Referee 1

Overall, a better description of the anthropogenic effects, in both the hydrological modeling of the inflow and in the SDM, is needed. This is particular important for future projections, where you need to emphasize how your hypotheses (no changes?) compare to the expected future behavior or potential SSP compatible with RCP4.5 and RCP8.5. These background information would be very useful to interpret the obtained results (i.e. differences between short and long terms and RCP4.5 vs. RCP8.5) and also to improve the discussion section.

Also, as suggested by one of the previous reviewer, and following the revision of the methodology section, I would downplay the use of terms as "innovative", "novel" and similar throughout the text (not limited to the examples reported in the specific comments).

We would like to thank the Referee for their constructive feedback. Please find our responses (in blue) to the main points raised (shown in black) below. Here reported the main improvements:

- We integrated information on the anthropogenic demand and assumption for future simulations
- We change the representation of future values for inflow, outflow and volume into boxplot for 30-years
- We downplayed the text on SDM making it simpler

Minor Comments

L21. please clarify the sentence "creation of statistical models into".
We modified and integrated the text with the outputs and aim for the statistical model.

L21. The study aims at simulating…
Thank you, we have corrected the text accordingly

L22. volume, as well as…
Thank you, we have corrected the text accordingly

L23. "whole 30-years of simulations…". It is not clear which 30-year period are you referring to here. Especially followed by the sentence on the results for 2021-2050 and 2041-2070. Do you mean something like "Average results on 30-year slices"?
We have corrected the text accordingly

L24-27. It is worth to mention here what is the reference (i.e. present) period for these differences. In any case, differences between RCP4.5 short-term and RCP8.5 long-term are very small. Why? If this is discussed in the text, it needs to me briefly mentioned here.
We added reference to the baseline (lines 26-27) and a brief sentence on factor driving the similar trend for RCP4.5 short-term and RCP8.5 long-term (lines 28-29 in the marked-up manuscript). This description was also added in section 3.2 with reference to climate projections from Bucchignani et al., 2016.

L27. January to April.
We have corrected the text accordingly (lines 30-31 in the marked-up manuscript)

L29. Report months in which the mins occur, similarly to the maxs.
We have corrected the text accordingly (line 32-34 in the clean manuscript version)

L30-31. Even if I understand that "increase in terms of low volume" and "decrease of high volume" refer to frequency, duration and severity (i.e. more frequent, longer and severe deficit), the sentence may be easily misunderstood. I suggest to reword.
We have moved this part closer to the initial metrics terms to be sure it is referring to them (lines 34-36 in the marked-up manuscript)

L54. It seems that a word is missing after snowpack (like dynamic or something).
We have added the missing word (line 61 in the marked-up manuscript)

L78. I suggest to reword point ii) to better highlight what is the limitation (since this a what the list is all about).
We have removed the last part of this sentence so to simplify and focus on the mentioned limitation (line 86 in the marked-up manuscript)

L84-87. Since stochastic SDM is not a new approach, I suggest to add few references and examples on how a stochastic component can be added to SDM, similarly to the overview that you did on SDM in general.
Thanks for your comment, we have integrated the text accordingly (lines 93-94 in the marked-up manuscript) adding reference to two papers dealing with this approach (Taylor et al., 2009; Ford, 2005).

L89. I suggest replacing effects with projections.
We have corrected the text accordingly (line 107 in the marked-up manuscript)

L112-119. "Temperature increase, …. (i.e. hydropower plants)." is more suited for the introduction, to explain the reasoning behind the study, than for methodology.
We understand the referee comment and we moved the suggested part into the introduction (lines 99-106).

L129-134. same as above. Not really relevant for this section, but it can be partially used in the introduction to better explain the context of the study.
Thanks for pointing this out, we moved some of this part into the introduction (lines 112-114) and changed it.

L136-137. Some wording here, such as "novel", "robust" and "particularly useful" is excessive. I suggest to keep the text as simple as possible.
We removed those words as suggested to keep the text simple.

L140. Similarly, "innovative" is not really needed here.
We removed it too.

Fig. 2. Clarify in the figure why there are two baselines in the SDM.
We substituted "&" with semi-columns in the baseline and added brackets in Figure 2 pointing to the use of a single baseline having a gap from 2005 to 2008 as added in Figure 2 caption and in line 212-213 in the marked-up version.

L170. Please provide some details on how the anthropogenic effects are modeled in GeoTransf (i.e. water use modeling). Is the basin highly regulated upstream of the reservoir? Is any of these anthropogenic effects accounted in the inflow? Similarly, is any of this modified for the projections?
We added additional information on water withdrawals in GeoTransf (lines 201-202 and in lines 290-291 of section 2.4 in the marked-up manuscript). Withdrawals upstream the reservoir were considered affecting the inflow and set unchanged to the maximum discharge for future simulations given the water infrastructure constrains in order to account for future increases of water demand from domestic and agricultural needs.

Please provide here a brief summary of the validation (i.e. accuracy, bias, other metrics).
We provided the resulting performance metrics (lines 197-198 in the marked-up manuscript) while referring to the main publications for the full explanation of the calibration and validation methods and results (Bellin et al., 2016).

L178. What are the hypotheses in the projections? Is climate the only thing that is changing? Any change in population, agricultural area, water used included? Please clarify (also in the context of expected future evolution of the area), since this is a key factor in interpreting the results.
Yes, future projections considered only climate as changing while the assumption on the maximum water withdrawals for water uses encompasses possible water demand increases due to tourists and agricultural needs. We added this information in lines 205-207 in the marked-up version.

L179. Other variables (drop input).
We have corrected the text accordingly

L181-184. This is not clear to me. Does it means that the SDM was forced (inflow) with a combination of modeled and observed data depending on the period? Please clarify and explicitly say what is used in which period.
Yes, we changed the sentence into: "the SDM was forced with past observations of inflow to S.Giustina (with data available for the period 1981-2016) and with GeoTransf values for future simulations due to the limited temporal overlap of past GeoTransf values (over 1981-2010) with observations of the S.Giustina stored volume (over 1999-2004; 2009-2016, Table 1)" (lines 213-216 in the marked-up version)

L199. Please quantify "low predictive performance" (i.e. xx statistical metric lower than yy).
We modified the sentence providing information while referring to the Supplementary material since the variables selection was not based on a single performance criterion, but it was based on the interpretation of three criteria (i.e. (i) data availability to the maximum target time period, (ii) correlations among the explanatory and response variables and (iii) the selection of the most parsimonious model). (lines 229- 233 in the marked-up version)

L202. The best models for each model types. Pleas reword, maybe something like "the best relationships/regressions".
We have corrected the text as suggested (line 235 in the marked-up version)

L204. I would invert this sentence, to emphasize that it was chosen thanks to being the best at monthly scale, and also the good performance at daily scale. Anyway, I do not want to argue with your choice,

but you should come up with a combined metric that show that model #2 is better than, let say, model #4. Right now this choice seems a little arbitrary.

We have corrected the text as suggested and we understand that presenting both daily and monthly results could confuse readers. For this reason, we decided to keep the monthly results while reporting daily results in brackets within the text and in Table 2 as well as adding this information in Table 2 caption.

L228. Please emphasize that here you are assuming that no change in how ecological flow are treated will occur in the future.

We have corrected the text as suggested (lines 264-266 in the marked-up version).

L250-252. This should be also mentioned above, in the model description. What does it means that this is a conservative assumption? Are population and water use expected to increase or reduce in the future in this area? Do you have any reference to back up this assumption?

Thanks for bringing this in the discussion and we added information in lines 205-207 in the marked-up version. Future water withdrawals were simulated at their maximum possible abstractions given the current infrastructure and water network (Bellin et al., 2016). This can be considered as a conservative assumption to account for potential future increases in water demand which is expected to rise for agriculture, due to the increase in temperature and hence evaporation and evapotranspiration, as well as for the domestic sector (La Jeunesse et al., 2016). Nevertheless, comprehensive analysis of the combined effects of water demands from multiple sectors (i.e. agriculture, domestic, etc) in future projections are currently missing for this area.

Table 3. I appreciate the extension of the analysis to additional metrics. However, what is the reason to include both absolute and relative frequency, if the relative values are derived by dividing for the same fixed values of 14 years everywhere? I would keep only the relative frequency and remove the absolute one. Similarly, you can divide the severity for 14 years and get a average annual severity (not necessary, but more consistent).

Thanks for the comment, we have removed absolute frequency and divided severity by 14 years in Table 3 and throughout the text for keeping a consistent reference to annual metrics as suggested.

L274. Maximum duration, rather than just duration.

Thank you and we have corrected the text accordingly (line 312 in the marked-up version)

L276. This last metric provides information…

We have corrected the text accordingly (lines 314-315 in the marked-up version)

L294-295. …to support reservoir… effects. This sentence is unnecessary, especially since you are discussing only the baseline run here.

We have removed this last part of the sentence accordingly (lines 332-333 in the marked-up version)

Figs. 3 and 4. There seems to be a limited capability to capture low values, such as the one in 2003 (a well known drought year in Europe). This needs to be highlighted and discussed.

We have highlighted and discussed two cases of very high turbined outflows in summer 2001 and very low stored volume in spring 2003 to cover two specific cases where the modelled values did not well represent real values (lines 344-349 in the marked-up version). For the very low volume in spring 2003, the S.Giustina reservoir was emptied due to construction works at the dam, while the 2003 drought

effects are still visible with monthly values of May and July ranking as the second lowest value for each month of these months over the available data.

Fig. 5. I still think that even 5-year slices are too small for such analyses on projections. Even if there are differences between two 5-year periods, what do they say about projections? The two 30-year periods are already overlapping, and the differences are quite small. Just report the same plot for the two 30-year periods (and two scenarios) and you will have a much easier to read figure with much more robust results.
Thank you for the feedback. We modified the figure into 30-year slice as suggested to simplify the representation.

Please clarify that colors are the same as in Fig. 4.
We have added this clarification in Figure 5 caption

L313. Please report just one digit in the percentages (as in the Table). Please apply throughout the text.
We have corrected the text accordingly leaving more digits only for predictive performance indicators (i.e. R2, RMSE and p-value).

L313-315. The way this concept is formulated is confusing. I would say changes in RCP4.5 are rather similar for short and long terms, whereas changes are bigger in the long term for RCP8.5 compared to short term.
We have corrected the sentence following your suggestions (lines 355-356 in the marked-up version), reporting percentage values only for RCP8.5 short and long term and referring to Table 4.

In general, small change in short term for RCP8.5 need to be discussed (just due to increase precipitation?).
Thanks for pointing this out and we have modified and provided information on RCP4.5 and 8.5 values coming from the COSMO-CLM projections referring to the publication of Bucchignani et al., 2016 where total precipitation show a consistent trend with results for the variables inflow, outflow and volume (lines 356-358 in the marked-up version).

L324-332. I'm not sure if looking at trend within the 30-year window is a good practice, since statistics on 5 years are not robust enough for climate projections. As I said above, I will focus on comparing short and long term.
Thanks for pointing this out and we have modified the text according to the new Figure representing boxplots over 30-years time slices (lines 373-382 in the marked-up version).

L360. This should be Figure 7.
We have corrected the text accordingly (line 409 in the marked-up version)

Figs. 8 and 9. remove absolute frequency, since the results are identical to relative frequency (as expected).
Thank you and we removed absolute frequency from Figures 8 and 9.

L384. …longer maximum duration…
Thank you and we have corrected the text accordingly (line 433 in the marked-up version)

L392. last longer in the case of the most extreme events (you are still looking at maximum duration only). Please carefully check the whole section for similar inconsistencies.

Thank you for the feedback and we modified this sentence to explicitly refer to the most extreme events as suggested and also modified lines 441, 455 and 508-509 in the marked-up version for consistency.

L426. I would replace trend with behavior/dynamic, since you are not really looking at trends here. Also, an adjective is needed in this sentence in my opinion (represents well/perfectly, how?).
We have corrected the text accordingly and moved it to the conclusions (line 550 in the marked-up version).

L426-433. This may fit better as the start of a conclusion section than of a discussion.
We moved this part at the beginning of the conclusion (line 550-554 in the marked-up version).

L449. It would be nice to simulate this, by changing the way water usage are modeled in the future, and having a quantification of the actual effect on summer deficits.
We agree on the importance of testing adaptation strategies to quantify their effects, but we believe this addition would change the current focus/objective of the manuscript.

L475. The climate model…. Are your referring to this specific climate model (COSMO-CLM)? Please clarify. In any case, the use of multiple models is important to account for uncertainty in the future, since even if a climate model reproduces well the climate of the past, this does not guaranty a good performance for the future.
We thank the reviewer for pointing at this and we have corrected the text accordingly (line 527 in the marked-up version).