Reply to R2

In this manuscript, the authors attempt to map snow droughts in Italian mountain regions onto a wide range of impacts (hydrological, ecological, sociological). They use a high-resolution Italian snow reanalysis product to calculate basin-wide snow water equivalent for 38 montane catchments, and identify snow droughts as water years in which the basin-wise December-June average SWE falls below the 30th percentile over the WY 2011-2023 period. They then identify how warm-season runoff and runoff efficiency in their catchments differ between snow drought and non-snow drought years; examine differences in vegetation phenology and productivity derived from remote sensing, as well as two in situ flux towers; and characterize societal impacts using reports of water restrictions and surveys of mountain hut managers. They find reduced runoff and runoff efficiency following snow droughts, modest increased gross primary productivity, and water supply impacts.

Overall, I found the manuscript very interesting, as understanding the impacts of snow droughts in a non-stationary climate is important to informing adaptation in snow-dependent regions. The effort to bring together the multifaceted impacts of snow-drought from a variety of different data sources was impressive. I also found the paper generally well-organized and well-written. I do, however, have a couple significant concerns about some of the methods and framing of results that I feel need to be addressed before I would consider the manuscript ready for publication. I detail each of these points below.

> We thank the Reviewer for their extensive review. Revisions are feasible, and we detail our revision strategy below.

Major Comment 1: Choice of snowpack measure. The authors characterize the winter snowpack by taking the average basin-wide value over the December-June period, including no snow days. My concern is that this measure does not capture just snowpack accumulation, but also aliases the length of the accumulation and melt seasons. For instance, a basin could reach the same peak SWE in year 1 and year 2, but if the snowpack begins accumulating at different points (say 1 December versus 1 January), the measure the authors use would report a very different value, even though the following warm season would likely be very similar hydrologically, at least from a snowpack perspective. Additionally, even if a fixed window were appropriate, the period seems much too wide. Looking at the histograms in Figure 2, the overwhelming majority of basin-years have snow durations, much less than 200 days, except at the highest elevations, which means that this window includes large chunks of time outside the snow season, especially at the lower elevations. I would strongly encourage the authors to use a more commonly, hydrologically-relevant measure such as basin-scale peak SWE to characterize the snow season. At the very least, the authors would need to validate their chosen measure by examining its relationship to warm-season streamflow vis-a-vis other snowpack metrics like peak SWE to justify its use as a hydrologically-relevant variable.

> This is a very good point. In the revised manuscript, we will also consider peak SWE and then decide on whether to use that or mean SWE based on the explanatory power vs. streamflow and other impacts. Thank you!

Major Comment 2: Snow drought definition. The authors define snow droughts as years in which their snowpack measure falls below the 30th percentile over the 2011-2023 period. Essentially, this choice imposes a fixed number of snow droughts (4) for each basin. There are numerous ways to define snow droughts, and the choice of method necessarily affects the set of events identified and thus the set of impacts one might ascribe to them (Gottlieb & Mankin, 2022). I would encourage the authors to consider a broader range of methods from the literature for identifying snow droughts and make sure that their results are robust to that analytical choice.

> Thanks for this comment. We agree that the definition of snow drought has implications in terms of how we interpret impacts. Our definition and the resulting snow droughts align very well with our knowledge of what years were characterized by low snow accumulation across the Italian mountains, which was an indirect validation of our choice (see, e.g., `https://essd.copernicus.org/articles/15/639/2023/` or `https://iopscience.iop.org/article/10.1088/1748-9326/acdb88`). We will mention this as well as review other methods, at least in the Discussion section.

Major Comment 3: Visualization of distributions and use of statistics. In Figures 1 and 2, it would be far more informative to show the distributions of local basin-year anomalies of temperature, precipitation, streamflow, etc., rather than the absolute values, as the current distributions largely show the differences in climatology across stations rather than the difference between snow drought and non-snow drought years. By calculating the departures from normal (e.g., degrees for temperature and percentages of normal for hydrologic quantities) and plotting those distributions, it would make those differences much more salient and easier to interpret. It would also make for a more appropriate statistical test, as it would compare within-basin anomalies in snow drought versus non-snow drought years, whereas now it also includes between-basin differences in climatology, which doesn't provide much information about the snow drought effect. Finally, the numbers the authors report about differences between snow drought and non-snow drought years (e.g., -50% length of snow season, -50% summer runoff, etc.) should reflect these local anomalies (if they do not do so already; I cannot tell where those values are coming from), rather than the pooled distribution.

> We agree with this comment, and we will recompute results in Figure 1 and 2 based on anomalies rather than absolute values.

Major Comment 4: Overinterpretation of GPP results. This is somewhat related to the last point, but the authors do not, as far as I can tell, apply any statistical tests to assess whether the vegetation metrics they examine are significantly different between snow drought and non-snow drought years. My guess is that they are not, as the authors show quite a conservative uncertainty range (inter-quartile range in Figures 5 and 6), and the differences between snow drought and non-snow drought years appear quite modest. Even in the cases with the largest departures above 1500m, the median in snow drought years seems to sit at about the 60th percentile of the non-snow drought distribution, which hardly seems like meaningful support for the claim that GPP is higher following snow droughts. This section either needs to be supported by strong statistical evidence if it exists, or the authors need to report this accurately as a null result — to be clear, it would be a very interesting null result! — if that is indeed the case.

We thank the reviewer for this important comment. To formally assess differences between snow drought and non-snow drought years, we conducted additional statistical analyses based on annual aggregates for each catchment and elevation class. Specifically, we performed paired comparisons across catchments and computed effect sizes and confidence intervals using a t-based approach for both GPP and green-up timing. For GPP, results indicate a clear elevation-dependent pattern. At lower elevations (0–1000 m), differences between drought and non-drought years were small and not statistically significant (e.g., 0–500 m: mean difference = 3.4 [gCm-2], 95% CI [13.0, 6.2], $p = 0.54$). At intermediate elevations (1000–1500 m), differences were modest but statistically significant (mean difference = 10.7, 95% CI [2.1, 19.3], $p = 0.019$). At higher elevations (>1500 m), GPP was consistently and significantly higher in drought years, with larger effect ranges (e.g., 2000–2500 m: mean difference = 33.4, 95% CI [25.8, 41.0], $p < 0.001$).

In contrast, green-up timing showed a consistent and statistically significant response across all elevation ranges. Green-up occurred earlier in drought years at all elevations, with mean advances of approximately 2–3 days at low and mid elevations (e.g., 0–500 m: 2.4 days, 95% CI [4.2, 0.7], $p = 0.016$) and larger advances at higher elevations (e.g., >2500 m: 6.7 days, 95% CI [10.7, 2.6], $p < 0.001$). The magnitude of this effect increased with elevation.

Together, these results clarify the patterns shown in Figures 4 and 5. While the visual differences in GPP are modest and not statistically significant at lower elevations, the statistical analysis confirms a significant increase in GPP following snow drought at higher elevations. At the same time, the consistent advancement of green-up across all elevations suggests that phenological responses to snow drought are more widespread, whereas productivity responses are more strongly constrained by elevation.

We will revise the manuscript to include these statistical results and to more accurately reflect the strength and elevation dependence of the observed effects.

The title is a bit awkwardly worded. I would drop the "Multisectorial impacts of" part, as mountain socio-ecohydrology already captures the fact that the authors are looking at a diverse set of impacts.

Based also on comments from Reviewer 1, We will simplify the title as "Impacts of Mediterranean snow droughts on mountain socio-ecohydrology".

l. 107-108: It would be useful context to provide the snowfall fractions of your mountain catchments, or report values from the literature.

Yes! We will take these from the literature, as computing them ourselves would require processing significant additional data.

l. 154-155: Are these error values for peak SWE/snow depth? They seem quite large. More detail would be useful here. What are those errors as a percentage of average? How do they vary across elevations?

Thanks for this comment. While they might appear large, we should always bear in mind that IT-SNOW is a national-scale, operational modeling chain at 200/500 m resolution. This necessarily degrades performance compared to a point-scale application of a snow model. As we discussed in Avanzi et al., 2023, this performance is in line with other similar large-scale reanalyses. We will add this to the manuscript.

l. 159: A bit more information about the BIGBANG dataset would be useful. Is it a reanalysis? Interpolated station data?

We will add this information in the paper. It is an interpolation of in-situ observations.

l. 240: I believe the Hatchett and McEvoy citation is incorrect. In that paper, they identify warm snow droughts based upon 1 April accumulated P > 100% and SWE < 100% of normal and dry snow droughts using 1 April accumulated P < 100% and SWE < 100% of normal. This is a very different snow drought definition than the one the authors use.

We apologize for this mistake, and we will fix this passage in the manuscript.

l. 341. It is unclear where the number of a ∼100% increase in melt-out events comes from, given that the increase at each individual elevation band is <100%.

We agree with this comment. As we wrote in the Abstract, we meant "up to ∼+100%" (precisely, 96% between 2500 and 3000 m). We will fix this.

ll. 348-349: It is also possible that warmer summertime temperatures in snow drought years (to which the snow drought itself may or may not causally contribute) drive increased evapotranspiration and thus lower streamflow and runoff efficiency (Goulden & Bales, 2014)

We totally agree and will add this to the manuscript.

Figures 5 and 6: shading the interquartile range (i.e., the middle 50% of the data, or 6-7 years) is an overly narrow portrayal of the variability in these measures that makes differences appear significant, even when they may not be. I would encourage the authors to either show all 13 individual years (thin and light) with the median heavier and darker, or more of the full range of the data. This would put the snow drought vs. non-snow drought differences in their fully context of the range of variability.

See our main comment above. Differences are indeed significant for intermediate - high elevation & GPP, and significant for all elevations when looking at greening date. In the revised manuscript, we will present the data in a way that more clearly highlights the differences between drought and non-drought years.

Figure 7: I don't believe a boxplot of elevations is appropriate here. Isn't the number of restrictions more relevant? A clustered or stacked bar plot showing the number of restrictions in different elevation bands would be more appropriate.

We agree and will revise this as suggested!

Figure 9: Again, I am not convinced that the results support the GPP/growing season length conclusions.

See our extensive reply above.

ll. 437-438: Your findings run counter to the Trujillo et al. (2012) paper, which finds increased vegetation productivity after snowier winters. Overall, this discussion section

could better position your findings in the context of the literature and attempt to explain possible discrepancies between them.

We agree, and we will incorporate this point into the Discussion. Broadly speaking, both winters with abundant snowfall and winters with snow drought can enhance the following summer's productivity, but through different mechanisms. Snow-rich winters increase water availability beyond typical levels, which will likely boost productivity, particularly in elevation bands that are otherwise water-limited. In contrast, snow-drought winters can enhance productivity at higher elevations that are usually energy-limited and do not require additional water, but instead gain from a shorter-than-normal snow season. Therefore, our findings and those of Trujillo are not necessarily inconsistent. We will add this clarification to the manuscript.

l. 441 Zhao et al., (2022) also show widespread increases in ET during droughts

Noted, we will add this!

l. 445: Pulliainen et al., (2017) also show that early snowmelt advances spring green-up across the northern hemisphere and enhances springtime carbon uptake, which feels like an important part of the literature to cite.

Noted, we will add this!

ll. 456-464: Much of the difference between the grassland and forested site likely has to do with the fact that soil moisture was measured at 30cm, while trees typically have much, much deeper rooting depths that allow them to access deeper soil water and/or groundwater or water in fractured bedrock, which respond more slowly to hydroclimatic variability (Baldocchi et al., 2021; McCormick et al., 2021)

We will include a comment on this in our discussion.