Response to reviewer#1 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, interest in our study and the insightful comments that helped us to improve the manuscript. Some points are shared with the other reviewers, especially on the description of the method used to develop the hydrometeorological reconstruction, which was not clear enough. We made major modifications to the description of the method, which is now in a specific section. A diagram is now added to bring more clarity. The introduction was also greatly improved. We also added a new figure in the section about the “Role of large-scale circulation and influence of ocean variability”, which improves and completes the insights of this section (now the 6th section). Please find below the answers to the comments point-by-point. For clarity, all reviewer comments are in **bold**.

<u>Summary:</u>
The authors present a hydrometeorological reconstruction established using a combination between a statistical downscaling method and data assimilation. This reconstruction is then used to study the multidecadal hydroclimate variability of the Seine basin, as well as the influence of the multidecadal variations on extreme events. The paper is well constructed and address relevant questions about the mechanisms of hydrological variability on the Seine basin. Figures are clear, and results and conclusion are relevant. I mostly have questions about some details in the methodological part.

Introduction:
p.2 l.22-32 – After reading this paragraph, we think that statistical downscaling method are really not appropriate and that these methods should be discard. But in fact, you use this method with a second step combining observations with the results of a statistical downscaling method to improve them. Maybe it would be worth to reformulate some sentences of this paragraph to explain that these methods are not enough to well characterize climate variability, and that we should use an ‘add on’ or an ‘evolution’ to take into account observations. This couldn’t be considered as a completely different method.

The third reviewer also highlighted the lack of clarity of this paragraph. Please, find below the new version of the paragraph added in the revised manuscript:

“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.

Given these limitations, following the same general idea as Kuentz et al. (2015), Bonnet et al. (2017) presented a new hybrid method that combines available long-term monthly observations of precipitation and temperature with the results of a statistical downscaling method applied to long-term atmospheric reanalyses. Compared to standard dynamical or statistical downscaling methods that only use large scale information and do not take advantage of local observations (e.g. temperature and precipitation) a more realistic representation of local hydroclimate variations may be obtained (Bonnet et al., 2017)"

p.2 l.33 to p.3 l.3 – Indeed, using observations could improve the results of a statistical downscaling method. But observations are also inhomogeneous so at the end, isn't it the same disadvantage than statistical downscaling methods? Inhomogeneous trends? Combining observations does not cancel the disadvantages of the downscaling (so we have inhomogeneous trend from the reanalysis + inhomogeneous trend from the observations?).

Indeed, uncertainties are also present in the observed precipitation and temperature series used to constrain the results of the statistical downscaling. They are however still interesting to use as they provide local information that are not given by large-scale reanalyses. The third reviewer also pointed out that a discussion on the different uncertainties associated with the reconstruction method was missing. We added a discussion about this in the conclusion, including your comments on the observation. Please find below the paragraph:

"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."

Why only talking about statistical downscaling methods? What other methods could be used to reconstruct the past (dynamical downscaling, weather generators, ...) and why is it better to choose statistical downscaling + data assimilation?

We think that using only dynamical downscaling or weather generators shares the same weaknesses associated with statistical downscaling in that context: only large scale information is used. We now make it clear in the revised manuscript:

"Given these limitations, following the same general idea as Kuentz et al. (2015), Bonnet et al. (2017) presented a new hybrid method that combines available long-term monthly observations of precipitation and temperature with the results of a statistical downscaling method applied to long-term atmospheric reanalyses. Compared to standard dynamical or statistical downscaling methods that only use large scale information and do not take advantage of local observations (e.g. temperature and precipitation) a more realistic representation of local hydroclimate variations can be obtained (Bonnet et al., 2017).

p.3 l.6 Why the Seine basin? Can you add some explanations?

We added a point to justify the choice of the Seine basin in the new version of the introduction. Please find below the explanations of this choice:

"Focusing on the Seine basin (Figure 1), one of the main French river basins, we are able to extend the reconstruction back to the 1850s. A major interest of this basin is indeed the existence of a few long and varied series of observations, which are useful either to develop or evaluate the reconstruction method"

P3 l.8 Extending until 1850 leads to using 20CR between 1850 and 1900 (quality - and very large dispersion over the 56 members) + using very few observations (so inhomogeneous trends). How can we drive conclusions over this period given the poor data quality?

This period has indeed to be interpret with caution. In our case, as the monthly observations used in the study are homogenized and as we considered only stations with no missing values, it is more likely that adding this information improve the quality of our reconstruction. Additionally, the combination of large scale atmospheric reanalysis and long-term observations has shown promising results in previous studies (Kuentz et al., 2015, Bonnet et al., 2017). Even if there are a lot of uncertainties between the years 1850 and 1900, the use of different kind of dataset allows to derive some common conclusion. As said 3 points before, we added a discussion about the uncertainties related to the reconstruction in the conclusion section.

p.3 l.16 – Are the observations independent for the evaluation?

All the hydrological variables except precipitation used in the evaluation (river flows, aquifer levels) are independent of the hydrometeorological reanalysis. This is why we focused the evaluation on these variables. Only observation of precipitation and temperature are combined with the results of the statistical downscaling method in the reconstruction method and are, therefore, not independent from the hydrometeorological reconstruction. We added this clarification in the revised version of the manuscript (below).

"These observations, which are used in particular to evaluate the hydrometeorological reconstruction developed in this study, are independent of it."

Data, models and methods

p.4 l.3 – The SMR developed by Moisselin, 2002, show significant inhomogeneous trends. It could be worth to add this fact the in text.

We are not aware of studies that demonstrate inhomogeneous trends in the SMR developed by Moisselin (2002). But it is true that the method is no longer state-of-the-art, and we now acknowledge it in the manuscript. Also, as now said in the discussion about uncertainties in the conclusion (your second comment), we made a comparison of low-frequency variations in precipitation between the SMR and the GPCC data, and both dataset show a good agreement over France. As we look at multi-decadal variability and we don't focus on long-term trends, this is not an important issue in our study. Note that in order to limit the potential influence of missing values on reconstructed long-term trends, we only use SMR stations with no missing values over the 1885-2005 period of interest.

p.5 l.1-2 – Is Safran really independent from the reconstruction as the same observations are certainly used in Safran and data assimilation?

Safran is not independent from the reconstruction. Some observation stations of precipitation and temperature used to constrain the statistical downscaling are also used in the Safran analysis. However, the only difference between the Safran-Surfex-AquiFR simulation and our hydrometeorological reconstruction is the quality of the meteorological reconstruction, as they share the same hydrological model. We clarified this point in the new version of the manuscript:

"A simulation based on the Safran-Surfex-AquiFR system is available over the 1958-present period. This so-called reference simulation in the following is used for the evaluation of the hydrometeorological on their common period. As they share the same hydrological model, potential differences between the reconstruction and the reference simulation only depend on the quality of the reconstructed meteorological forcing."

p.5 l.16 – At this point two questions:
- What about the spatial and temporal coherence? I suppose the spatial is respected as the same domain is used for the entire basin, but the temporal one?

The spatial coherence is respected as all the meteorological variables come from the same analogue day. As the analog days are selected from the same atmospheric reanalysis, a temporal coherence is also present in the meteorological reconstruction. We added a paragraph on this point in the new section on the development of the Seine reconstruction in the revised manuscript. Note also that evaluating river flows is an indirect way to assess that the spatial and temporal coherence are correct: without a good representation of spatial and temporal coherence, the simulated river flows would not be realistic.

"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."

- Isn't it possible, for each day, to constrain the results of the downscaling to create trajectories in function of observations, instead of creating all the 56 trajectories independently and them choose the closest to the observations? In the first case, the final trajectories would be different than the trajectories from the downscaling method without data assimilation. -> I believe this is done by the process explained l.17-32, and with only 3 trajectories instead of 56. Maybe reverse the explanations in the text, explaining the daily constrain before talking about the monthly one (not mention the monthly one before).

Yes, it is possible in theory. However, to be effective, the daily constraint needs a large sample of analogue days to find a good match with observed temperature and precipitation. But in practice the quality of analogue days rapidly decreases: with the limited sample sizes allowed by observations, it is impossible to find for example 100 very good analogue days (i.e. with large scale predictors close to the target). The interest here to downscale the 56 members of the reanalysis is to create an ensemble of possibilities large enough while maintaining the quality of these analogue days. As mentioned by the other reviewers, this section was not clear and difficult to follow. We made important modifications in this section (now a particular section), which is now much clearer. A description of the method used in Bonnet et al., 2017 was added (as asked by the two other reviewers). We also added a diagram to make this section easier to follow, as suggested by the third reviewer. Please find this new 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).

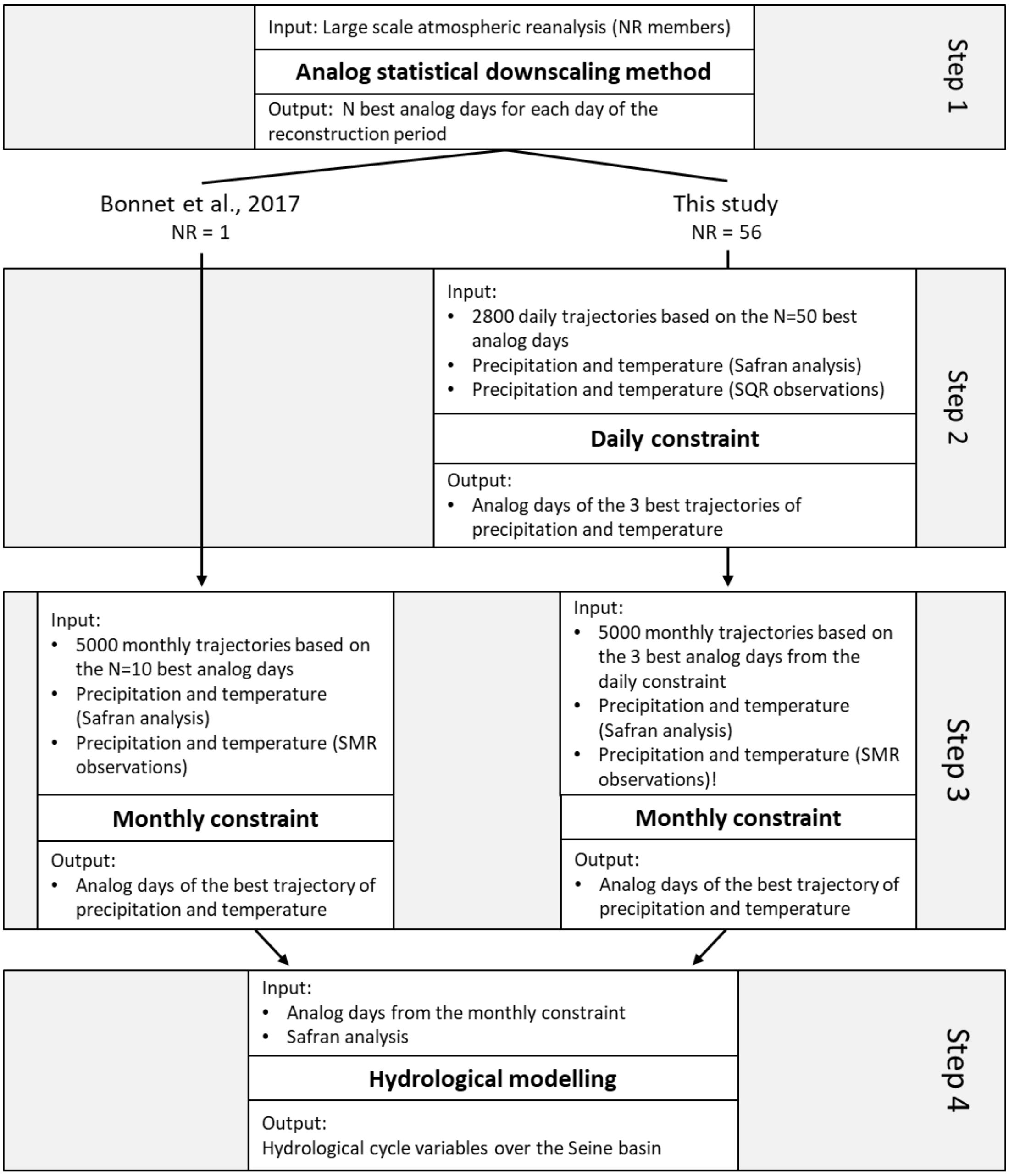

Figure 2. Schematic representation of the methods used to obtain the hydrological reconstruction, with on the right the method used in this study and on the left the method used in Bonnet et al. (2017), which doesn't include the daily constraint."

p.5 l.25 – Indeed, all the variables from Safran are obtained but the predictors for the downscaling method are only optimized for precipitation and temperature, isn't it? So what about the quality of the reconstruction for other variables, did you assess it? Is it good enough to used them in the hydrological model? It is for example possible to compare the reconstructed signals to Safran on the recent period.

With our hybrid approach, the quality of other variables (specific humidity, wind etc.) is not worse than the direct results we could obtain with statistical downscaling alone. As predictor for the statistical downscaling part of the method, we use 4 variables: precipitation, temperature, specific humidity at 850hPa and the pressure at sea level. With these predictors, we suppose that we have a good representation of the large-scale atmospheric conditions which influence the meteorological conditions over the Seine basin. Then, the reconstruction is optimized for temperature and precipitation by constraining them with the available observations. Several sets of predictors were tested with Safran for variables of interest such as the precipitation, the temperature, but also wind speed, relative humidity and solar radiation. We don't show that in the article to keep it short. Note that the evaluation of river flows and aquifer levels indirectly shows (with independant data) that the representation of all forcing variables is "good enough" to reproduce river flows reasonably for our purposes.

p.6 l.7 – Is it possible to give a little more details about these tests? Comparison to what?

A methodological choice had to be done here about the weight given to each of the two variables. We chose to give a little more weight to precipitation, which is an essential variable in the representation of river flows. A test carried out on the impact of the weight given to the temperature variable is illustrated Figure R1 below.

Adding temperature to the daily constraint greatly improves the representation of temperature with Safran compared to the daily constraint based on precipitation only, with a correlation gain of about 0.2 at the daily time scale and 0.07 at the monthly time scale. However, with a greater weight for temperature the correlations for precipitation logically slightly decrease.

A balance must therefore be found to improve the daily correlations of temperature without affecting too much the daily correlations of precipitation, for which a good representation is essential to study the high and low flows of the Seine basin, as well as extreme hydrological events. A weight of 0.5 is finally assigned to the temperature variable for the daily constraint.

We added a paragraph to illustrate some tests realized during the development of the reconstruction in the new section:

"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."

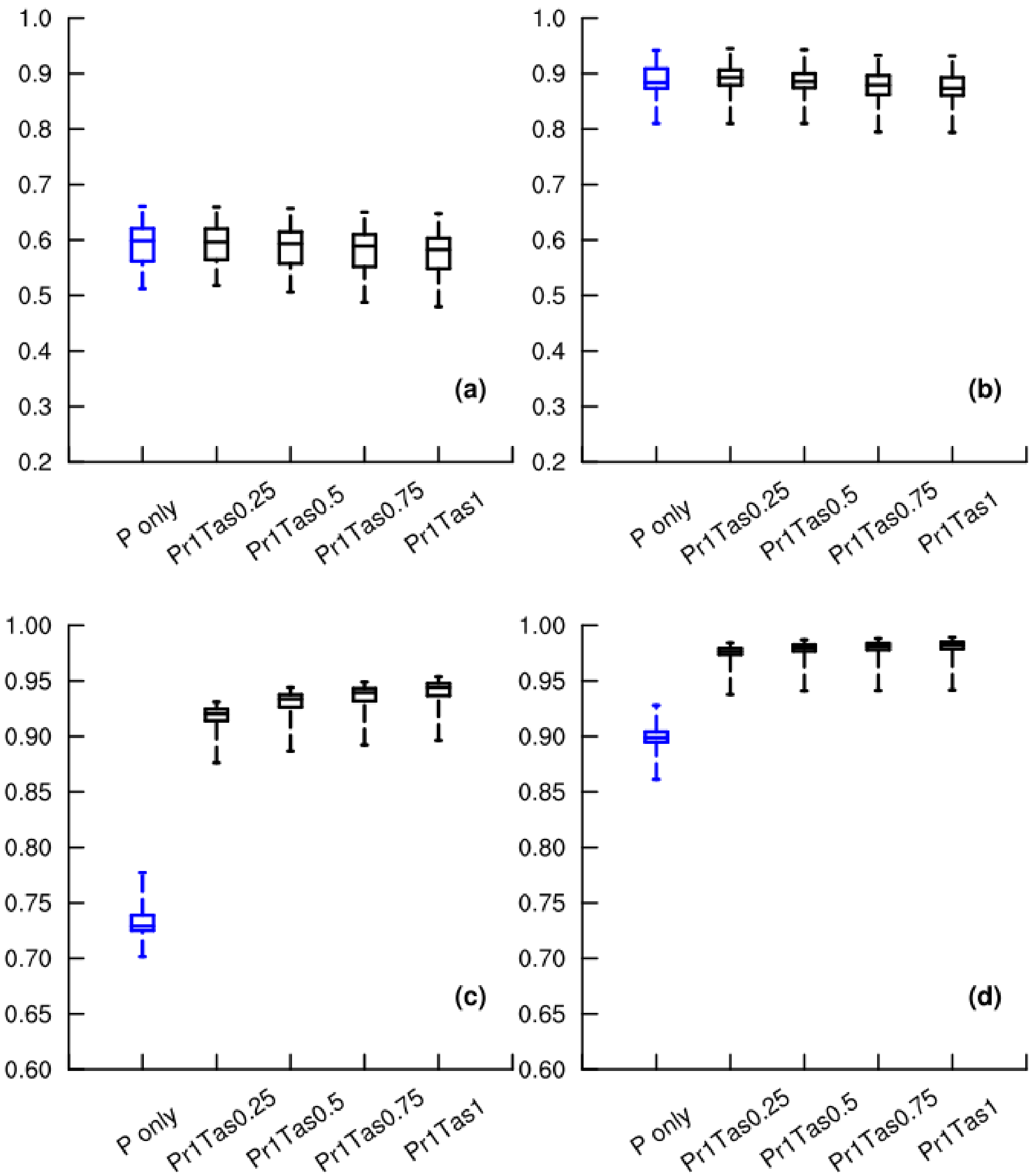

Figure R1: Spatial distribution of the correlations between (a-c) daily and (b-d) monthly (a-b) precipitation and (c-d) temperature over the Seine basin. Correlations are calculated between the Safran analysis (considered as observations) and the precipitation and temperature derived from the statistical downscaling method only constrained at daily time scale by (blue) precipitation only and (black) precipitation and temperature considering different weights for the temperature (indicated in 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.

p.6 l.8 – Why 3 analogue days?

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. The greater the number of analogue days is, the farther some analogue days 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 (as there is less spread to find a good monthly trajectory with fewer analogue days). We therefore made different tests in order to find the best number of analogues to retain at the different steps.

This is illustrated in figure R2, 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 increases (Figure 1a 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 to be effective.

Figure R2 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).

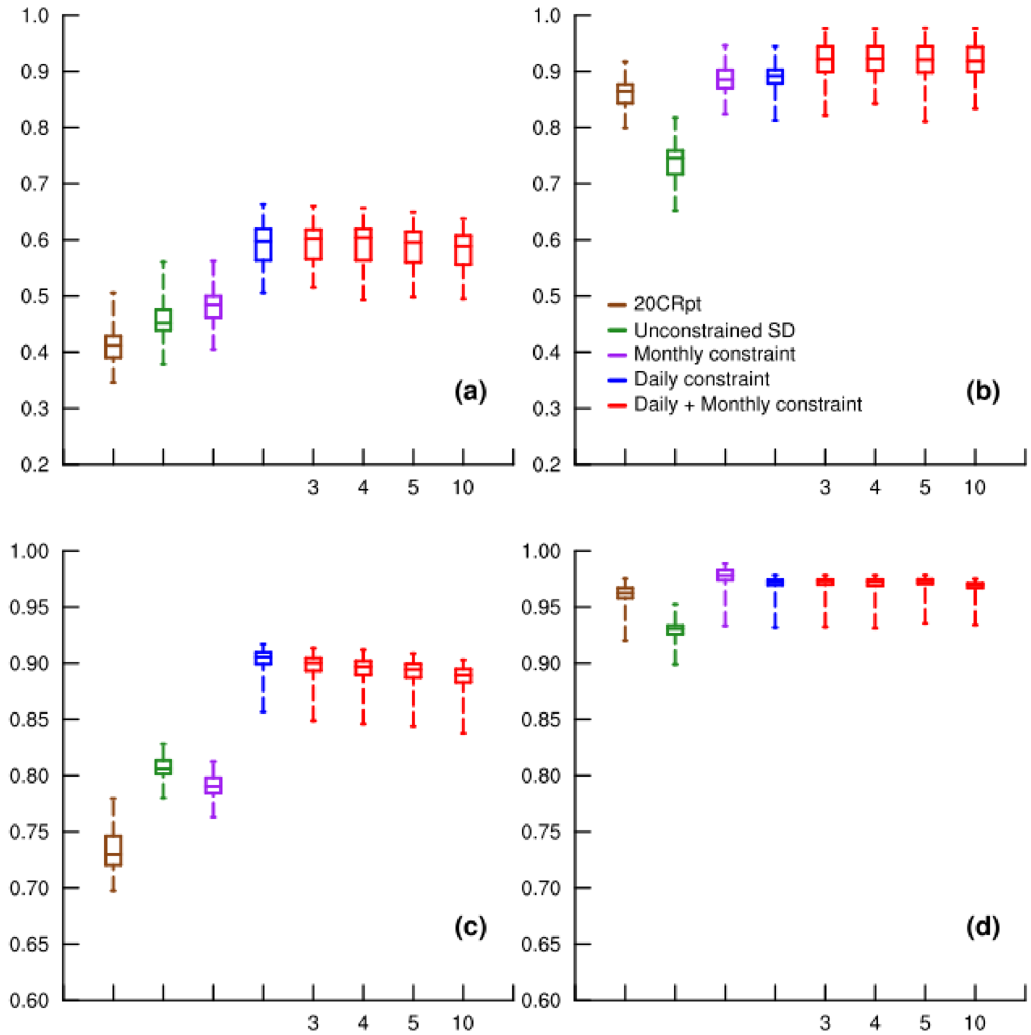

Figure R2: Spatial distribution of the correlations between (a-c) daily and (b-d) monthly (a-b) precipitation and (c-d) temperature over the Seine basin. Correlations are calculated between the Safran analysis (considered as observations) and the precipitation and temperature derived from (brown) the 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.

p.6 l. 10 – Is it possible to sum up the method in Bonnet, 2017 in a few lines?

We added a summary of the method used by Bonnet et al., 2017 in the new section about the development of the hydrometeorological reconstruction. Please find the description of this method four points above.

p.6 l.15 – What is the influence of using different types of observations for the data assimilation on different periods?

The quality of our reconstruction may be reduced for periods constrained only by one monthly series of precipitation (1850-1885 or the 2005-2010) in comparison to the 1885-2005 period, constrained by daily and monthly precipitation and temperature. The results, therefore, have to be interpreted with caution on these periods. We added a sentence about this point at the end of the new method section:

"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."

p.6 l.22 – Is the spatial (temporal?) coherence conserved after the correction of these biases?

As these are only climatological biases, this doesn't influence the coherence of the reconstruction.

p.6 l.22 – Maybe change the title "Method" in "Extraction of multidecadal variability" as the previous paragraph was also talking about methodological facts?

A large part of this section is about the way we extracted the multidecadal variability, but it is also about the way we deseasonalized the series before calculating the daily and monthly correlation as well as the acronyms used for the seasons. We therefore preferred to keep "Method" as title.

Evaluation of the Seine reconstruction
p.7 l.20 – On Figure 3, we see that before 1900, reconstructions are not close to observations. This is not a surprise as 20CR has a poor quality before 1900 and the network of observations is less dense. The signal at Paris seems more "flat" than observations, with under and over estimations.

We think that it is difficult to disentangle the respective impacts of errors in the reconstruction (due to 20CR or the low-density of the resolution network) and potential non anthropogenic influences, or measurement errors. We know that they exist and are important for the Seine at Paris. Note also that even if the magnitude of these variations is uncertain, the signals are however in phase, which shows that a large part of these variations can still be reproduced by our reconstruction. Note that we now better discuss the limits and uncertainties of the reconstruction in the conclusion.

p.8 l.4 Which type of correlation is used?

The Pearson correlation coefficient is used here. We added the precision in the revision of the manuscript.

Multidecadal hydroclimate variations
p.9 – Isn't it difficult to drive conclusions about other variables than P, T or Q as they are difficult to model in hydrological models?

We use a state-of-the-art physically-based hydrological model with a detailed representation of water exchanges in the soil and resolution of water exchanges at the surface, therefore we have a reasonable confidence in the representation of evapotranspiration and soil moisture. The evaluation shows that we can have a good confidence in the representation of precipitation and river flows. From a surface water budget perspective, it suggests that the other variables of the surface water budget are correctly represented, although it is still possible that some error compensations exist, for example between evapotranspiration and variations in soil moisture.

Conclusion
Wouldn't it be more logical to reverse the conclusion part (for now, first) and the discussion part (second) in the conclusion?

Agreed. The discussion part on the limit and uncertainties has been improved and moved before the results in the new version of the manuscript. Some perspectives are discussed at the end of the conclusion.

Orthographic corrections
P.1 l.4 – Reformulate the sentence "This method improves the representation of daily flow as well as at longer time step"
P.1 l.4 – Provides
P.1 l.8 – Maybe "regulate" instead of "modulate" would be better?
p.1 l.12 – to influence the drought intensities
p.1 l.16 – Missing a "," after "for example"
p.1 l.22, p.2 l.11 – Same remark for modulate / regulate
p.2 l.1 – Verify the expression "internal variability in climate and/or…"
p.2 l.20 – "It" not necessary in "which makes …"
p.2 l.25 – "." In the middle of the sentence, between "downscaling" and "of"
p.6 l.2 – Maybe add "spatial" before "error". The sentence is not really clear, maybe there is a way to rephrase it, talking about spatial errors for both precipitation and temperature.
p.12 l.27 – The SSTs there are

Thank you, modifications made.