the length and density of your paper, Figure 17 and its related §do not appear necessary for me ;

Figure 17 is the only figure providing an ensemble characterisation of a spatio-temporal extreme low-flow event. Figure 13 will be removed as it is partly redundant with figures 10 and 11.

p29, §6.4 : have you compared the Safran Hydro and SCOPE Hydro analyses on the 1958-2012 period, where hydrological simulations are both available ? A scatterplot of duration, severity or spatial extent by year could be interesting ? ;

We actually did this analysis manually for a few exceptional events. Drawing a scatterplot as suggested would require a formal temporal comparison of spatio-temporal extreme low-flow events across different datasets (which is difficult for now -- see section 6.4). To give a concrete example, the x-axis of Fig. 10 and Fig. 11 corresponds to the name of the events, itself corresponding to a spatial center date, independent for each dataset. If we do not link events across datasets, two events occurring during the same period (and that should be linked together) will be identified differently in the x-axis. This would most likely generate more questions than answers and this could hardly be done in an automated way.

p30, §6.6 : considering drought simulation, my experience is that conceptual RR models could be strongly biased. In a future work, you could consider a very simple method, using a bias correction of streamflow simulations by quantile classes, as proposed by F. Bourgin in its PhD at IRSTEA.

As shown in the response to referee #1 and more specifically the figure plotting biases between SCOPE Hydro and Safran Hydro, a slightly negative bias may be detected, but remains largely under 10%. Moreover, we didn’t want to implement any streamflow quantile-quantile bias correction as it would add some more temporal transferability hypotheses in the hydrometeorological modeling chain.