Response to Reviewer 1

First of all, we thank the reviewer for his careful reading of our study and the many suggestions that helped us to improve the manuscript. In particular, we improved the presentation of the logical framework of the paper following the reviewer's advices.

The article assesses the role of large-scale circulation in long-term variability of river runoff in France. The study is based on large set of meteorological, oceanological and hydrological data, including reanalysis, observations and model simulations. The article proves the strong impact of the Atlantic Multidecadal Variability (AMV) on spring river flow decadal variability. The summer low frequency discharge variation is explained by AMV control of spring soil moisture conditions. The article is quite interesting and important for understanding of processes, improving of hydrological predictions and risk assessment.

At the same time, several parts of the manuscript suffer from an absence of well-described logical framework and clear writing, making it difficult to read. The authors process numerous data that they describe in dedicated section, but the necessity to use such a huge dataset is not explained in the beginning. The presentation of the results and article logic is quite specific: the big part of explanations is given in figures' captions and scattered in the text. The conclusions of the sections are often a good step for the next work/section, but they are not clearly related to the initial objective of the section. Some questions/explanations are only addressed in following sections.
I recommend to enlarge the description of each section in the Introduction, clearly explaining the logical schema of article. The section 3. "Multi-decadal variability in observed French hydro-climate" is written in the manner to prove the representability of use Gave d'Ossau gauging station for all France. And on the way, other issues are studied (spatio-seasonal patterns, role of evapotranspiration, modification of the statistical distribution of low and high discharges, use spatially-averaged precipitation for analysis etc). It needs the structural improvement. The Conclusion section is well written but actually represents the Discussion. I suggest to divide it and put in separate section.

Following the reviewer's advices, we improved the presentation of the logical framework of the paper.
-As suggested by the reviewer, the description of each sections at the end of the introduction has been enhanced in order to give a general view of the paper. The objectives of the study have also been clarified in the introduction.
-In section 2, the interest of each dataset for our study is now better justified.
-We added a presentation of the objectives at the beginning of some sections (when they were missing) and added the conclusions directly related to these objectives at the end of some sections (when they were missing).
-The material presented in the conclusion of the first version of the manuscript has been divided in two parts, one for a new discussion section, one for conclusion.
-We modified the presentation of some results in section 3. Our idea was not really to argue that the Gave d'Ossau is representative of the behavior of all gauging stations in France. We just wanted to begin with a visual illustration for a particular case, before tackling the issue from a more general point of view.
-As suggested by Reviewer 2, we added sub-section titles in order to improve the clarity of the logic of the paper.

Specific comments: pp-lines

11863-28: "Historical period" has very large interpretation. What do you mean exactly?

We mean: from late 19th century to present. Precision added.

11864-18: "current literature" do you mean your article? And is it your suggestion or this idea was already proposed by other researchers?

No, we meant “other studies”. The other studies are given afterwards (Giuntoli et˜al., 2013, Sutton and Hodson, 2005; Sutton and Dong, 2012). We added “indeed” to make it clear that by “other studies” we mean the studies given afterwards.

These studies do not deal specifically with potential links at multi-decadal time-scales between the AMV and river flows per se, but their results themselves suggest that a link might exist. For example, Sutton and Dong (2012) describe a link between the AMV and precipitation in France. Based on this result, one can think that a signal on river flows may also exist.

11864-25-29: In spite of good reasoning of the study overall, the formulation of the objective of the article is quite vague.

The objectives of the study were stated more clearly p11863-26-28 in the previous version of the manuscript. As it was probably too early in the introduction to give the objectives, they have been moved near the end of introduction, just before the improved description of the different sections.

11865-06: " Series at all selected gauging stations start before 1940 and cover at least 70 yr (the median length is 90 yr)". This sentence is confusing.

The sentence has been rewritten to be clearer.

"These stations have been selected so that measurements cover at least 70 years. Series at all selected gauging stations start before 1940. Their median length is 94 years."

11866-5-10: Four precipitation data-set were used. Argue please the necessity of each one in "Data description" section.

Modification made. First, we have a high-quality dataset: HPS. Data have been homogenized and therefore HPS is particularly well adapted to the study of low-frequency variations. Unfortunately, this dataset ends in 2000 and its spatial coverage of France is poor (the number of precipitation observations in France in the early 20th century is intrinsically limited).

We also use a second dataset, GPCC, a priori less adapted to the study of low-frequency variations because of the way it has been designed (e.g. the stations within each grid cell vary with time), but that ends in 2011. As the spatial averages of GPCC and HPS over France are extremely well correlated at multi-decadal time-scales on their common period (Fig 7), we consider that GPCC is useful to extend our knowledge of averaged precipitation over France until 2011.

Precipitation from the 20CR reanalysis is also studied. The same level of accuracy cannot be expected from precipitation in a reanalysis compared to observations as HPS, especially in the case of 20CR which only assimilates sea level pressure, and don't use precipitation observations. But as no precipitation observations are used in 20CR, it provides an estimation of average precipitation in France that is totally independent from the other ones, which makes it interesting for our study.

Finally, we also use a high resolution precipitation dataset for hydrological modelling, SAFRAN. A high-resolution for hydrological modelling over France is necessary and the other precipitation products used in this work cannot deliver that high resolution.

But, mainly because the spatial coverage of precipitation observations in France is poor on the first half of the 20th century, this dataset only covers the 1961-2012 period.

The text has been modified to better justify the use of each data-set.

11867-27: Explain why did you applied exactly 19yr window for filtering.

The choice of the filter lies in a trade-off. We want to study multi-decadal variations in river flows and therefore it is necessary to filter-out higher-frequency variations and especially inter-annual variations. A large window is therefore necessary. But as we don't want to loose accuracy at the beginning and end of the series, we don't want to use tapering. Without tapering, the larger the window is, the greater is the number of points lost by applying the filter, and we don't want to loose a too large chunk of data.

With an half-amplitude point of 18 years and the complete filtering-out of variations with a period smaller than 9 years, the 19yr Hamming window is sufficient to extract the multi-decadal variations from the interannual noise, an we only loose 9 years of data at the beginning and end of the series.
Note that we tested the sensitivity of our results to different filters, and especially filters with larger windows and greater half-amplitude point. The results are not crucially dependent on this choice.

Precision added

11869 -6 : "magnitude of decadal variations in river flows is generally greater than in precipitation". What does it mean? Develop please your idea.

Modification made. We added:
“Indeed, the median over France of the ratio of multi-decadal standard deviation to inter-annual standard deviation is for example of 0.28 (0.25) in JJA (DJF) for river flows and only 0.19 (0.20) for precipitation (Fig.~3).”

11869-27: "as spring evapotranspiration in France tends to be energy-limited rather than water-limited" the reference is needed.

We added a reference (but unfortunately, it is in french.)

11870-8-13: "which could have important practical impacts" - which impact and on what? Develop please this phrase.

Modification made.

11870-16: "spring precipitation averaged over France" Which data-set was used for averaging and why?

For this analysis GPCC is used. The information was given in the caption of Fig 6. Now, it is also in the text. We use this dataset as it covers the early 21st century and therefore allows correlations with river flows to be computed on a longer period. As GPCC is very consistent with HPS (our reference data-set) on their common period, as shown in Fig 7, we are confident that GPCC provides a meaningful estimation of averaged precipitation in France. We added this precision to the article to justify our choice.

"Note that the average of precipitation over France is used here because of the lack of precipitation data specific to each river basins on the whole period of interest." In "Data" the

authors cited 4 data-sets providing precipitation. Why is it impossible to calculate the average for each basing at least from one data-set ? Next, on fig.6. authors give more details, but this manner of presentation of results does not facilitate article reading. Fig. 6b is necessary to prove the use of average over whole France value of precipitation. How "different locations (HPS data)" were selected. What was the principle? In the context as it is described in text the use of average over whole France value just simplifies the work but not justifies.

The highest quality dataset regarding temporal variations in precipitation in this study, and our reference, is HPS because data have been homogenized to remove spurious temporal variations. However, its spatial coverage of France is very poor (mainly because 100-year precipitation series over France are rare). As a result, there is not even one station from HPS in many watersheds analyzed in our study.

The two other long precipitation datasets are gridded and provide (in theory) a complete coverage of France. However, their resolution is low (0.5° for GPCC and more than 1.5° for 20CR) given the spatial variability of precipitation and the small size of many watersheds analyzed in this study. It is not possible to estimate meaningfully precipitation on these watershed based on those low-resolution precipitation datasets.
There are other important issues in this context. 20CR is an atmospheric reanalysis (that only assimilates sea level pressure and doesn't use precipitation observations) and therefore one must remain very cautious when using its estimation of precipitation. Even with the low resolution of GPCC, the number of effective observations in each grid point in France is often very small, sometimes nil, especially in the early period (it can be checked at http://kunden.dwd.de/GPCC/Visualizer by looking at “GPCC landsurface full data product version 6 0.5°”, and “numbers of stations” as product). Moreover, the GPCC dataset used here has not been optimized for temporal variability study. In particular, the number of effective observations in each grid cell varies in time. Therefore, some errors may exist in the decadal variations at each grid point in GPCC (the producer of this dataset notes that one must be careful with temporal variability study with this dataset.) It is true that we show that on average over France GPCC is very close to HPS (Fig. 7) and therefore that the spatial average is suitable. But it does not mean that GPCC is reliable for each individual grid point.

Precipitation can be estimated on each watershed with the SAFRAN dataset (and it is done in the paper for Fig 12a) but only on a shorter period (1961-2012), when the spatial coverage of precipitation observations in France is good enough.
As it is much better to study multi-decadal variability on a longer period when it is possible, the SAFRAN dataset is not used at this point of the study.

We didn't select different locations for HPS: we used all the data available. The spatial coverage is poor because (i) 100-year precipitation observations in France are rare. (ii) data have been homogenized: if a precipitation serie exists in a isolated area, without other series nearby, it is not possible to homogenized it (as the comparison with other series is necessary in this context) and it cannot be used.

For all those reasons it impossible to obtain a meaningful estimate of precipitation on the watersheds in France on the entire period studied in this paper.

As noted by the reviewer, Fig 6 justifies the use of the average of precipitation in France, as the multi-decadal variations in spring precipitation in France are large-scale.

11874-9: "...decades with below normal river flows are characterized by non-significant and moreover generally larger precipitation in summer". The sentence is not exact " non-significant and moreover generally larger"?

Modification made.

11876: In what degree the runoff/precipitation relations obtained from simulated runoff and observed are different? Why did you selected the ratio obtained from simulated runoff? Taking into account the length of your in situ data-set, it is natural to work with observed discharges.

The ratio between observed runoff and precipitation cannot be computed on the entire period of interest, as precipitation corresponding to each watersheds cannot be estimated on the entire period (see our response to a previous remark). We can compute this ratio only on the SAFRAN period (1961-2012). In Figure 12, precipitation comes from SAFRAN: it was not clear and a precision has been added). Moreover, the SWI can only be estimated with the SIM hydrological model, therefore on the limited period that we study in this section. Therefore, using observed river flows would not help to extend the period for this analysis.
Here we use the hydrological model to test the plausibility of an hypothesis. It seemed to us more logical to use all the variables from the model (SWI, runoff) in order to maintain a maximum consistency between the variables.

Note moreover that in the paper we compare simulated river flows to observed river flows (for annual, spring, and summer means). As the precipitation used to compute the observed and the simulated runoff to precipitation ratios would be the same in the two cases (SAFRAN), the comparison of the observed and simulated ratios would not add new information to the already present comparison of simulated and observed river flows.

Note that we have a good confidence in the spatial variations of SAFRAN precipitation, as it has been evaluated against observations and it uses a large number of observations.

11877-24: "This is a physically plausible mechanism, consistent with the decadal variations simulated at several stations..." Please detail how much stations? "Several stations" is not most part?

This discussion was not clear and failed to convey our point. It has been rewritten.

11877-28: It is not clear if this paragraph is an attempt to explain the absence of relation between SWI and summer discharges for some (how much?) stations or it is already the general discussion.

This paragraph is not an attempt to explain the absence of relation between the SWI and summer discharges for some stations nor is part of the general discussion. It is meant to highlight another difficulty in interpreting the results described in the section. Part of the multi-decadal variations in (spring) evapotranspiration on the period studied might be related to anthropogenic forcing and in particular sulfate aerosols. Because the temporal variations of the sulfate aerosols forcing over Europe are somewhat in phase (probably coincidentally) with the AMV, it might not be easy to disentangle the impact of AMV and the impact of aerosols on evapotranspiration during the period studied in this section. As said in the response to the previous remark, the entire discussion has been rewritten in order to be clearer. A new reference has been added to this discussion.

11878 -1: "GHG could be associated with..." The sentence is very long and general. Please rephrase and better explain your idea.

This part has been rewritten and clarified.

11880-17-20: This conclusion seems dubious as the "direct anthropogenic effects" on river flow can be also much lower that of expected

We are not sure to understand what is meant by the reviewer. Because of the lack of clarity of our original text at this point he may not have understood what we tried to say. The text has been modified to make it clearer.