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?

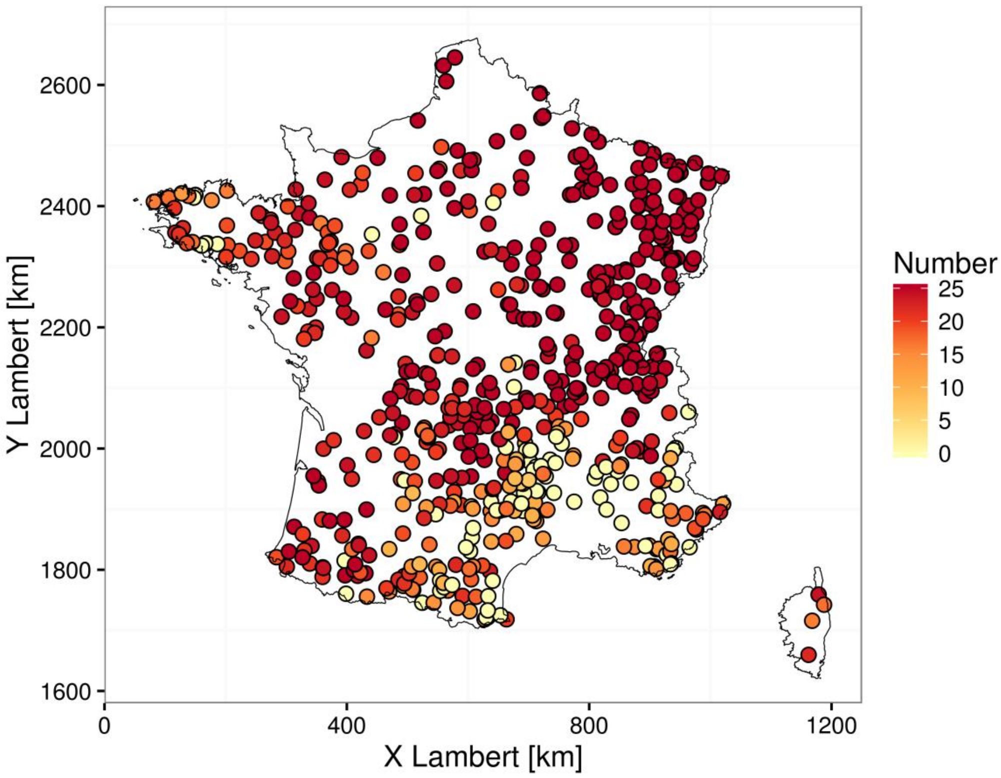
Dayon's study is based within France, this will be specified.

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?

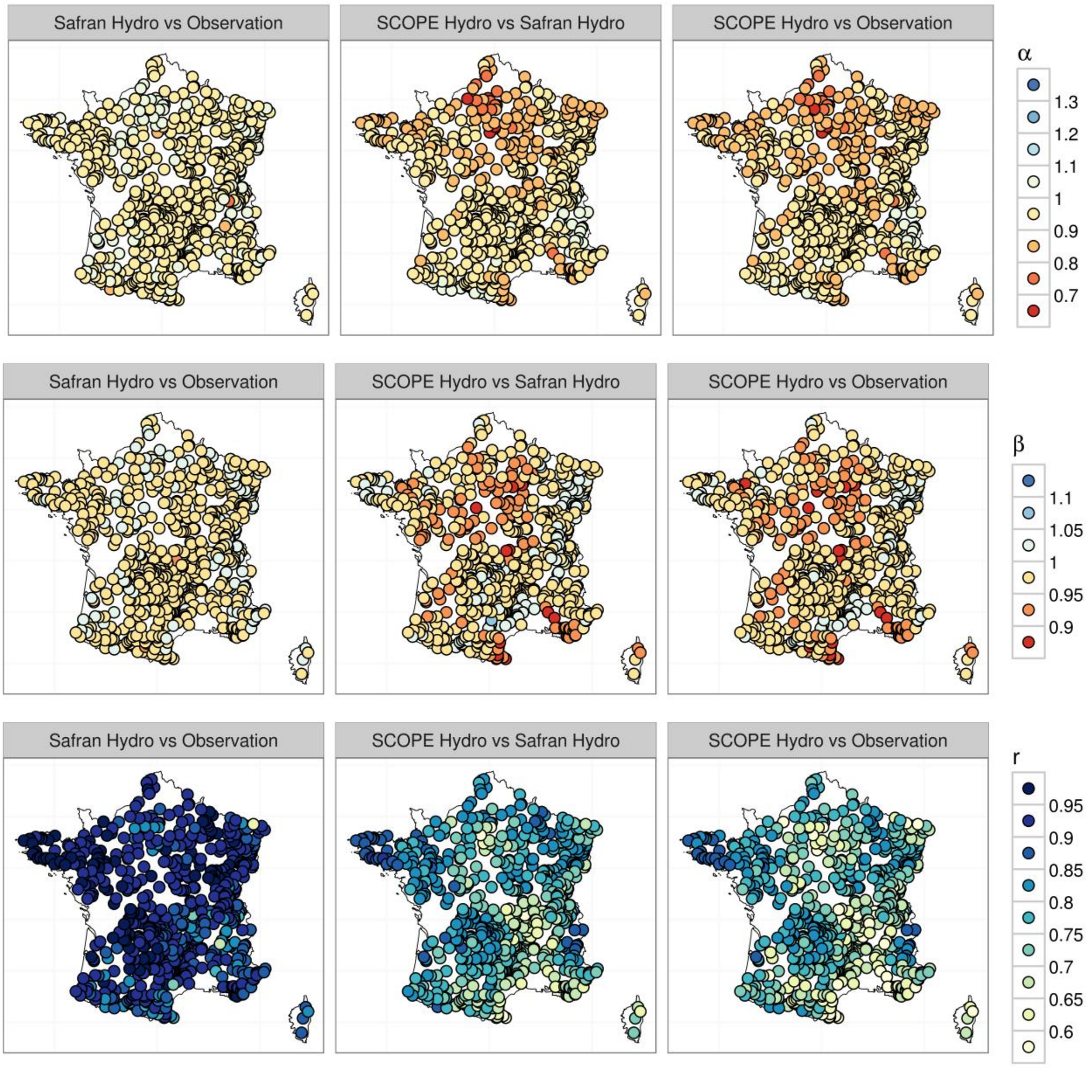
We actually had the same problem (even if the file size was the same than for the other figures). The figure has been created differently and it now works correctly.

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.

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 :

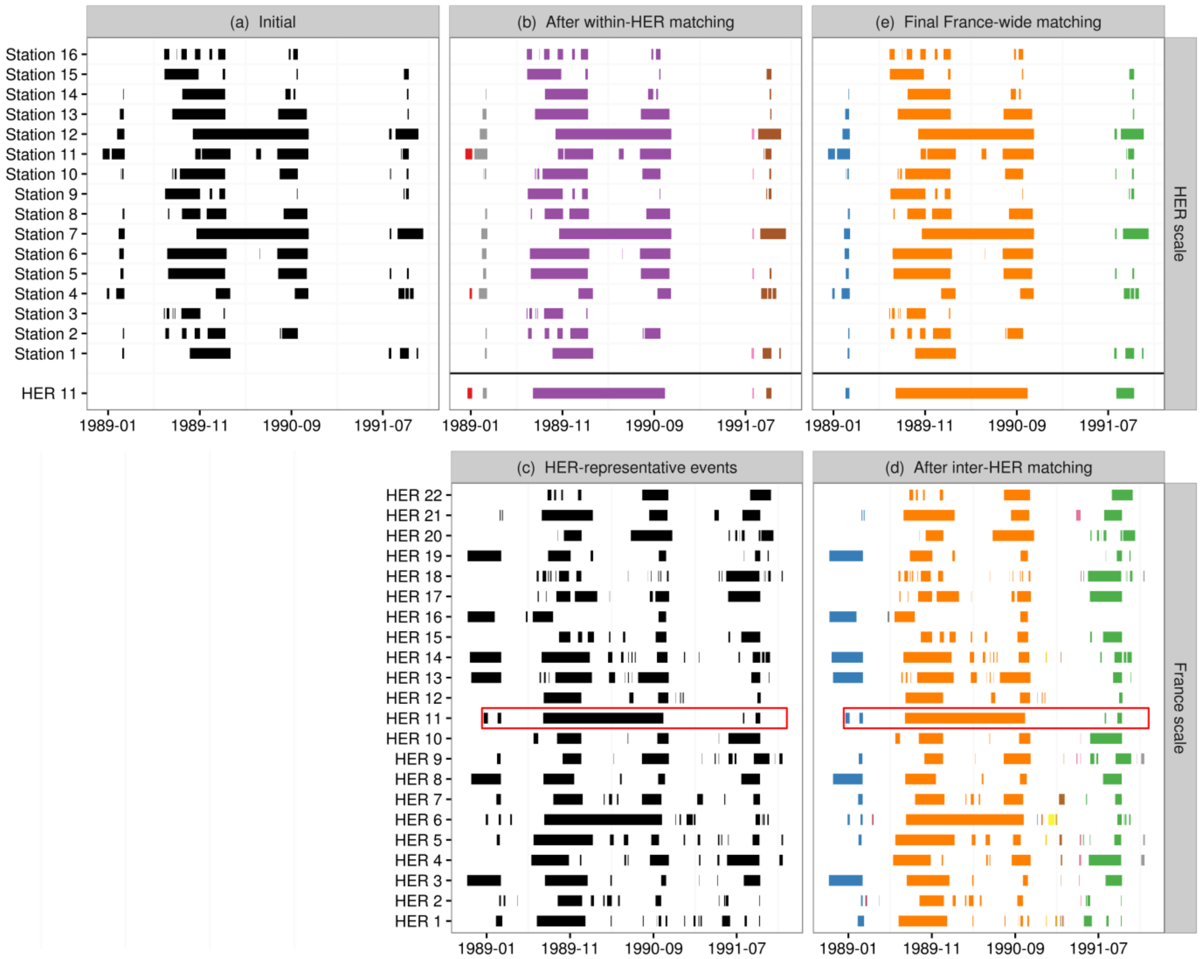
§2.2.2 SCOPE Climate : this paragraph presenting the bias correction via a resampling-based correction approach and improvement of spatial coherence via Schaake Shuffle should be improved in order to be easily understood ;

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 (...) ;

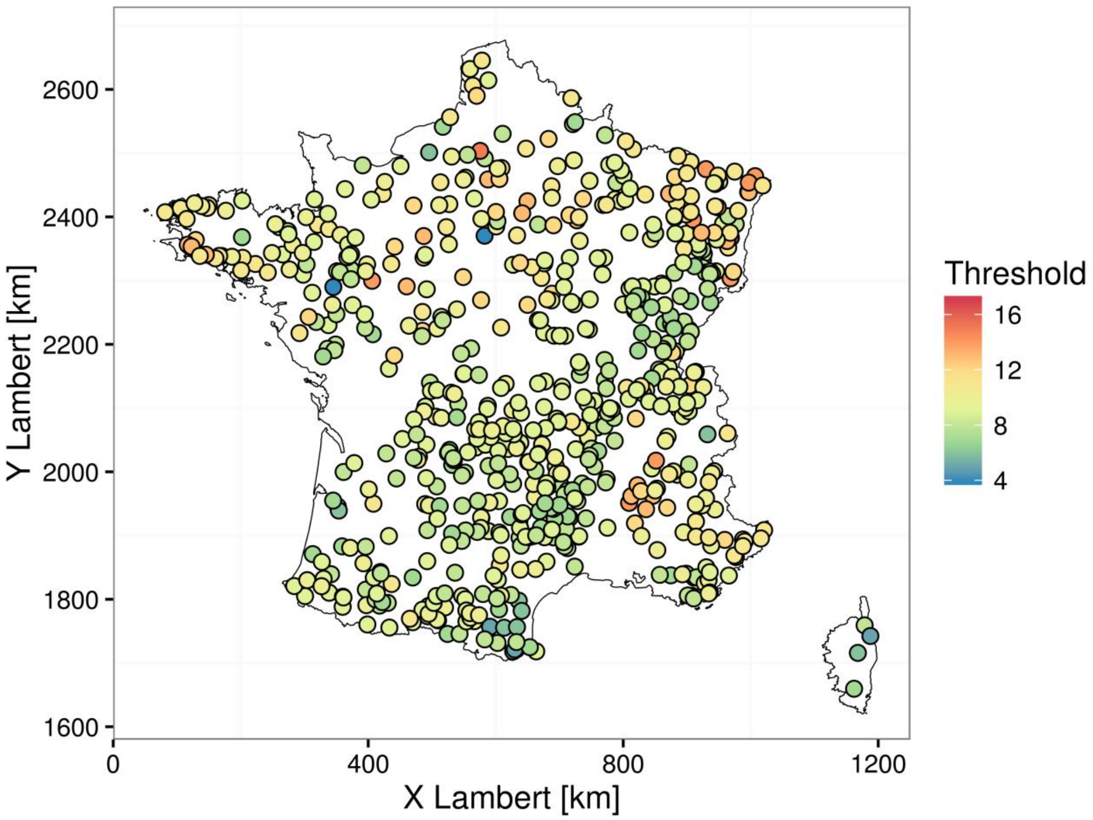
Indeed, this will be corrected.

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:



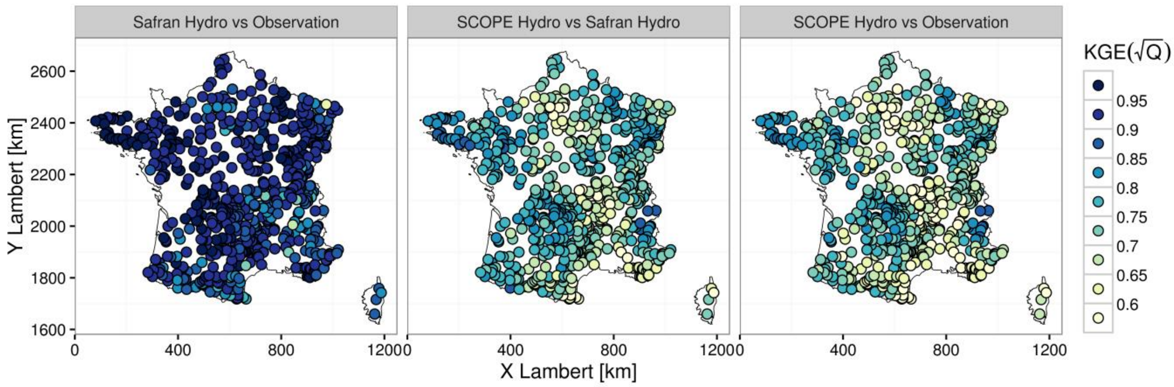
A unique value has not been tested as having a different value for each station allows a re-calibration against Safran to improve low-flow event identification. But except for specific stations, values are kept between 8 and 12.

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.

Synthesis of my review :
General comments Overall I find this an excellent paper which makes an important contribution to understanding French low flows but also in providing a highly transferable method for reconstruction of daily river flows in other countries/locations. The authors are to be congratulated; this paper provides a benchmark study. The paper would however benefit from some additional work on grammar and tightening the communication. Some aspects are a little hard to follow and would benefit from some careful thought on how to present in a clearer way to help the reader follow what is going on. I think that a lot of the appendix information should be integrated – e.g. the work flow figure is very useful to understanding what is going on. All of my specific comments below are minor and try to be constructive. Finally I apologise from my delay!

The authors would like to thank Referee 3 for his positive comments on the manuscript. The detailed answers to the specific comments (in italic) are presented below.

Specific comments In section 2.2.2 SCOPE climate it would be useful if the authors could introduce the data, then outline the steps in its use and then deal systematically with the additional treatments...for example it would be helpful to the reader of the final sentences in the section appeared earlier. These steps could then be used to organise the section

For a better understanding, we will detail the SCOPE method in appendix. Only the SCOPE Climate dataset will be detailed in the main text. See also the related responses to referee #1 and referee #2.

Given the objective of creating an ensemble reconstruction, why was only one hydrological model used and why is no consideration given to the uncertainty in the GR6J parameters. The model is calibrated for the period Jan 1973- sept 2006. Was there a reason for choosing this period based on variation in flow conditions? I ask as the model is expected to reconstruct conditions that are potentially very different from the calibration period. If the focus is on low flows – was consideration given to how the model performed for different duration/intensity events during the period of observations.

There is a work in progress concerning the hydrological modeling structural uncertainty (see section 6.6). This will be presented in a future paper. The calibration period has been chosen as 90% of the observed date were available during this period. Moreover, it indeed includes three important extreme low-flow events (1976, 1990, 2003).

I note that validation across the full set of catchments is not shown but is done – what were the salient points – it would be useful to summaries these in a couple of sentences. "It has to be noted that thorough validation experiments not shown here – out-of-sample experiments, split-sample experiments – have been performed to carefully quantify the overall hydrological modelling performance."

Thank you for this suggestion. A few salient points will be added to the manuscript.

I find the spatial mapping procedure difficult to follow and its communication would benefit from more clearly laying out the steps and then showing the example application

The principle is first explained (simply an overlap of dates). Then the definition of the required parameters is developed (spatial domain for the matching). Finally, the example only provides a better understanding of the use of two spatial domains (by HER and then over France). We do not think it would be possible to explain these two steps without an example. Nevertheless, we will try to improve the understanding of this step.

Is it possible to make a conclusion around which aspect of drought – severity or duration – uncertainties are greatest?

It would be difficult to establish such a statement as severity and duration are two different variables with different units (mm and days). It would be possible to compute the range of values using return periods but it would add uncertainties related to the distribution fitting. Moreover, these two aspects highly depend on the specific event considered. Figure 16 shows for example that there are more stations detecting a long event rather than a severe event in 1990. This is the contrary for the 1893 event.

When reporting seasonality you mention no visible trend – is there evidence of trend in the other parameters- severity/duration? There would seem to be for severity in the Correze catchment.

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

Please dont start section 4 with figures – text first.

This will be corrected

On page 22 the text states "More generally, this figure highlights the fact that the only events having hit more than 70% of France occurred after 1940." How confident can you be that this is a real trend or an artefact of the quality of the underlying data. While 20CR and such reanalysis data are hugely valuable confidence will reduce in time. Just a thought to consider which might be mentioned in the discussion.

As mentioned previously, no statistical tests have been performed. We will mention uncertainties deriving from 20CR in the discussion (already discussed in Caillouet et al., 2016).

I dont think there is a need to have so many sub sections in 6.2 – these would be better consolidated.

Thank you for the suggestion. As this paper covers many subjects, the subsections allow the reader to quickly find a specific item.

I think it would be more helpful to the reader to have the workflow image and other material in the appendices integrated into the text. This would not lengthen the paper and increase its readability.

Indeed, thank you for the suggestion, we will integrate this figure into the main text.

The next generation 20CR gets back to 1850 if I am not mistaken. It would be useful to indicate this here with the potential to extend a further 20 years.

You're right (even if there is less confidence in 20CR before the 1870s)!

Minor points

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

Page 4 line 27 – are these daily observation since gauge commencement?

Yes, they are.

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.

Dear Kerstin Stahl,

Please, find attached the revised manuscript.
According to the reviewer's comments, the manuscript has been changed following these main points:

- Introduction has been slightly reduced,
- The SCOPE Climate dataset is described in the main text and the SCOPE method is detailed in appendix,
- The spatial matching procedure has been improved by several sentences and – more important – three figures have been modified: Figure 4 (added), Figure 5 (more detailed and some mistakes corrected) and Figure 8 (different colours).
- Figure 13 (in the previous version) has been removed.

Minor revisions:

- KGE formulation has been corrected,
- Caption in Figure 17 has been added,
- All the clarifications asked by the referees have been added,
- The errors pointed out by the referees have been modified.

Best regards,
Laurie Caillouet

Ensemble reconstruction of spatio-temporal extreme low-flow events in France since 1871

Laurie Caillouet[1], Jean-Philippe Vidal[1], Eric Sauquet[1], Alexandre Devers[1], and Benjamin Graff[2]

[1]Irstea, UR HHLY, Hydrology and Hydraulics Research Unit, 5 rue de la Doua, BP32108, 69616 Villeurbanne Cedex, France
[2]Compagnie Nationale du Rhône (CNR), 2 rue André Bonin, 69004 Lyon, France

Correspondence to: Laurie Caillouet (laurie.caillouet@gmail.com)

Abstract.

The ~~historical depth~~ length of streamflow observations is generally limited to the last 50 years even in data-rich countries like France. It therefore offers too small a sample of extreme low-flow events to properly explore the long-term evolution of their characteristics and associated impacts. In order to overcome this limit, this work first presents a daily 140-year ensemble reconstructed streamflow dataset for a reference network of near-natural catchments in France. This dataset, called SCOPE Hydro (Spatially COherent Probabilistic Extended Hydrological dataset), is based on (1) a probabilistic precipitation, temperature and reference evapotranspiration downscaling of the Twentieth Century Reanalysis over France, called SCOPE Climate, and (2) a continuous hydrological modelling using SCOPE Climate as forcings over the whole period. This work then introduces tools for defining spatio-temporal extreme low-flow events. Extreme low-flow events are first locally defined through the Sequent Peak Algorithm using a novel combination of a fixed threshold and a daily variable threshold. A dedicated spatial matching procedure is then set up to identify spatio-temporal events across France. This procedure is furthermore adapted to the SCOPE Hydro 25-member ensemble in order to characterize in a probabilistic way unrecorded historical events at the national scale. For the first time, extreme low-flow events are described and compared in a spatially and temporally homogeneous way over 140 years on a large set of catchments. Initial results bring forward well-known recent events like 1976 or 1989-1990, but also older and relatively forgotten ones like the 1878 and 1893 events. These results contribute to improve our knowledge ~~on~~ of historical events and provide a selection of benchmark events for climate change adaptation purposes. ~~This study moreover~~ Moreover, this study allows for further detailed analyses of the effect of climate variability and anthropogenic climate change on low-flow hydrology at the scale of France.

1 Introduction

Hydroclimate projections for the 21st century generally agree on an increase in low-flow severity in France (see e.g. Chauveau et al., 2013; Vidal et al., 2015) – and more generally in Southern Europe (see e.g. Forzieri et al., 2014; Prudhomme et al., 2014; Giuntoli et al., 2015) – that could undermine current water management practice and require drastic measures for adapting water uses and for sharing resources among different economic sectors (irrigation, hydropower production, etc.). In this context of adaptation to climate change, a ~~deep~~ comprehensive knowledge of major historical events that affected France

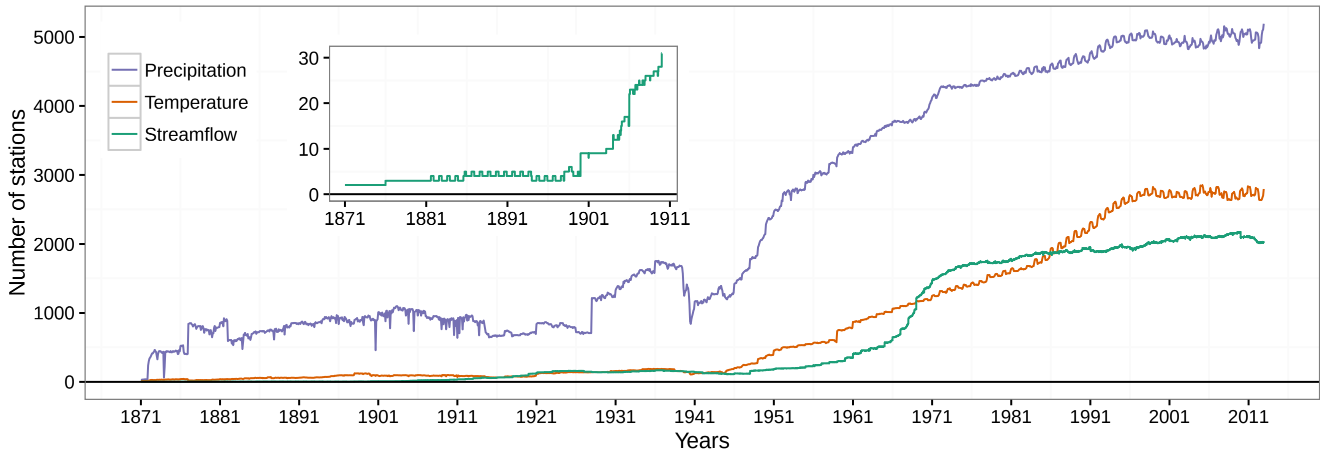


Figure 1. Evolution of the monthly averaged number of available precipitation and temperature stations in the Météo-France database (as of March 2015) since 1871 as well as the daily number of discharge stations in the Banque HYDRO French database since 1871.

constitutes a reference basis for anticipating future severe low-flow events. Studying these events requires a dense network of long and reliable streamflow series to derive robust characterizations at the national scale. However, even in a data-rich country like France, few local observations are currently available in databases before the 1950s, as shown by Fig. 1. Data from less than 10 hydrometric stations were available before 1900, and their number shows a slow increase after ~~1950 only.~~ 1950. Consequently, country-scale low-flow studies using observations usually focus on recent periods when enough data are available ~~(see e.g. Hannaford and Marsh, 2006; Giuntoli et al., 2013) or on~~ (see e.g. Giuntoli et al., 2013) or on a few stations available ~~on longer periods (see e.g. Hisdal et al., 2001; Pfister et al., 2006). Recent hydrometric data rescue actions for specific events (Auffray et al., 2011) or from specific historical sources (Le Gros et al., 2015) contributed to increase the breadth and depth of streamflow data over France, but they are still insufficient to allow studying extreme events at the national scale over long periods.~~ for longer periods (see e.g. Hisdal et al., 2001; Pfister et al., 2006, for studies in Europe).

Reconstruction of streamflow time series through hydrological modelling, and using precipitation and temperature data as inputs, have been developed to overcome the limited historical ~~depth~~ extent of hydrometric records (see Jones et al., 2006, and references therein). Crooks and Kay (2015) and Lennard et al. (2015) for example reconstructed daily streamflow series for specific locations in the UK, using ~~reconstituted~~ recovered historical meteorological data. Spraggs et al. (2015) extended meteorological time series back to 1798 to derive daily streamflow reconstructions and study the implications of extreme low-flow events for the Anglian region in the UK. In France, comprehensive reconstructions of droughts and low flows from 1958 onwards have been carried out with a hydrometeorological modelling chain (Vidal et al., 2010b; Soubeyroux et al., 2010), providing a reference basis for future spatio-temporal drought projections (Vidal et al., 2012). The latter reconstructions were however limited by the availability of upper-air and surface meteorological data before the 1950s. Figure 1 illustrates this limited availability of surface temperature and precipitation observations in France. It ~~has to~~ should be noted that climate proxies like dendrochronology data may also provide relevant information for reconstructing streamflow with low – seasonal to annual – temporal resolution (see e.g. Meko et al., 2012; Nicault et al., 2014). Dendrochronology data is most often used to derive meteorological drought indicators rather than hydrological drought indicators (see e.g. Cook et al., 2015; Labuhn et al., 2016).

A way to reconstruct meteorological data at high small spatial and temporal scales – and then catchment-scale hydrological data – is to use climate downscaling methods from large-scale atmospheric and oceanic data. Auffray et al. (2011) for example reconstructed hydrometeorological conditions that led to the 1859 flood of the Isère river (French Alps) based on purposely rescued pressure data and statistical downscaling with an analogue method. The recent release of two extended global reanalyses – the Twentieth Century Reanalysis (20CR, Compo et al., 2011) and the European Reanalysis of the Twentieth Century (ERA-20C, Poli et al., 2016) spanning the entire twentieth century (respectively from 1871 and 1900) – prompted studies aimed at deriving meteorological reconstructions on of the 20th century through downscaling methods, and at using these to reconstruct streamflow series. Kuentz et al. (2013) reconstructed the 20th century hydrological variability of the Durance catchment (Southern French Alps) based on a an analogue downscaling method (Kuentz et al., 2015) and a lumped conceptual hydrological model. More recently, Brigode et al. (2016) applied the same downscaling method and a hydrological model to reconstruct streamflow variability in a Northern Québec catchment. Few studies have been performed at a country scale, with the exception of Dayon (2015) who used a deterministic statistical downscaling approach combined to with a physically-based hydrological model to reconstruct hydrometeorological data in France. Caillouet et al. (2016) recently provided a reconstruction of meteorological fields at a 8 km spatial resolution and daily temporal resolution in France through a probabilistic precipitation and temperature downscaling of 20CR.

When streamflow time series are available from observations or hydrological simulations, numerous methods allows allow defining and characterizing extreme low-flow events. Two main approaches can generally be distinguished (Tallaksen and Van Lanen, 2004). The first one considers low-flow characteristics like annual minimum n-day discharge (see, e.g. Smakhtin, 2001)(see e.g. The second one focuses on characteristics of events temporally defined by deficits under a given threshold (see e.g. Fleig et al., 2006; van Huijgevoort et al., 2012; Van Loon and Van Lanen, 2012). The latter approach allows characterizing events at the local scale but has yet barely been used to study their spatial aspect. Some studies indeed looked at the areal extent of droughts as a spatial characteristic, but they did not identify independent drought events (e.g. Bonaccorso et al., 2013; Tallaksen and Stahl, 2014). Most of these studies focused on meteorological droughts and used a gridded and spatially homogeneous index to characterize the spatial extent of drought events (often the Standardized Precipitation Index, McKee et al., 1993). Algorithms defining spatio-temporal events as a sequence of spatially contiguous and temporally continuous areas where the index is under a given threshold value have even been developed (Andreadis et al., 2005; Vidal et al., 2010b). However, to the authors' knowledge no study proposed a spatio-temporal definition of low-flow events based on station data, i.e. on data neither homogeneous nor continuous in space.

This paper proposes an ensemble reconstruction of spatio-temporal extreme low-flow events in France since 1871, through an ensemble downscaling of the 20CR extended reanalysis – based on the 20CR-SANDHY-SUB meteorological downscaled dataset developed by Caillouet et al. (2016) – and the subsequent hydrological modelling of a large number of near-natural catchments. Compared to previous low-flow studies, this study has three main combined specificities: (1) local events are characterized by a combination of a fixed threshold and a daily variable threshold, (2) they are defined in a probabilistic way to take account of the uncertainties of the downscaling step in the streamflow reconstruction process, and (3) a spatio-temporal characterization of extreme low-flow events is provided at the scale of France.

The main objective of this work is to develop a method for identifying spatio-temporal extreme low-flow events based on probabilistic hydrometerological reconstructions. This work therefore presents beforehand the ~~refinement steps applied to the 20CR-SANDHY-SUB dataset (Caillouet et al., 2016) to create the~~ *SCOPE Climate* (Spatially COherent Probabilistic Extended Climate) dataset, derived from the statistical downscaling of 20CR, and introduces the *SCOPE Hydro* (Spatially COherent Probabilistic Extended Hydrological) dataset, the hydrological dataset derived from SCOPE ~~-~~Climate. The last objective of this work is to present some examples of how characteristics of spatio-temporal extreme low-flow events may be analysed and exploited to extract relevant information from the SCOPE Hydro historical reconstructions running from 1871 onwards. The paper is structured as follows. Section 2 introduces the different reanalysis and observation datasets used. Section 3 describes the hydrological modelling step as well as the new spatial and probabilistic definition of extreme low-flow events. Section 4 shows some examples of reconstructed streamflow time series. Section 5 provides some characteristics of the extreme low-flow events that occurred in France since 1871. Results are finally discussed in Sect. 6.

2 Data

2.1 Observed streamflow

A set of 662 near-natural catchments in France has been selected for the hydrological modelling and the study of extreme low-flow events (see Fig. 2). It is a combination of two existing datasets:

- The French low flow reference network consisting in 236 gauging stations, selected by Giuntoli et al. (2013) and meeting the following criteria: (a) at least 40 years of daily records, (b) the gauging station controls a catchment without direct human influence on river flow, (c) data quality is suitable for low flow analysis.
- A set of 632 gauging stations selected by Catalogne et al. (2014) and meeting the following criteria: (a) at least 26 years of daily records on the 1970-2005 period, (b) with no or few anthropogenic ~~influence~~influences, (c) with ~~a~~ good quality during low-flow periods.

Streamflow data for these catchments were extracted from the French HYDRO database (http://hydro.eaufrance.fr/). Two case study catchments with contrasted regimes and long observation records will be used to exemplify local-scale reconstruction results in Sect. 3.1 and Sect. 5: the Corrèze@Brive-la-Gaillarde with an oceanic dominated regime, and the Ubaye@~~Barcelonnette~~Barcelonnette, with a snowmelt dominated regime. Figure 18 in Annex A shows their daily interannual regimes. Observations are available since 1 January 1918 for the Corrèze and since 1 January 1904 for the Ubaye.

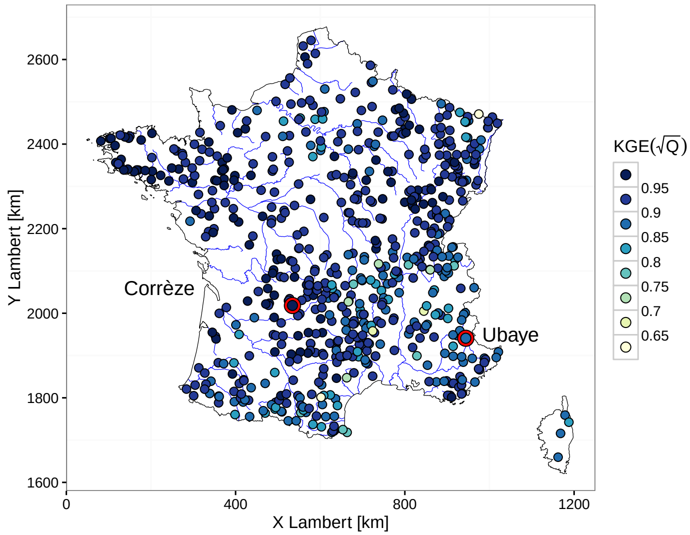


Figure 2. Location of the 662 hydrometric stations, and KGE calibration score for the hydrological model. The ideal value of KGE is 1. Red circles highlight the two case study catchments: the Corrèze@Brive-la-Gaillarde (Western slope of the Massif Central mountain range) and the Ubaye@Barcelonnette (Southern Alps).

2.2 Meteorological forcing datasets

2.2.1 Safran near-surface reanalysis

Safran is the French near-surface meteorological reanalysis (Quintana-Seguí et al., 2008; Vidal et al., 2010a). It provides meteorological variables such as precipitation and temperature ~~onto~~ on a 8-km resolution grid over France ~~and~~ at a daily temporal resolution from 1 August 1958 to present. Daily reference evapotranspiration is computed with the Penman-Monteith formulation (Allen et al., 1998). Safran data have been extensively used for hydrological studies, and notably for some ~~of them~~ dedicated to low flows (Vidal et al., 2010b; Soubeyroux et al., 2010; Pushpalatha et al., 2012; Nicolle et al., 2014). Catchment-scale data are computed with a weighted mean (for temperature) or sum (for precipitation and evapotranspiration) of each contributive cell of the 8-km grid to the catchment surface, based on the river network defined by Sauquet (2006).

2.2.2 SCOPE Climate

SCOPE Climate is a daily 8-km spatially coherent reconstruction of precipitation, temperature, and reference evapotranspiration fields ~~since 1871~~ over France. ~~It is based on the 20CR-SANDHY-SUB dataset that provides an ensemble of 25 analogue dates for each day during~~ This dataset is available as a 25-member ensemble of Safran-like daily gridded time series over the period 1 January 1871 to 29 December ~~2012, independently for 608 climatically homogeneous zones paving France (Caillouet et al., 2016). These analogues are~~ 2012. SCOPE Climate is obtained through the statistical downscaling of the

Twentieth Century Reanalysis (20CR, Compo et al., 2011) with the SANDHY method (Stepwise Analogue Downscaling Method for HYd by two additional analogy levels (Caillouet et al., 2016). Analogues dates from the period 1 August 1958 to 31 July 2008 are converted to meteorological variables by resampling Safran reanalysis data20CR (Compo et al., 2011) with the SCOPE method (Spatially COherent Probabilistic Extension method), detailed in Annex B, as an extension of the work by Caillouet et al. (2016).

The median of annual precipitation between Safran and 20CR-SANDHY-SUB for the 1959-2007 periodshows a dry bias of Validation experiments of SCOPE Climate have been performed with both dependent and independent data at different time steps (not shown). SCOPE Climate presents very little bias in comparison to Safran on their common period. There is a slight overestimation of precipitation in spring (around 10% on average over France (see top right panel of Fig. 7 in Caillouet et al., 2016). This bias, due to too many dry days resampled during the analogue selection, may call into question any interpretation of derived hydrological features – and especially extreme low flows –, and a bias correction step had to be set up. In a similar context, Timbal et al. (2006) for example introduced a correction factor to adjust the reconstructed rainfall series but this technique – of Safran precipitation) as well as common bias correction techniques – does not allow to retain the physical consistency and multivariate correlation structure inherent in the analogue approach. A resampling-based correction approach similar to the one adopted by Sippel et al. (2016) is therefore considered here: for each day between 1871 and 2012, the N analogues giving the lowest precipitation are removed. N analogues are then randomly resampled among the (25-N) left to keep a 25-member sample size. N, which is defined so that the bias with respect to Safran data is minimized, increases with the precipitation underestimation , with a maximum of 3 over France. This process is independently done on each climatically homogeneous zone. Importantly, this resampling-based correction for precipitation does not affect the temperature bias and interannual correlation described in Caillouet et al. (2016) (not shown). a slight underestimation (resp. overestimation) of temperature in summer (resp. winter), that does not exceed 1°C in absolute terms. The seasonal variability of both precipitation and temperature is well reconstructed and SCOPE Climate is able to capture the increasing trend in temperature since the 1990s. SCOPE Climate shows an overestimation of reference evapotranspiration, in particular in autumn and winter as well as an underestimation of its seasonal variability.

In order to build gridded time series from the 20CR-SANDHY-SUB dataset, Caillouet et al. (2016) independently combined analogues dates from one zone to another. The resulting lack of spatial continuity (see the discussion by Caillouet et al., 2016) could be heavily detrimental to the identification of spatio-temporal low-flow events. This issue is here addressed through the Schaake Shuffle procedure, initially developed to reconstruct space-time variability in forecast meteorological fields (Clark et al., 2004). This procedure has been widely used as a post-processing of ensemble meteorological forecast fields for streamflow forecasting (see e.g. Robertson et al., 2013; Verkade et al., 2013; Demargne et al., 2014; Šípek and Daňhelka, 2015). Vrac and Friederichs (2015) also adapted it recently for multivariate bias correction of downscaled climate simulations. In the Schaake Shuffle approach, which can be seen as an empirical copula on rank correlation (Wilks, 2014), the ensemble members are reordered so that their rank correlations across both space and variables match the ones from a randomly-picked sample of observed multivariate fields. In the present application, rank correlations are considered across the 608 climatically homogeneous zones and across the three variables, and observed fields are taken from the Safran reanalysis. For each target date, 25 dates are randomly selected within a 120-day window around the corresponding Julian day and among the period 1 August 1958 to 31 July 2008, a period

~~consistent with the archive period for analogue dates in the SANDHY downscaling step (Caillouet et al., 2016). Observed rank correlations are derived from the Safran multivariate meteorological fields from these dates and applied to the reconstructed ensemble, thus ensuring a spatial and inter-variable coherence of any single ensemble member.~~

~~The succession of steps (1) SANDHY (Radanovics et al., 2013; Ben Daoud et al., 2016), (2) subselection with additional analogue levels (Caillouet et al., 2016), (3) Bias Correction (see above), and (4) Schaake Shuffle (see above) is called SCOPEand summarised by the 19 diagram in Annex B. The use of 20CR as the large-scale reanalysis input provides the SCOPE Climate dataset. SCOPE Climate is available as a 25-member ensemble of Safran-like daily gridded time series of temperature, precipitation and reference evapotranspirationover the period 1 January 1871 to 29 December 2012.~~ Catchment-scale data used for the hydrological modelling are computed as for ~~the~~ Safran data.

3 Methods

This section first presents the hydrological modelling step. The method specifically developed to characterize extreme low-flow events is then described in three steps: the local definition of events, the spatial matching of events and the application of this matching to the ensemble case.

3.1 Hydrological modelling

The hydrological model selected for this work is GR6J, a daily lumped continuous rainfall-runoff model developed specifically for low-flows (Pushpalatha et al., 2011). It derives from the widely used GR4J model (Perrin et al., 2003) – which has been used recently in a reconstruction context by Brigode et al. (2016) – with a 5th parameter added to better model exchanges between surface and groundwater (Le Moine, 2008) and a 6th parameter added to better model low-flow periods (Pushpalatha et al., 2011). GR6J has been intensively used in France, in particular for low-flow studies (see Pushpalatha et al., 2012). It is here combined with ~~Cemaneige~~ CemaNeige (Valery, 2010; Valéry et al., 2014), a semi-distributed snow-accounting routine with 2 parameters.

GR6J is calibrated over the period 1 January 1973 to 30 September 2006~~–~~, called CAL in the following~~–, when~~, where more than 90% of the discharge stations have ~~data, and which~~ available data. Moreover, this period includes three well-known extreme low-flow events: 1976 ~~(Bremond, 1976; Vivian, 1977; Zaidman et al., 2002, see e.g.)~~(see e.g. Bremond, 1976; Vivian, 1977; Zaidma 1989-1990 (see e.g. Mérillon and Chaperon, 1990) and 2003 (see e.g. Moreau, 2004). A warm-up period of 3 years before the calibration period is considered to initialise the levels of the different water stores. The calibration uses Safran data as inputs and the Kling-Gupta-Efficiency (KGE, Gupta et al., 2009) on the square root transformed streamflow ~~(to give equal weights to high- and low-flows)~~ – to reduce the bias towards high flows – as objective function. The ~~KGE is~~ perfect value of KGE is 1 and this score is computed as follows:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (1)$$

where r is the linear correlation coefficient between simulation and observation, α the ratio between simulated and observed variance and β the ratio between simulated and observed mean. The two ~~Cemaneige~~ CemaNeige parameters are calibrated only for ~~stations~~ catchments with a snowfall/rainfall ratio over 10% – representing 187 stations out of 662 – in order to prevent unrealistic values for stations without enough snow episodes. For ~~these~~ the remaining 475 stations, the median values of the calibrated ~~Cemaneige parametersare adopted~~CemaNeige parameters, from the 187 snow-influenced catchments, are adopted to run CemaNeige. Calibrated models are then run with the two meteorological datasets described in Sect. 2.2 on the 662 stations to obtain:

Safran Hydro: 662 daily streamflow series over the 1 August 1958 to 31 July 2014 period using Safran as input,

SCOPE Hydro: 662 x 25 daily streamflow series over the 1 January 1871 to 29 December 2012 period using SCOPE Climate as input.

3.2 Characterizing extreme low-flow events

3.2.1 Local definition of extreme low-flow events

An approach based on deficit characteristics under a given threshold is adopted here to identify extreme low-flow events (Tallaksen and Van Lanen, 2004; Fleig et al., 2006). The first two panels of Fig. 3 present two commonly used thresholds. A fixed threshold characterizes the low-flow season which depends on the hydrological regime of the catchment (see e.g. Tallaksen et al., 1997). A variable threshold characterizes any deviation from the normal seasonal pattern (see e.g. Van Loon and Van Lanen, 2012). A period of water deficit – filled in red in Fig. 3 – is considered when discharge falls below the threshold level and continues ~~so~~ until the threshold is exceeded again. As the aim of the study is to characterize extreme low-flow events, the third panel proposes a mixed threshold as the daily minimum values of the two thresholds described above. This mixed threshold thus allows identifying events deviating from the normal seasonal pattern only during the low-flow period.

The mixed threshold is computed here using the 90th percentile of daily streamflow over the CAL period: (1) from all days for the fixed threshold, and (2) independently from each Julian day for the daily variable threshold. The daily variable threshold is then smoothed through a 10-day moving average to improve the day-to-day consistency. Note that alternative daily variable threshold estimates using percentiles over moving windows around the Julian date may be equally suitable (Van Loon and Laaha, 2015). The choice of percentile will highly condition the identification and characterization of extreme low-flow events. This aspect will be further discussed in Sect. 6.1.

The Sequent Peak Algorithm method (SPA, Vogel and Stedinger, 1987; Tallaksen et al., 1997) is then applied to pool extreme low-flow events. ~~This method if often assimilated to a procedure for preliminary reservoir design. For a daily inflow to a reservoir~~ For a period i, a deficit volume is computed by cumulating the difference between the daily discharge Q_i and ~~a desired yield Q_0~~ the threshold Q_0 – here corresponding to the mixed threshold defined above~~, the required storage at the beginning of a~~. Only the periods with a positive deficit are considered as periods with extreme low flows. The daily deficit

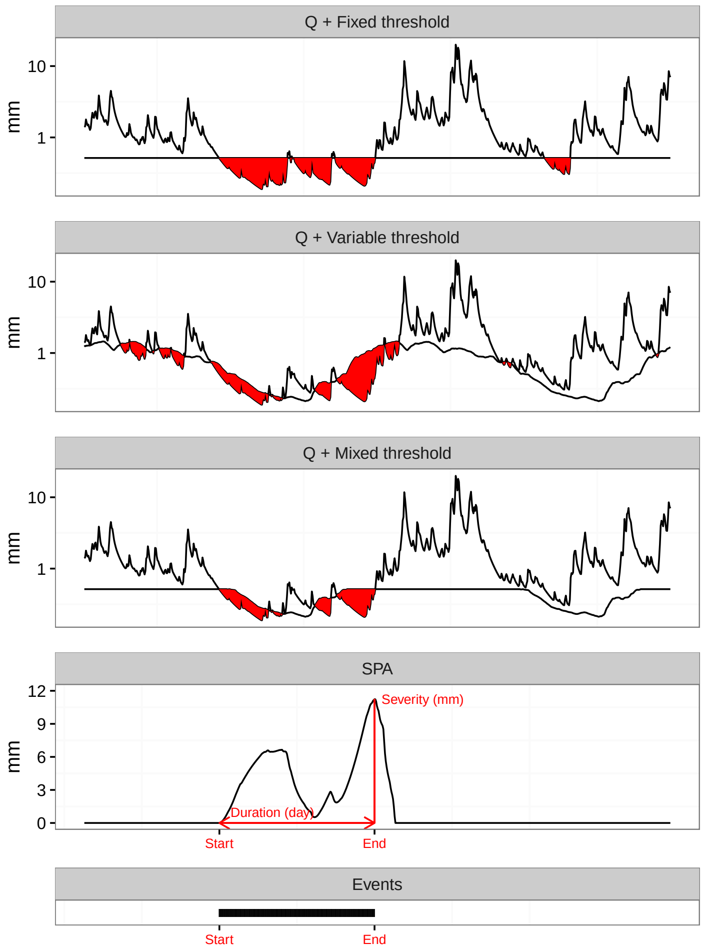


Figure 3. Extreme low-flow events characterization method. First panel: Fixed threshold. Second panel: Variable threshold. Third panel: Mixed threshold. Fourth panel: Sequent Peak Algorithm output using mixed threshold, and extreme low flow events characteristics. Fifth panel: Event definition (Start date to End date).

volume for the period i, S_i, is defined by computed following equation 2:

$$S_i = \begin{cases} S_{i-1} + Q_0 - Q_i, & \text{if positive} \\ 0, & \text{otherwise} \end{cases}$$

$$S_i = max(0, S_{i-1} + Q_0 - Q_i) \quad (2)$$

The fourth panel of Fig. 3 presents an example SPA curve, corresponding to the required storage deficit volume time series, S_i. An uninterrupted sequence of positive S_i defines a period with storage depletion and a subsequent filling upan extreme low-flow event. The event characteristics are then defined by a *severity* (required storage in the periodlargest deficit volume, max(S), in mm), a *start date* (the beginning of the depletion deficit period), an *end date* (time of the maximum depletiondeficit), and a *duration* (number of days between these two dates). The bottom panel of Fig. 3 presents the formalisation of a single event used in Sect. 3.2.2 and Sect. 3.2.3 as a bar running between the start and end date.

The above procedure is applied for identifying and characterizing extreme low-flow events for each gauging station, independently on (1) observed time series, (2) Safran Hydro simulations, as well as (3) each of the 25 SCOPE Hydro reconstructions.

3.2.2 Spatial matching of extreme low-flow events

The development of a spatio-temporal event definition is required to precisely characterize a specific event across France. Indeed, events should be defined across the 662 stations to studyfor example the spatial extent of an event , for example, their spatial extent or the worst-affected region. The spatial definition developed in this section relies on a concept similar to the one developed by Uhlemann et al. (2010) for assessing trans-basin flood events. The definition is presented below for a single set of independent events by station (deterministic case). The adaptation to the probabilistic ensemble case is developed in Sect. 3.2.3.

The spatial definition developed here only uses the start and end dates as represented in the bottom panel of Fig. 3. When an overlap of event dates occurs between different stations, the local events are considered as parts of the same spatio-temporal event. The spatial domain, i.e. the set of stations over which the matching is done, has next to be chosen. The matching is here first done within French Hydro-ecoregions (HERs, Wasson et al., 2002). The zoning of France into 22 HERs shown in Fig. 20 of Annex C has been developed for the implementation of the EU Water Framework Directive. It is based on geology, relief and climate criteria to define biological, physico-chemical and hydromorphological reference conditions. The France subdivision in HERs have been widely used in recent hydrological studies (see, e.g., Sauquet and Catalogne, 2011; Cipriani et al., 2012; Snelder et al., 2013). The spatial matching is then done a second time across HERs to derive spatio-temporal extreme low-flow events at the national scale.

This two-step approach originates in the large variability of extreme low-flow periods across catchments in France. Indeed, a single-step spatial matching gathering events from all 662 stations would lead to too much aggregated events in both space and time, possibly merging local events that do not proceed from a common meteorological driver (not shown). It may thus

~~lead to spatially match an event occurring in the North-West of France with an event occurring in the South-East of France 6 months later, only because of the late recovery of a single North-West station and an outlier early start for a single South-East station.~~ give too much influence to a single catchment.

~~Spatial matching procedure for the local events reconstructed from Safran Hydro over the 1989-1991 period, through the example of the 16 stations in HER 11. (a) Independent events represented by black bars (see bottom of Fig. 3). (b) Matched events after within-HER matching. Each colour defines a spatio-temporal event at the HER scale. HER 11 representative events are represented at the bottom of the panel. (c) HER-representative events for the 22 HERs. Each black bar represents one event. (d) Matched HER-representative events after inter-HER matching. (e) Spatial HER-representative events matching reported to local events after France-wide matching. Each colour defines a spatio-temporal event at the France scale.~~

Figure 4 ~~in Annex ??~~ summarises the different steps described here. Figure 5 illustrates the whole spatial matching process for the 16 stations in HER 11 ("Causses aquitains", South-Western fringe of the Massif Central mountain range) during the 1 January 1989 to 31 December 1991 period. The corresponding procedure is detailed below:

1. Figure 5(a): Local events resulting from the SPA procedure are represented by black bars for each station.

2. Figure 5(b) corresponds to results from the *within-HER matching* process in Fig. 4: Five spatio-temporal events~~built from temporally overlapping bars are~~ , identified here through different colours, are built from the temporal overlap of local bars. This overlapping procedure ~~may~~ often lead to pool formerly independent ~~events~~ local events into a single spatio-temporal event for a given station ~~within a single one:~~ (see for example Station 16 with the single purple spatio-temporal event gathering several local events identified by different bars). In such a case, the new local characteristics of this extreme low-flow event are computed from the former ones in the following way: the start date is the ~~earlier~~ earliest start date, the end date is the latest end date, the duration is the sum of durations, and the severity is the maximum of severities. For each spatio-temporal event, a *HER-representative event* is defined using the median of start dates and end dates from all stations concerned. This HER-representative event is shown at the bottom of Fig. 5(b) with label "HER 11".

3. Figure 5(c): HER-representative events are shown for all 22 HERs with black bars. The HER 11 representative events obtained above are framed in red.

4. Figure 5(d) corresponds to results from the *inter-HER matching* process in Fig. 4: All HER-representative events are spatially matched together using the same overlapping process than in the (b) subfigure. The HER 11 representative events – again framed in red – are here pooled among only 3 distinct spatio-temporal events at the scale of France.

5. Figure 5(e) corresponds to results from the *France-wide matching* process in Fig. 4: The inter-HER matching for HER 11 representative events – bottom of Fig. 5(e) – is finally reported to each station through the *local allocation* HER_i process in Fig. 4, leading to some additional local pooling, and notably the merging of early 1988 events (red and grey in Fig. 5(b)) and of the late 1991 events (pink and brown in Fig. 5(b)). New local event characteristics – start and end dates, duration, severity – are computed using the same procedure as above.

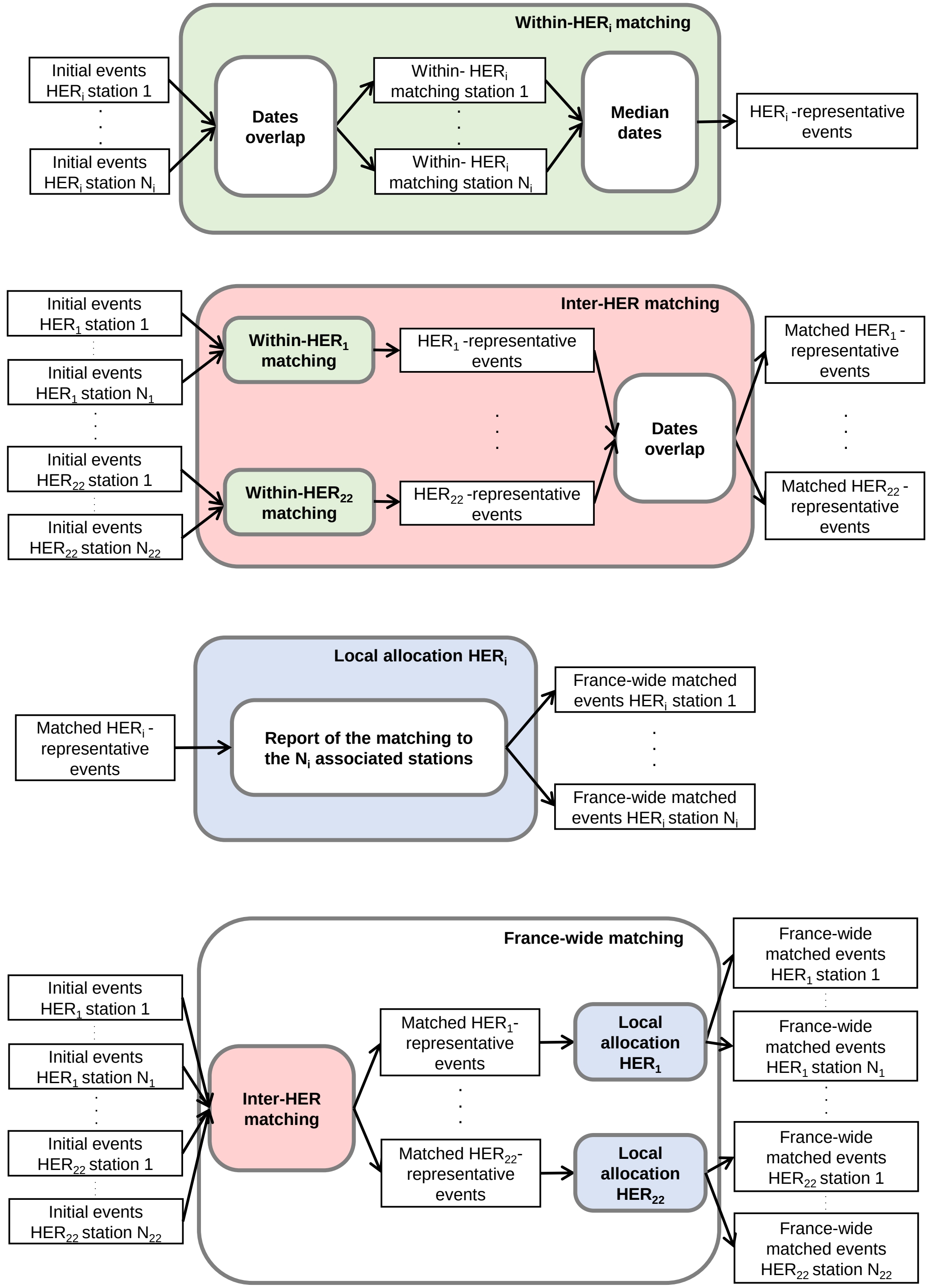


Figure 4. Synthetic diagram detailing the spatial matching procedure**12**

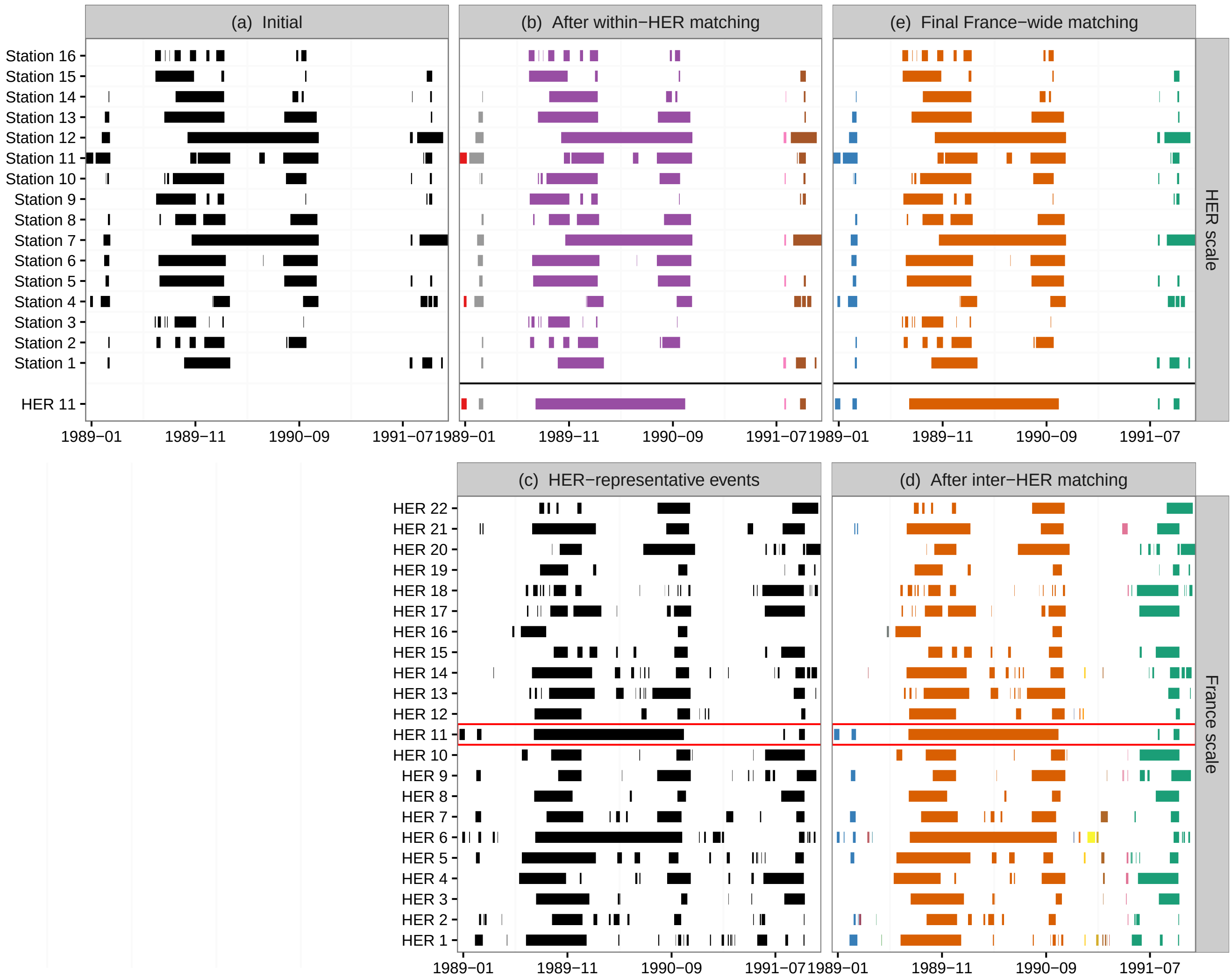


Figure 5. Spatial matching procedure for the local events reconstructed from Safran Hydro over the 1989-1991 period, through the example of the 16 stations in HER 11. (a) Independent events represented by black bars (see bottom of Fig. 3). (b) Matched events after within-HER matching. Each colour defines a spatio-temporal event at the HER scale. HER 11 representative events are represented at the bottom of the panel. (c) HER-representative events for the 22 HERs. Each black bar represents one event. (d) Matched HER-representative events after inter-HER matching. (e) Spatial HER-representative events matching reported to local events after France-wide matching. Each colour defines a spatio-temporal event at the France scale.

Twenty catchments with major aquifers – most of them in HER 9 – for which extreme low-flow events can last several years are sidelined from the whole spatial matching procedure, including the building of HER-representative events. This prevents once again too much aggregation in both space and time. The France-wide definition of spatio-temporal events matching obtained with all other stations is reported to the local events for these stations by overlapping with their HER-representative events. This may therefore lead to local events in these catchments being identified as belonging to several distinct spatio-temporal events defined at the scale of France, and consequently to prevent any meaningful comparison of characteristics.

This spatio-temporal definition of events is applied to events from Safran Hydro between 1958 and 2012.

3.2.3 Adaptation to the ensemble case

Deriving average events for Station 1 in HER 11. (a): Daily number of members simulating an event between the 1 August 1958 and the 29 December 2012. (b): Zoom on the 1989-1991 period. Examples are taken with 4 possible minima, set at 5, 10, 15 or 20 members.

Choice of the minimum number of members for creating the average event series for Station 1 in HER 11. (a): Events from Safran Hydro and average series created with a threshold between 1 and 25 members. (b): Computation of the number of days simulating an extreme low-flow event for Safran Hydro and the 25 series of average events. The number of members providing the closest average events total duration to the Safran Hydro events total duration – here 10 – is selected.

Ensemble spatial matching for Station 1 in HER 11 during the 1989-1991 period. Events from Safran Hydro, average event series and the 25 members of SCOPE Hydro are represented. (a): Each color defines one common spatial event for events from Safran Hydro and average series. Black bars are independent local events from the 25 members of SCOPE Hydro. (b): Matching between the 25 event members from SCOPE Hydro and average events. Black events not overlapping average events are considered as probabilistic noise and are removed from the dataset. Dashed lines are set at the start and end dates of each average event.

Characterizing The local characterization of extreme low-flow events following (see Sect. 3.2.1and) applied to SCOPE Hydro leads to 25-member ensemble of local event definition an ensemble of 25 local event definitions, for the 662 stations in France over the 1871-2012 period. The ensemble aspect leads to several questions: How many ensemble members out of 25 detecting an event are necessary to consider that it actually occurred? How to match events across the 25 ensemble members at the station scale? How to link spatially local ensembles of events that are detected, or not detected, or partially detected? This section attempts to provide responses to these questions by adapting the spatial matching described above to the ensemble case. The proposed approach goes along three steps: (1) reducing the ensemble reconstructions to a single deterministic series of extreme low-flow events to get back to the simpler Safran Hydro case, (2) applying the spatial matching described above to the resulting these deterministic series, and (3) reporting the spatio-temporal definition of events to each ensemble member.

The For the first step, the overall idea is to derive a single deterministic series of low-flow events from the 25-member ensemble that matches approximately the behaviour of the one derived from Safran Hydro, and that matches more specifically the total time in extreme low-flow conditions given by the Safran Hydro simulation. The reduction to a single deterministic series is done independently for each station following the procedure below~~and as~~, exemplified in Fig. 6 and Fig. 7 for Station 1 in HER 11.

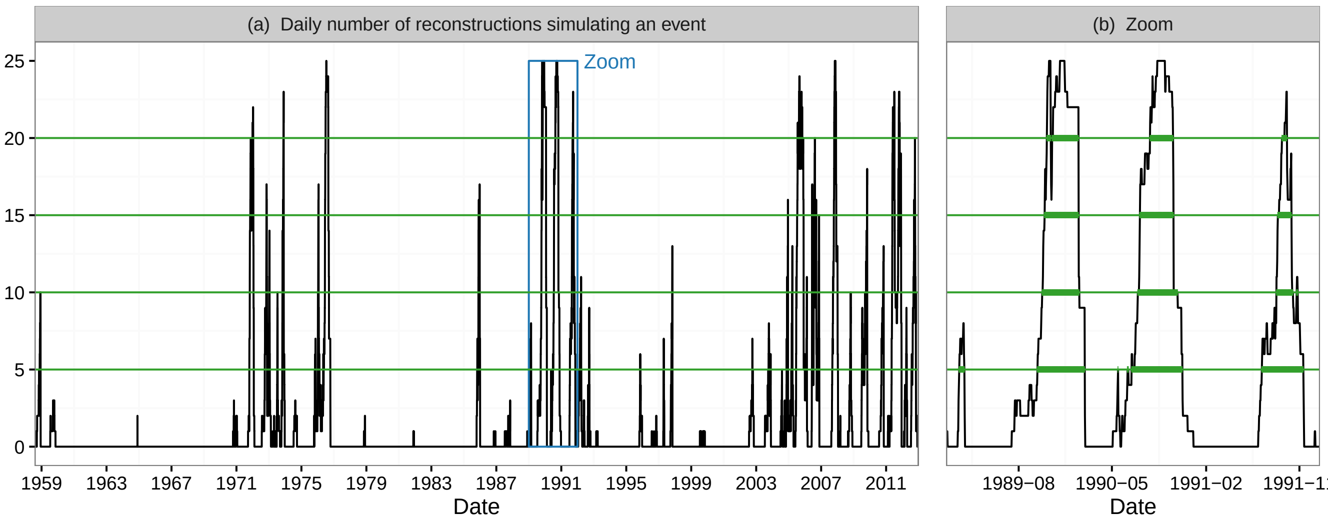


Figure 6. Deriving average events for Station 1 in HER 11. (a): Daily number of members simulating an event between the 1 August 1958 and the 29 December 2012. (b): Zoom on the 1989-1991 period. Examples are taken with 4 possible minima, set at 5, 10, 15 or 20 members.

1. Figure 6(a): The daily number of ensemble members detecting an extreme low-flow event between 1 August 1958 and 29 December 2012 – the common period between Safran Hydro and SCOPE Hydro – is computed and represented by a curve (between 0 and 25 members). For a given number of members between 1 and 25, an *average event* is created each time this number intersects the curve. This event starts when the curve exceeds the number and the event continues until the curve falls below it. Fig. 6(a) plots 4 different possible number of members and associated average events.

2. Figure 6(b): For a number of members set at 5 (resp. 10, 15, 20), 6 (resp. 4, 3, 4) average events are created between 1989 and 1991.

3. Figure 7(a): The 25 possible series of average events are created between the 1 August 1958 and the 29 December 2012, using the 25 possible number of members. They represent 25 deterministic series of events that could summarise the events obtained from the 25 initial series. Actual events from Safran Hydro are also represented (in red).

4. Figure 7(b): The total duration of average events over the whole period is computed for each of the 25 series. The ~~chosen~~ final number of members is the one giving the closest value to the total duration of Safran Hydro events over the same period (10 for this specific station).

~~The~~ For the second step, the spatial matching developed in Sect. 3.2.2 is ~~then~~ applied to these average events considered as a single deterministic series of reconstructed events, similar to this from Safran Hydro.

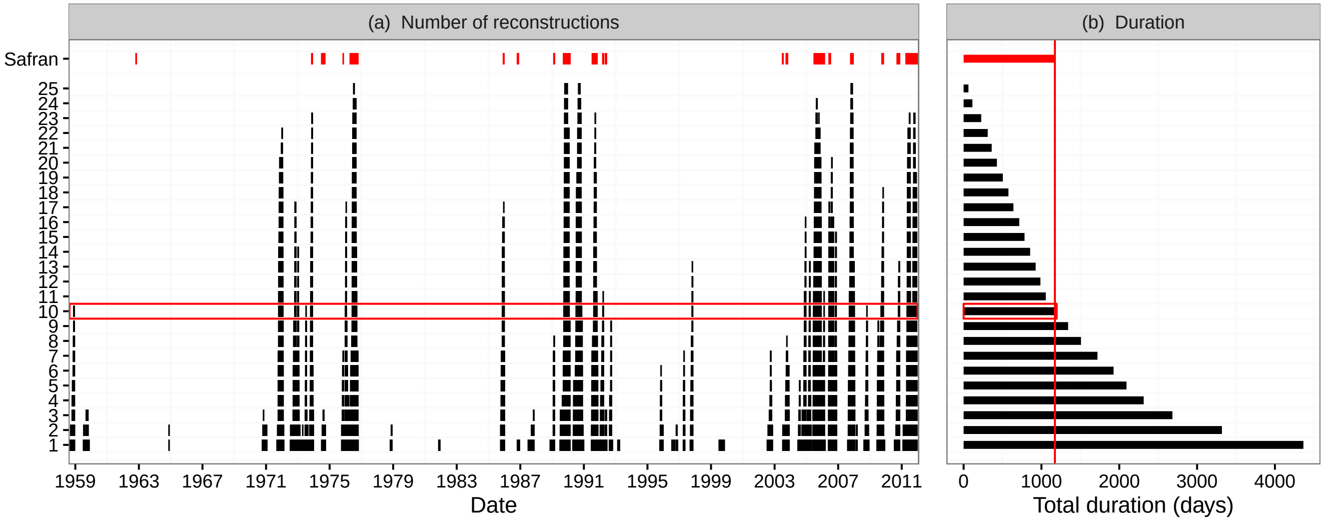


Figure 7. Choice of the minimum number of members for creating the average event series for Station 1 in HER 11. (a): Events from Safran Hydro and average series created with a threshold between 1 and 25 members. (b): Computation of the number of days simulating an extreme low-flow event for Safran Hydro and the 25 series of average events. The number of members providing the closest average events total duration to the Safran Hydro events total duration – here 10 – is selected.

The final step deals with reporting spatio-temporal (deterministic) events to each ensemble member of SCOPE Hydro. It is illustrated in Fig. 8 with the same station as above over the 1989-1991 period~~, and goes~~. Spatio-temporal events obtained from Safran Hydro are presented in Figure 8 only for a discussion in Sect. 6.3. They do not take part into the spatial matching procedure for the probabilistic case. The third step is as follows:

1. Figure 8(a): Events from all 25 SCOPE Hydro ensemble members are represented as well as average events. ~~Safran Hydro events are also represented, but only for a discussion in Sect. 6.3. Considering a number of 10 ensemble members, four~~ Four independent average events are available ~~on~~ in this period (see also Fig. 6(b)). After the spatial matching (which is not shown), they are pooled into two spatio-temporal events (~~orange and green~~red and mauve).

2. Figure 8(b): For each of the 25 ensemble members, events temporally overlapping with average events are matched together. Ensemble member events not overlapping with an average event (in grey) are removed. ~~This actually concerns~~ These are the events detected by less than the ~~chosen~~ final number of ensemble members (here 10 for this station).

The choices made during the spatial and probabilistic matching will be further discussed in Sect. 6.2 and 6.3.

Characteristics of extreme low-flow events – start date, end date, duration, and severity – are then computed independently for each ensemble member following the rules detailed in Sect. 3.2.2. A duration of 0 days and a severity of 0 mm are attributed to an ensemble member that does not detect an event, e.g. the ensemble member #22 for late 1991 event in Station 1

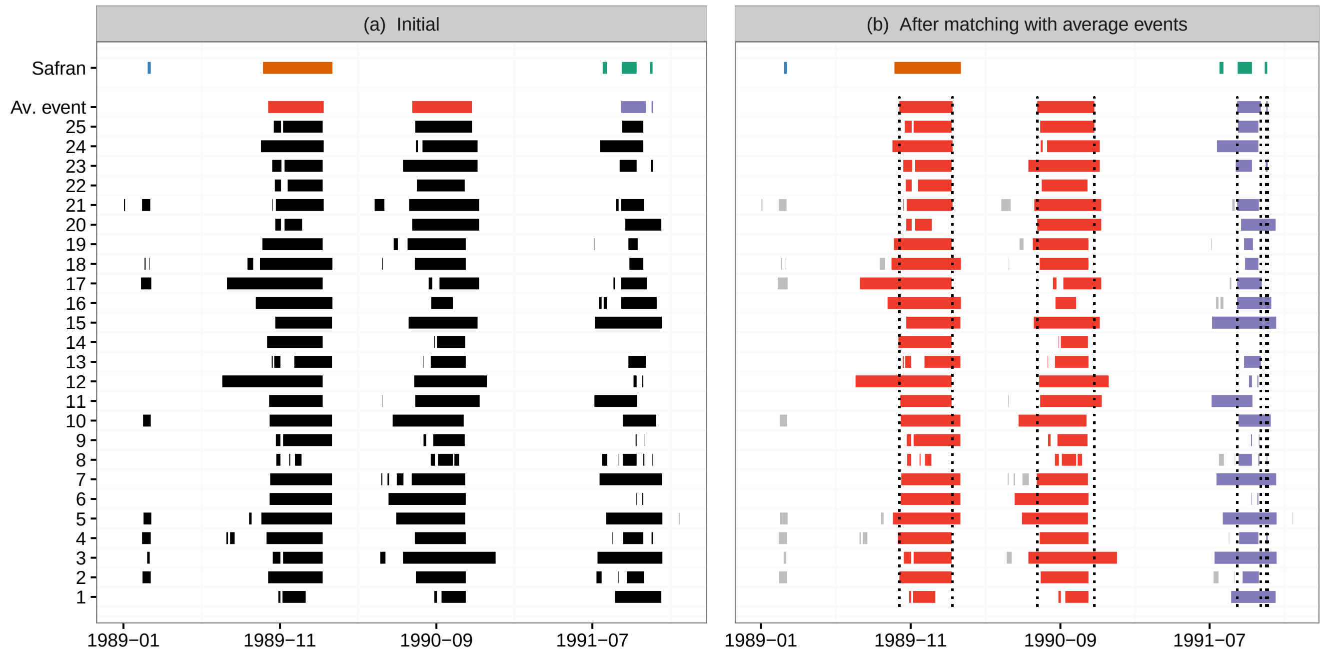


Figure 8. Ensemble spatial matching for Station 1 in HER 11 during the 1989-1991 period. Events from Safran Hydro, average event series and the 25 members of SCOPE Hydro are represented. (a): Each color defines one spatio-temporal event for Safran Hydro and average series. Events from average series and Safran Hydro are independently spatially matched, leading to different colours (not the same datasets). Black bars are independent local events from the 25 members of SCOPE Hydro. (b): Matching between the 25 event members from SCOPE Hydro and average events. Black events not overlapping average events are considered as probabilistic noise and are removed from the dataset. Dashed lines are set at the start and end dates of each average event.

of HER 11 (see Fig. 8). This allows keeping ~~an~~ a homogeneous ensemble definition of events, which is particularly relevant when considering ensemble statistics like the median value of extreme low-flow event characteristics (see Sect. 5).

The spatial matching procedure adapted to SCOPE Hydro finally leads to spatio-temporal extreme low-flow events that are characterized locally for each station by ~~an~~ a 25-member ensemble of start date, end date, duration, and severity.

3.2.4 Conventions for representation purposes

Spatio-temporal events from SCOPE Hydro are assigned a specific *spatial ~~centre~~ center date* for plotting corresponding local summary characteristics along a temporal dimension. For each station where at least one ensemble member detects the ~~events~~event, a local median start date is computed. The spatial start date is then computed as the median ~~dates~~ date over all stations detecting the event. The same procedure is applied for end dates, and the spatial ~~centre~~ center date is taken as the middle date between the spatial start date and the spatial end date. It thus allows plotting characteristics – start and end dates, duration, and severity – of a specific spatio-temporal event reported locally for different stations in a homogeneous way along the temporal axis.

The daily number of stations concerned affected by a given spatio-temporal low-flow event can be computed. An estimate of the spatial extent of this event in terms of percentage of France surface area is computed based on the surface area of each HER and the percentage of stations hit affected within each HER weighted by catchment areas. This estimates estimate is used to compute the maximum spatial extent reached during each spatio-temporal event.

Duration and severity may be expressed in terms of return period in order to compare event characteristics among themselves and across different catchments. As extreme low-flow events are not a regular phenomenon occurring every T year, the approach developed by Shiau and Shen (2001) is used here: a mean interarrival time between two events is considered as the basis for computing return periods. Severity and duration variables can further be equal to zero. This is accounted for in the following way. Let X denote the target random variable and F_X denote its cumulative distribution function. The Then for all $x > 0$, the cdf can be computed as follows using the total probability law:

$$\forall x > 0, F_X(x) = p_0 + G(X) \times (1 - p_0) \quad (3)$$

where p_0 is (the probability that $X = 0$and) is simply estimated as the empirical frequency of zeros in the observed sample, and $G(x)$ is the cdf of the distribution estimated on the non-zero values in the observed sample. A Generalized Pareto Distribution with 3-parameters (GPD3) for the duration and a Pearson type III (P3) distribution for the severity are selected based on a an L-moment diagram (not shown). These distributions are calibrated for each reconstruction ensemble member and each catchment. Only ranges in return periods are considered in order not to put too much confidence in fitted values.

4 SCOPE Hydro: Examples of reconstructed streamflow time series

Corrèze and Ubaye daily discharge time series during the extreme low-flow event of 1972, from observations, Safran Hydro and SCOPE Hydro. Note the different root scales for the y-axis.

Corrèze and Ubaye annual discharge time series during the 1871-2012 period, from observations, Safran Hydro and SCOPE Hydro. Note the different scales for the y-axis.

This section briefly gives some quantitative results on the hydrological model calibration, and then shows SCOPE Hydro reconstructed streamflow time series for the two case-study catchments to briefly illustrate the hydrological inputs used to study extreme low-flow events.

Figure 2 shows the KGE calibration values for all 662 modelled catchments. More than 70% of the catchments have values over 0.9. KGE values are lower for mountainous areas (Massif Central, Alps, Pyrenees) and the minimum value is 0.61. It has to be noted that thorough validation experiments not shown here – out-of-sample experiments, split-sample experiments – have been performed to carefully quantify the overall hydrological modelling performance. Out-of-sample experiments have shown overall good reconstructions, with a slight underestimation of streamflow. Split-sample experiments across the 1973-1989 and 1990-2006 periods have shown a good stability of calibrated parameters when they were validated on independent periods.

Detailed investigations have shown that the decrease in performance identified for a few stations was related to anthropogenic influences.

Figure 9 presents the observed and simulated daily streamflow time series for the Corrèze and Ubaye catchments during the extreme low-flow event of 1972 (see e.g. Duband, 2010; Chauveau et al., 2014). The temporal dynamics at both the seasonal and daily scales are fairly well simulated by SCOPE Hydro for both catchments. SCOPE Hydro but also Safran Hydro tend to underestimate spring streamflow for the Ubaye catchment for these specific years, possibly due to a non-satisfactory modelling of the early snowmelt in the hydrological model (see also Fig. 18). Safran Hydro and observations are is most of the time included in SCOPE Hydro ensemble range, showing a good reasonable reliability of the ensemble reconstructions for this period. This statement is also valid for the observations even if they are more often in periphery of the SCOPE Hydro ensemble range. Additionally, the low-flow periods – Summer 1972 for the Corrèze and Winter 1971-1972 for the Ubaye – are reasonably well simulated.

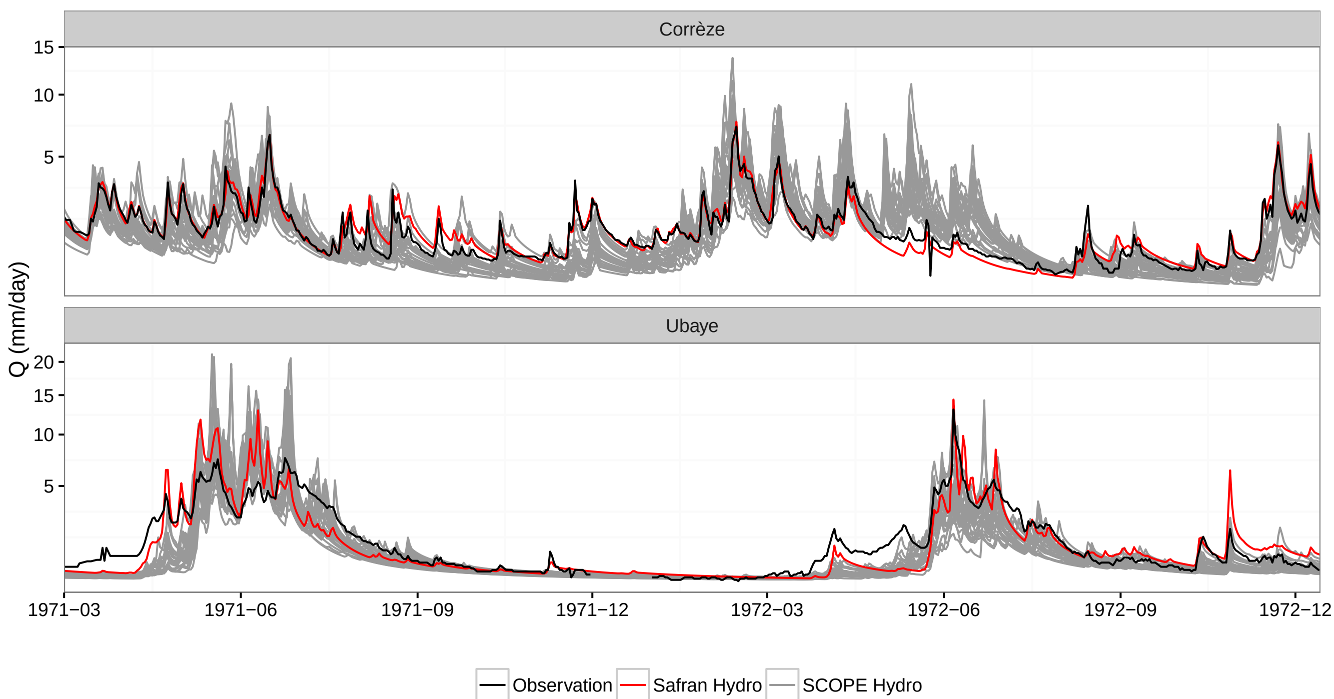


Figure 9. Corrèze and Ubaye daily discharge time series during the extreme low-flow event of 1972, from observations, Safran Hydro and SCOPE Hydro. Note the different root scales for the y-axis.

Figure 10 shows the annual streamflow time series for the two catchments over the whole period. The observed interannual variability is well simulated by both Safran Hydro and SCOPE Hydro. Here again, Safran Hydro and observations are generally included in the SCOPE Hydro ensemble range. SCOPE Hydro allows to fill data gaps for years with missing observations (for example around 1925 for the Corrèze and 1918 for the Ubaye) and to extend our knowledge back until the end of the 19th century.

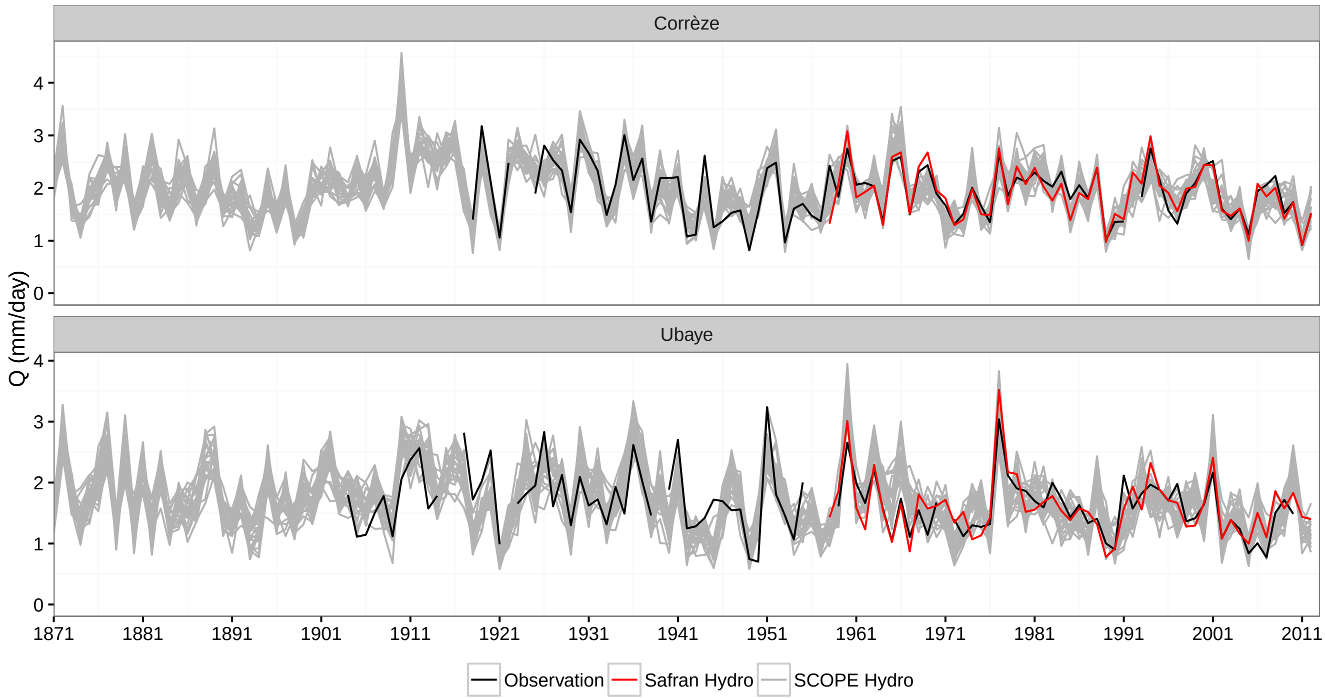


Figure 10. Corrèze and Ubaye annual discharge time series during the 1871-2012 period, from observations, Safran Hydro and SCOPE Hydro. Note the different scales for the y-axis.

5 Extreme low-flow events since 1871

This section provides example results that can be derived from the SCOPE Hydro reconstruction dataset after the definition of extreme low-flow events (Sect. 3.2.1) and the ensemble and spatio-temporal matching (Sect. 3.2.2). The first part focuses on spatio-temporal events examined only at the local scale for the two catchment case studies (Corrèze and Ubaye). The second part then provides summary results on spatio-temporal events drawn at the scale of France.

5.1 Temporal ensemble reconstruction of extreme low-flow events for the Corrèze and Ubaye catchments

~~Extreme low-flow event duration (in days) for the Corrèze and Ubaye catchments over the 1871-2012 period. The x-axis represents the spatial center dates of the events (see Sect. 3.2.4). The 25 ensemble members are represented with a boxplot for each event. The lower and upper hinges correspond to the first and third quarters, respectively. Wiskers extend to 1.5 times the~~

interquartile range. Events with remarkable values from any given ensemble member are labelled with the year of the spatial center date. The dashed line shows the 1-year duration for information.

As for Fig. 11 but for severity (in mm).

Extreme low-flow event start date (Julian days) for the Corrèze and Ubaye catchments over the 1871-2012 period. The x-axis represents the spatial center dates of the events (see Sect. 3.2.4). The dashed line is set on the first of January. Grey points represent the start dates of individual ensemble members detecting the event. Black points represent the average date of the grey points.

Duration of extreme low-flow events versus their severity for the Corrèze and Ubaye catchments over the 1871-2012 period. Grey points represents individual ensemble members. Black points represents the median characteristics of the 25-member ensembles. Events remarkable given their median duration or severity are labelled with the year of their spatial centre dates. Note the square root scale for both axes.

Figure 11 shows the ensemble durations for each spatio-temporal extreme low-flow event as experienced by the Corrèze and Ubaye catchments over the 1871-2012 period. For both catchments some long events occurred at the end of the 19th century – for example 1893 and 1898 for the Corrèze, and 1878, 1888, and 1897 for the Ubaye – followed by a wet period until the beginning of the 1940s, interrupted only by the 1921 event. Several clusters of events can be identified: during the 1940s and 2000s for the Corrèze, and 1940s and 1980s for the Ubaye. Other long events also occurred in-between these clusters like the 1955, 1962 and 1972 events for the Corrèze, and the 1957, 1962 and 1972 events for the Ubaye. Even if long events are for a large part different from one catchment to another, some common ones may be identified in 1945, 1962, 1972 and 1990. The majority of events are not detected by all 25 ensemble members – as shown by the bottom of boxplots reaching 0 days – with the exception of some specific events like the 1990 event for the Corrèze and the 1921 and 1972 events for the Ubaye. This illustrates the large uncertainty in the characterization of events, as each of the 25 ensemble members is a potential reconstruction of the event.

Figure 12 shows in a similar way the severity obtained for each spatio-temporal extreme low-flow event for both catchment case studies. As for the duration, a wet period occurred between 1910 and 1940, here again interrupted by the 1921 event. Some events appear as exceptionally severe like the 1976, 1978 and 1990 events for the Corrèze, as well as the 1921 and 1972 events for the Ubaye. Moreover, these events are detected by all 25 ensemble members.

Based on Fig. 11 and Fig. 12, a higher number of events occurred for the Corrèze catchment than for the Ubaye catchment. A higher number of extreme events and higher severity values are simulated after 1940 for the Corrèze. These reconstructions also identify long and/or severe low-flow events during periods for which no streamflow observation is available, notably the last decades of the 19th century.

Figure 13 presents the start date of each spatio-temporal extreme low-flow event for the Corrèze and Ubaye catchments over the 1871-2012 period. An average date representing the ensemble central tendency is computed using a seasonal index (Burn, 1997). As for Fig. 11 and Fig. 12, there is a high variability among ensemble members, with start dates of individual members lying within a range of +/- 2 to 3 months around the average start date. The seasonality of the start dates illustrates the different regimes in the two catchments. Indeed, extreme low-flow events start between April and November for the oceanic-

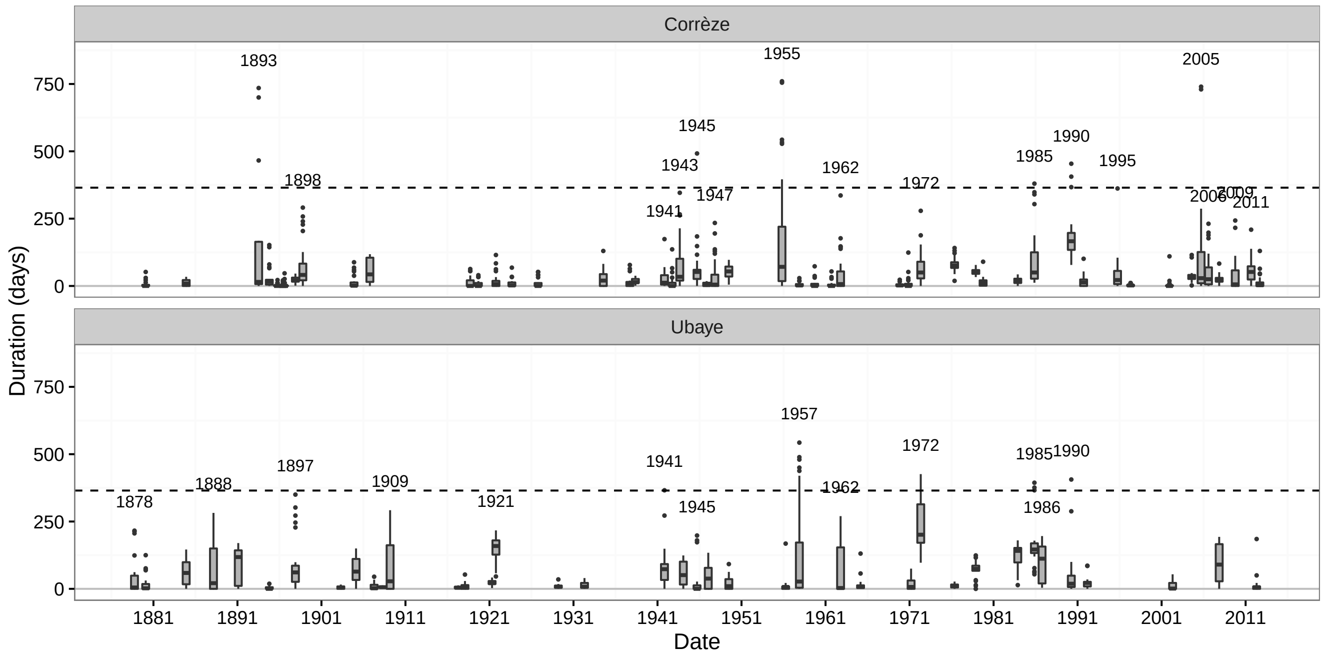


Figure 11. Extreme low-flow event duration (in days) for the Corrèze and Ubaye catchments over the 1871-2012 period. The x-axis represents the spatial center dates of the events (see Sect. 3.2.4). The 25 ensemble members are represented with a boxplot for each event. The lower and upper hinges correspond to the first and third quartiles, respectively. Whiskers extend to the smallest and largest value still within 1.5 times the interquartile range. Events with remarkable values from any given ensemble member are labelled with the year of the spatial center date. The dashed line shows the 1-year duration for information.

dominated regime (Corrèze), and between October to April for the snowmelt-dominated regime (Ubaye). The only exception is the particular event of 1880 which began in December 1879 for the Corrèze river. There is no visible trend on the seasonality of start dates.

~~Figure ?? provides a two-way summary of spatio-temporal extreme~~

~~5.2 Spatio-temporal characterization of extreme low-flow events since 1871~~

~~Benchmark events selected from median characteristics over the 1871-2012 period. Filled white circles identify stations which share their most extreme event with less than 4 other ones. These events are not listed in the legend.~~

~~Median value of return period (in years) for duration (left) and severity (middle), and median of the number of days after the earliest start date (right) for the 1878, 1893, 1976 and 1990 extreme low-flow events. The earliest start date is shown on the top right corner for each event. Filled grey circles represent stations that did not detect the event.~~

5.2 Spatio-temporal characterization of extreme low-flow events since 1871

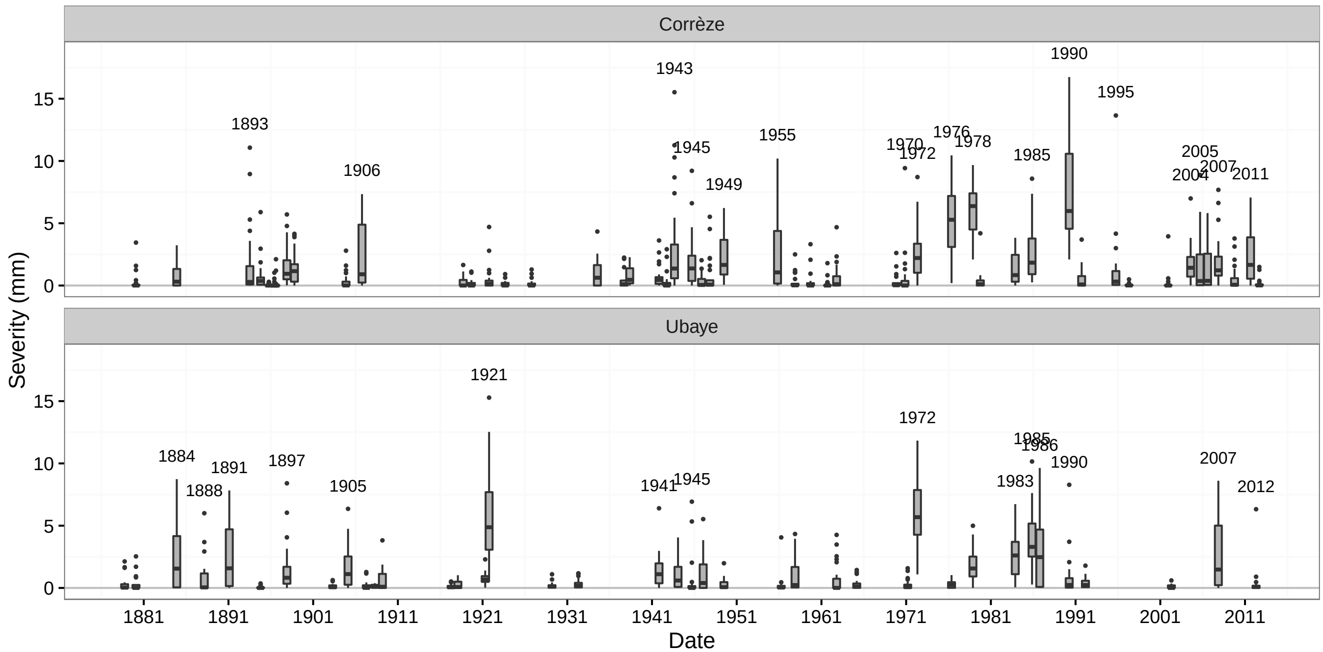


Figure 12. As for Fig. 11 but for severity (in mm).

~~Return period in severity for the 25 ensemble members of the 1893 event. Filled grey circles represents stations that did not detect the event.~~

The probabilistic and spatial matching done in Sect. 3.2.3 ensures the spatial consistency of extreme low-flow events. This critically allows studying these events at the national scale. Unless mentioned otherwise, plotted values in the following figures are median values over the 25 ensemble members.

Figure 14 introduces the spatial extent of the spatio-temporal extreme low-flow events identified over the 1871-2012 period. This spatial extent represents the maximum percentage of France ~~hit~~ affected by the event in terms of surface area, regardless of the intensity or duration of the event. Some events already mentioned in Sect. 5.1 are emphasized such as the 1893, 1921, 1945, 1976 or 1990 events. This last event is the only one having affected almost the whole of France. More generally, this figure highlights the fact that ~~the only events having hit~~ events covering more than 70% of France have only occurred after 1940.

Figure 15 highlights the longest and most severe events that ~~hit~~ affected individual catchments. Two events are highlighted in terms of duration: the 1976 event for the north of France and the 1990 event for the centre of France. The 1972 (resp. 1945) event was the longest one having ~~hit~~ affected the east (resp. south). The pattern is more patchy for the most severe events. The 1976 event is still predominant for the north of France. The ~~1878~~ 1893 event was the most severe in the north-east whereas the ~~1893~~ 1878 event still has the record for some parts of the Mediterranean coastal area. The 1949 and 1990 events share the first place for the centre of France. This 1921 event is highlighted for different stations scattered in France, from the north-

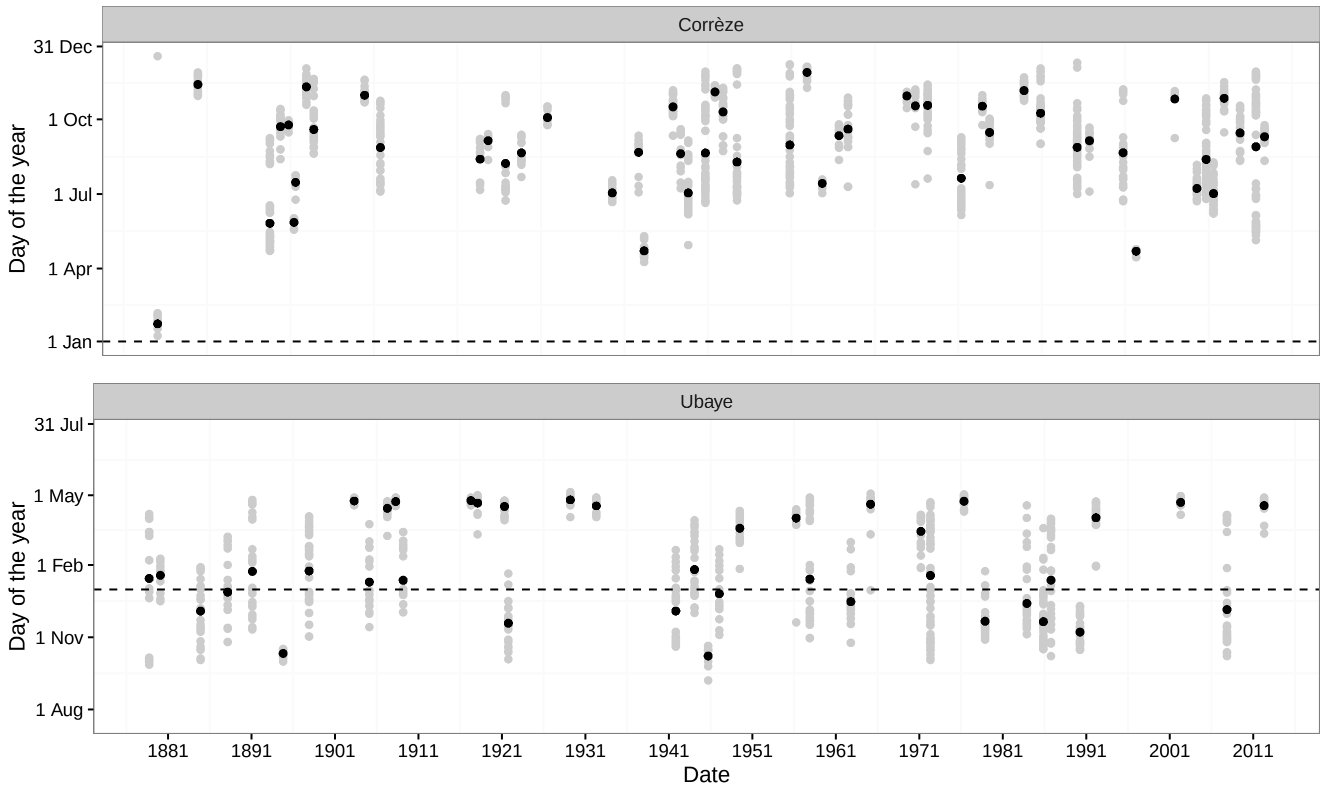


Figure 13. Extreme low-flow ~~events in terms of duration vs severity for both catchments, by plotting characteristics from individual ensemble members, but also as the median duration of the ensemble vs the median severity. These plots first show the unsurprisingly high correlation between these two characteristics. The 1976, 1978 and 1990 events for the~~ event start date (Julian days) for the Corrèze ~~as well as the 1972~~ and ~~1921 events for the Ubaye appear to be exceptional over the~~ Ubaye catchments over the 1871-2012 ~~period. Figure ?? allows identifying reconstructed ensemble members with specific combinations of duration and severity, an information not provided by Fig. 11 and Fig. 12. Some ensemble members generate for example very long events, but without exceptional severity values. Median value of~~ period. The x-axis represents the ~~maximum~~ spatial ~~extent~~ center dates of ~~spatio-temporal~~ the events ~~from individual ensemble members of SCOPE~~(see Sect. ~~Hydro over the 1871-2012 period~~3.2.4). ~~See text for details. X-axis coordinates are~~ The dashed line is set on the ~~median~~ first of January. Grey points represent the start dates ~~when~~ of individual ensemble members detecting the ~~maximum spatial extent is reached~~event. ~~Years and months of~~ Black points represent the ~~median dates are labelled for events with a spatial extent over 60%~~ average date of ~~France~~the grey points.

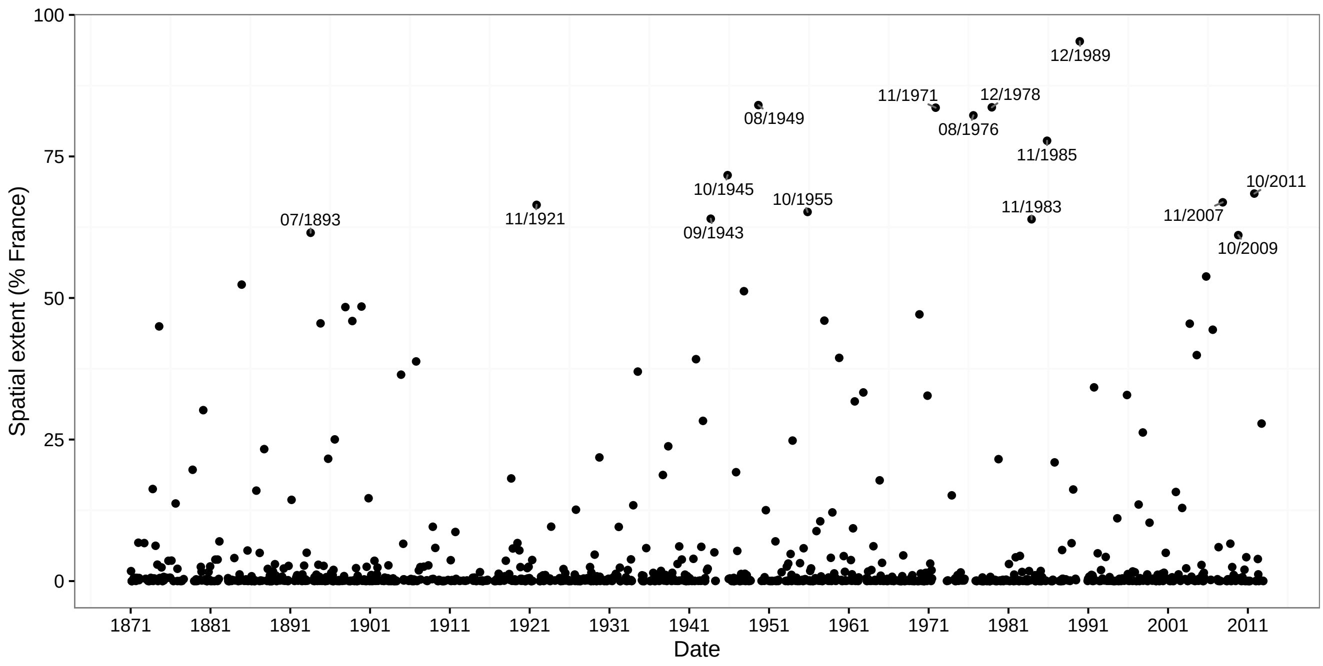


Figure 14. Median value of the maximum spatial extent of spatio-temporal events from individual ensemble members of SCOPE Hydro over the 1871-2012 period. See text for details. X-axis coordinates are the median of the dates when the maximum spatial extent is reached. Years and months of the median dates are labelled for events with a spatial extent over 60% of France.

west to the south-east. It may have been an exceptionally severe event for the entire France even if not the most severe for all stations. Note that the 2003 event, which is often used as a reference at the European scale (see e.g. Laaha et al., 2016) is not highlighted as the most severe event for any station in France. ~~These events~~ After further investigation, the 2003 event seems to be underestimated by SCOPE Hydro in comparison to Safran Hydro or the observations, both in terms of severity and duration (not shown). This may be due to the lack of soil-atmosphere retroactions which are important for this event and are not taken into account in our hydrometeorological reconstruction chain. The events highlighted by Fig. 15 are identified as *Benchmark events* and their characteristics can be drawn for France as a whole.

Four of these benchmark events are further studied in Fig. 16 in terms of median characteristics. The 1878 and 1893 events have been selected because of the lack of related quantitative or even qualitative hydrometric observations available. Information on their meteorological drivers – or their socio-economic impacts – have however been documented elsewhere: see Moureaux (1880b) for the 1878 event, and Vimont (1893), Plumandon (1893), Dehérain (1893) or Cook et al. (2015) for the 1893 event. ~~On the contrary~~In contrast, the 1976 and 1990 droughts have been ~~largely~~ widely studied in terms of low flows in France: see e.g. respectively Bremond (1976) and Mérillon and Chaperon (1990). The first two columns of Fig. 16 show the local return period of duration and severity of these spatio-temporal events. They show that the four selected events had very different spatial patterns. The 1878 event only ~~hit~~ affected the south of France, and particularly the Mediterranean coast. ~~On~~

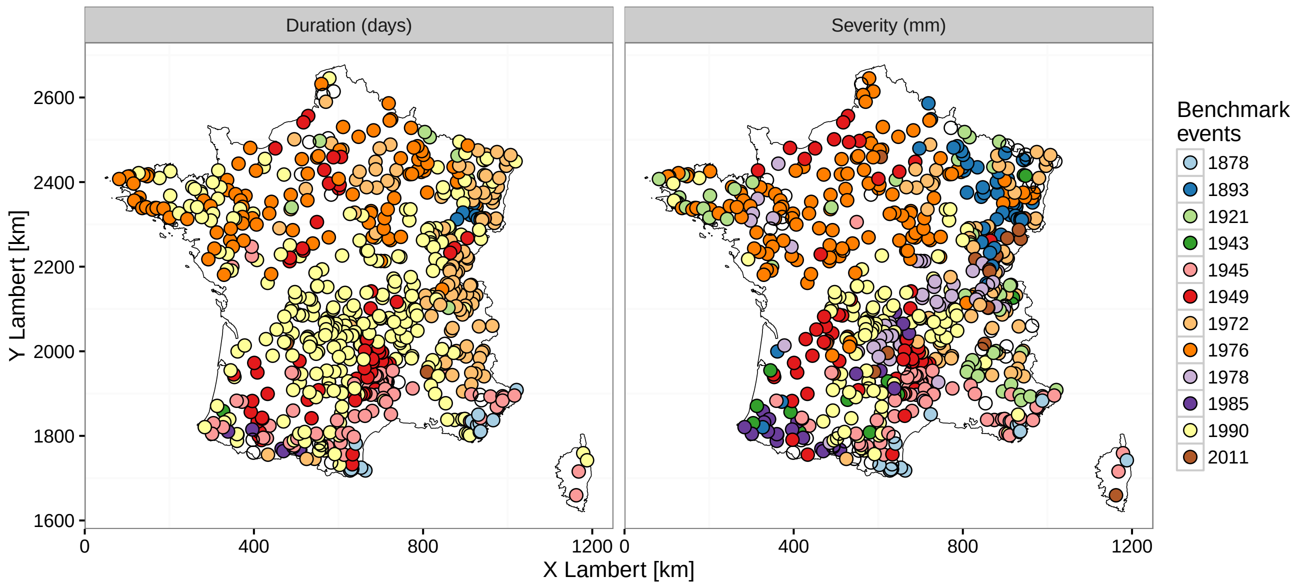


Figure 15. Benchmark events selected from median characteristics over the 1871-2012 period. Filled white circles identify stations which share their most extreme event with less than 4 other ones. These events are not listed in the legend.

~~the contrary~~In contrast, the 1976 only spared this coastal area. The 1893 and 1990 events both ~~hit~~ affected the entire France ~~excepted~~ except the high-elevation snow-influenced alpine stations – as well as some lower elevation stations in the south for the 1893 event. Return periods for duration and severity highlight regions with the most extreme values reached during each event: north-east for the 1893 event, and northern half of the country for the 1976 event. The 1990 event ~~on the contrary~~ was exceptionally long over most of France, as already suggested in Fig. 14. Differences between the ~~two return periods~~ return periods in duration and severity may also be spotted: the 1893 and 1976 events were more outstanding in terms of severity – with return periods higher than 50 years for many stations – than duration, whereas the 1990 event was exceptional in terms of duration, with return periods higher than 20 years or even 50 years for the majority of stations. The third column presents the start date for each of ~~the four extreme low-flow~~ these benchmark events. Both the 1893 and 1990 events' start dates lies within a common 6-month period for ~~all stationsconcerned. On the contrary~~the vast majority of stations. In contrast, two groups of stations are distinguished for the 1878 and 1976 events, with one of them displaying start dates occurring more than a year later than the first reconstructed one. This may be a consequence of choices made for the spatial matching and will be further discussed in Sect. 6.2.

All above results are drawn from median characteristics of the 25-member ensemble reconstruction. Generally, the more severe or the longer the event, the higher the number of members detecting the event. To give an idea of the variability of spatio-temporal characteristics within this ensemble, Fig. 17 presents the 25 individual reconstructions of the 1893 event in

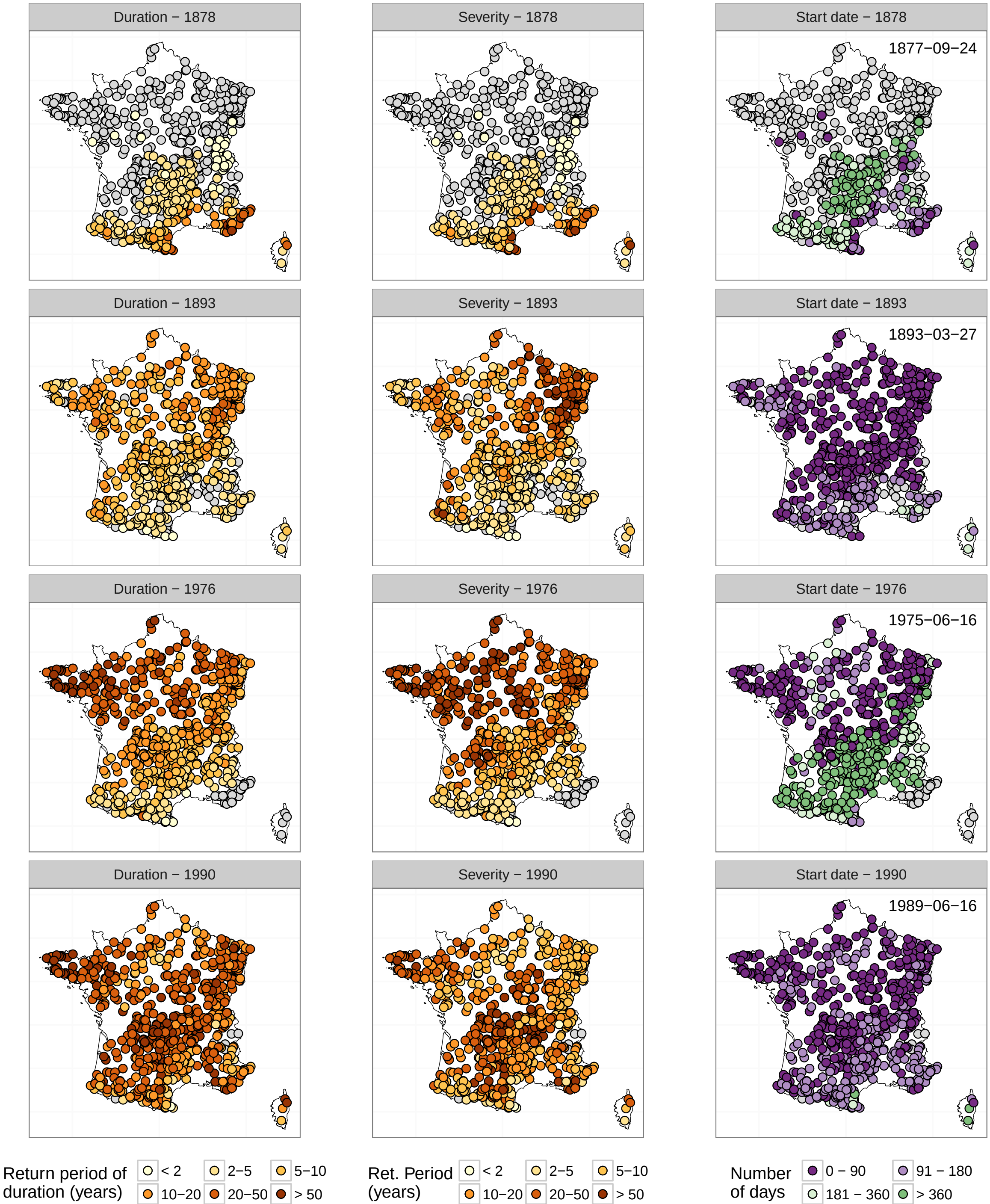


Figure 16. Median value of return period (in years) for duration (left) and severity (middle), as well as median start date, as the number of days after the earliest start date (right) for the 1878, 1893, 1976 and 1990 extreme low-flow events. The earliest start date is shown on the top right corner for each event. Filled grey circles represent stations that did not detect the event.

terms of severity. Even if most of the members display an intense drought in North-Eastern France, some of them also simulate an extreme event in the centre (e.g. member 11 and 24), the west (e.g. member 1 and 21) or the south-west (e.g. member 4 and 8). All members however indicate that the Mediterranean coast was not severely ~~hit~~ affected by this event. This high variability among the ensemble members has already been noticed for the meteorological reconstructions (Caillouet et al., 2016). The 25 maps in Fig. 17 represent 25 plausible reconstructions of the 1893 extreme low-flow event that may have occurred under the synoptic conditions from the 20CR reanalysis for this period. This aspect will be further discussed in Sect. 6.5.

6 Discussion

This section discusses some issues previously broached. The sensitivity of the definition of local extreme low-flow events through the Sequent Peak Algorithm to the threshold is presented in Sect. 6.1. Different issues raised by the spatial matching and the ensemble matching are then discussed in Sect. 6.2 and Sect. 6.3, respectively. Section 6.4 touches upon issues that may arise when comparing spatio-temporal events from two datasets. Some limitations of summary statistics derived from the ensemble values is discussed in Sect. 6.5. Lastly, Sect. 6.6 discusses additional sources of uncertainties not considered in this study. Uncertainties derived from 20CR have already been discussed in Caillouet et al. (2016) and will not be further developed here.

6.1 Threshold sensitivity for the local definition of events

This study is based on a quantile threshold defined in Sect. 3.2.1 for identifying local extreme low-flow events. Changing the threshold – even within the 70-95th percentile range commonly used for defining low-flow thresholds (Hisdal et al., 2001; Fleig et al., 2006; Van Loon and Van Lanen, 2012) – would lead to a different definition of events and different results in Sect. 5. A higher threshold value, e.g. based on the 80 percent exceedance probability (Q80), would for example lead to a higher number of events, but also to a new definition of the events studied here. A sensitivity analysis showed for example that the 1921 event would be pooled with an event occurring in 1922 with a Q80 threshold for the Corrèze river, which is not the case with the chosen Q90. The choice of the threshold also somewhat conditions the appropriate spatial domain for the spatial matching applied in Sect. 3.2.2. This is further discussed ~~in Sect. 6.2.2.~~ below.

6.2 Spatial matching sensitivity

Section 3.2.2 introduced the method developed to spatially match local events, in the deterministic case. This section aims at summarising the sensitivity of this matching process to different parameters.

6.2.1 Sensitivity to the spatial domain

The spatial matching is done considering a specific set of gauging stations as spatial domain. Different spatial domains (for example entire France, or by HER, or by a specific set of stations) would lead to different spatio-temporal event definitions. Consequently, this matching should be preferably done on a consistent set of stations along time to have ~~an~~ a homogeneous

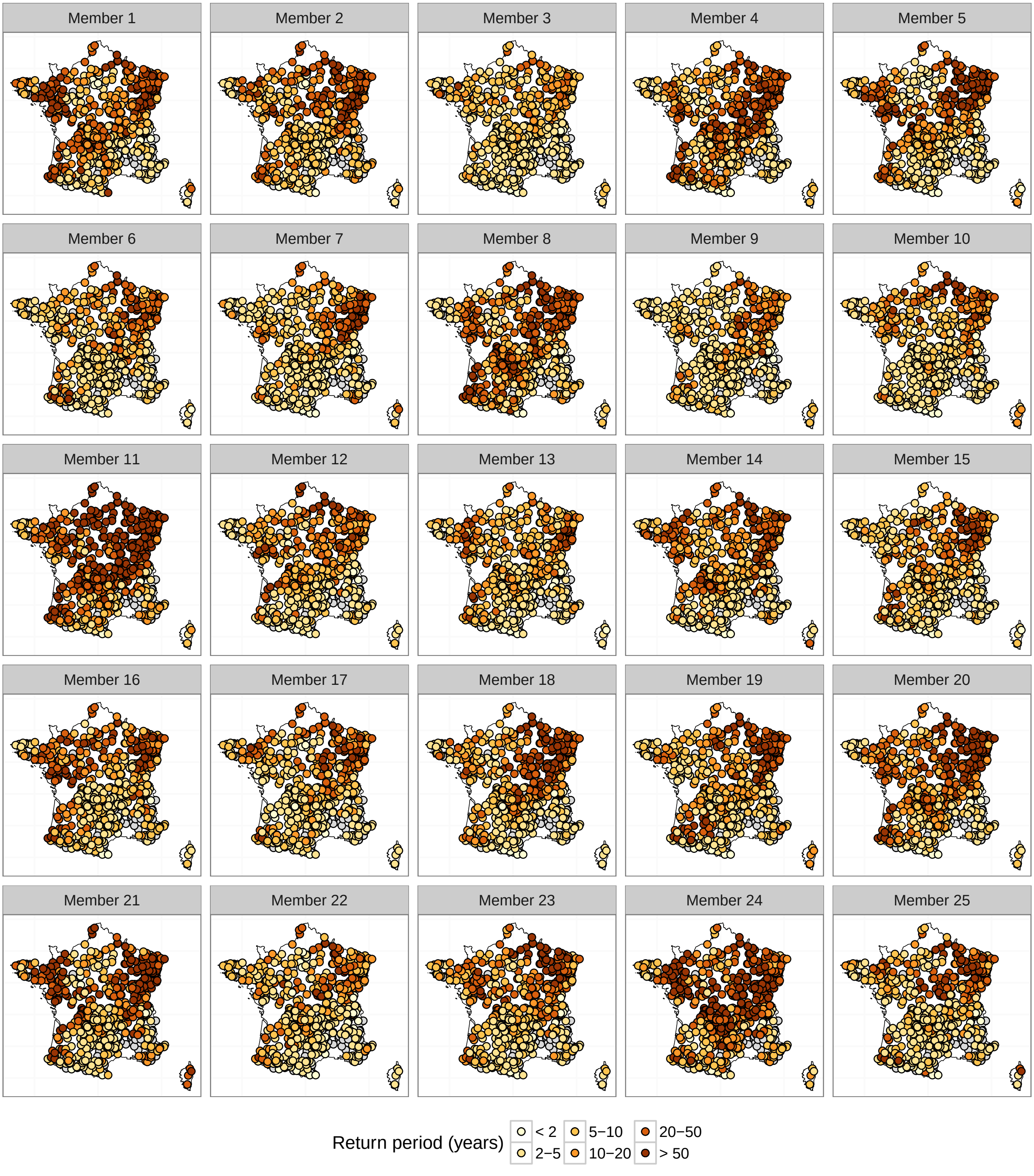


Figure 17. Return period in severity for the 25 ensemble members of the 1893 event. Filled grey circles represents stations that did not detect the event.

definition of events. This is the reason why the spatial matching had not been applied to streamflow observations that have an evolving network – less than 30 stations before the 1910s and more than 600 stations after the 1970s – but also missing data.

6.2.2 Sensitivity to the threshold defining local events

The spatial matching procedure tends naturally to aggregate in time consecutive and overlapping local low-flow events from different stations. Moreover, the higher the threshold value – in mm or m^3/s – used for defining local events, the higher the number of such events (See Sect. 6.1 above). Consequently, the higher the threshold value~~in mm or m^3/s~~, the greater the risk to aggregate potentially independent spatio-temporal events.

6.2.3 Consequences ~~on~~ of spatial matching limitations

Figure 16 (right) shows that the choices made here – Q90 threshold + two-step (HERs then France) spatial matching – might lead to too much spatio-temporal aggregation for specific events like 1878 and 1976. For these two events, a possible overconfident aggregation might indeed reveals itself through the differentiation of stations in two groups by their start dates.

When looking more into details in the reconstructed 1976 event, it appears that it had two main components – one in summer 1975 and one in summer 1976 – that were linked together throughout winter 1975/1976 for stations mainly located in the northern half of the country (not shown). This very dry winter in Northern France has been ~~largely~~ extensively documented (Brochet, 1977; Vidal et al., 2010b), and has led to extreme winter and spring low-flows in snow-dominated catchments – visible as pale green filled circles in Fig. 16 (third row, right) – notably in the Alps (Vivian, 1977). This also prevented severely ~~hit~~ affected northern catchments from late summer 1975 ~~on~~ onwards to fully recover during the regular recharge season and thus to experience an extreme low-flow event spanning more than a year (Bremond, 1976; Gazelle, 1977). This analysis also helps interpreting the north-south difference in duration and severity shown in Fig. 16. It thus validates the spatio-temporal aggregation of this event found here with the specific choices of threshold and spatial domain(s).

The 1878 event also has two temporal peaks, but available references and hydrometric data related to this specific event are much too scarce to finely assess the spatio-temporal aggregation from SCOPE Hydro. Precipitation observation summaries however suggest that the second half-year of 1877 was very dry all around the Mediterranean and especially in the Eastern Pyrenees (Moureaux, 1880a, p. 22). This region remained relatively dry throughout 1878, and especially during February (Moureaux, 1880b, p. 10). The month of September 1878 was then exceptionally dry over a large part of France, including not only the Mediterrean area, but also the Pyrenees and the Massif Central (Mascart, 1880, p. 33), leading to a second component of the spatio-temporal extreme low-flow event.

Beyond these two specific cases, it has to be noted that both the 1893 and the 1990 events appear quite satisfactorily defined in space and time by the spatial matching procedure proposed here (see Fig. 16, rows 2 and 4.).

6.3 Ensemble matching sensitivity

Section 3.2.3 introduced *average events* as a way of summarizing local events from all ensemble members in a single deterministic way for applying the spatial matching procedure. Events from individual ensemble members that do not overlap temporally with average events are removed from the final dataset.

In Fig. 8, very short events detected by only a few ensemble members during May 1990 for Station 1 in HER 11 (shown in grey) are for example adequately removed in the final spatio-temporal event definition. However, events detected by 8 ensemble members before July 1989 are also removed – the threshold for deriving an average event being set at 10 for this station – while being detected in the Safran Hydro dataset (in blue). Even if the threshold definition is calibrated against Safran Hydro summary statistics (total duration of events during a given period), it may thus introduce some limited discrepancies between the two datasets. It has to be noted that this particular discrepancy is only local: Fig. 5(e) shows that an event occurred in early 1989 for 12 out of 16 stations in HER 11 in the Safran Hydro dataset, and an event concurrently occurred in 8 stations for average events in the SCOPE Hydro dataset (not shown). The two datasets therefore agree well in the occurrence of an event during this period over a large part of this HER, even if differences may occur locally due to the local sensitivity of ensemble matching choices.

Fig. 8(b) shows another case of local discrepancy between the two datasets for Station 1 in HER 11: the late 1990 event is clearly detected locally by SCOPE Hydro ~~– by 23 out of 25 ensemble members –~~ (in red) but not by Safran Hydro (in orange). However, Fig. 5(e) shows that Station 1 is the only one (out of 16) in HER 11 where no event is detected during this period in the Safran Hydro dataset. Moreover, 15 stations in this HER are ~~hit~~ affected by an event around this period in the SCOPE Hydro dataset (not shown). This shows once again that the two datasets agree well when considering the regional occurrence of low-flow events.

Figure 8(b) also features a favourable case where the late 1991 event (shown in ~~green~~mauve) is detected by 23 SCOPE Hydro ensemble members as well as by Safran Hydro (in green). The cases discussed above thus perfectly illustrate the compromise reached here to derive a deterministic series of events from an ensemble of series and to then apply the spatial matching procedure.

6.4 Comparison of spatio-temporal events from two different datasets

The previous section briefly touched upon a comparison of extreme low-flow events derived from the Safran Hydro dataset and the SCOPE Hydro dataset, at the local scale, but also ~~in a~~ from a larger spatio-temporal ~~view~~perspective. One may then want to ~~draw~~ make a more formal comparison of spatio-temporal events between the two ~~data sets~~ datasets (and possibly with a spatial matching of ~~event~~ events derived from observations, but see comments in Sect. 6.2.1). This section discusses further the limitations for automating such comparisons.

The simplest approach would consist ~~in~~ of identifying locally previously defined spatio-temporal events through a date overlapping procedure. However, this can lead to ~~identify~~ the identification of multiple spatio-temporal events from one dataset, but to only one event from the second dataset, calling into question the initial spatial matching done (independently) for each dataset. Moreover, this approach would identify events only locally and through a temporal dimension, not considering the spatial extent of the event at the scale of France, which would be too restrictive in practice as shown by the above examples. One response would be to formally identify a spatio-temporal event from one dataset to a spatio-temporal event from another dataset. Some further technical developments would thus have to be set up and tested for comparing any two spatio-temporal extents of extreme low-flow events in both ~~a~~ temporal and spatial ~~dimension~~dimensions.

When some temporal overlapping rules may rather easily be set up at the local scale, corresponding rules for allowance in spatial extent would be much harder to formalise due to the irregular location of hydrometric stations (and upstream catchments). One way forward would be to start the identification based on HER-representative events, which carry less spatio-temporal variability, and report this identification to the local scale. Lastly, one would also need to deal with the ensemble property of the SCOPE Hydro dataset. Again, average events as defined in Sect. 3.2.3 may be ~~the right~~ a promising way forward. Nevertheless, wrongly identifying multiple spatio-temporal events from one dataset to only one event from the second dataset is still a possibility.

6.5 Use of median event characteristics

Figures 14, 15 and 16 display median characteristics – duration and severity – in order to show summary results at the scale of France. Nevertheless, Fig. 17 showed a strong variability between events from each individual ensemble member. Taking maximum characteristics instead of median characteristics would lead to very different results in Sect. 5. ~~This is illustrated in Fig. **??** by grey and black points, that represent characteristics from individual ensemble members and median characteristics, respectively. Extreme grey points do not necessarily correspond to stand-out median event characteristics. For example, the grey point plotted for the Ubaye river with a duration of approximately 500 days and a severity of approximately 3.5 mm corresponds to the 1958 spatio-temporal event, which is not labelled as an extreme event based on its median characteristics.~~ Benchmark events shown in Fig. 15 could alternatively be determined based on these maximum characteristics and used for adaptation purpose scenarios. The ensemble characteristics of each spatio-temporal event should therefore be kept as far as possible along the analysis. Indeed, this ensemble reconstruction provides 25 plausible spatio-temporal events corresponding to specific synoptic conditions as given by the 20CR reanalysis. This can be used to better understand the relationship between meteorological conditions and extreme low-flow events.

6.6 Uncertainties in reconstructed streamflow

The above section dealt with the uncertainty related to the statistical downscaling step in the SCOPE approach. Several other sources of uncertainty in streamflow reconstructions are not considered in this work. First, all results presented here are conditional on the large-scale information provided by the 20CR atmospheric reanalysis, and using an alternative extended reanalysis like ERA-20C (Poli et al., 2016) may lead to different outputs. Second, the hydrometeorological modelling chain used here for deriving the SCOPE Hydro dataset – SCOPE and ~~GR6J-Cemaneige~~ GR6J+CemaNeige – is only one of the many choices for reconstructing high-resolution meteorological fields and catchment-scale streamflow. In order to asses the uncertainty related to this choice, a follow-up study will compare the SCOPE-Hydro dataset (1) to 20CR-driven reconstructions made by Dayon et al. (2015) with a different downscaling method and the physically-based Isba-Modcou hydrological chain, and (2) to reconstructions mixing their downscaling method and the GR6J+~~Cemaneige~~ CemaNeige hydrological model. Lastly, both above mentioned reconstructions assume a constant land cover and land use, while the wooded area for example is known to have strongly ~~evolved~~ expanded in France since the 19th century (see e.g. Koerner et al., 2000). Model parameter uncertainty could also be assessed by modifying the optimisation process for the calibration step.

7 Conclusions

This paper describes an ensemble reconstruction of spatio-temporal extreme low-flow events since 1871 over 662 near-natural catchments in France based on reconstructed climate and streamflow from the SCOPE Climate (Spatially COherent Probabilistic Extended Climate) and SCOPE hydro (Spatially COherent Probabilistic Extended Hydrological) datasets. SCOPE Climate builds on a probabilistic downscaling of the 20CR reanalysis over France as described by Caillouet et al. (2016). This ensemble high-resolution daily meteorological dataset is used as forcings for the GR6J hydrological model – together with the ~~Cemaneige~~ CemaNeige snow-accounting model – to derive the SCOPE Hydro 25-member ensemble daily reconstructed streamflow dataset over a large set of catchments. These two consistent datasets may thus be used for various hydrometerological studies~~–~~, and not only for low flows~~–~~, by extending the limited historical ~~depth~~ amount of surface meteorological and streamflow observations currently available in databases (see examples in Fig. 9 and Fig. 10 for streamflow and Caillouet et al. (2016) for precipitation and temperature).

This work proposes an innovative analysis of extreme low-flow events from the SCOPE Hydro dataset based on two distinct features. First, extreme low-flow events are defined locally using the Sequent Peak Algorithm with a combination of a constant threshold and a daily variable threshold. This allows retaining periods with both absolute low-flow values and values significantly lower than the average seasonal cycle, therefore combining the advantages of the two types of thresholds. Second, a spatial matching procedure is developed to identify events both in time and ~~across a set of stations~~space, thus extending previous approaches that could only be applied to data continuous in space like gridded meteorological drought indicators. This spatial matching procedure is then adapted to the ensemble case as provided by SCOPE Hydro.

The above low-flow analysis features are then applied to the SCOPE Hydro dataset to derive and intercompare characteristics of individual spatio-temporal extreme low-flow events at the country scale. For the first time, these events are qualified, quantified and compared in a spatially and temporally homogeneous way over 140 years on a large set of catchments. Results bring forward well-known recent events like 1976 or 1989-1990, but also older and relatively forgotten ones like the 1878 and 1893 events. These results contribute to ~~improve our knowledge on~~ improving our knowledge of historical events by taking advantage of the ~~higher historical depth of~~ more abundant historical upper-air atmospheric data compared to hydrometric data, and to derive benchmarks events over France. Results also highlight that ~~the worst events in terms of local severity and duration, or spatial extent often belong to the pre-1950 period~~benchmark events may be found as early a the 19th century for specific regions, and often before the second part of the 20th century. This strongly suggests to reconsider the common use of only very recent events like 2003 as ~~references~~ reference for building worst-case scenarios for climate change adaptation strategies.

Such worst-case scenarios could instead be derived from historical benchmark events, and more specifically from events in any individual ensemble member in the SCOPE Hydro reconstructions, taken as a plausible hydrological consequence of specific large-scale atmospheric situations. They may also help to put recent events and their socio-economic impacts like the 2015 event into a ~~deeper~~ more comprehensive historical perspective, without resorting only to meteorological drought indicators as ~~proxys~~ proxies for historical low-flows (Van Lanen et al., 2016). This study moreover allows for further detailed analyses of the effect of climate variability and anthropogenic climate change on low-flow hydrology at the scale of France over the last 140 years, by for example extending recent works on the influence of ~~decadal~~ multidecadal variability on trends (Hannaford et al., 2013).

Author contribution

All co-authors collectively designed the experiments. L. Caillouet and A. Devers developed the model code and L. Caillouet performed the simulations. L. Caillouet prepared the manuscript with contributions from all co-authors.

Appendix A: Observed and simulated interannual regimes for the two example catchments

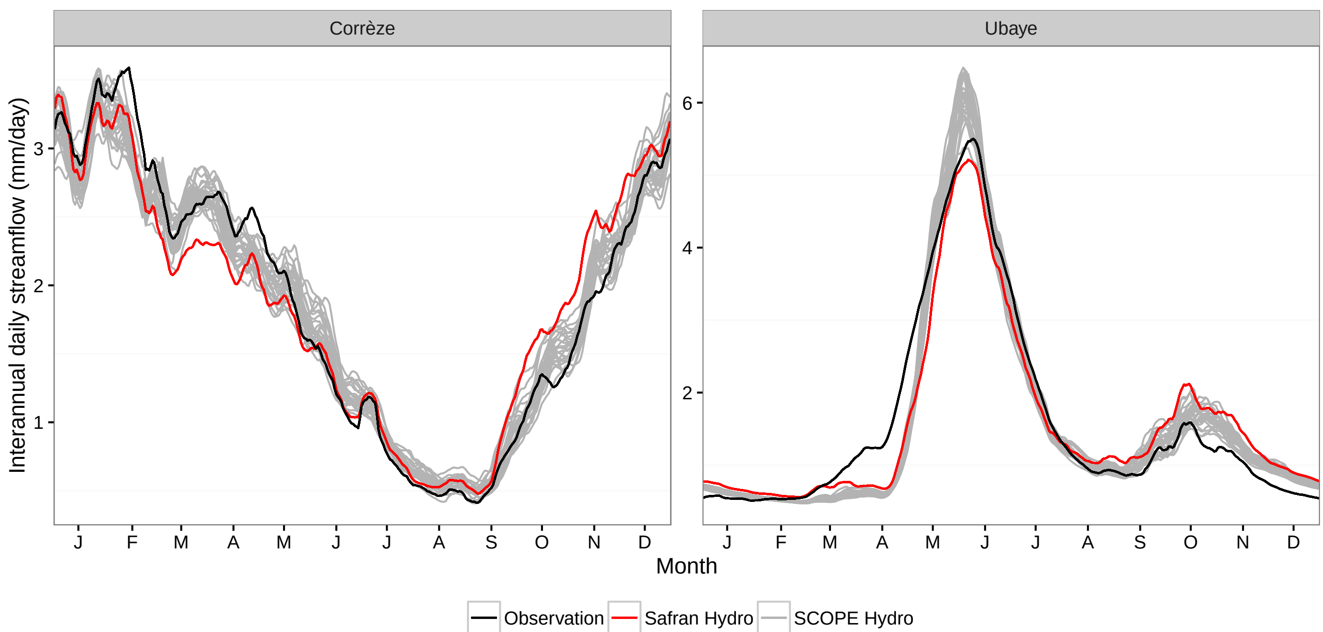


Figure 18. Daily interannual average streamflow – from observations, Safran Hydro and SCOPE Hydro (see Sect. 3.1) – over the CAL period for the Corrèze and Ubaye rivers, smoothed by a 10-day moving average.

The interannual regime of the two catchment case-studies are presented in Fig. 18. The Corrèze catchment with an oceanic-dominated regime has a summer low-flow period. The Ubaye catchment with a snowmelt-dominated regime has a main low-flow period in winter and a second one at the end of the summer. The simulated seasonal cycles ~~match generally well~~ generally match the observed ones well, with some discrepancies for the Ubaye catchment during the snowmelt period, probably due to remaining upstream reservoir operation management.

Appendix B: ~~Synthetic diagram for the~~ The SCOPE method

~~Synthetic diagram showing the sequence of steps for the SCOPE method and its use for (20CR-driven) SCOPE Climate reconstructions. See Sect. 2.2.2 for details.~~

The SCOPE method ~~summarised by Fig. 19 includes the SANDHY-SUB method developed by Caillouet et al. (2016), the~~ (Spatially COherent Probabilistic extension method) is a statistical downscaling tool based on the SANDHY method (Stepwise Analogue Downscaling Method for HYdrology, Ben Daoud et al., 2011; Radanovics et al., 2013; Ben Daoud et al., 2016), improved by different steps.

The SANDHY method is a probabilistic statistical downscaling method following an analogue approach based on the idea introduced by Lorenz (1969) that similar atmospheric situations lead to similar local effects. The analogue approach uses two concurrent data sets over an archive period, from a large-scale reanalysis and a local-scale meteorological data set. Large-scale predictors (like atmospheric circulation patterns) from any date in the target period are compared to those of the archive period and dates with the most similar predictors are chosen as analogues. Local-scale variables (or predictands) like precipitation or temperature from the analogue dates are taken as plausible values for the target date. The SANDHY method is described by four analogy levels optimized by Ben Daoud et al. (2011, 2016) on the Seine and Saône basins. The predictand considered is daily precipitation, as the initial aim of this method was quantitative precipitation forecast. Large-scale 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. The spatial domain where the analogy is sought has been optimized by Radanovics et al. (2013) on 608 zones covering France for the analogy level on geopotential.

The application of SANDHY using predictors from 20CR and predictand from Safran provides an ensemble of 125 analogue dates for each day during the period 1 January 1871 to 29 December 2012, independently for 608 climatically homogeneous zones covering France. Analogues dates from the period 1 August 1958 to 31 July 2008 are converted to meteorological variables by resampling Safran reanalysis data. The analogues obtained from SANDHY are also used for temperature and reference evapotranspiration even if the predictors have been chosen for their link to precipitation. SANDHY has been improved by two new analogy steps by Caillouet et al. (2016). A first level on the large-scale sea surface temperature (or SST) selects 80 analogues from the 125 initial analogues derived from SANDHY. It allows to improve the seasonal cycle of precipitation. A second level on the large-scale 2-meter temperature selects 25 analogues from the 80 previous analogues. It allows to reduce the bias in temperature and to improve its seasonal variability. The dataset obtained with the improvement of

SANDHY by two additional analogy levels – called 20CR-SANDHY-SUB – is further described and validated by Caillouet et al. (2016).

The median of annual precipitation between Safran and 20CR-SANDHY-SUB for the 1959-2007 period shows a dry bias of around 10% on average over France (see top right panel of Fig. 7 in Caillouet et al., 2016). This bias, due to too many dry days resampled during the analogue selection in comparison to Safran, may call into question any interpretation of derived hydrological features – and especially extreme low flows –, and a bias correction step ~~,~~ had to be set up. In a similar context, Timbal et al. (2006) introduced a correction factor to adjust the reconstructed rainfall series but this technique – as well as common bias correction techniques – does not allow to retain the physical consistency and multivariate correlation structure inherent in the analogue approach. A resampling-based correction approach similar to the one adopted by Sippel et al. (2016) is therefore considered here: for each day between 1871 and ~~the Schaake Shuffle step.~~ 2012, the N analogues giving the lowest precipitation are removed. N analogues are then randomly resampled among the (25-N) left to keep a 25-member sample size. N is independently defined for each of the 608 zones in France. It is chosen between 0 and 3, so that the bias with respect to Safran data is minimized. By construction, this number increases with the precipitation underestimation. Importantly, this resampling-based correction for precipitation does not affect the temperature bias and interannual correlation described in Caillouet et al. (2016) (not shown). This technique allows to retrieve a bias in precipitation around 0%. Considering only summer, autumn, and winter, the mean bias over France is decreased in comparison to 20CR-SANDHY-SUB as most of the zones were affected by a negative bias. Only the spring season – where the precipitation was sometimes overestimated in 20CR-SANDHY-SUB – is affected by the same bias in absolute terms but with an opposite sign.

In order to build gridded time series from the 20CR-SANDHY-SUB dataset, Caillouet et al. (2016) independently combined analogues dates from one zone to another. The resulting lack of spatial continuity (see the discussion by Caillouet et al., 2016) could be heavily detrimental to the identification of spatio-temporal low-flow events. This issue is here addressed through the Schaake Shuffle procedure, initially developed to reconstruct space-time variability in forecast meteorological fields (Clark et al., 2004). This procedure has been widely used as a post-processing of ensemble meteorological forecast fields for streamflow forecasting (see e.g. Robertson et al., 2013; Verkade et al., 2013; Demargne et al., 2014; Šípek and Daňhelka, 2015). Vrac and Friederichs (2015) also adapted it recently for multivariate bias correction of downscaled climate simulations. In the Schaake Shuffle approach, which can be seen as an empirical copula on rank correlation (Wilks, 2014), the ensemble members are reordered so that their rank correlations across both space and variables match the ones from a randomly-picked sample of observed multivariate fields. In the present application, rank correlations are considered across the 608 climatically homogeneous zones and across the three variables (precipitation, temperature, and reference evapotranspiration), and observed fields are taken from the Safran reanalysis. For each target date, 25 dates are randomly selected within a 120-day window around the corresponding Julian day and from the period 1 August 1958 to 31 July 2008, a period consistent with the archive period for analogue dates in the SANDHY downscaling step (Caillouet et al., 2016). Observed rank correlations are derived from the Safran multivariate meteorological fields from these dates and applied to the reconstructed ensemble, thus ensuring a spatial and inter-variable coherence of any single ensemble member. As the set of analogues remains the same for each day ans is only shuffled across ensemble members, but with different positions, local characteristics such as bias or correlation do not change.

The succession of steps (1) SANDHY (Radanovics et al., 2013; Ben Daoud et al., 2016), (2) subselection with additional analogue levels (Caillouet et al., 2016), (3) Bias Correction (see above), and (4) Schaake Shuffle (see above) is called SCOPE and is summarised by Fig. 19. The use of 20CR as the large-scale reanalysis input provides the SCOPE Climate dataset, described in Sect. 2.2.2.

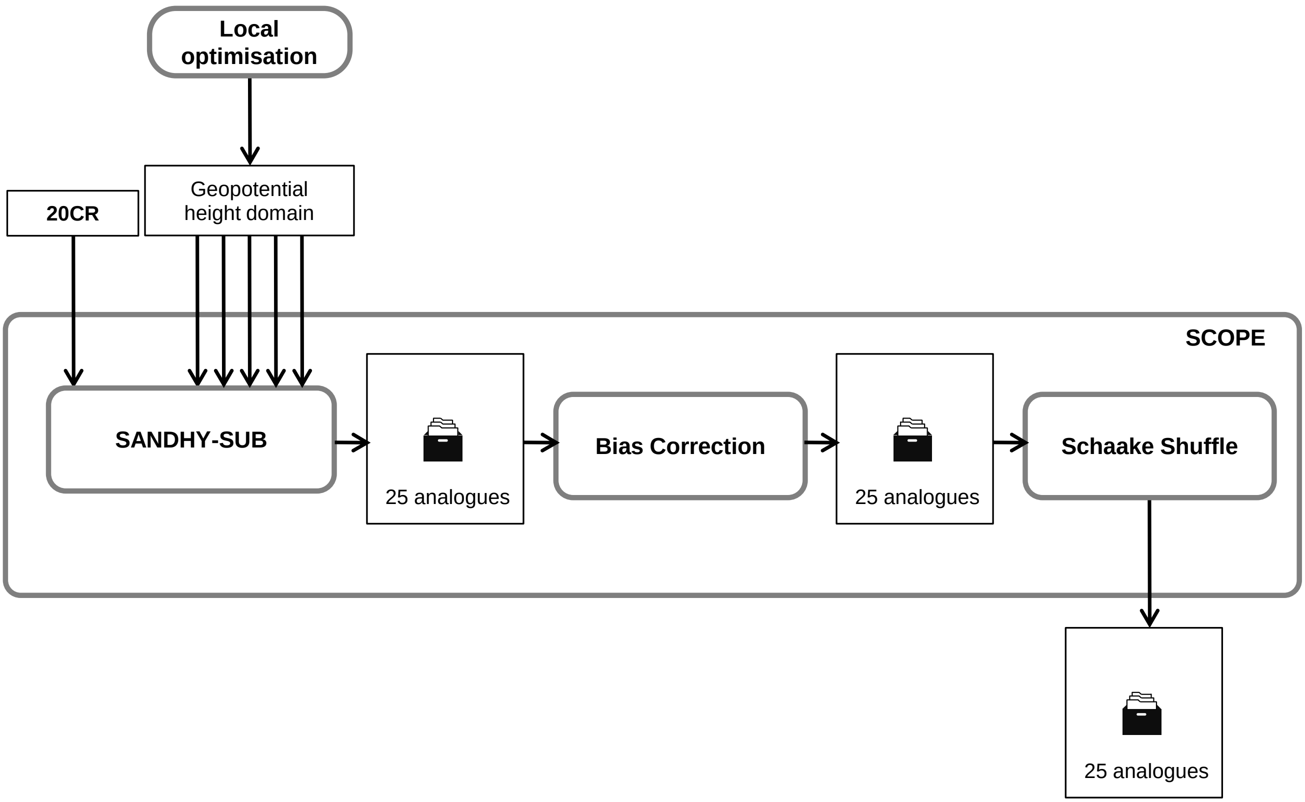


Figure 19. Synthetic diagram showing the sequence of steps for the SCOPE method and its use for (20CR-driven) SCOPE Climate reconstructions. See Sect. 2.2.2 for details.

Appendix C: Hydro-Ecoregions

Figure 20 shows the 22 HERs used for in the spatial matching procedure. HER 1 (“Pyrénées” and HER 2 (“Alpes internes”) include high mountainous areas whereas HER 3 to HER 5 (“Massif Central sud”, “Vosges”, “Jura-Préalpes du Nord”) include mountainous areas with lower elevations. HER 6 to HER 8 (“Méditerranéen”, “Préalpes du Sud”, “Cévennes”) include Mediterranean areas where convective precipitation events often occur in Autumn.

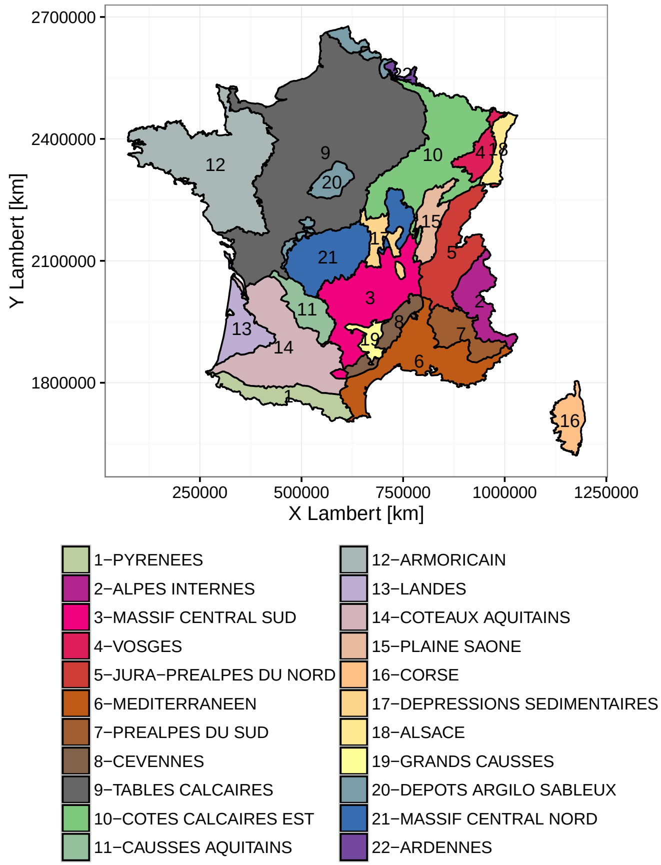


Figure 20. Map of the 22 Hydro-ecoregions, adapted from Wasson et al. (2002).

Appendix D: ~~Synthetic diagram for the spatial matching procedure~~

~~Synthetic diagram detailing the spatial matching procedure presented in Sect. 3.2.2.~~

~~The overall spatial matching procedure described in Sect. 3.2.2 is synthesised in Fig. 4. A first matching is done within HERs considering an overlap of dates across stations inside each HER. This matching is applied to each HER to derive HER-representative events computed using the median start and end dates of each spatio-temporal event defined within each HER. A second matching is then done considering an overlap of dates across the 22 HER-representative events to identify spatio-temporal events at the scale of France. The inter-HER matching is then reported to the local stations, giving the final France-wide matching.~~

Acknowledgements. The authors would like to thank Météo-France for providing access to the Safran database. The authors would like to thank Guillaume Thirel for his constructive advices provided during the hydrological modelling step. Analyses were performed in R (R

Core Team, 2016a), with packages dplyr (Wickham and Francois, 2015), ggplot2 (Wickham, 2009), ggrepel (Slowikowski, 2016), cowplot (Wilke, 2016), RColorBrewer (Neuwirth, 2014), reshape2 (Wickham, 2007), sp (Pebesma and Bivand, 2005; Bivand et al., 2013), grid (R Core Team, 2016b), gridExtra (Auguie, 2016), scales (Wickham, 2016) and lubridate (Grolemund and Wickham, 2011). L. Caillouet PhD thesis is funded by Irstea and CNR.

References

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop Evapotranspiration – Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper 56, FAO, Rome, 1998.

Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P.: Twentieth-century drought in the conterminous United States, J. Hydrometeorol., 6, 985–1001, doi:10.1175/JHM450.1, 2005.

Auffray, A., Clavel, A., Jourdain, S., Ben Daoud, A., Sauquet, E., Lang, M., Obled, C., Panthou, G., Gautheron, A., Gottardi, F., and Garçon, R.: Reconstructing the hydrometeorological scenario of the 1859 flood of the Isere river, Houille Blanche, pp. 44–50, doi:10.1051/lhb/2011005, 2011.

Auguie, B.: gridExtra: Miscellaneous Functions for "Grid" Graphics, https://CRAN.R-project.org/package=gridExtra, r package version 2.2.1, 2016.

Ben Daoud, A., Sauquet, E., Lang, M., Bontron, G., and Obled, C.: Precipitation forecasting through an analog sorting technique: a comparative study, Advances in Geosciences, 29, 103–107, doi:10.5194/adgeo-29-103-2011, 2011.

Ben Daoud, A., Sauquet, E., Bontron, G., Obled, C., and Lang, M.: Daily quantitative precipitation forecasts based on the analogue method: Improvements and application to a French large river basin, Atmos. Res., 169, 147–159, doi:10.1016/j.atmosres.2015.09.015, 2016.

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V.: Applied spatial data analysis with R, Second edition, Springer, NY, http://www.asdar-book.org/, 2013.

Bonaccorso, B., Peres, D. J., Cancelliere, A., and Rossi, G.: Large Scale Probabilistic Drought Characterization Over Europe, Water Resour. Manag., 27, 1675–1692, doi:10.1007/s11269-012-0177-z, 2013.

Bremond, R.: L'incidence du déficit de pluie sur l'écoulement des rivières, Ponts et Chaussées Magazine, 12, 49–58, 1976.

Brigode, P., Brissette, F., Nicault, A., Perreault, L., Kuentz, A., Mathevet, T., and Gailhard, J.: Streamflow variability over 1881-2011 period in northern Quebec: comparison of hydrological reconstructions based on tree rings and on geopotential height field reanalysis, Clim. Past Discuss., doi:10.5194/cp-2016-5, 2016.

Brochet, P.: The 1976 drought in France: climatological aspects and consequences, Hydrological Sciences Bulletin, 22, 393–411, doi:10.1080/02626667709491733, 1977.

Burn, D. H.: Catchment similarity for regional flood frequency analysis using seasonality measures, J. Hydrol., 202, 212–230, doi:10.1016/S0022-1694(97)00068-1, 1997.

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.

Catalogne, C., Sauquet, E., and Lang, M.: Using spot gauging data to estimate the annual minimum monthly flow with a return period of 5 years, Houille Blanche, pp. 78–87, doi:10.1051/lhb/2014042, 2014.

Chauveau, M., Chazot, S., Perrin, C., Bourgin, P.-Y., Sauquet, E., Vidal, J.-P., Rouchy, N., Martin, E., David, J., Norotte, T., Maugis, P., and de Lacaze, X.: What will be impacts of climate change on surface hydrology in France by 2070?, Houille Blanche, pp. 1–15, doi:10.1051/lhb/2013027, 2013.

Chauveau, M., Garcia, J., Mahé, M., and Chazot, S.: Étude de la gestion quantitative et des débits du Rhône en période de « basses eaux », Rapport BRLi. Phase 2. étude des étiages historiques, reconstitution des débits désinfluencés et évaluation de l'empreinte des influences anthropiques sur les débits du Rhône. Document A – Rapport principal de mission 1 : Étude des étiages historiques, Agence de l'Eau Rhône-Méditerranée-Corse, 2014.

Cipriani, T., Toilliez, T., and Sauquet, E.: Estimating 10 year return period peak flows and flood durations at ungauged locations in France, Houille Blanche, pp. 5–13, doi:10.1051/lhb/2012024, 2012.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.: The Schaake Shuffle: A Method for Reconstructing Space-Time Variability in Forecasted Precipitation and Temperature Fields, J. Hydrometeorol., 5, 243–262, doi:10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2, 2004.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Q. J. Roy. Meteor. Soc., 137, 1–28, doi:10.1002/qj.776, 2011.

Cook, E. R., Seager, R., Kushnir, Y., Briffa, K. R., Büntgen, U., Frank, D., Krusic, P. J., Tegel, W., van der Schrier, G., Andreu-Hayles, L., Baillie, M., Baittinger, C., Bleicher, N., Bonde, N., Brown, D., Carrer, M., Cooper, R., Čufar, K., Dittmar, C., Esper, J., Griggs, C., Gunnarson, B., Günther, B., Gutierrez, E., Haneca, K., Helama, S., Herzig, F., Heussner, K.-U., Hofmann, J., Janda, P., Kontic, R., Köse, N., Kyncl, T., Levanič, T., Linderholm, H., Manning, S., Melvin, T. M., Miles, D., Neuwirth, B., Nicolussi, K., Nola, P., Panayotov, M., Popa, I., Rothe, A., Seftigen, K., Seim, A., Svarva, H., Svoboda, M., Thun, T., Timonen, M., Touchan, R., Trotsiuk, V., Trouet, V., Walder, F., Ważny, T., Wilson, R., and Zang, C.: Old World megadroughts and pluvials during the Common Era, Science Advances, 1, doi:10.1126/sciadv.1500561, 2015.

Crooks, S. M. and Kay, A. L.: Simulation of river flow in the Thames over 120 years: Evidence of change in rainfall-runoff response?, J. Hydrol., 4, 172–195, doi:10.1016/j.ejrh.2015.05.014, 2015.

Dayon, G.: Evolution du cycle hydrologique sur la France au cours des prochaines décennies, Ph.D. thesis, Université Joseph Fourier, Grenoble, 2015.

Dayon, G., Boé, J., and Martin, E.: Transferability in the future climate of a statistical downscaling method for precipitation in France, J. Geophys. Res-Atmos., 120, 1023–1043, doi:10.1002/2014JD022236, 2015.

Dehérain, P.-P.: La sécheresse en 1893 et la disette de fourrages, Revue des Deux Mondes, 119, 850–881, 1893.

Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., and Zhu, Y.: The science of NOAA's operational hydrologic ensemble forecast service, B. Am. Meteorol. Soc., 95, 79–98, doi:10.1175/BAMS-D-12-00081.1, 2014.

Duband, D.: Rainfall-run-off retrospective of extremes droughts since 1860 in Europe (Germany, Italia, France, Rumania, Spain, Switzerland), Houille Blanche, pp. 51–59, doi:10.1051/lhb/2010041, 2010.

Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Demuth, S.: A global evaluation of streamflow drought characteristics, Hydrol. Earth Syst. Sc., 10, 525–552, doi:10.5194/hess-10-535-2006, 2006.

Forzieri, G., Feyen, L., Rojas, R., Flörke, M., Wimmer, F., and Bianchi, A.: Ensemble projections of future streamflow droughts in Europe, Hydrol. Earth Syst. Sc., 18, 85–108, doi:10.5194/hess-18-85-2014, 2014.

Gazelle, F.: La sécheresse de 1976 en Aquitaine orientale et dans le sud du Massif Central, Revue Géographique des Pyrénées et du Sud-Ouest, 48, 245–268, doi:10.3406/rgpso.1977.3514, 1977.

Giuntoli, I., Renard, B., Vidal, J.-P., and Bard, A.: Low flows in France and their relationship to large-scale climate indices, J. Hydrol., 482, 105–118, doi:10.1016/j.jhydrol.2012.12.038, 2013.

Giuntoli, I., Vidal, J.-P., Prudhomme, C., and Hannah, D. M.: Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models, Earth Syst. Dynam., 6, 267–285, doi:10.5194/esd-6-267-2015, 2015.

Grolemund, G. and Wickham, H.: Dates and Times Made Easy with lubridate, Journal of Statistical Software, 40, 1–25, http://www.jstatsoft.org/v40/i03/, 2011.

Gupta, H. V., Kling, H., Yilmaz, K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Hannaford, J. and Marsh, T.: An assessment of trends in UK runoff and low flows using a network of undisturbed catchments, Int. J. Climatol., 26, 1237–1253, doi:10.1002/joc.1303, 2006.

Hannaford, J., Buys, G., Stahl, K., and Tallaksen, L. M.: The influence of decadal-scale variability on trends in long European streamflow records, Hydrol. Earth Syst. Sc., 17, 2717–2733, doi:10.5194/hess-17-2717-2013, 2013.

Hisdal, H., Stahl, K., Tallaksen, L. M., and Demuth, S.: Have streamflow droughts in Europe become more severe or frequent?, Int. J. Climatol., 21, 317–333, doi:10.1002/joc.619, 2001.

Jones, P. D., Lister, D. H., Wilby, R. L., and Kostopoulou, E.: Extended riverflow reconstructions for England and Wales, 1865-2002, Int. J. Climatol., 26, 219–231, doi:10.1002/joc.1252, 2006.

Koerner, W., Cinotti, B., Jussy, J.-H., and Benoît, M.: Évolution des surfaces boisées en France depuis le début du XIX[e] siècle : identification et localisation des boisements des territoires agricoles abandonnés, Revue Forestière Française, 52, 249–269, doi:10.4267/2042/5359, 2000.

Kuentz, A., Mathevet, T., Gailhard, J., Perret, C., and Andréassian, V.: Over 100 years of climatic and hydrologic variability of a mediterranean and mountainous watershed: The Durance River, in: Cold and mountain region hydrological systems under climate change: towards improved projections, edited by Gelfan, A., Yang, D., Gusev, Y., and Kunstmann, H., vol. 360 of *IAHS Red Books*, pp. 19–25, IAHS, 2013.

Kuentz, A., Mathevet, T., Gailhard, J., and Hingray, B.: Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM method, Hydrol. Earth Syst. Sc., 19, 2717–2736, doi:10.5194/hess-19-2717-2015, 2015.

Laaha, G., Gauster, T., Tallaksen, L. M., Vidal, J.-P., Stahl, K., Prudhomme, C., Heudorfer, B., Vlnas, R., Ionita, M., Van Lanen, H. A. J., Adler, M.-J., Caillouet, L., Delus, C., Fendekova, M., Gailliez, S., Hannaford, J., Kingston, D., Van Loon, A. F., Mediero, L., Osuch, M., Romanowicz, R., Sauquet, E., Stagge, J. H., and Wong, W. K.: The European 2015 drought from a hydrological perspective, Hydrol. Earth Syst. Sci. Discussions, doi:10.5194/hess-2016-366, 2016.

Labuhn, I., Daux, V., Girardclos, O., Stievenard, M., Pierre, M., and Masson-Delmotte, V.: French summer droughts since 1326 AD: a reconstruction based on tree ring cellulose δ^{18}O, Clim. Past, 12, 1101–1117, doi:10.5194/cp-12-1101-2016, 2016.

Le Gros, C., Sauquet, E., Lang, M., Achard, A.-L., Leblois, E., and Biton, B.: The hydrological yearbooks published by the Société Hydrotechnique de France: a valuable source of information on hydrology in France, Houille Blanche, pp. 66–77, doi:10.1051/lhb/20150048, 2015.

Le Moine, N.: Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles pluie-débit ?, Ph.D. thesis, Université Pierre et Marie Curie, 2008.

Lennard, A. T., Macdonald, N., Clark, S., and Hooke, J. M.: The application of a drought reconstruction in water resource management, Hydrol. Res., 47, 1–14, doi:10.2166/nh.2015.090, 2015.

Lorenz, E.: Atmospheric Predictability as Revealed by Naturally Occurring Analogues, Journal of Atmospheric Sciences, 26, 636–646, 1969.

Mascart, E., ed.: Annales du bureau central météorologique de France – Année 1878 – Tome II : Bulletin des observations françaises et revue climatologique, pp. 25–36 (Planches), Gauthiers-Villars, 1880.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency andduration to time scales., in: 8th Conference on AppliedClimatology, Am. Meteorol. Soc., Anaheim, Calif, 1993.

Meko, D. M., Woodhouse, C. A., and Morino, K.: Dendrochronology and links to streamflow, J. Hydrol., 412-413, 200–209, doi:10.1016/j.jhydrol.2010.11.041, 2012.

Mérillon, Y. and Chaperon, P.: Drought in France in 1989, Houille Blanche, pp. 325–340, doi:10.1051/lhb/1990025, 1990.

Moreau, F.: Drought crisis management: Loire basin example, Houille Blanche, 4, 70–76, doi:10.1051/lhb:200404010, 2004.

Moureaux, T.: Sur le régime des pluies en France pendant l'année 1877, in: Annales du bureau central météorologique de France – Année 1877 – Pluies en France – Observations publiées avec la coopération du ministère des travaux publics et le concours de l'association scientifique, edited by Mascart, E., Gauthiers-Villars, 1880a.

Moureaux, T.: Sur le régime des pluies en France pendant l'année 1878, in: Annales du bureau central météorologique de France – Année 1878 – Tome III – Pluies en France – Observations publiées avec la coopération du ministère des travaux publics et le concours de l'association scientifique, edited by Mascart, E., Gauthiers-Villars, 1880b.

Neuwirth, E.: RColorBrewer: ColorBrewer Palettes, https://CRAN.R-project.org/package=RColorBrewer, r package version 1.1-2, 2014.

Nicault, A., Boucher, E., Bégin, C., Guiot, J., Marion, J., Perreault, L., Roy, R., Savard, M. M., and Bégin, Y.: Hydrological reconstruction from tree-ring multi-proxies over the last two centuries at the Caniapiscau Reservoir, northern Québec, Canada, J. Hydrol., 513, 435–445, doi:10.1016/j.jhydrol.2014.03.054, 2014.

Nicolle, P., Pushpalatha, R., Perrin, C., François, D., Thiéry, D., Mathevet, T., Le Lay, M., Besson, F., Soubeyroux, J.-M., Viel, C., Regimbeau, F., Andréassian, V., Maugis, P., Augeard, B., and Morice, E.: Benchmarking hydrological models for low-flow simulation and forecasting on French catchments, Hydrol. Earth Syst. Sc., 18, 2829–2857, doi:10.5194/hess-18-2829-2014, 2014.

Pebesma, E. J. and Bivand, R. S.: Classes and methods for spatial data in R, R News, 5, 9–13, http://CRAN.R-project.org/doc/Rnews/, 2005.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, doi:10.1016/S0022-1694(03)00225-7, 2003.

Pfister, C., Weingarter, R., and Luterbacher, J.: Hydrological winter droughts over the last 450 years in the Upper Rhine basin : a methodological approach, Hydrolog. Sci. J., 51, 966–985, doi:10.1623/hysj.51.5.966, 2006.

Plumandon, J.-R.: La sécheresse du printemps 1893, La Nature, pp. 27–28, 1893.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C: An atmospheric reanalysis of the twentieth century, J. Climate, 29, 4083–4097, doi:10.1175/JCLI-D-15-0556.1, 2016.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, Proc. Natl. Acad. Sci. U. S. A., 111, 3262–3267, doi:10.1073/pnas.1222473110, 2014.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, J. Hydrol., 411, 66–76, doi:10.1016/j.jhydrol.2011.09.034, 2011.

Pushpalatha, R., Perrin, C., Le Moine, N., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, J. Hydrol., 420-421, 171–182, doi:10.1016/j.jhydrol.2011.11.055, 2012.

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchistéguy, L., and Morel, S.: Analysis of near-surface atmospheric variables: validation of the safran analysis over france, J. Appl. Meteorol. Clim., 47, 92–107, doi:10.1175/2007JAMC1636.1, 2008.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2016a.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2016b.

Radanovics, S., Vidal, J.-P., Sauquet, E., Ben Daoud, A., and Bontron, G.: Optimising predictor domains for spatially coherent precipitation downscaling, Hydrol. Earth Syst. Sc., 17, 4189–4208, doi:10.5194/hess-17-4189-2013, 2013.

Robertson, D. E., Shrestha, D. L., and Wang, Q. J.: Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting, Hydrol. Earth Syst. Sc., 17, 3587–3603, doi:10.5194/hess-17-3587-2013, 2013.

Sauquet, E.: Mapping mean annual river discharges: geostatistical developments for incorporating river network dependencies, J. Hydrol., 331, 300–314, doi:10.1016/j.jhydrol.2006.05.018, 2006.

Sauquet, E. and Catalogne, C.: Comparison of catchment grouping methods for flow duration curve estimation at ungauged sites in France, Hydrol. Earth Syst. Sc., 15, 2421–2435, doi:10.5194/hess-15-2421-2011, 2011.

Shiau, J. T. and Shen, H. W.: Recurrence analysis of hydrologic droughts of differing severity, J. Water Res. Pl. Man.-ASCE, 127, 30–40, doi:10.1061/(ASCE)0733-9496, 2001.

Sippel, S., Otto, F. E. L., Forkel, M., Allen, M. R., Guillod, B. P., Heimann, M., Reichstein, M., Seneviratne, S. I., Thonicke, K., and Mahecha, M. D.: A novel bias correction methodology for climate impact simulations, Earth System Dynamics, 7, 71–88, doi:10.5194/esd-7-71-2016, 2016.

Slowikowski, K.: ggrepel: Repulsive Text and Label Geoms for 'ggplot2', https://CRAN.R-project.org/package=ggrepel, r package version 0.5, 2016.

Smakhtin, V. U.: Low flow hydrology: a review, J. Hydrol., 240, 147–186, doi:10.1016/S0022-1694(00)00340-1, 2001.

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, Hydrol. Earth Syst. Sc., 17, 2685–2699, doi:10.5194/hess-17-2685-2013, 2013.

Soubeyroux, J.-M., Vidal, J.-P., Baillon, M., Blanchard, M., Céron, J.-P., Franchistéguy, L., Régimbeau, F., Martin, E., and Vincendon, J.-C.: Characterizing and forecasting droughts and low-flows in France with the Safran-Isba-Modcou hydrometeorological suite, Houille Blanche, pp. 30–39, doi:10.1051/lhb/2010051, 2010.

Spraggs, G., Peaver, L., Jones, P., and Ede, P.: Re-construction of historic drought in the Anglian Region (UK) over the period 1798–2010 and the implications for water resources and drought management, J. Hydrol., 526, 231–252, doi:10.1016/j.jhydrol.2015.01.015, 2015.

Tallaksen, L. and Stahl, K.: Spatial and temporal patterns of large-scale droughts in Europe: Model dispersion and performance, Geophys. Res. Lett., 41, 429–434, doi:10.1002/2013GL058573, 2014.

Tallaksen, L., Madsen, H., and Clausen, B.: On the definition and modelling of streamflow drought duration and deficit volume, Hydrolog. Sci. J., 42, 15–33, doi:10.1080/02626669709492003, 1997.

Tallaksen, L. M. and Van Lanen, H. A. J.: Hydrological Drought: Processes and estimation methods for streamflow and groundwater, vol. 48 of *Developments in Water Science*, Elsevier, 2004.

Timbal, B., Arblaster, J., and Power, S.: Attribution of the late 20th century rainfall decline in South-West Australia, J. Climate, 19, 2046–2062, doi:10.1175/JCLI3817.1, 2006.

Uhlemann, S., Thieken, A. H., and Merz, B.: A consistent set of trans-basin floods in Germany between 1952–2002, Hydrol. Earth Syst. Sc., 14, 1277–1295, doi:10.5194/hess-14-1277-2010, 2010.

Valery, A.: Modélisation précipitations - débit sous influence nivale. Élaboration d'un module neige et évaluation sur 380 bassins versants., Ph.D. thesis, AgroParisTech, 2010.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simple': What is useful in a temperature-based snow-accounting routine? Part 2 - Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, J. Hydrol., 517, 1176 – 1187, doi:10.1016/j.jhydrol.2014.04.058, 2014.

van Huijgevoort, M. H. J., Hazenberg, P., van Lanen, H. A. J., and Uijlenhoet, R.: A generic method for hydrological drought identification across different climate regions, Hydrol. Earth Syst. Sc., 16, 2437–2451, doi:10.5194/hess-16-2437-2012, 2012.

Van Lanen, H., Laaha, G., Kingston, D. G., Gauster, T., Ionita, M., Vidal, J.-P., Vlnas, R., Tallaksen, L. M., Stahl, K., Hannaford, J., Delus, C., Fendekova, M., Mediero, L., Prudhomme, C., Rets, E., Romanowicz, R. J., Gailliez, S., Wong, W. K., Adler, M.-J., Blauhut, V., Caillouet, L., Chelcea, S., Frolova, N., Gudmundsson, L., Hanel, M., Haslinger, K., Kireeva, M., Osuch, M., Sauquet, E., Stagge, J. H., and Van Loon, A. F.: Hydrology needed to manage droughts: the 2015 European case, Hydrol. Process., doi:10.1002/hyp.10838, 2016.

Van Loon, A. F. and Laaha, G.: Hydrological drought severity explained by climate and catchment characteristics, J. Hydrol., 526, 3–14, doi:10.1016/j.jhydrol.2014.10.059, 2015.

Van Loon, A. F. and Van Lanen, H. A. J.: A process-based typology of hydrological drought, Hydrol. Earth Syst. Sc., 16, 1915–1946, doi:10.5194/hess-16-1915-2012, 2012.

Verkade, J. S., Brown, J. D., Reggiani, P., and Weerts, A. H.: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales, J. Hydrol., 501, 73–91, doi:10.1016/j.jhydrol.2013.07.039, 2013.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Clim., 30, 1627–1644, doi:10.1002/joc.2003, 2010a.

Vidal, J.-P., Martin, E., Franchistéguy, L., Habets, F., Soubeyroux, J.-M., Blanchard, M., and Baillon, M.: Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite, Hydrol. Earth Syst. Sc., 14, 459–478, doi:10.5194/hess-14-459-2010, 2010b.

Vidal, J.-P., Martin, E., Kitova, N., Najac, J., and Soubeyroux, J.-M.: Evolution of spatio-temporal drought characteristics: validation, projections and effect of adaptation scenarios, Hydrol. Earth Syst. Sc., 16, 2935–2955, doi:10.5194/hess-16-2935-2012, 2012.

Vidal, J.-P., Hingray, B., Magand, C., Sauquet, E., and Ducharne, A.: Hierarchy of climate and hydrological uncertainties in transient low flow projections, Hydrol. Earth Syst. Sc. Discussions, 12, 12 649–12 701, doi:10.5194/hessd-12-12649-2015, 2015.

Vimont, E.: La sécheresse en 1893, L'Astronomie, 12, 309–311, 1893.

Vivian, H.: L'hydrologie nord-alpine et la sécheresse de 1976, Revue de Géographie de Lyon, 52, 117–151, doi:10.3406/geoca.1977.1199, 1977.

Vogel, R. M. and Stedinger, J. R.: Generalized Storage-Reliability-Yield Relationships, J. Hydrol., 89, 303–327, doi:10.1016/0022-1694(87)90184-3, 1987.

Vrac, M. and Friederichs, P.: Multivariate – Intervariable, Spatial, and Temporal – Bias Correction, J. Climate, 28, 218–237, doi:10.1175/JCLI-D-14-00059.1, 2015.

Šípek, V. and Daňhelka, J.: Modification of input datasets for the Ensemble Streamflow Prediction based on large-scale climatic indices and weather generator, J. Hydrol., 528, 720–733, doi:10.1016/j.jhydrol.2015.07.008, 2015.

Wasson, J. G., Chandesris, A., Pella, H., and Blanc, L.: Typology and reference conditions for surface water bodies in France: the hydro-ecoregion approach, in: Typology and ecological classification of lakes and rivers, edited by Ruoppa, M. and Karttunen, K., no. 566 in TemaNord, pp. 37–41, Nordic Council of Ministers, Copenhagen, Denmark, 2002.

Wickham, H.: Reshaping Data with the reshape Package, Journal of Statistical Software, 21, 1–20, http://www.jstatsoft.org/v21/i12/, 2007.

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, http://ggplot2.org, 2009.

Wickham, H.: scales: Scale Functions for Visualization, https://CRAN.R-project.org/package=scales, r package version 0.4.0, 2016.

Wickham, H. and Francois, R.: dplyr: A Grammar of Data Manipulation, https://CRAN.R-project.org/package=dplyr, r package version 0.4.3, 2015.

Wilke, C. O.: cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2', https://CRAN.R-project.org/package=cowplot, r package version 0.6.2, 2016.

Wilks, D. S.: Multivariate ensemble Model Output Statistics using empirical copulas, Q. J. Roy. Meteor. Soc., 141, 945–952, doi:10.1002/qj.2414, 2014.

Zaidman, M. D., Rees, H. G., and Young, A. R.: Spatio-temporal development of streamflow droughts in north-west Europe, Hydrol. Earth Syst. Sc., 6, 733–751, doi:10.5194/hess-6-733-2002, 2002.