Reply to referee 1

We would like to thank you for your constructive comments and feedback on this manuscript. We think that the suggested revisions based on the Referee’s comments will certainly improve the article. Please find our responses (in blue) to the main points raised (shown in black) below.

Introduction

The SDM methodology needs to be better introduced. Section 1 fulfill this role, but in my opinion it needs to be integrated directly in the introduction instead of being a separate section.

> → Thank you for your suggestion and we agree that bringing the current section 1 into the introduction would provide more information on SDM up in the manuscript while making it more synthetic. We will implement it in the revised version.

Methodology

This section needs to clearly highlight where the stochastic component of the model plays a role. If I have understood correctly, only the modeling of the reservoir level and outflow is done in SDM, whereas the input flow is modeled with a deterministic approach. Since human activities (and climate change) may influence also the upper basin, some considerations on the assumption made may be useful for a reader. E.g. How much anthropogenic activities are in the upper basin?

> → Yes, reservoir level and outflow modelling are modelled through regression models and analysed following a stochastic approach. We acknowledge we should bring up earlier in the methodology some of the assumptions on human activities that are now presented in the future projections and limitations sections (line 235 and 374) and we will implement these suggestions in the revised version.

Results

I found very limiting the lack of tests on the statistical significance of the differences from the baseline. This is particularly relevant since you are using just a single climate projection model.

Additionally, the analysis on the extremes (30th and 80th percentile) needs to be expanded. Limiting the analysis on the number of events may severely bias the analysis. You should integrate with the duration of such events, number of days under a threshold and the severity (in the case of drought).

It can happen that a smaller number of events in the future is due to the occurrence of less event but much longer.

→ Thanks for the constructive feedback. We see the need to introduce statistical tests and further expand the analysis of extreme conditions. We will integrate statistical tests for baseline vs future projections as well as integrating the number of events in the future with other metrics in the revised version, such as the maximum number of consecutive months below/above the selected thresholds (duration) and the changes in the number of events below/above different thresholds (magnitude).

Specific Comments

P2 L51-57. Here a more detailed explanation of SDM is need, with a clear description of the advantages and disadvantages of deterministic vs. stochastic approach (move and reword from Section 1).

→ Similarly to the first comment, we agree that bringing the current section 1 into the introduction would provide more information on SDM up in the manuscript and we will implement it in the revised version.

P3 L72-73. Can you please better clarify the role of connectors and the differences from converters? (maybe using two different examples rather than the same).

→ Yes, we can rephrase the sentence providing a different example, as follow: "converters - parameters influencing the flow rates (e.g. temperature variable acting to alter evaporation from a water body) and (iv) connectors – as arrows transferring information within the model (e.g. linking the monthly effects on reservoirs' water discharges) (Sterman et al., 2000)."

P4 L94-99. This paragraph would fit better in Section 3 in my opinion. In general, Section 1 seems unnecessary, and I will move its content in either introduction (first part) or Section 3 (second part).

→ We agree with your suggestion to move this paragraph in section 3.

P5 L125. It would be nice to have these numbers contextualized in comparison to e.g. average flow.

→ Thanks for the suggestion and we can report in brackets the percentage with reference to the average annual flow.

P6 L149. Fig. 2. It is not clear to me the role of soil (type?) and land-use as vulnerability factors. Please briefly clarify in the text.

→ Soil and land-use are reported in Fig2 for completeness and as background information coming from the GeoTransf hydrological model output. Both these issues were considered as vulnerability factors affecting run-off and its effect on the reservoir stored volume and water turbined. Although they were both kept constant in case of future conditions, we thought their representation could provide a wider picture on those factors affecting the final risk conditions. We can specify this role in the text (lines 145-146) of the revised version.

P6 L150. part of the caption in Fig. 2 seems missing.

→ Thanks for pointing that out and here reported the full caption that we will report in the revised version: “Causal loop diagram used to describe the risk variables and their interactions leading to critical states of S.Giustina reservoir operations. Climate variables are in green font, blue font for hydrological-related components, yellow font variables are those involved in S.Giustina operations leading to critical states”.

P7 L158. Fig. 3. I understand the need to have different “baseline” period in different components. However, is there any way to homogenize those references or at least use a different name for each one of them? At the moment, when you refer to baseline in the text it is really difficult in some cases to understand to which period are you referring to.

→ We understand such difficulty, although the use of different baselines is constrained by the data availability. As suggested, we can explicitly add the variable name and its baseline period in the revised version in order to prevent potential misunderstanding.

P9 L197. Table 2. It is not 100% clear to me what are the other models reported here. Are those the best for each model type (among the ones in supplement materials)? Please clarify.

→ Yes, the model reported here are the best for each model type. The whole list of tested model are then reported in the supplement material. We agree that this can be clearer and we can improve the text and the table caption.

P9 L197. Table 2. Please highlight in the table the models selected as “best”.

→ Thanks for the suggestion and we will make bold the selected model in the revised version in table 2.

P10 L216. Eq. (2). Is the random effect still just the month as in Eq. (1), or is it in the first term as well through V? Please clarify an eventually emphasize this key difference.

→ Thanks for pointing that out as in both equations the random effect is represented by the variable “month” and there is indeed a nested effect, which is due to the monthly effect on both inflow and volume variables. We acknowledge and agree on the need to emphasize the presence of nested models in the revised version of the text.

P10 L227. I'm not familiar with this procedure, but I'm assuming that you ends with 58 calibrations. which one is used at the end, or how are them combined? Please clarify.

→ Yes, all the 59 calibrations are used to compute a mean value of each model performance in predicting Volume and water turbined (R2 and RMSE). By doing so, we aimed to avoid common limitations of arbitrary choice of splitting a dataset into training and testing, which can provide biased estimation of model’s performances. We agree that this needs to be clearer and we will specify the computation of a mean value from the forward time-window approach.

P11 L242. Why this asymmetric choice instead of 20 and 80? This needs to be justified.

→ Thanks for pointing that out. This assumption was based on previous studies reported in the text (Majone et al., 2016; Yilmaz et al., 2008) characterizing low and high flow in flow duration curves where the threshold of 0.2 exceedance probability was selected for high flows and the threshold of 0.7 exceedance probability was selected for low flows. Such assumption was reported here as the two corresponding thresholds of 30th and 80th percentiles to characterize critical conditions. Following your suggestion, we can integrate these thresholds adding the 10th and 90th percentiles in order to provide an additional description of critical conditions in a symmetric way.

P11 L244. Which baseline period? It may worth to specify here.

→ We agree that we need to specify in the text that we are referring to volume values from the baseline period (1999-2004 & 2009-2017). We will report this information in the revised version.

P11 L247. It is not clear to me where the Monte Carlo approach is used. Is it used only for the percentile analysis?

→ The Monte Carlo approach is used to analyse future values of volume stored below and above the selected 30th and 80th quantiles in order to provide a more robust assessment (i.e. having a larger set of future values based on the regression models) of critical conditions in the future. We agree this can be clarified in the current paragraph.

P11 L254. This sentence is a little confusing. I'm assuming that the storage for "flood prevention" is higher than the maximum due to the additional volume stored to prevent a potential downstream flood, but the way the sentence is presented is confusing. Please reword.

→ Thanks for pointing that out and we can rephrase it in the revised version into: "Figure 4 shows the modelled and real values, with volume ranging from 0 (i.e. no usable volume) and a maximum level of 151.20 Mm3. Only in case of emergency flood prevention additional volume can be stored up to 159.30 Mm3."

P12 L257. Figure 4. Please clarify the meaning of the dotted lines (percentiles?). Same in Fig. 5.

→ Thanks for pointing that out. We will integrate the use of dotted lines for percentiles in figure 4 and 5 in the revised version.

P12 L265. If I understand correctly, this means that the most important component (inflow) is the one that is not affected by the stochastic modeling. This needs to be further discussed and emphasized in order to clarify the value of using the stochastic approach in similar conditions.

→ Yes, inflow is used as an (deterministic) input variable to predict water turbined and volume stored. While deterministic assumptions of variables interactions are often implemented to describe dynamics of a system, the use of statistical regressions allows to replicate past conditions, predict and stochastically test future ones accounting for its uncertainty range. We agree that this can be emphasized in the text and we can implement it in the revised version.

P13 L270. In the whole section, statistical significance of the changes need to be reported. Even changes with different sign may not be that different if the changes are not statistically significant (also accounting for natural fluctuation within the 30-year window). See Welch test or similar.

→ Thanks for the constructive feedback. We see the need to introduce statistical tests between baseline and future values and we can integrate them in the revised version.

P 14. Figure 6. I'm not sure that showing each single year is relevant, since these are projections. I would find a way to show time slices rather than annual/sub-annual values.

→ We see the need to better emphasize trends rather than single projected values. Following your suggestion, we will replace them with (e.g. 10-year average) time slices.

P15 L302. It would be interesting to integrate the analysis with the result in terms of contribute to the annual storage. Differences in rainy months may be much more relevant for the total storage that during the dry months, even if the percentage differences are comparable.

→ Thanks for bringing this in the discussion as in our initial discussions we computed Fig 7 and 9 rescaling them to the total annual water turbined and total annual storage (hence providing information on average month effect over the whole years). However, since these results have a similar trend to the monthly percentage variation figure (current fig 7 and 9), we opted to only provide this latter.

P16 L312. Number of events is not the only relevant metric. For drought/water scarcity: are those events longer? what about the number of deficit days per year? Is the total/average deficit of these events increasing? For flood: is the max surplus increasing? Those are examples of simple analyses that can be added without much effort.

→ Thank you for the suggestion, this is an interesting analysis that we can integrate into the revised version to improve the characterization of future critical conditions. For this purpose, we can compute the changes in the number of months below/above different thresholds (i.e. magnitude) as well as the maximum number of consecutive months below/above the selected thresholds (i.e. duration).

P16 L322. You should add also the same statistics for the reference period. How those number differ from the reference values? Are the differences in the short term (2021-2050) in line with what is already observed? Is the current condition (2021) very different from the reference period? Some insights on that would be useful to understand the reliability of these projections.

→ Following the previous comment, we can compute different metrics for the volume baseline period and future scenarios of RCP4.5 and 8.5 both 2021-2050 and 2041-2070 in order to explore trends and differences.

P17. L340. Please check the use of the term “negatively” here (and in the rest of the discussion as well). This can be interpreted as either mathematically (sign minus) or qualitatively (to worse). I suggest to rephrase.

→ Thanks for pointing that out. We agree and will rephrase it.

P18 L353-354. What about flood? Can increasing storage during winter be a problem in case of flood? Please elaborate.

→ Thanks for pointing out such an interesting side-effect due to the implementation of adaptation strategies such as an earlier water accumulation. We see the point of clarifying how such strategy can affect flood events. In particular, we will clarify that our discussion is based on:

- The trend of volume decreases in spring and summer showing higher percentage values than the increases in winter (from figure 7)
- The results on the number of events higher than the 80th quantile (Figure 8), which are expected to decrease for all scenarios except for the RCP8.5 2021-2050 (showing higher values than for the baseline volume).
- The additional storage for “flood prevention”, which availability needs to be always ensured and managed together with the Civil Protection to prevent a potential downstream flood, even in case strategies of earlier reservoir accumulation would be considered.

P18 L356. Increasing in high and low volumes? I’m not sure that this sentence is in line with your results. Please clarify.

→ Thanks, we will rephrase it in order to refer to the number of events with volume below than the 30th percentile and above the 80th percentile.

P18 L359. I think that the Monte Carlo results may give an idea on the robustness of such changes, but this is not discussed at all at the moment.

→ Thanks, we see the need to bring the Monte Carlo results into the discussion. We can integrate this into the revised version.

P18 L364-365. This is more a conclusion that a discussion of the results.

→ Thanks for pointing that out and we will move such a sentence into the discussion section in the revised version.

P18 L375. This is a very important point that needs to be highlighted in the methodology as well.

→ Thanks and we agree that having such information earlier in the methodology can help readers to better understand such an assumption on human activities.