Reviewer 2

We would like to thank Reviewer 2 for the constructive comments. Please find below a point-by-point response to the review. Most of the minor comments have been addressed directly in the manuscript (highlighted in red). In the following, we address (in italic font) the major and moderate comments of the Reviewer (in bold font).

The manuscript 'Snow data assimilation for seasonal streamflow supply prediction in mountainous basins' by Metref et al. provides an interesting study regarding the big challenge of improving streamflow predictions in mountainous, snow-dominated regions. The authors investigate the additional value of directly assimilating streamflow, fractional snow cover (FSC) and SWE measurements (taken by cosmic ray sensors (CS)) and their combinations in terms of improving seasonal forecast in three French basins (one in the Pyrenees and two in the Alps) applying a conceptual semi-distributed hydrological model (MORDOR-SD) as basis. They test their results during reanalysis (assimilation) and forecast periods and found that (not surprisingly) the assimilation with streamflow improved the estimates during both, reanalysis and prediction. Including CRS and FSC to the assimilation process could further improve the seasonal prediction in two of the three catchments.

In general, this topic is interesting to the readers of the journal. However, the manuscript in its current state needs major improvements before considering for publication. The authors should add important additional information and clarifications at several points (see comments below) as the paper currently produces several question marks in the eyes of the reader at some points. I agree with the points raised by reviewer 1. In addition, I have further points, which are listed below. English language should be improved.

General comments:

- **In your study, you are focusing on snow-dominated catchments where the simulation of snow processes plays an important role. However, the description of the snow module of the MORDOR-SD model is entirely missing here (e.g., I guess it is a simple day-degree approach to describe snow melt). This should at least be described (Section 2) and discussed (Section 5 or 6) carefully.**

A description of the snow module has been added in Section 2.1.

- **In general, assimilation can lead to good results regarding streamflow predictions (as you have shown). However, it should also be discussed in the paper, if adding more physical realism in describing snow cover processes could also lead to improved results regarding streamflow predictions.**

Adding more physical realism to the snow cover processes would perhaps improve the streamflow predictions. However, this question is not in the scope of the paper and, since the experiments we perform do not provide information on that topic, adding a discussion would be mainly speculation on our part.
For the reviewer's information, in the context of operational prediction, EDF teams do work on improving streamflow predictions by simultaneously upgrading the physical realism of their models and enhancing their assimilation capabilities.

- **As reviewer 1 already stated, the introduction is difficult to read and a mix of state of the art, presentation of some results, objectives, research questions, outline, and some methods. The Introduction should be carefully improved including a solid state of the art paragraph.**

The introduction has indeed been restructured.

- **What is the reason for selecting the three chosen basins Verdon, Naguilhes, and Gui? Are they very different in terms of topography, meteorology, geology, etc. to learn different behaviours regarding catchments response? Do you expect to gain additional information, if you would select further catchments out of the 50 catchments operated by EDF?**

These 3 catchments were selected according to two criteria: (i) the quality of the hydrometric data (to avoid assimilating poor quality data); (ii) the presence of CRS data on the basin. Moreover, they offer an interesting variety of hydro-climatic contexts. This explanation has been added to the article (l. 121-123).
This selection would obviously deserve to be extended, but it already allows us to clearly identify the potential and limits of our assimilation strategies.

- **You tried out the settings of assimilating i) Q, ii) Q and FSC, iii) Q and CR, and iv) Q, FSC and CR. Why didn't you show the results of just assimilating FSC (regarding CR you stated it would deteriorating the system estimation (l.42ff – this however, would fit rather in a discussion section instead of the intro))?**

As previously stated, the introduction has been modified. The FSC-only assimilation provides very poor results every year and for every diagnostics. This result is not *surprising, as this variable is only indirectly correlated to the snow water equivalent in the basin. We have decided not to show these results for the sake of clarity and to lighten the article. This behavior is now noted in the Summary and Conclusions section.*

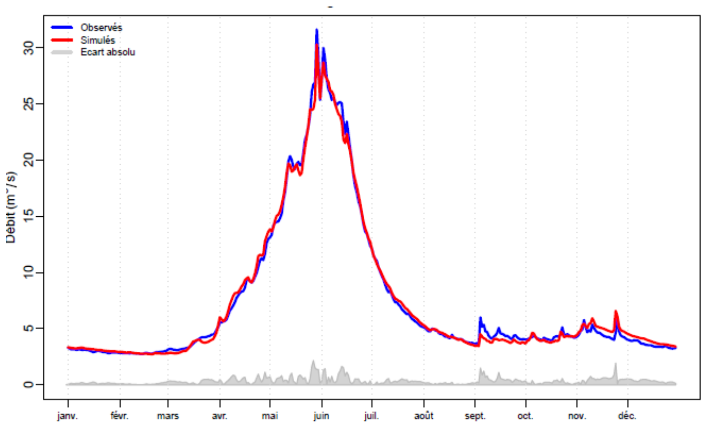
- **How well does the MORDOR-SD model perform in general in your catchments regarding calibration and validation periods (e.g., according to objective functions such as Nash Sutcliffe Efficiency)?**

Modeling performances are pretty good in these basins. For example, hereafter the values of Kling-Gupta and Nash-Sutcliffe Efficiencies on available periods (calibration periods are respectively: 1998-2013 for Verdon, 1987-2012 for Naguilhes, 2004-2013 for Guil)

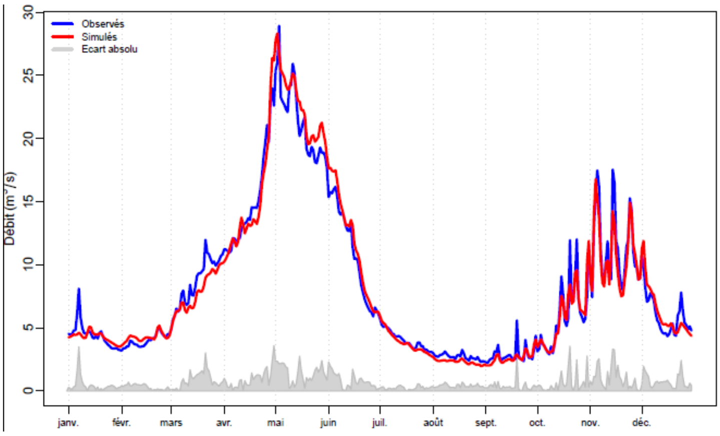
	KGE	*NSE*
Verdon (1997-2017)	*0.921*	*0.846*
Naguilhes (1987-2017)	*0.880*	*0.760*
Guil (2004-2017)	*0.961*	*0.926*

Observed (blue) and Simulated (red) long-term mean daily streamflows are also illustrated:

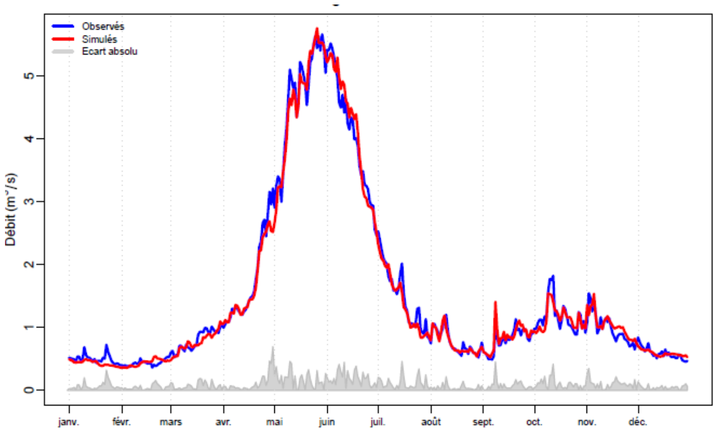
Guil



Verdon



Naguilhes



- **I miss the link of meteorological and snow conditions at certain years in the three catchments and your results (especially in the discussions in Sections 5 and 6). Snow and meteorology conditions can be quite different throughout the years and might affect the quality of your streamflow predictions. What is the impact on e.g. a lot of snow vs. shallow snowpack during single winters in the three catchments?**

For one catchment, the performance of the forecasting system is indeed very variable from year to year. Some meteorological situations (very heterogeneous or localized episodes) can lead to significant precipitation estimation errors and degrade the forecasts. The assimilation of snow observations will only have a positive impact if they are spatially representative. On the other hand, we have not detected any systematism associated with the amount of snowfall in a given year.

- **In the lines 28-30 you raise three questions on the relevance of using in satellite and situ snow observations to improve seasonal streamflow predictions in mountainous catchments. However, I have the feeling that these questions are not properly answered in the course of the manuscript.**

We agree that the third question (How do existing analysis methods perform in estimating the snowpack from the snow observations?), as formulated, may appear to go beyond this study. We decided to remove it from the introduction.

Considering the two first questions:

(i) How much is the SSS prediction sensitive to the snowpack? We attempt to answer this question using the Sobol analysis presented in section 3, which demonstrates that SSS prediction is primerely sensitive to snowpack water content.

(ii) Do the snow observations contain enough information to estimate the snowpack accurately to impact the quality of predictions? Results of the study demonstrate that assimilating snow observations (FSC and SWE) in combination with streamflow observations improve SSS prediction, compared to assimilation of streamflow only.

Specific comments (chronologically):

- **86: The paper is actually divided into six parts (including the introduction). -> Better reformulate: 'The paper is structured as follows:'**

  Corrected in the text.

- **Figure 1 and Section 2.1: At least a basic introduction to the model and its components should be given. Please add information (e.g., as a legend in Figure 1) on the variables and parameters shown in the graph.**

  As discussed in Reviewer 1 comments, Figure 1 has been removed. A model description has been added in Section 2.1 and the reference to Garavaglia et al. (2017) is provided for a detailed model description.

- **93: How many metres does one elevation band encompass? Regarding Figures 5-7, this seems to be 250 m for each catchment!?**

  Classically, the number of elevation zones is optimized depending on the hypsometric curve of the catchment according to the following criteria: (i) the relative area of each elevation zone has to be greater than or equal to 5% and less than or equal to 50 %, and (ii) the elevation range of each zone has to be lower than 350 m. It leads to 8 elevation bands for Verdon and Guil and 4 for Naguilhes.

- **95: What are the orographic gradients (lapse rates) applied in this study for temperature and precipitation?**

The orographic gradients gpz and gtz for precipitation and temperature respectively are now defined in the text (lines 95-96 and 106-108).

- **99: What are the 5 state variables (I guess S, G, U, L, Z and N?) and the 2 global variables? Please add this information at least in the the methods descriptions and the legend of Figure 1. In additions, why do you write the span of 10-12 free parameters? How many did you have in your setup?**

The mention of five variables in each elevation band was a reference to the 4 storage water levels U,S,L and Z and the snowpack bulk temperature TST. There is only 1 global variable N representing the deep storage water level. The number of free parameters refers to the model calibration process, ranging from 10 to 12 depending on the site-specific calibration strategy. This is now made clearer in the article (l. 97-101).

- **106-111: Not entirely clear; Please improve the descriptions in these lines. / Figure 2: Is this a catchment averaged meteorological data set shown here or is it representative for one elevation band? In general, not sure if Figure 2 is really needed. In addition, the Verdon basin is actually introduced one Section later and the reader might be wondering why you already mention it here.**

This paragraph has been reformulated (line 106-113) and should now be clearer. / Figure 2 shows the catchment averaged meteorological data set which is the only input of the MORDOR-SD model. Introducing stochastic perturbations is crucial for the following experiments, we have hence decided to use an illustration (Figure 2) even though the Verdon basin is not yet described.

- **Figure 3: This graph just shows the location of the basins in France. The graph misses topographic information as well as important information such as at least the location of its capital and the name of the mountain ranges (Pyrenees, Alps). In addition, I suggest giving a more detailed overview on the three selected basins in the Figure.**

Figure 3 (now Figure 2) has been updated. Topographical and geographical information has not been added to the figure to keep it clear, but is given in the text (l. 124-129).

- **140f: Please add information on the expected footprint of the CRS as well as limitations of this sensor type.**

The expected CRS footprint is classically about 5m, and this measurement technique is known to provide accurate SWE estimations, except for very shallow snow-depth. This was added in the text (l. 146-148).

- **142ff: Please describe in more detail how FSC was derived. Did you look at basin-averaged values or did you consider elevation band based FSC values. I think just taking FSC values for the entire catchments (with elevation ranges of approx. 2000 m) is not sufficient and might not be representative for the application of assimilation data.**

FSC data used in this study are basin-averaged values. We agree that elevation band-averaged values are potentially more relevant, but they have not been used because they are too incomplete (limiting cloud cover over small areas).

- **Figures 5-7: Not introducing U, L, Z, S, TST before makes the figures questionable (see comments above). Information regarding elevation bands is missing in the y-axis. The chosen (linear) colour representation is not very meaningful. In addition, I would suggest to add a row in showing the average Sobol indices for the entire time period to get a clearer overall picture. Interpretation why the Sobol indices are higher for some elevation bands as well as distinct years is missing in the text.**

The differences between elevation bands is mainly due to the differences of their absolute snow content. For example, high elevation bands have smaller areas (by definition of the elevation bands) hence they have less snow content which leads to less uncertainty.

Similarly, differences between years and between basins are most likely due to differences in snowfall since the perturbations are prescribed relative to the state variables (in percent) but the sensitivity of the streamflow is absolute.

- **170-174: Please avoid repetitions – was already introduced before.**

This experimental setup information is important. We decided to mention it in the introduction and give it in more detail in the Protocol and Diagnostics section.

- **Figures 9 and 10: Please insert for a better readability legends. Why do you show the selected years, assimilation configurations (assim. of Q in Fig. 9, assim. of Q and Q&CRS), and the selected catchment as an example in those Figures as examples? Are other seasons/years similar in their quality?**

Legends have been added to Figures 9 and 10 (now Figure 8 and 9).

These figures are only illustrations. Figure 9 (now 8) illustrates ensemble of simulation with and without streamflow assimilation which relates to Section 5.1. And Figure 10 (now 9) illustrates the improvement brought by the assimilation of (Q, CRS) in comparison to Q assimilation only.
The exhaustive study, over the different years and different basins, is provided by the following diagnostics (Figures 10 to 15).