Response to Reviewer 2

General Comments
This paper investigates the existence of multi-decadal variability (MDV) in annual and seasonal flows in French rivers, and explores some possible driving mechanisms (precipitation, large-scale atmospheric circulation, oceanic variability, hydrological processes). The analyses presented in this paper are of interest for the hydro-climate community and are quite convincing. I would therefore recommend publication, subject to the following moderate revisions:

First, we thank Reviewer 2 for the careful reading and evaluation of our paper and the many suggestions that greatly helped us to improve the manuscript. Here a summary of the main changes made to the paper following the reviewer's comments:
-We added sub-section titles in order to improve the clarity of the logic of the paper.
-We improved the discussion on the detrending issue and moved it to the section Data and Methods. We also modified Figure 3 to show the impact of detrending.
-We added a new figure to show the decadal variations in river flows outside the 1938-1985 period (New figure 5).
-The following changes have been made to the figures:
Figure 1: we added the histogram of basin sizes.
Figure 2: we added the results for all the seasons.
Figure 3: we added the analysis for temperature. We added the analysis for non-detrended series.
Old Figure 5: It has been removed following the reviewer's suggestion.
Figure 6: we added the analysis for temperature.
Figure 7: we added the correlation map in spring between SLPI and precipitation in France.

Our detailed answers to the reviewer's comments are given bellow.

1. I agree with Reviewer 1 that the paper would benefit from improvements in its structure. An easy-to-implement improvement would be to use level-2 or 3 subtitles to make the overall structure more apparent to the reader. For instance, section 2 could be split into 2.1 station data (Q then P), 2.2 reanalysis data (P, SLP, SST, hydrology), and 2.3. statistical tools (filtering and testing). Section 3 could be split into evidence of MDV / link with P and T, section 4 into P / SLP / SST, and section 5 into motivation / model efficiency / impact of soil moisture on summer flows / MDV in soil moisture. Some other reorganization might also be valuable, for instance the presentation of the data is a bit confusing in the present state, mostly due to the 3 distinct precipitation datasets. Rather than an organization based on "data product" as currently done, I would recommend an organization based either on the data type as suggested above (station data vs. reanalyses), or maybe on the hydrometeorological variable (=> all 3 P precipitation datasets in the same section).

We implemented all the suggestions made by Reviewer 1 to improve the structure of the paper (see our response to Reviewer 1's comments). As suggested by Reviewer 2, the different sections have been split in different sub-sections to improve the clarity and the logic of the paper. The presentation of data has been modified following the suggestions of Reviewer 2.

2. The existence of MDV is convincingly demonstrated for spring and annual flows, based on the visual inspection of the series and the MTM spectral analysis. However, the demonstration is less convincing for other seasons, because these two specific analyses are not shown for winter, summer and autumn flows. The authors are mostly using the results of figures 3 and 4 to demonstrate the existence of MDV for these seasons, but I don't think these results are sufficient for this purpose. In Figure 3, the ratio is primarily controlled by the properties of the low-pass filter. As a very rough approximation, considering that the filter used by the authors is similar to a 18-years moving average, one would expect that for an iid series (hence displaying no MDV whatsoever) the ratio would be close to 1/sqrt(18) = 0.24. While the boxplots for MAM and YEAR are clearly larger than this value, suggesting some form of autocorrelation that may be multi-decadal, the boxplots for other seasons are actually quite close to 0.24. Maybe the authors could implement a Monte-Carlo analysis, by applying their filter to iid series, to evaluate a "critical value" for this ratio? In a similar vein, the fact that two periods yield significant differences (Fig 4) is not a proof of MDV. I guess the best way to convincingly demonstrate the existence of this MDV for other seasons would be to show figure 2 for all seasons.

-Thanks for the suggestion: the test with iid series has been added to the paper. 10,000 random series of gaussian white noise have been generated. Their length (before filtering) has been chosen to be equal to the median length of observed series (94 years for river flows). The results have been added to figure 3, with lines showing the 25th, 50th, 75th, 95th quantiles of the distribution of the ratios obtained with white noise series. Strong shift of the distribution compared to white noise are seen in spring and annual mean; small shift are seen in fall and winter.

-We added the results for the other seasons in Fig 2. The low-frequency variations seen in summer, winter and fall generally correlate with what is seen in spring and annual average, but are much clearer in summer. Significant multi-decadal signals in the MTM spectrum are only seen in summer. Note that the large signal seen for this river in summer may not necessarily be representative of what is seen for all gauging stations (we don't argue in the paper that it is the case). For example, given the importance of snow for this catchment, snow might be involved in the summer multi-decadal fluctuations.

-We are not sure to understand exactly what is meant by Reviewer 2 regarding Fig 4. "the fact that two periods yield significant differences (Fig 4) is not a proof of MDV". Maybe by MDV, the reviewer thinks about a long-term quasi-periodic multi-decadal signal of internal origin. We use it in a somewhat looser sense. As the long-term linear trends are

removed before computing the differences, the 21-year differences cannot be due to long-term trends. Therefore, we think that Figure 4 shows that significant decadal variations exist during the 20th century (between 1938-1958 and 1965-1985) and that one can speak of multi-decadal variability. Obviously, from this result it cannot be extrapolated that multi-decadal variations necessary exist in other periods. For example, in theory they could be caused by specific multi-decadal variations in the external forcing specific to this period.

Given our understanding of the remark, and because we thought that it was an interesting addition to the paper, we added a figure showing the river flows anomalies (with boxplots) on four 21-year periods (except for the last one that is a little bit shorter) on the historical period: 1910-1930, 1938-1958, 1965-1985 and 1995-2012. The choice of the periods is based on the examination of the Gave d'Ossau serie. This graph shows that important multi-decadal anomalies are not limited during the 20th and early 21st century to 1938-1958 and 1965-1985.

Note that we agree that the evidence shown for the existence of MDV in French river flows are stronger for spring and annual average compared to the other seasons, in particular because the physical mechanisms supporting those variations are clearer for spring variations, and the signal is more consistent over France. It may not have been clear in the previous version of the manuscript. We improved this point in the discussion.

3. **I think a discussion on detrending would be useful. Firstly, I think it is introduced much too quickly in section 2 (p. 11865 lines 25-29): the authors need to explain the motivation behind it. The authors actually discuss this issue later in the paper (p. 11876 lines 21-29), but I would recommend moving this discussion in section 2.**

OK. Done.

Secondly, this discussion also clearly highlights potential caveats of detrending, and I'm personally not convinced that detrending is a good thing in the MDV context of this paper. According to the authors, "it is necessary to remove the potential effect of long-term anthropogenic climate change (...) by detrending the data". Somehow, this implicitly assumes that the computed trend is indeed caused by anthropogenic CC! This sounds like a massive assumption in the absence of any attribution study.

We didn't mean that. By the use of "potential" we meant that if anthropogenic forcing impacts river flows, it is expected to lead to a long-term monotonic trend rather than to multi-decadal variations (on a sufficiently long period), and removing a trend will (mostly) get rid of it. The sentence was not clear and has been modified. Note also that independently of its cause, whether or not it is caused by anthropogenic forcing, if long-term (secular) trends exist they will be problematic for some analyses in the study.

Moreover, in the presence of MDV, a trend computed on raw data is likely to be spuriously created by the MDV behavior – ideally, removing the MDV would be needed before computing the trend! This is obviously a chicken-and-egg situation

that has no easy solution as far as I know, but I think it deserves some discussion.

We agree with that, but we also think that the issue is more or less serious depending on the length of the period studied (compared to the period of oscillation). For the long period studied in the paper with observations (~100 years), we think that the issue of MDV creating spurious trends is limited. But, on the shorter period of hydrological modelling, it is a major issue, as discussed in the paper (hence our choice to show the results obtained with or without detrending in that case). The discussion on this point has been improved.

Note that for Fig. 2 no detrending was applied, as there is no reason to detrend for this particular analysis. We added new subfigures to figure 3, the result of the same analysis, but with the trends included.

Lastly, removing a linear trend on such long periods is also questionable, as discussed by the authors for SST (p. 11867 line 1). My feeling is that detrending makes sense for temperature variables, which indeed display consistent long-term trends which have been attributed to anthropogenic CC. I'm much more circumspect for hydrologic variables (precipitation and runoff), since the literature generally shows weaker and much less consistent trends.

For SST averaged on a large domain as the North-Atlantic, that exhibits a large trend, there are many reasons (statistical, theoretical, from modelling) to think that a straight line is not a very good model of the impact of GHG forcing on a long period (even if it is still commonly used). For hydrological variables, it is much less clear that a non-linear trend would be more adequate.
We agree with the fact that the literature shows weaker and much less consistent trends, but the vast majority of studies (of river flows) to our knowledge have focused on relatively short period (less than 50-year and understandably so, because much more observations are available on shorter period) when it is especially difficult to separate the impact of multi-decadal variability and of potential long-term trends, which may result in weak and inconsistent (relatively to the start date) trends, which leads to the same chicken-and-egg situation as previously noted**.**

I'm not suggesting the authors should redo all analyses without detrending, but rather that they should justify their choice more precisely, and discuss potential caveats. Lastly, it is sometimes unclear whether or not detrending has been implemented (e.g. in figure 2). In some cases, detrending is specified in the figure caption but not in the text (e.g. for figure 4). In order to avoid ambiguity, I would recommend systematically writing "detrended" whenever needed (and even maybe "undetrended"!).

We added a better discussion to justify our choice of detrending in section 2. We added "detrended" and "undetrended" where relevant in the text. We also modified Fig 3 to

show the results for both detrended and undetrended variables to show the impact of this methodological choice when long series are involved.

Specific Comments

p. 11865 line 9: Are you using civil or hydrological years for the annual mean? Please make it explicit (although I don't think it would change anything given the low-pass filtering).

We use civil years. Modification made. We also added a precision at the end of section 2, where the seasons are defined.

p. 11865 lines 11-14: Did the authors consider simply leaving these values as missing? I think correlations and filtering can accommodate missing data quite easily – I'm less sure for MTM. The authors should elaborate a bit on this (what technical difficulty with missing data led to this decision of filling?), because I agree that the filling approach is a bit crude (and temporal interpolation would probably not be better indeed).

Using our filter with series with missing data is not strictly impossible: it is indeed easy to compute the 19-year weighted averages with missing values (we think it is the solution the reviewer has in mind; we don't see another way to use the filter with missing data).
But computing 19-year averages with missing values is in the end equivalent to using an infilling approach: the missing years are supposed implicitly to have the same mean as the years with values in the 19-year window (not exactly here because the window is not rectangular in our case but it is the same idea) which could enhance artificially multi-decadal variability. Our approach is therefore more conservative we think, as it is the less likely to enhance artificially MDV. Moreover, window averaging with missing value may result in high-frequency noise in the filtered series. We added a better justification of our choice.

p. 11865 lines 19-21. To be honest, I find the decision to keep strongly influenced stations disputable. It's probably acceptable at the annual scale, but at the seasonal scale, the effect of dams may be to transfer flows between seasons. Moreover, although these stations are indeed flagged in figure 1, the interpretation of the results in the following sections does not really use this information. Lastly, it seems that some of these strongly influenced stations are not present in all figures. For instance, the point near Paris (is it the Seine@Paris?) seems absent in figures 4-5, appears again in Figures 6-10, then disappears again from Figure 11. This needs to be clarified.

Yes, it is the Seine@Paris. We only use the serie that corresponds to the period before the creation of the lakes on the Seine (1885-1973) as it is sufficiently long to be included in our study. We didn't want to merge the pre-1973 serie with the post-1973 serie to avoid

inhomogeneities. For Figure 4-5, it is not possible to compute a robust 1965-1985 average, because 12 years of data are missing. But when correlations are involved (that are computed on specific periods adapted to each gauging stations,) the Seine@Paris can be used. We should have mentioned this point in the paper. A precision has been added in Figure 4's caption.

We added a new figure (new Figure 5, see also our response to major comment 2) with the 21-year differences in river flows for 4 periods. The results obtained with all the stations and with only the stations with no or little influences are plotted separately. We can see that the use of influenced stations has little impacts on the distribution of decadal anomalies in France. A corresponding discussion has been added.

p. 11866 line 6. Please indicate the spatial and temporal resolutions of this reanalysis (and of other subsequent reanalyses as well).

OK. Done

p. 11866 lines 11-24: the precise definition of SLPI should be given here rather than in a subsequent "results" section.

Done

p. 11866 line 27: please indicate which low-pass filter is used.

The same low-pass filter as used for precipitation, river flows etc. is used here. Precision added.

p. 11867 line 7-9: this is a bit confusing – I understand that a linear regression between SST and CO2 is used; as a result, the evolution with respect to time is not linear any more. Is it correct? Please reword to avoid ambiguity.

Yes. Done.

p. 11867 lines 10-12: this sounds a bit speculative – unless of course the authors actually trialed other de-trending approaches and saw little difference, but in this case I would slightly reword this sentence to make it more explicit.

Yes, we did it. Precision added.

p. 11867 lines 26. Maybe the authors could summarize which variables will actually be used in the case study?

Precision added.

p. 11869 lines 4-5: I disagree that Fig. 3b alone shows the existence of MDV (see general comment 2)

See our response to general comment 2.

p. 11870 lines 8-13: I think that this paragraph (and the corresponding figure) is a bit too short to bring any significant additional information: it could therefore be removed. Moreover, the wording is a bit confusing: a change in the mean is already a modification of the distribution of river flows! Do the authors mean a change in the shape of this distribution? But in this case, the changes reported in Figure 5 do not seem inconsistent with a simple translation of the distribution (shape and variance remaining unchanged).

We agree with that. This paragraph and the corresponding figure have been removed.

p. 11870 lines 23-24: As noted by the authors, Mediterranean France is indeed quite distinct from the rest of the country – but in this case, wouldn't it make sense to use precipitation averaged over non-Mediterranean France as an explanatory variable?

Yes, but precipitation in the Mediterranean France is not very interesting for our study as it has little impact on river flows (at least at multi-decadal time-scales) for the vast majority of the gauging stations studied. In Fig 6a, we can see that taking the France average is relevant for river flows for the vast majority of stations in France, even the (few) ones located in the southeast (which is logical as they may correspond to large catchment, the Rhone for example.). Therefore we don't think it is really useful in this paper to do an analysis dedicated to precipitation in the southeast.

p. 11872 line 14: except maybe in the Mediterranean area?

We did the correlation map between SLPI and precipitation and it is negative even in the Mediterranean area, even if the correlations are sometimes weaker (in absolute value) and very small for two departments.
This comment made us realize that showing the correlation map between SLPI and spring precipitation in France is useful. It has been added to Fig 7.

p. 11873 lines 13-17: I agree that this would be needed to formally established causality. However, given the difficulty of AOGCMs to reproduce low-frequency variability and its impacts (as mentioned by the authors in the introduction), is it really feasible? A short reminder of this difficulty could be added here.

Right, we agree this approach may fail. Precision added.

p. 11873 lines 23-29: this is a bit confusing – maybe because the exact meaning of "direct causality" is unclear to me. In my eyes, the fact that negative precipitation anomalies in spring may impact summer flows through the catchment memory sounds like a direct causality link. Maybe "synchronous" or something similar would be better than "direct"?

We agree that "synchronous" is better than "direct". Modification made.

p. 11874 line 10: the differences in Figure 4h are so small that I wouldn't use the adjective "larger" (especially in South-Western France where many streamflow stations are located)

Agreed. Modification made.

p. 11875 line 20: good correlations may hide systematic biases – is it the case here? Maybe reporting Nash-Sutcliffe efficiencies would be more appropriate?

Some systematic biases exist for some stations. We agree that generally efficiency is a better variable to evaluate a hydrological model as it provides a more complete view of the performance of the model. But in our study we are not especially interested with systematic biases: our main objective is too see if SIM is able to capture the low-frequency variations in rivers flows and this is why we look at this particular metric. Note that a study has already evaluated SIM using NS efficiencies. We added the reference (Habets et al. 2008).

p. 11876 lines 1-2: I have to confess that I find this sentence very optimistic... there are many examples in the literature of models "having the good answer for the bad reason" (and not only in hydrology!). So I wouldn't say that a good agreement with observed flows is sufficient to claim that other components of the continental water cycle are very likely well simulated. I would suggest moderating this statement.

Yes, we agree that models can have the good answer for the bad reason. But we also think that the risk is more or less important depending on the nature of the model and of the physical problem. For example, the fact that SIM is mostly a physical model and has mostly not been fitted to reproduce river flows evolution (compare to more conceptual model) is worth noting in that context. Moreover, because of water conservation, the evolution of evapotranspiration is physically strongly constrained when river flows and precipitation are known. Only errors in the evolution of soil moisture can make the evapotranspiration wrong if river flows and precipitation are right. That's being said, we agree that our statement was too strong. It has been modified.

p. 11876 line 8: Isn't it a bit risky to use a single-day SWI (31/05)? It probably depends on what is exactly the "soil" represented is SIM, and therefore what is the

temporal dynamics of daily SWI. Can it react quickly to some moderate precipitation event? (therefore yielding a "high" SWI on 31/05 but that would decrease to a much lower SWI value in just a couple of days?). A few words of explanation would be valuable.

From a physical point view, for the reason mentioned by the reviewer, it is true that the impact of the SLPI (and AMV) on soil moisture may be more easily seen on the seasonal average rather than on the value on the 31/05. But in any case, the value of SWI on 31/05 is always what matters the most for summer, not the spring average. What matters for evapotranspiration is soil moisture in summer, that is more related to SWI on 31/05 rather than on the entire spring (actually for a given value of SWI on 31/05, what happened before has no importance for summer soil moisture).

It is true that soil moisture may respond quickly to a specific precipitation event and it is likely one of the main reason for the soil moisture memory to be limited in a climatological sense. But, again in a climatological sense, soil moisture memory exists, at least in SIM (as seen indirectly in our analysis, but also as it can been seen by studying the autocorrelation of SWI).
We added an explanation.

p. 11879 lines 1-14. I find the distinction between interannual variability and MDV a bit confusing. I have the feeling that the authors consider that using raw annual (or seasonal) data corresponds to studying interannual variability, while low-pass filtering is necessary to studying MDV. I would disagree with this: a MDV signal can be seen in unfiltered data, as nicely illustrated in Figure 2. Moreover, establishing a link between a climate indice with a marked MDV component (such as the AMO index) and unfiltered data is also a sensible way of studying MDV. Filtering is indeed very useful to make MDV behaviors much more visible in figures; however, significance is more difficult to reach due to the large autocorrelation induced by filtering. Consider for instance figure 9: correlations of about 0.8 are needed before reaching significance (with roughly 100-year long series). I wouldn't be surprised if the same figure based on unfiltered data yielded similar results in terms of significance, although with much lower correlation coefficients! Maybe the authors should try to clarify this distinction between interannual variability and MDV.

Our point was not clear. We consider that using raw (seasonal) data corresponds to studying inter-annual, decadal, multi-decadal (etc.: everything) variability while using low-pass filtered data corresponds to studying variability only at multi-decadal time-scales. Therefore with raw data it is not possible to really be sure whether the correlation is due to the low-frequency component of the signals or to the high frequency component of the signals.
It is known that the NASST index has an important multi-decadal component (the AMV)

and it can therefore be supposed that MDV is likely to be at least partly responsible for the link between NASST and a given variable. But we think that using filtered series is necessary to really prove it. Inter-annual variations exist in the NASST index, partly due to the North Atlantic oscillation (NAO) that can also impact precipitation in France. Therefore, part of the relation between NASST and precipitation (or river flows) could be the result of the impact of NAO on both precipitation and SST, at inter-annual time scales. We haven't quantified to what extent it is true, but at least it is physically possible.

We added a (shorter) discussion on this point (and also slightly modified the introduction where the same paper is cited). Moreover, the previous discussion was not clear and led to the idea that our results are not consistent with Guiontoli et al. (2013), which is not true, as they mostly are.

p. 11879 lines 15-16: see general comment 2.

OK. We modified the discussion in order to better reflect the level of uncertainties associated.

p. 11880 line 6: I find this statement quite optimistic: if a large reservoir that transfers a significant part of the flow from one season to the next is built in the middle of the study period, the MDV assessment could be significantly impacted. This statement needs a reference in my opinion.

Our phrasing was ambiguous. We didn't mean that direct influences cannot create multi-decadal variations. It clearly can. But there is generally no reason to suppose that those "artificial" multi-decadal variations are in phase over France (except if dams are built at the same time on all rivers, which does not seem to be the case here). We modified the text.

p. 11881 lines 23-29: references would be useful.

OK. We added some references.

p. 11882 lines 5-8: Haven't such assessments already been reported in the hydrologic or climate literature?

To the best of our knowledge, no complete assessment exists in the context our particular study. But our sentence was probably misleading because it gave the impression that nothing has been done on the subject, which is not the case. A relevant reference was

given in the introduction and has been added to the conclusion.

Figures, general comment: axes labels are sometimes a bit small, please scan all figures and increase font size if necessary. Moreover, in figures with multiple maps sharing the same colorbar, it would be better to show a single large colorbar (in the present state, some replicated colorbars are very difficult to read, e.g. figure 4).

We agree that in the HESS discussion some figures are hard to read, especially fig 4, but it is due to the resizing of the figure to fit the (unusual) format of HESSD. We are sorry for that. In a "normal" format, this figure is readable. We made some changes to figures in order to improve readability.

Figure 1: A histogram of catchment sizes would be useful, or at least the range of catchment sizes should be mentioned in the text.

A histogram of catchment size has been added to Fig 1 and discussed briefly in the text.

Figure 3: I think it's worth showing the same figure for temperatures, given the comparison made in the following figure 4. Same comment for figure 6.

Done.

Figure 5: Is it after detrending?

Yes. But the figure has been removed following the reviewer's suggestion

Technical corrections
p. 11869 line 23: "or" rather than "nor"
p. 11873 line 7: robustly estimate
p. 11882 line 5 : Whether or not climate models are able to capture...
caption of figure 10: automn <-> autumn

Thanks. Modification made.