Response to reviewer#3 of: “Influence of the multidecadal hydroclimate variations on hydrological extremes: the case of the Seine basin” by R. Bonnet, J. Boé and F. Habets.

First, we would like to thank the reviewer for his carefully reading and his interest in our study. The comments greatly helped us to improve the manuscript. Some points are shared with the other reviewers, especially regarding the lack of clarity of the description of the reconstruction method. We made major modifications to the description of the method, which is now in a dedicated section. A diagram was added to help the readers to follow the main steps of the method. Please find below our point-by-point answer to the comments. For clarity, all reviewer comments are in **bold**.

Summary:
Overall this is an interesting paper that develops reconstructed flows for the Seine basin back to the 1850s. Variability in the reconstructions is then assessed and linked to SSTs. While this work will be of interest to the readers of the journal I do not think it is ready for publication. It is difficult to read at present and clarification is needed at times as to what is done. A fuller discussion of the limitations, uncertainties and assumptions of the work is needed. Discussion of other modes of climate variability that could be influential would also be welcome. I highlight some issues below which the authors should address. I hope that these are seen as constructive, as I believe this could be a nice addition to the literature.

We added a paragraph in the conclusion to talk about the limitations, uncertainties and the assumptions made in this work. We also added some discussions about other modes of climate variability in the introduction, but also we better justify why we are more interested in the Atlantic multidecadal variability (AMV). These points are discussed in more details with the comments below.

In addition hydro appears twice in the title, while the paper really only deals with SSTs as a source of variability. Suggest title change to “Influence of multi-decadal variability on hydrological extremes: the case of the Seine basin. You could include SST before variability if you wish.

Thank you for the suggestion. We changed the title to: **“**Influence of multi-decadal variability on high and low flows: the case of the Seine basin**”.** We use “high and low flows” instead of hydrological extremes because as you have noted below, we do not directly study the influence of multidecadal variability on floods or droughts, but rather on high and low flows of the Seine basin, even if two droughts are analyzed in details.

Overall the manuscript needs to be thoroughly edited. There are lots of problems with the use of plurals throughout eg. Precipitations , which should always be singular.

We apologize for these language errors; a much more in-depth proofreading was carried out.

It is not clear to the reader what you mean by multi-decadal phases. This appears throughout the paper and I strongly suggest you define use of it early on or use a different term. I assume it means wetter than average or drier than average periods.

Indeed, a multidecadal phase mean a wetter than average or a drier than average period of at least 20 years. Based on river flows, the positive multi-decadal phases correspond to a succession of years with on average higher flows and conversely. We added this precision in the article, please find below the correction:

“Multidecadal river flows variations at Paris and at Aisy-sur-Armançon are generally in phase. A strong positive multidecadal phase, which corresponds to a succession of years with higher than average river flows, is visible around 1920. On the contrary, a negative phase is present around 1890 and 1960.”

“To better understand the mechanisms at the origin of the multidecadal hydroclimate variations on the Seine basin, a composite analysis between the negative multidecadal phases, which correspond to the drier than average periods identified on the Seine river flows at Poses after low-frequency filtering, and positive multidecadal phases, which correspond to the wetter than average periods, is conducted for the main hydrological variables (Figure 7).”

The abstract states the obvious a little too much, eg. Wet periods are conducive to more flooding, dry periods to droughts. I would be more interested to learn what specifically your reconstructions offer and the new insights they provide. In addition, the final sentence of the abstract gives the impression that this paper looks at dynamical chains in how ssts influence atmospheric circulation. It does not.

The abstract has been rewritten to better highlight the usefulness of the hydrometeorological reconstruction and the new insights provided by it. Please, find the new abstract below:

“The multidecadal hydroclimate variations of the Seine basin since the 1850s are investigated. Given the scarcity of long term hydrological observations, a hydrometeorological reconstruction is developed based on hydrological modelling and a method that combines the results of a downscaled long-term atmospheric reanalysis and local observations of precipitation and temperature. This method improves previous attempts and provides a realistic representation of daily and monthly river flows. This new hydrometeorological reconstruction, available over more than 150 years while maintaining fine spatial and temporal resolutions, provides an interesting tool to improve our understanding of the multidecadal hydrological variability in the Seine basin, as well as its influence on high and low flows. This long term reconstitution allows analysing the strong multidecadal variations of the Seine river flows. The main hydrological mechanisms at the origin of these variations are highlighted. Spring precipitation plays a central role by directly influencing the multidecadal variability in spring flows, but also soil moisture and groundwater recharge, which then regulate summer river flows. These multidecadal hydroclimate variations in the Seine basin are driven by anomalies in large scale atmospheric circulation, which themselves appear to be influenced by sea surface temperature anomalies over of the North Atlantic and the North Pacific. The multidecadal hydroclimate variations seem also to influence high flows and low flows over the last 150 years. The analysis of two particularly severe historical droughts, the 1921 and the 1949 events, illustrates how long-term hydroclimate variations may impact short-term drought events, with in particular an important role of groundwater-river exchanges. The multidecadal hydroclimate variations described in this study, probably of internal origin, could play an important role in the evolution of water resources in the Seine basin in the coming decades. The way in which the associated uncertainties are accounted for in future projections remains to be addressed”

The literature review mentions the AMV as an important source of variability, but what about others modes of variability eg. The NAO, East Atlantic Pattern etc. There is more than just the AMV at play and it is my view that this should be reflected in the literature review. These other modes of variability could also be examined for their relationships with the extremes in the reconstructed flows or discounted if the literature has previously addressed these issues.

We agree that there are other modes of variability influence river flows and hydrological extremes over France, but not necessarily at multidecadal time-scales, which are the center subject of the paper. As our paper focuses on multi-decadal hydrological variations, we limited the literature review to modes of variability important at multidecadal time scales: basically, over France, to the AMV. Indeed, the NAO and EAP are first and foremost interannual modes of variability. Even if Folland et al. (2015) suggest a link between the NAO and the AMV, this link does not seem totally clear (Guan & Nigam 2009). Following your suggestions and the ones of reviewer 2, we added a quick discussion about other modes of variability and we now make it clearer that we are interested in multidecadal variations. We also largely improved the discussion about the link between the AMV and hydrometeorological variations over Europe.

“A large body of work has dealt with river flows variability over Europe. At interannual time scales, the large scale atmospheric circulation over the North-Atlantic plays a major role in hydrological variations over Europe (Kingston et al., 2006b, a; Bouwer et al., 2008; Steirou et al., 2017). The North Atlantic Oscillation (NAO) (Cassou et al., 2004; Hurrell and Deser, 2009) in particular is known to influence river flows over Europe, mainly in winter (Kingston et al., 2006b, a; Bouwer et al., 2008; Steirou et al., 2017).

Regarding the uncertainties in impact projections due to internal variability, variations at longer time-scales, i.e. at decadal and multidecadal time-scales, are far more important than interannual variations, as they may modulate the hydroclimate state on several decades, i.e. at climate time scales. They may temporarily reinforce or reduce, or even reverse, especially over the coming decades, the long-term impacts of climate change.”

In your critique of 20CR on page 2 line 30 you explicitly state that that “this approach is far from optimal, as it does not make use of the long-term meteorological observations that may exist”. I would strongly recommend a citation for this or remove.

The "it" of "it does make use" refers to the reconstruction methods only based on 20CR and does not refer to 20CR. We were talking about local meteorological observations (e.g. precipitation or temperature) that can be used, in addition to 20CR, to obtain the local meteorological forcing necessary for hydrological modelling. Please, find below the new version of the paragraph:

“To move forward, long-term hydrometeorological reconstructions based on hydrological modelling have been developed (e.g Kuentz et al. 2015; Caillouet et al. 2016). Due to the scarcity of meteorological observations in the early 20th century (Minvielle et al., 2015), the meteorological forcing needed for hydrological modelling must first be reconstructed. The recent release of long-term global atmospheric reanalyses (e.g. Twentieth Century Reanalysis (20CR, Compo et al. 2011) from the National Oceanic and Atmospheric Administration (NOAA)) opens great opportunities in that context. Statistical downscaling methods, typically used in climate change impact studies, can be applied to derive the high resolution meteorological forcing necessary for hydrological modelling from these global atmospheric reanalyses, as in Caillouet et al. (2016). This approach presents two main limitations. First, the quality of the reconstruction depends on the quality of the reanalyses. As the density of assimilated observations (e.g. surface pressure in NOAA 20CR, Compo et al. 2011) strongly evolves over time, potential unrealistic trends and/or low frequency variations may exist (Krueger et al., 2013; Oliver, 2016; Bonnet et al., 2017). Second, this approach does not take advantage of the long-term local meteorological observations that may exist.”

There are uncertainties not considered in this paper related to input data, hydrological modelling, other datasets used etc. It would be welcome if the discussion could, in a paragraph, flesh these out.

Indeed, there is a chain of uncertainty associated with the development of our hydrological reconstruction. We added a long discussion on the uncertainties and limits of the method in the conclusion:

“Although the reconstruction developed in this study is an interesting tool for studying the past variability of the hydrological cycle over the Seine basin, it is obviously not perfect. Uncertainties are present throughout the modelling chain. The statistical downscaling method used at the first step of the reconstruction method assumes that the learning period, over which the large-scale reanalysis and the Safran analysis overlap (1959-2010), is representative of the meteorological conditions of the 1851-2010 period. The consistent performances of the reconstruction over the entire period shown by several analyses in this study suggest that this hypothesis has no major impact on our results. Important uncertainties are associated with the 20CRv2c reanalysis at the beginning of the period, due to the smaller number of assimilated observations (Krueger et al., 2013). We use monthly homogenized local precipitation and temperature observations to constrain the results of statistical downscaling in order to improve the temporal homogeneity of the reconstruction, but the homogenization method is not state-of-the-art. The good agreement between the low-frequency variations of the homogenized monthly precipitation series and of the Global Precipitation Climatology Centre dataset (Schneider et al., 2008) from 1901 to 2011 (not shown) still gives good confidence in the overall realism of the multidecadal variations described in this study.
The hydrological model used in this work is also not perfect, with potential consequences on our results. The amplitude of the variations in the piezometric levels at Toury in particular is strongly underestimated. Although this is a one-point measurement, it is possible that this underestimation exists for the entire Beauce aquifer, which could imply an underestimation of the multidecadal variability of reconstructed river flows, especially in summer when the role of groundwater is particularly important. This underestimation is likely due to hydrological modelling, and could come from a too simple representation of the limestone aquifer that neglects some confined parts. The river channel dimensions and conveyance capacity are also supposed constant over time, whereas they could be affected by multidecadal climate variability (Slater et al., 2019).
Some important uncertainties are also associated with the observations used for the evaluation of the reconstruction. They may indeed be influenced by non-climatic anthropogenic influences such as dams and pumping, not taken into account in the hydrological model, or changes in measurement methods for example. We have evaluated the reconstruction against multiple observations in an effort to ensure that our results are not too dependent on a particular source of information and to better understand the potential limitations of the reconstruction.”

As mentioned, the methodology is complex and difficult to follow. The use of analogues to derive daily precipitation seems to move from 2800 samples to 60 to 3. Why three, why such a decrease in sample size and are you left with a large enough sample?

We totally rewrote a large part of the method section in the revised manuscript, by adding some explanations, in particular on the method used in Bonnet et al. (2017) (as asked below in a comment consistently with the other reviewers), on which we based the new method developed here. It should be clearer now. See also our answer to following comments below.

To respond to your comment, an important objective of our hydrometeorological reconstruction is to improve the representation of daily river flows. As a result, a balance has to be found for the number of analogue days to be used during the monthly constraint. The greater the number of analogue days used for the monthly constraint is, the farther some analogues are from the target day, with likely in the end a degradation of the representation of precipitation and temperature, and therefore of river flows. On the other hand, too few analogue days could limit the improvement in low frequency variations expected from the monthly constraint. We therefore made different tests in order to find the best number of analogues to retain at the different steps.

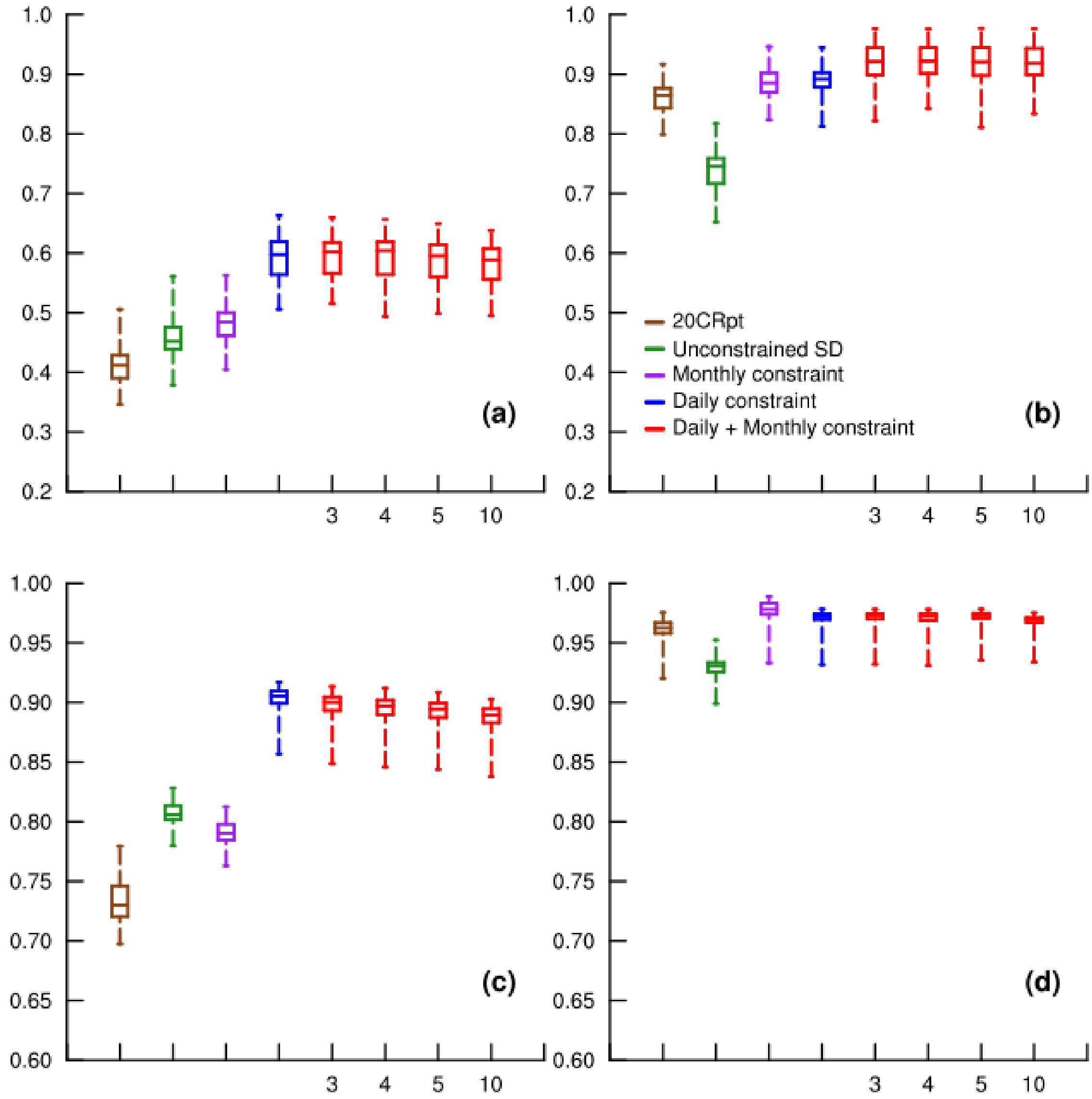

Figure R1: Spatial distribution of the correlations between (a-c) daily and (b-d) monthly (a-b) precipitation and (c-d) temperature over the Seine basin. Correlations are calculated between the Safran analysis (considered as observations) and the precipitation and temperature derived from (brown) the reconstruction developed in Bonnet et al., 2017, (green) the downscaling method alone, (purple) the downscaling method only constrained by monthly precipitation and temperature, (blue) the downscaling method only constrained by daily precipitation and (red) the downscaling method constraint by daily and monthly precipitation and temperature, based on different tests for the number of analogs used for the monthly constraint (X axis). The correlations are calculated on the 1958-2005 period and the series have been deseasonalyzed beforehand. The boxplots show the minimum/25th percentile/median/75th percentile and the maximum.

This is illustrated in figure R1, which shows the results of one of these tests. It shows that the daily correlations between reconstructed and observed temperature and precipitation decrease when the number of analogue days increase (Figure 1 a and c). After testing different possibilities, we decided to keep the 3 best analogue days from the daily constraint. These 3 analogue days are then used to apply the monthly constraint. With 3 analogues, the ensemble is large enough for the monthly constraint.

Figure R1 also shows that the double constraint method, at daily and then monthly time scales, greatly improves the daily correlations of precipitation and temperature compared to the statistical downscaling method alone, or to the downscaling method only constrained at monthly time scale (Figure 4.11 a and c).

Similarly there are assumptions made that things like channel dimensions and conveyance capacity remain stationary over time. Recent research has shown that they also react to variability Slater, L.J., Khouakhi, A. and Wilby, R.L., 2019. River channel conveyance capacity adjusts to modes of climate variability. Scientific reports, 9(1),pp.1-10.

Thank you for this interesting publication. Indeed, our hydrological model, as most hydrological model used in the literature, doesn't take into account the possible changes in channel dimensions or conveyance capacity due to multi-decadal climate variations. We added this information in the paragraph that now discusses the different uncertainties (see two points above).

In terms of your methods does your use of analogs constrain the variability to that of the shorter record? The methodology is complex and hard to follow as it is written. Previously published work on which this data is based needs to be shortly summarised. In addition it would be helpful to the reader to provide a flow diagram of the key steps in the workflow. For example, I have read the paper multiple times and at the start of the results I am unsure what the reference simulation is.

We acknowledge that our method section was not clear enough and needed to be improved. Indeed, the final variability is limited by that of the shorter record, as in the end, an analogue is used. We added a paragraph that summarizes the main ideas of our method in order to clarify the method section (below). We also added a diagram to clarify the explanation of the method. Please find the new method section below.

"3 Development of the Seine hydrometeorological reconstruction

A new hydrological reconstruction, based on hydrological modelling, is developed over the Seine basin, improving the method presented in Bonnet et al. (2017), with two main objectives: (i) to extend the study period to the 1850s, in order to characterize more robustly multidecadal hydroclimate variations, and (ii) to improve the representation of river flows, particularly at the daily time scale, in order to obtain a better representation of high and low flows and study their multidecadal variations. Figure 2 describes the main steps of the method developed in the present study and highlights the improvements over the one used in Bonnet et al. (2017).
To obtain the meteorological forcing necessary for hydrological modelling the main idea of the Bonnet et al. (2017) method is to use the analog method (Lorenz, 1969), a stochastic statistical downscaling method, to downscale a long-term atmospheric reanalysis such as NOAA 20CRv2c and produce an ensemble of trajectories of precipitation and temperature over France (Step 1, Figure 2). Then, local long-term monthly precipitation and temperature observations are used to select the best trajectory.

The analog method is based on the hypothesis that two days with similar large scale atmospheric states (e.g. large scale atmospheric circulation over the North Atlantic) are characterized by similar local weather conditions. In its most basic form, for each day D of the reanalysis, the day Da (the so-called analog day), with the closest large scale atmospheric state is searched in the learning period, defined as the common period between the reanalysis and the observational database with the local variables necessary for hydrological modelling, e.g. here the Safran analysis. The local variables of interest of the day Da in the observational database are selected as an estimate of the local weather conditions for the day D. To quantify the similarity between large scale atmospheric states, four predictors are used in the present work: precipitation, surface temperature, sea level pressure and specific humidity at 850 hPa. An Euclidean distance is computed for each predictor, except for sea level pressure, for which the Teweles and Wobus score (Teweles Jr and Wobus, 1954; Obled et al., 2002) is calculated. The distances and the score are then combined after standardization to give the same weight to each predictor. Two domains of analogy are used. The domain for sea level pressure is delimited by the following coordinates: 44°N, 56°N, -11°E, 16°E. The domain for the three other predictors is defined by 46°N, 51°N, -2°E, 7°E.

In Bonnet et al. (2017), instead of searching only for the best analog day Da for each day D of the reanalysis, the N best analog days were selected Da1, Da2, ... DaN, with N = 10. The corresponding maps of precipitation Pr(Da1), Pr(Da2), ... Pr(DaN) and temperature Tas(Da1), Tas(Da2), ... Tas(DaN) from Safran constituted different estimates of precipitation and temperature for the day D. Multiple trajectories of precipitation and temperature over the domain of interest were then created by repeatedly selecting randomly one of the 10 analog days for each day D of the reconstruction. In practice, 5000 trajectories were created. The monthly averages of precipitation (temperature) for these trajectories were computed. From the 10 different maps of precipitation (temperature) over France obtained with the analog method (as N = 10) for each day D of the reanalysis, 5000 different monthly maps of precipitation (temperature) were obtained with this procedure (Bonnet et al., 2017).

For each month of the reconstruction, the 5000 maps of precipitation (temperature) obtained on the Safran grid were regridded and compared to the actual observed precipitation (temperature) map, using the long-term homogenized precipitation (temperature) series over France (see section 2) as reference. Regridding simply consisted in selecting the Safran grid point the closest to the long-term homogenized precipitation (temperature) stations. The spatial root mean square errors (RMSE) were computed for temperature and precipitation. The sum of the RMSEs corresponding to precipitation and to temperature was then computed, after the temporal standardization of the series of RMSEs in order to give the same weight to each variable. In the end, for each month of the reconstruction, the daily series of analog days among the 5,000 ones that leads to the lowest sum of RMSEs was selected. The term "monthly constraint" used in this study refers to this last step (it corresponds to Step 3 in Figure 2).

This approach benefits from the advantages of the analog statistical downscaling method. From the analog days, all the meteorological variables from Safran necessary to force the Surfex-AquiFR hydrological model were obtained. The spatial and inter-variable consistencies were maintained after this procedure, because for each day of the reconstruction the entire map of precipitation (and temperature, humidity etc.) over France from Safran was selected based on a single analog day. Compared to a basic statistical downscaling method, this approach allows additionally taking into account local observations in the downscaling process and not simply large scale information. This approach is therefore more accurate, as shown in Bonnet et al. (2017). Note that the temporal consistency of the meteorological forcing is ensured by both the temporal consistency of the predictors and of the local observations.

In the present study, to extend the study period, the long-term NOAA 20CRv2c atmospheric reanalysis (Compo et al., 2011), which begins in 1851, is used. This reanalysis is based on a global atmospheric model, using observed sea ice and sea surface temperature as boundary conditions, and with the assimilation of surface and sea level pressure observations. 56

members, sampling the reanalysis uncertainties, are available. Compared to Bonnet et al. (2017) where only one member of the long term reanalysis is downscaled, here we statistically downscale with the same analog method as described above, the 56 members of NOAA 20CRv2c. It leads for each day D of the reconstruction period to a much larger pool of analog days, which allows adding a new step: a daily constraint with local observations (Step 2, Figure 2). The objective of this additional daily constraint is to obtain a better representation of the daily variations of the meteorological forcing.

As previously, for each day D of the reconstruction period (1852-2008) of a given member, the N best analog days Da1, Da2 ... DaN in the learning period (1958-2008, limited by the availability of Safran) i.e. with the most similar large-scale atmospheric states are searched. In the present method, N = 50. As the 56 members of NOAA 20CRv2c are downscaled, in the end 2800 potential analog days are obtained for each day D of the reconstruction period (with potentially similar analog days for the different members). As each analog day corresponds to a day of the learning period, the corresponding daily maps of precipitation and temperature from Safran are selected and compared to the daily station observations (SQR, see section 2.1) after regridding. Regridding consists in selecting the Safran grid point the closest to each observation station over the Seine basin. Note that the number of stations varies on the 1852-2008 period. The comparison is therefore done each day of the reconstruction with the available stations.

The daily comparison is based on the following approach:

(i) The average daily bias in mean precipitation averaged over the Seine basin is calculated for the 2800 analog days, and the 60 analog days with the lowest bias are selected.

(ii) The spatial root mean square errors for the 60 analog days are calculated for temperature. For precipitation, the error to the cubic power rather than to the square power is used, in order to give more weight to strong values of precipitation, and the absolute value is used.

(iii) The daily series of spatial errors obtained for precipitation and temperature are then standardized based on the statistics of the entire period and added, with a weight of 1 for precipitation and 0.5 for temperature.

(iv) Finally, each day of the reconstruction period, the 3 best analog days (out of 60), i.e. with the smallest errors, are selected.

Based on these 3 selected analog days, a monthly constraint is then applied as in Bonnet et al. (2017) and described above, except that the number of analog days is different (3 versus 10) (Step 3, Figure 2). Multiple trajectories are created by repeatedly randomly selecting one of the 3 analog days for each day D of the reconstruction. The monthly averages are computed over the Seine basin and then long-term monthly homogenized precipitation and temperature series are used to select the best overall trajectory. The interest of using a monthly constraint after the daily constraint is that monthly data are homogenized contrary to daily data (Section 2) and therefore it allows for a better representation of low-frequency variations.

Multiple tests have been conducted to set-up the different ad-hoc aspects of the method, trying to obtain the best overall hydrometeorological reconstruction. These tests concern, for example, the best combination of weights given to precipitation and temperature errors, the number of analogs selected at each steps etc. For example, selecting only the 3 best analog days leads to best overall performance in capturing daily and monthly variations. Using more analog days may allow for a better representation of monthly variations but degrade the representation of daily variations.

To sum up, the hydrometeorological reconstruction developed on the Seine basin is constrained on a daily basis over the period 1885-2003 by observations of precipitation and temperature (SQR), on a monthly basis over the period 1885-2005 by homogenized observations of precipitation and temperature (SMR), and over the 1852-1884 and 2005-2008 periods by the monthly series of precipitation at Paris (Slonosky, 2002) (see section 2.1 for more details). The results, especially at the daily time scale have therefore to be interpreted with more caution over the period only constrained by the monthly series of precipitation.

During the development of the reconstruction, mean climatological biases were found on reconstructed precipitation and incoming shortwave radiation with comparison to Safran on their common period. These mean climatological biases are simply corrected based on Safran as reference before forcing the hydrological model.

The meteorological forcing obtained on the 1852-2008 period with the approach described in this section is finally used to force the Surfex-AquiFR hydrogeological model to obtain the hydrological reconstruction over the Seine basin (Step 4, Figure 2).

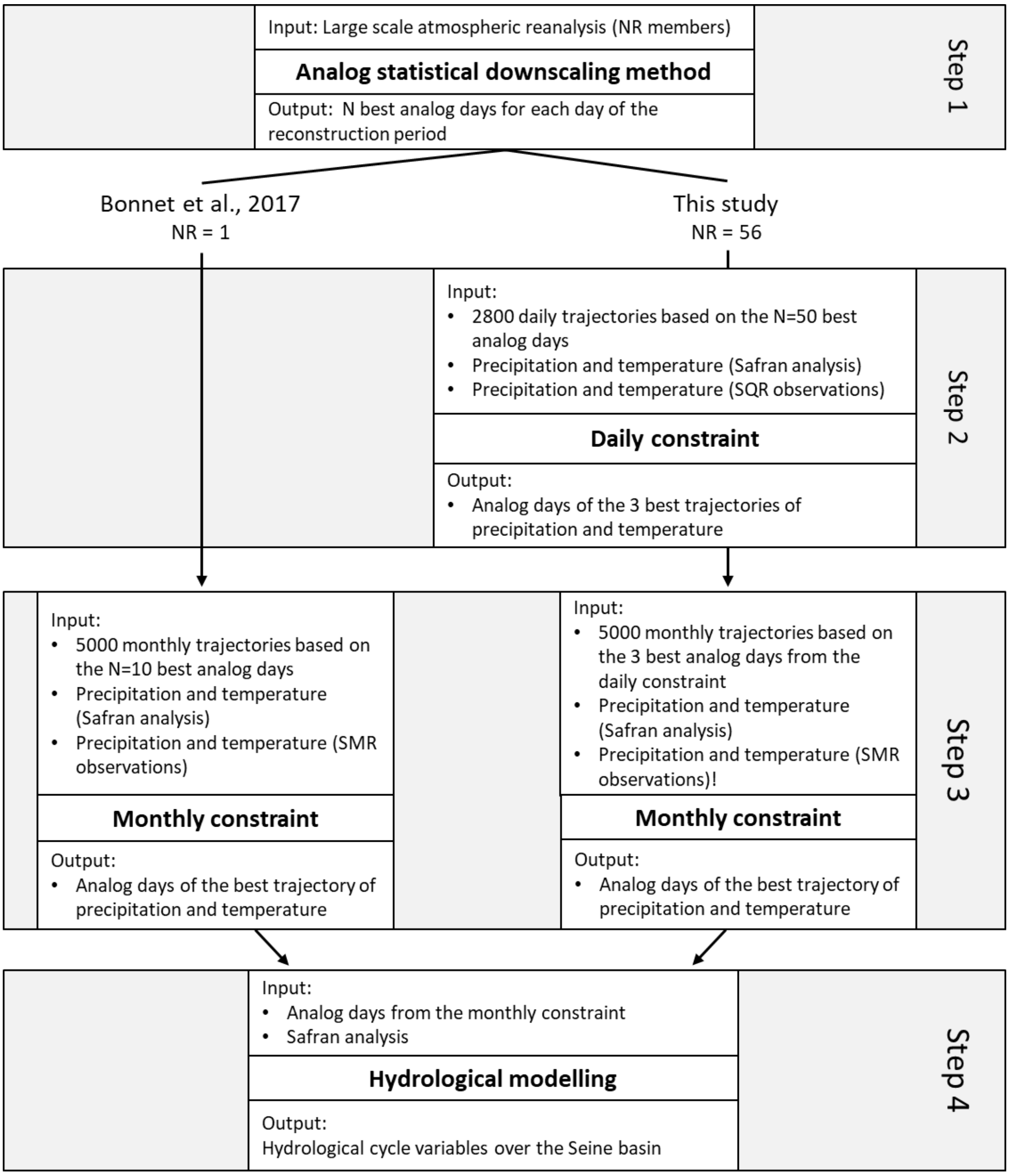

Figure 2. Schematic representation of the methods used to obtain the hydrological reconstruction, with on the right the method used in this study and on the left the method used in Bonnet et al. (2017), which doesn't include the daily constraint."

I don’t like the word anthropized – heavily impacted would work just as well.

We changed the word anthropized to “heavily impacted by human activities”.

It appears that the reconstructions persistently underestimate the annual maxima (fig 4). What does this say about your reconstruction, does it affect your results and how might this be improved?

The annual maxima are very similar in the reconstruction and in the reference simulation based on the same Safran-Surfex-AquiFR hydro-meteorological system over their common period (1959-2014). Therefore, the differences of the annual maxima are not related to the quality of the reconstructed meteorological forcing. They may be due to the hydrological model, which would cause a general underestimation of the annual maxima (it is true the annual maxima are closer to the observations after 1960, but it is likely the result of the construction of flood-controlling dams, not taken into account by the hydrological model).

A mean bias does not affect much our results as we are mostly interested by the variability and the variability of the annual maxima is well reproduced, over the entire period. Note also that the selection of high flows is based of percentiles rather than in terms of absolute river flows. The historical flood of 1910 for example is the biggest event in our reconstruction and in the observations.

Have the piezometric level data been quality assured?

A fairly high level of confidence is placed in the piezometric observations of the Beauce groundwater at Toury. Indeed, the piezometric station at Toury is managed by a sugar beet factory consistently over decades. As one of the oldest dataset of piezometric head over France, it is used as a reference by water management agencies, especially the one that manage the beauce aquifer. It is, therefore, validated by experts. Additionally, the observations at Toury in the recent period are consistent with the ones observed at nearby stations, which give us also confidence on this series. If some uncertainties on the change of measurement method could exist, it hasn’t a strong influence on the observations quality as the several meters variations observed are far above the order of magnitude of traditional measurement error.

It is stated on page 8 line 29 that “The length of the reconstruction allows to show that multidecadal variations are also present before the 20th century.....” Is this not expected. In the same sentence what do you mean by negative phase.

By negative phase, we mean a succession of years with lower than average river flows. The fact that the reconstruction shows multi-decadal variations before the 1900s was not obvious and is an interesting result. The long-term available river flow observations show the beginning of a negative multidecadal phase in this period (Figure 3c-d), but it is difficult to affirm something regarding this period based only on observations, due to observational uncertainties. Moreover, the internal climate multidecadal variability (that is assumed here as the main driver of these variations) in the North Atlantic region might not to be constant over time (Qasmi et al., 2017). The context, which justifies the interest of this result, was not well described in the first version of the manuscript. We improved that point in the new version.

“Annual river flows of the Seine at Poses (Figure 1) from the reconstruction show strong multidecadal variations during the 20th century (Figure 7), consistently with the observed variations over France described in Boé et Habets (2014). As said in the introduction, major sampling uncertainties exist when dealing with multidecadal variations on the short observational record. Additionally, since the mid-20th century, the climate has been strongly impacted by anthropogenic forcings, making it difficult to disentangle the respective role of internal variability and external forcings in observed hydroclimate variations.

It is therefore very interesting to note, thanks to the extended length of our reconstruction compared to previous works, that multidecadal variations also exist before the 20th century, with lower than average river flows around the 1885-1905 period (about 15\% lower than the average of 446m3/s for annual river flows). It reinforces our confidence in the reality of multidecadal hydrological variations in France observed over the 20th century and in the idea that they are at least partly of internal origin. Phased multidecadal variations also exist for the seasonal averages, with the strongest absolute variations seen in spring and winter. Note however that as climatological river flows are smaller in summer and early fall, the multidecadal variations seen in these seasons are also important."

Throughout you could quantify these above and below average phases in terms of the magnitude of their anomaly.

A quantification of the intensity of these phases was carried out in section 4.1 of the manuscript, where the relative difference between the negative and positive multidecadal phases for the four seasons are calculated. Variations ranging from 30 to 40% are visible in summer, for example, which can have serious impacts in terms of water resources. Therefore, the positive and negative multidecadal phases broadly correspond to variations of 15 to 20% in relation to the anomaly of the Seine river flows.

Indeed what can your work tell us about an appropriate baseline for assessing changes in the Seine flows – one that accounts for the types of variability you see. Thirty year baselines often only sample one component of multi-decadal variations.

This work indicates that considering a baseline of only 30 years, to dimension structures such as drinkable water plants or irrigation dams, is clearly insufficient and even dangerous from a practical perspective. The use of periods of at least 60 years would be much better. We added this interesting point on the new version of the conclusion.

"The results described in this paper illustrate how dangerous it can be for practical purposes to rely on short periods of few decades (e.g. three) to characterize hydrological hazards."

There are a number of statements that require clarification or a more precise wording, such as: multi-decadal phases? P7 line 10 over what period are correlations derived? Page 8 line 23 A partially captive part of the aquifer is not represented in the model – what is a partially captive part? Pg 11 line 11 a large number of stations have a ratio between one and two. . ..with a ratio of 1.1 consistent with the reconstruction in that region. I can't interpret this. Pg 11, line 24 strong positive multidecadal phase?

Clarifications have been made in the text on these different points. For the first point, which is probably the most important one, a more precise definition of "multidecadal phases" is now given (see also a previous comment).

For the second point, the period considered for the correlation is indicated in the legend of the figure, but we added it in the text to be clearer:

"The medians of the correlations with observations over the 1958-2005 period are respectively of 0.7 and 0.97 for daily and monthly river flows (Figure 3)"

Regarding the third point, a part of the Beauce aquifer contains lenses of clay, which divide the aquifer in a confined and unconfined part, and modify the hydrodynamic of the aquifer flow. These lenses of clay are not represented yet within the model because of his too coarse spatial resolution. This too simple representation of the limestone aquifer that neglects these confined parts could induce the underestimation of the piezometric levels variations. We added this clarification in the article.

“A partially captive part of the limestone aquifer, which contains lenses of clay, is not represented in the model due to his insufficient resolution. This part of the aquifer amplifies the flow time and, therefore, the memory of the aquifer.”

For the fourth point, we changed the second part of the text, which could be confusing:

p11, l24: “The ratio of low flow days between negative and positive multidecadal phases is indeed between one and two for a large number of stations (Figure 13b). For the observations at Aisy-sur-Armançon (triangle located in the south east of the basin) the ratio is 1.1, consistent with the reconstruction.”

For the last point, as for the first point, the strong positive multidecadal phase mean a wetter than average period. We choose here two major historical droughts, which occur during opposite phase of the multidecadal variability of the Seine river flows. This is interesting to look at the impact of the multidecadal variability describe before on particular extreme events.

Is it fair to call a day with flow above the 95th percentile a flood day. Is it fairer to say a high flow day? The same issue applies to drought days, are these more fairly described as low flow days? In reality you are not strictly looking at floods and droughts in this part of the analysis.

It is true that the flows above the 95th of below the 5th percentile are not rare enough to be qualified as flood or drought days. We changed these terms “floods” and “droughts” to high and low flows in the revised version of the manuscript.

It would be useful to outline a potential or plausible chain of causation as to how SSTs in the North pacific relate to Seine flows.

The potential mechanisms linking SST anomalies are still not well understood. We added some information on that point in the new version of the manuscript (below).

“Significant SST anomalies between the negative and positive multidecadal phases of the Seine river flows are also observed in the North Pacific. SSTs there are significantly warmer during the negative multidecadal phases of Seine river flows compared to the positive phases. The potential mechanisms linking the SST anomalies in the North Pacific and multidecadal hydroclimate variations over France are not clear. Ding et al. (2017) suggest that the Pacific decadal variability (PDV) (Mantua and Hare, 2002) combined with La Niña events could result in precipitation and temperature anomalies over Europe by favoring NAO-like circulation patterns. The phase of the NAO and therefore the sign of precipitation and temperature anomalies is dependent on the type of La Niña. Note also that the apparent link between the North Pacific SSTs and Seine river flows might not be causal. Indeed, the SST anomalies in the North Pacific might be to a certain extent also driven by the AMV. Ruprich-Robert et al. (2017) indeed suggest that the AMV may influence North Pacific SSTs, mainly through an atmospheric teleconnection originating from the tropical Atlantic.”

Do your extended flows add additional insight into AMV related variations in flows than done in previous work. Has the magnitude of anomalous periods been the same or different in your reconstructions relative to previous work?

Indeed, our hydrometeorological reconstruction allows adding additional insights to our comprehension of the link between the AMV and the river flows. The first contribution of this work is that it allows to confirm on a longer period the hypotheses made by Sutton and Dong (2012) and Boé and Habets (2014). By using a longer period of study, we are able to use two more multidecadal phases in our analysis, which makes it more robust. Our work also highlights that this influence is not limited to river flows over France, but also concerns high and low flows over France.

Note that a new analysis has been added to section 6 on the "Role of large-scale circulation and influence of ocean variability", with interesting new results.

Please find below the new analysis and the figure associated:

“In order to further investigate the relationship between the AMV and the multidecadal hydroclimate variations over the Seine basin, the lagged correlations between the AMV and spring river flows and precipitation are computed. Significant anti-correlations (around -0.8 and -0.9) are found between the AMV index calculated with the SSTs from the 20CRv2c reanalysis and low-frequency filtered spring precipitation over the Seine catchment, with a lag of about 10 years, consistent with the lag found by Boé and Habets (2014) (Figure 12a). Similar results are obtained using the AMV index from Wang et al. (2017). This paleoclimate reconstruction, based on multiple proxies, is largely independent of the first AMV index. Given the length of the paleoclimate AMV index of Wang et al. (2017) and the availability of a very long observed series of precipitation at Paris (Section 2), it is possible to investigate the temporal stability of the AMV / spring precipitation relationship. Very interestingly, similar results are obtained on the 1779-1889 and 1890-1990 periods, highlighting the robustness of the relationship.

The influence of the AMV on spring precipitation is reflected on spring river flows. Significant anti-correlations are found between the AMV and the river flows of the Seine at Paris at multidecadal time scales on the 1876-1985 period, with a lag of about 10 years, as for precipitation (Figure 12b). The results obtained with the observations and the reconstruction are very similar.

As said in the introduction, the 20th century is a very short period to characterize multidecadal climate variations. On such a short period, the sampling uncertainties are extremely large, which may question the robustness of the results. Additionally, non-stationnarities may exist (Cassou et al., 2018) and the strong influence of anthropogenic forcings on the climate of the second half of the 20th century may complicate the interpretation of observed multidecadal variations. The results previously described are therefore important because they confirm those obtained in previous works, but on a period considerably longer, at least from the end of the 18th century. They show the robustness of the teleconnection between the AMV and spring precipitation over France described in Sutton and Dong (2012) and confirm the existence of the lag suggested in Boé and Habets (2014), even if its physical origin is still unknown. These results also confirm that the multidecadal hydroclimate variations observed over France are likely not mainly the result of anthropogenic forcings.

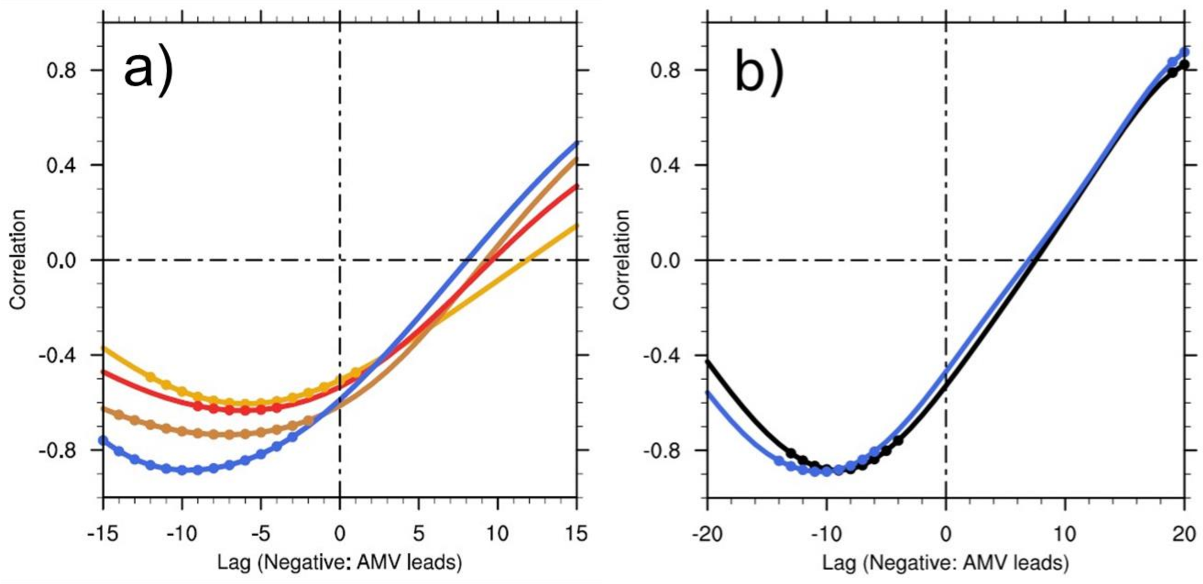

Figure 12. (a) Correlations between the paleoclimate AMV index of Wang et al. (2017) and the long series of annual precipitation observed at Paris (Slonosky, 2002) over the period 1780-1889 (yellow), 1890-1989 (brown) and 1779-1989 (red). In blue, the correlations between the AMV index calculated from the SST of the 20CRv2c reanalysis and the precipitation from the reconstruction over the period 1882-1979. The AMV index is defined as the average of the North Atlantic SSTs, from which the forced signal has been removed (Deser et al., 2010). Here, the forced signal is estimated with the trend calculated from the EEMD method (see section 2.4). (b) Correlations between the AMV index calculated with the SST of the 20CRv2c reanalysis and the spring river flows from (blue) the reconstruction at Paris and (black) the observations at Paris Austerlitz over the period 1882-1979. The 20CRv2C reanalyses SSTs use SODAsi.2 (Giese et al., 2016) for latitudes between 60N and 60S and COBE-SST2 (Hirahara et al., 2017) for the highest latitudes. The trends in precipitation and river flows series are calculated with the EEMD algorithm and subtracted from the original series before the calculation of the correlations. The series are filtered with a Lanczos filter with a 21-year window. The points represent significant correlations with p<0.05 (at least 95% of the null hypothesis is rejected) according to the "phase-scrambled bootstrapping test" (Davison and Hinkley, 1997).”

Section 4.1 is hard to follow, it might be worthwhile organising by season. At times the text jumps from season to season

Agreed. Modification made.

I think section 6 needs to come earlier in the paper, given it importance to the message of the paper.

Agreed. The section 6 and the section 5 have now been switched.