In this paper, the authors set out to reconstruct low-flow events in a reference network of catchments across France from 1871 to present. Due to the lack of observed streamflow records prior to the 1950s, downscaled climatological data (20CR) is used to reconstruct streamflow using hydrological models. Using knowledge of more recent low flow events over the past 50 years, the authors validate their novel approaches of low-flow event identification, spatial matching, and hydrological modelling, which succeed in identifying the well-known events of 1976 and 1989-1990. The approach highlights two additional extreme low-flow events in the 19th century: 1878 and 1893 and the authors conclude that many severe, long and widespread drought events occur prior to 1950.

This paper provides a valuable contribution to science in both methodology and results. The methods are many and varied and though multiple subjective decisions are included, they are well thought out and discussed. Whilst the many steps of the spatial matching procedure are quite challenging to decipher, the results are well presented. The combined threshold level is a particularly interesting concept. This is a very long paper, which might daunt the average reader, but the content is valuable and in my opinion the many maps and graphs all convey interesting information. The maps in Figures 15 and 16 are of particular merit, though Figure 15 would benefit from some slight adjustments (see comment below).

The authors would like to thank Referee 1 for his positive comments on the manuscript. We also thank him/her for 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.

If the editor deems the paper too long (which would be understandable), there are several ways in which the main body of this manuscript could be cut down:

1. The introduction could be reduced.

We believe the introduction as it is introduces the reader with the numerous different concepts covered in this paper. A shortened introduction might decrease the understanding of all facets of the topic. Nevertheless, we will check again if it can be shortened.

2. The description of the derivation of the SCOPE climate data from 20CR-SANDHY-SUB could be stated much more succinctly in the main text, and a detailed description provided in the appendix.

Thank you for the suggestion. We will indeed only present the SCOPE Climate dataset in this section and detail the entire SCOPE downscaling method (with more details) in the appendix.

3. Similarly the description of the spatial matching procedure is lengthy, and could be summarised with a longer description in an appendix. Those who wish to reproduce your methods would be willing to read the appendices, whereas others, who are mainly interested in the results would not need to understand that level of detail.

The aim of this paper is to focus on the methodology developed for identifying spatio-temporal events, a methodology that could be transposed in other contexts. In this way, the results are only shown as basic example results of the method. As a consequence, we would like to keep the entire method description in the main text.

4. The day of the year analysis in figure 12 could be cut out, however it does show the spread of the start dates.

As this figure is the only example of low-flow seasonality, we would prefer to keep it in the main text.

5. Furthermore, Figure 13 could be removed, as the majority of the information is given in figures 10 and 11. Again however, it does show the linear relationship between severity and duration of events.

Thank you for the suggestion. This figure will indeed be moved to the appendix or to some supplementary material.

Comments that must be addressed:

1. The derivation of the SCOPE climate data is not well explained. Page 6 needs a lot of attention. It appears to be a significant amount of work that is not published elsewhere. This procedure needs to be clearly described, and moved to the appendix. For example, it is not clear what variables from the 20CR are downscaled (500hPa geopotential height?)

We agree that this procedure would require a lot more details to be fully understandable, and we also agree that having a deeper description in the appendix would be the way forward. As mentioned above, only a description of the SCOPE Climate dataset would be kept in the main text. Considering the question above: 20CR output variables are only used as predictors (and not downscaled). As detailed in Caillouet et al. (2016), 20CR large-scale predictors (geopotential height, air temperature, humidity and vertical velocity at different pressure levels, as well as sea surface temperature) are used to derive local-scale predictands (precipitation and temperature) in the SANDHY-SUB statistical downscaling method. SCOPE adds two more steps (which will be detailed in the appendix) to SANDHY-SUB.

2. Please indicate the ideal value of KGE (1 I assume?) on both the description of the metric on page 7 and the map given as Figure 2. Furthermore, please check the equation provided on page 7, as this indicates that a low value should be a good score. Gupta (2009) provide the equation you present as the Euclidian Distance (ED), whilst KGE is 1-ED.

Indeed, the equation is incorrect. This will be corrected. The ideal value of KGE (1) will also be specified.

3. Your description of the spatial matching procedure needs some work to improve its clarity. I suggest it is lengthened and sent to an appendix.

As the wish of the authors is to produce a methodological paper with some example results, we want to keep the spatial matching procedure in the main text. We will try to work on the clarity of the method.

4. Page 25 Figure 17 – add a scale/legend to this figure

Indeed, it has been cut, this will be corrected.

5. Page 31 lines 3 and 4 – You state "Result also highlight that the worst events in terms of severity and duration or spatial extent often belong to the pre-1950 period". This statement is not backed up by your graphics. Figure 14 indicates the majority of the most widespread events were post 1940, Figure 15 indicates that 1990 and 1976 were the longest, and similarly in figure 13 only 1945 is picked out among the most severe in the Corrèze catchment. Furthermore, figures 10 and 11 do indicate some long and severe droughts prior to 1950, but the latter period on these graphs seem to show significantly more events and more serious events than the earlier period. I think what you mean to say is that many severe, long and widespread events occur prior to the 1950s, but I don't think it's true that the majority of the most severe events are in the early part of your reconstructions. Please amend this sentence accordingly.

Indeed, this is a mistake in the sentence writing. This is not pre-1950 but actually post-1940. This will be changed accordingly.

Further recommended minor adjustments and comments:

1. Page 2 lines 2 to 6 – references here may be assumed to be based in French catchments due to the outlining of the lack of available data in France, then using the word consequently to start the next sentence. Rephrase to indicate these are not all French studies.

It will be clarified.

2. Page 3 lines 3 and 4 – "high spatial and temporal scales" and "large-scale atmospheric and oceanic data" high scale means different things to different people. Do you mean high resolution or not?

Indeed, this is a mistake. The right formulation is "small spatial and temporal scales", this will be corrected.

3. Page 3 line 13 – within which country was Dayon's study based?

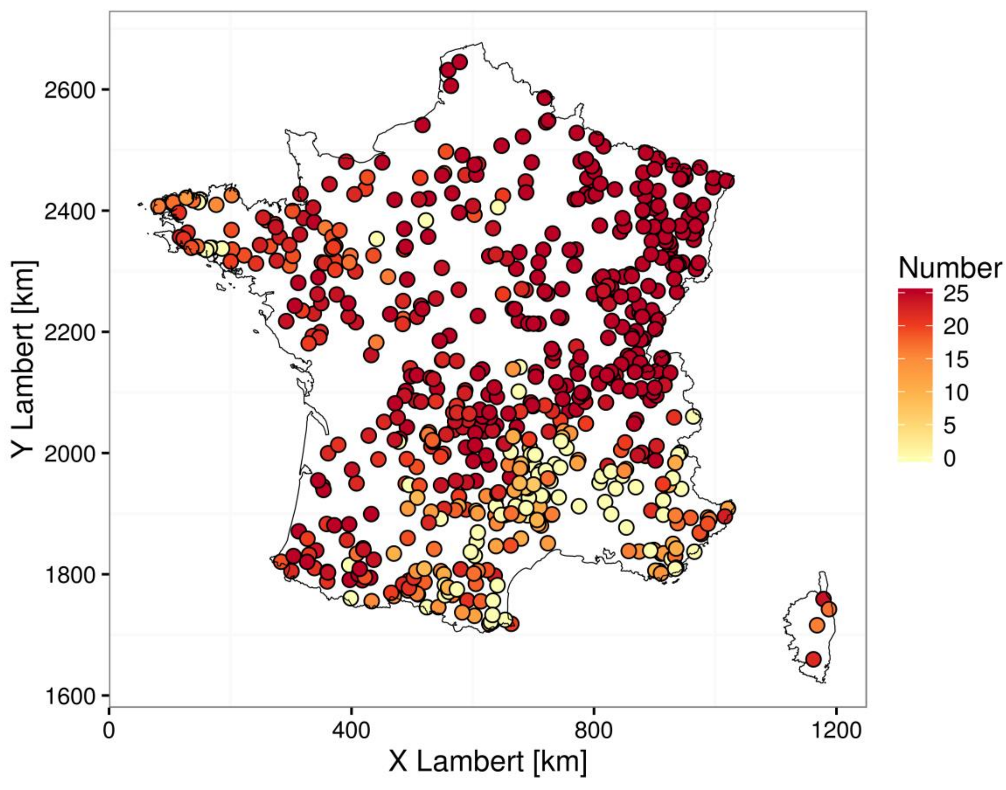
Dayon's study is based within France, this will be specified.

4. Page 3 line 16 – downscaling of 20CR – what variables did you downscale and at what resolution?

As mentioned previously, this will be added in the details of the SCOPE method in appendix. Large-scale (2.5°) predictors are (1) temperature at 925 hPa and 600 hPa, (2) geopotential height at 1000 hPa and 500 hPa, (3) vertical velocity at 850 hPa and (4) humidity as a combination of the relative humidity at 850 hPa and precipitable water content in the entire column.

5. Page 3 line 34 to page 4 line 1 – you mention the use of probabilistic data to account for uncertainty, but very few of your figures display the uncertainty from the ensemble in the results (only really Figures 10 and 11). It would be nice to see some spatial mapping of uncertainty – perhaps a map of one of your events, with density of colour according to how many of the 25 ensemble members identify that event, this map could accompany figure 17.

Thank you for this suggestion, this would indeed be a nice figure. For a specific event, the spatial pattern would be similar to the one obtain for duration or severity. Indeed, there is a higher probability that a high number of members detects the event if the latter is particularly long or severe for a station. Below is an example of the number of members detecting the 1893 event (to be compared to corresponding maps of return period in duration and severity in Fig. 16):



As this paper is a methodological paper with already a lot of example results, this will not be added. But a sentence such as: “The more severe or the longer the event, the higher the number of members detecting the event.” will be added.

6. Page 4 line 19 – what does “without direct human influence” mean?!

This means without abstractions, derivations and reservoir operations, i.e. human influence of catchment processes

7. Page 4 line 27 – we are referred to Annex A which mentions Safran Hydro and SCOPE Hydro before the datasets are introduced. I suggest you add “see section 2.2.1 and 2.2.2” to the caption of Figure 18.

This will be added.

8. Page 7 line 20 – I don’t think it can be said to give “equal weights”, better to say “reduces the bias towards high flows”

This will be corrected.

9. Page 7 line 24 – how many stations used CemaNeige then? State this. Could you highlight them somehow on Figure 2?

All stations used for CemaNeige. CemaNeige is calibrated locally on 187 stations with high snow influence. Median values of the calibrated parameters are then used for the simulations on all 475 remaining stations. This will be rephrased.

10. Page 7 line 26 – so all the catchments were given the same values for the 2 CemaNeige parameters? Is this realistic? Or were the catchments calibrated several times, and then the median of those given to each individual catchment? This needs to be clearer.

Only catchments with little snow influence (475 catchments) have the same CemaNeige parameters. Using fixed parameters for these catchments allows keeping realistic values as there were not enough snowfall episodes during the calibration period to obtain realistic calibrated CemaNeige parameters. This will be rephrased.

11. Page 8 lines 17-24 – I appreciate that this algorithm comes from reservoir design, but it would be more appropriate to explain it in the context of flow deficit here.

This will be adapted to this context.

12. Page 8 line 25 - Your end date is defined as the time of maximum depletion. Do you think this is representative of the end of an extreme low flow/drought period? What about the time it takes for the stream to return to “normal” flows? See Parry et al (2016)

http://ppg.sagepub.com/content/early/2016/06/02/0309133316652801.refs

Our end date corresponds to the maximum depletion, so actually to the return to the normal flow, i. e to the date where the flow exceeds the threshold again (but not the normal deficit). But indeed, the date corresponding to the return to 0 deficit would be useful for assessing drought recovery. This would be a nice additional study which is out of the scope of this paper.

13. Page 11 Figure 4 – Station 11 shows a single black bar in (a), but a red and a grey bar in (b), is this actually one event or is fig (a) not quite clear enough?

In fact, these are two very close but independent events with only a few days between them, hence the two colors. The resolution of the figure will be increased in order to clearly distinguish the two separate events.

14. Page 14 Figure 7 – why is the late 1990 event not picked up by Safran?

This event is not picked up for this particular station but is actually picked up for all other stations of the HER (see Discussion section 6.3, p. 28, l.27-30). This may happen locally, due to the sensitivity of the method to different parameters like the local threshold.

15. Page 14 Figure 7 – this figure takes a very long time to load on my PC – it interrupts scrolling significantly, and almost crashes my browser. Is its file size much larger than Figure 4? If so why? Can it be reduced without compromising its quality?

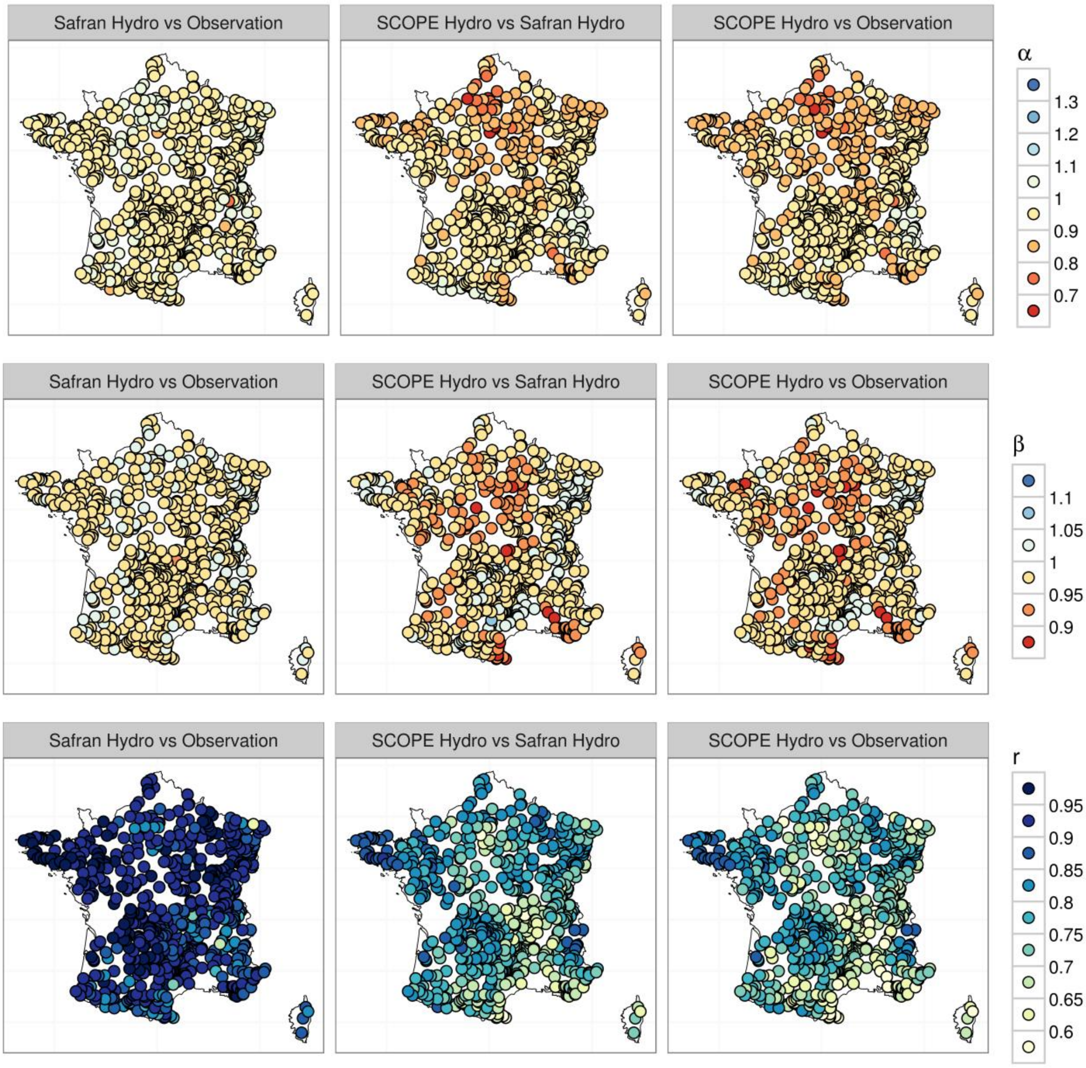
We actually had the same problem (even if the file size was the same than for the other figures). The figure has been created differently and it now works correctly.

16. Captions for Figures 8 and 9 – don't think you need to point out the different scales on the yaxes.

Thank you for the suggestion.

17. Page 16 – provide an assessment of the bias in SCOPE Hydro (median) compared with observed flows, for low and high flow seasons. Calculate the percentage of time the obs are within the SCOPE Hydro range for all stations.

SCOPE Climate and SCOPE Hydro will be made available in forthcoming data papers. They will also provide a deeper description of these two datasets, including median bias, KGE or reliability. As this paper is already very dense, we do not wish to add further validation results. For information, the maps of the KGE decomposition (in median) are the following (alpha as variance, bias as beta and r as linear correlation on the calibration period):



18. Page 17 line 12 – not sure I agree that the observations are "most of the time" included in the SCOPE Hydro range – the obs seem to be on the periphery most of the time – address with comment above.

This will be rephrased. Further analyses on the reliability of the SCOPE Hydro ensemble (that will not be shown here) actually support this statement.

19. Page 18 Figure 10 and page 20 Figure 12 – the dashed lines are visible on the computer, but not on the print out for me. This could easily be a problem my end, but please double check it prints correctly for you.

We did not encounter this problem.

20. Page 18 figure 10 – the "Whiskers" (note spelling!) do not extend to 1.5 times the IQR, they extend to the largest (and smallest) observations still within 1.5 times the IQR

Indeed, this will be corrected.

21. Page 19 line 17 – "A higher number of extreme events and higher severity values are simulated after 1940 for the Corrèze" – do you have any stats to prove this? Mann-Whitney U test for step change? Is there statistical significance?

These assumptions are only based on the figures and detailed statistic trend tests have not been considered as the aim of the paper was to provide basic examples of the method.

22. Page 20 line 9 – "There is no visible trend on the seasonality of start dates" – did you do any statistical tests?

Cf. answer to above comment.

23. Page 22 Figure 14 – You have displayed months and years on this plot whilst all other plots just categorise by year. You could remove the months from this plot to make the plot clearer.

Other plots are categorised by event names (actually a year) on the contrary to this plot which is categorised by date of maximum spatial extent. Keeping the months allows a better understanding of the seasonality of the event and a coherence with the x-axis. Moreover, removing the month might confuse the reader as the names of the events do not always correspond to the year of maximum spatial extent (for example the 1990 event corresponds to 12/1989).

24. Page 23 Figure 15 – this is a great Figure, but it really isn't colour-blind friendly. Remember 1 in 10 men is colour-blind. I know this will be difficult to take into account with a map like this, but I suggest you make sure that the 3 or 4 main events you are picking out are in contrasting colours that colour-blind people can differentiate from the others. At the moment, for those with Protanope colour-blindness (the most common) 1990 is clearly visible, 1893 and 1985 look the same as each other, 1878 and 1978 look the same as each other, 1943 and 1949 look the same as each other, and the rest are all very similar. Try using the mobile app "CVSimulator" to test your images (I've run your image through this app, see graphic at end of document). On this note – Figures and 4 and 7 use orange and green which are indistinguishable to the colour-blind. Reconsider this colouring if possible. Figures 16 and 17 are OK, Figure 20 isn't but I don't think the issue can be avoided here sadly.

Thank you for this remark and this very interesting app. We tried to use colorblind safe color scales even if the result is not always perfect. Figure 15 requires 12 different colours and it is not possible to find an entire adapted color scale. Nevertheless, we will indeed try and change the colors to distinguish the major events from each other.

25. Page 23 line 4 – I think you have 1878 and 1893 the wrong way round in this sentence

You are right, this will be corrected.

26. Page 23 lines 7 to 9 – comment on the 2003 event – this is a major finding, I suggest you add this to the conclusion!

This result should be taken carefully. As mentioned, it is not possible to automatically compare events from Safran Hydro and SCOPE Hydro but it is possible to compare exceptional events by manually selecting the events in each dataset. A comparison showed that the 1976 and 1990 events are very well reconstructed by SCOPE Hydro but the duration and severity of the 2003 event are understimated by SCOPE Hydro. This may be due to the lack of soil-atmosphere retroactions which are important for this event and not taken into account into our hydrometeorological reconstruction chain. A sentence will be added to the paper to prevent the reader from any over-interpretation.

27. Page 26 lines 4 to 9 – this discussion seemed to suggest that 1990 wasn't a particularly severe event, whereas Figure 16 indicates it really was just as severe as the 1893 and 1976 events, especially in the massif central regions.

The sentence should be rephrased with the removal of "on the contrary". The discussion focused on the fact that this event was exceptionally long for a majority of France, more than exceptionally severe.

28. Page 26 lines 10 and 11 – you state that the 1990 event start dates lie within a common 6 month period for all stations concerned – however I can definitely see some green dots on that lower right map in Figure 16! I suggest rewording "all stations concerned" as "for the vast majority of stations"

This will be corrected.

29. Page 30 section 6.6 – no mention of parameter uncertainty and model structural uncertainty – e.g. GR4J and GR5J.

Model structural uncertainty is mentioned with the use of the Isba-Modcou hydrological model. A sentence will be added on the parameter uncertainty.

30. Page 30 conclusion – no mention of the use of Safran Hydro

Indeed, the aim was to mention only the main datasets and results in order to make the conclusion more understandable.

Spelling/Grammatical Errors

Thank you for your very attentive reading, all your corrections will be taken into account.

References

Caillouet, L., Vidal, J.-P., Sauquet, E., and Graff, B.: Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France, Clim. Past, 12, 635-662, doi:10.5194/cp-12-635-2016, 2016.