Reply to Anonymous Referee #2

The discussion paper "Seasonal forecasting of snow resources at Alpine sites" by Terzago et al. comprises a thorough analysis of the capabilities of a new framework to predict snow depths over lead times covering the entire winter season. To this end, the authors ran the physically based snowmodel SNOWPACK using input data from seasonal weather forecasting systems by ECMWF and Météo France that were downscaled and debiased to the location of three stations in the Western Italian Alps. The analysis shows some prognostic skill of the framework relative to a climatology-based reference guess.

While an interesting topic and a well-prepared manuscript, my main concern arises from the fact that no (trusted) local precipitation measurements were available. This necessitated the use of ERA5 data as ground truth or target, which likely derailed some of the findings, in particular those that deal with the effect of downscaling and debiasing. I acknowledge that not having accurate precipitation data is the norm, in particular in such a use-case driven study. Yet, the absence of local precipitation data is critical to a point, where I wonder how useful certain sections of this study are at all (4.2 – 4.4). In this situation, I recommend to infer local (solid) precipitation from snow depth using data assimilation and use this in lieu of measurements. This approach would be in line with the authors' underlying premise that the snow model is the strongest members in their system (see line 582). I think SNOWPACK even has an in-built data assimilation mode that allows to use snow depth instead of precipitation as input.

Reply: We acknowledge that the use of ERA reanalysis as a reference for the bias-correction of precipitation can lead to uncertainties in the bias-corrected data. For this reason, following the suggestion of the reviewer, we have inferred total precipitation from snow depth records using the parameterizations already available in the SNOWPACK model. The assimilation of snow depth data in the SNOWPACK model allowed us to obtain a new daily total precipitation time series, which has been compared to the ERA5 total precipitation data in terms of seasonal cycle and probability density function (PDF) of daily precipitation. The results are shown in Figure 2.

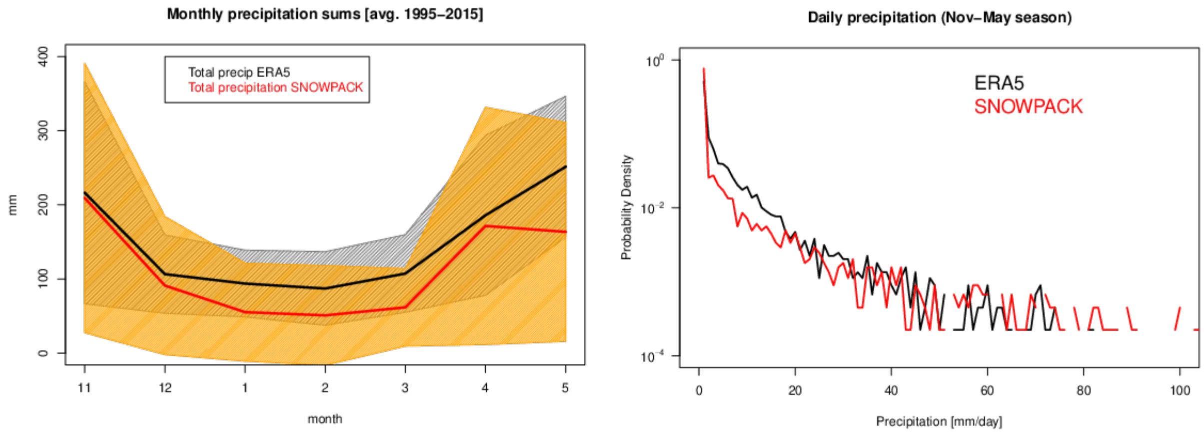

Figure 2: Comparison between ERA5 (black) and SNOWPACK-derived (red) total precipitation: a) seasonal cycle of total precipitation from November to May expressed in terms of multiannual mean and corresponding standard deviation, and b) probability density function (PDF) of daily precipitation in the two datasets.

Figure 2a shows the seasonal cycle of total precipitation from November to May in the two datasets, and in particular the multi-annual mean and the standard deviation of monthly values, the latter being an indicator of interannual variability. ERA5 and SNOWPACK seasonal cycles are quite similar, with ERA5 laying in the range of the SNOWPACK mean ±1 standard deviation. Looking at the differences between the two datasets, ERA5 shows slightly higher monthly means in late winter (JFM) and late spring (May), and less interannual variability (smaller standard deviation) compared to SNOWPACK. More in detail, SNOWPACK allows for lower monthly precipitation values and lower daily precipitation values in the range between few mm/day and 20 mm/day (Figure 2b), while above this threshold the PDFs of the two datasets are similar. In conclusion, although with these differences, both ERA5 and SNOWPACK total precipitation datasets can be considered valid approximations of the observed precipitation amount, in absence of more accurate observational data. From this analysis we do not expect that the use of ERA5 as a reference for bias correction introduces large discrepancies in the bias corrected data. This is also confirmed by the fact that precipitation bias correction successfully adjusts the huge bias of the MFS6 snow depth climatology and makes it very close to the observed snow depth climatology (Figure 5b of the paper). We discuss this topic more in detail in Section 4.2 on the impact of the bias correction.

Another concern is the lack of discussion about the strength of the forecasting skills. In particular for readers who do not juggle with BBS, CRPSS, or AUCSS on a daily basis, wording such as "demonstrates skill", "show skill", or "surprisingly good skill" is not really meaningful. Is "skill" just slightly better than guessing? The authors have quantitative data, but need to put them in context. I found an interesting statement in line 490: "seasonal forecasts of snow depth appear more robust than streamflow forecasts". This is where quantitative reference should be made to corresponding score data from the evaluation of seasonal streamflow forecasts.

Reply: The evaluation of the forecast characteristics through the various scores and skill scores is a standard in forecast verification, even though not immediately understandable to the non-expert readers. Further explanation on these metrics can be found for example in Calì Quaglia et al., 2021. To help the reader, in our manuscript we employ skill scores only. Skill scores measure the improvement (or the worsening) of a given forecast method compared to a trivial forecast based on the climatology, the persistence of the observed anomaly, etc. The more positive is the skill score, the better is the quality of the forecasts; the more negative is the skill score, the worse is the forecast quality and hence the quality of the forecast method. The added value of the new forecast method is directly understandable from the sign and the absolute value of the skill score. Regarding the statement mentioned by the reviewer, i.e. seasonal forecasts of snow depth seems to be more robust than streamflow forecast: the statement is probably intuitive but should be taken with caution since it is only qualitative and not quantitative. Statements based on skill scores are instead quantitative and more rigorous, so they are the preferred way to convey the results of our analysis. To facilitate the understanding we revised the text of the manuscript improving the explanation of how to interpret skill scores.

Calì Quaglia, F., Terzago, S., and von Hardenberg, J.: Temperature and precipitation seasonal forecasts over the Mediterranean region: added value compared to simple forecasting methods, Climate Dynamics, pp. 1–25, 2021

Further specific comments:

Line 86 / 603: In the context of this study "multi-model" arguably suggests that there is more than one snow model involved, which is not the case. Please revise.

Reply: “Multi-model” has been changed with “multi-system” throughout the text.

Section 2.3.3: Was local terrain accounted for in the downscaling (sky view for longwave, and terrain shading for shortwave)?

Reply: We did not take into account local sky view and terrain shading in the downscaling of radiation. We performed a basic linear interpolation to the station coordinates without considering local factors. This simple approach has been better clarified in the text in Section 2.3.3.

Section 2.4: Even if the authors use soil temperature boundary conditions, I would recommend a spin-up to allow for a realistic initial soil temperature profile for all simulations, i.e. also if snow depth was zero on Nov-1.

Reply: In the SNOWPACK model set-up which we adopted, we do not need to provide the initial soil temperature profile to the model. All SNOWPACK simulations are performed providing the ground temperature, i.e. the soil temperature in the topmost part of the soil at the snow–soil interface, as an input variable. The ground temperature is the unique soil “boundary condition”, and we assume that deep soil layers do not affect the snowpack dynamics, so no soil layer is considered. This assumption is based on the fact that in all the three sites considered the soil is usually continuously covered by snow from early November to the end of the simulation period in May (see Figure 3 below with the

climatological mean values). In general, we can say that from November to May the ground is continuously covered by snow and that i) the ground temperature remains close to 0°C during that period, ii) the soil temperature at lower depths does not fall well below 0°C (i.e. no significant soil freezing occurs during the simulation period) due to the isolating effect of the overlying snowpack. This condition is documented for example in Wever et al., 2015 (Figure 7 of that manuscript) for the Alpine site of Weissfluhjoch, 2540 m a.s.l., located at similar elevation as the stations considered in this study for which no soil temperature measurements are available.
Given this SNOWPACK model setup, we do not need to provide the initial conditions of the soil layers, so we do not need to perform the spin-up to generate them. The spin-up is performed only when snow depth is already present at the beginning of the forecast period (i.e. November 1st): in this case we perform the model spin-up to reconstruct the snow profile, in terms of number of snow layers, layers' thickness, temperature, ice/liquid water content, snow density, etc. This has been better explained in Section 2.4.

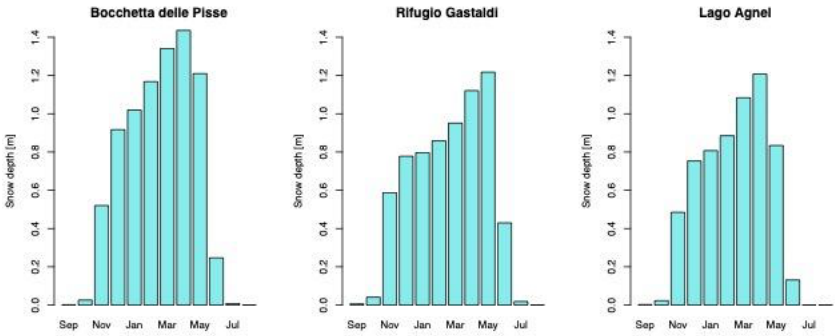

Figure 3: Observed monthly snow depth climatology over the period 1995-2015 for the three stations considered in this study

Section 4.5: I generally appreciate the approach of using a complex and trusted snow model to focus the uncertainty analysis to the forcing data and the models / methods used to derive them. However, at the same time I wonder, if a temperature index model would actually provide better results, because a) it only uses two input parameters and avoids deteriorated performance due to uncertainties associated with the other forcing fields (wind, radiation, ...); b) local calibration of a temperature index model is simple and fast, and can compensate for remaining systematic biases, e.g. arising from the lack of local precipitation data. It's probably not realistic to expect the authors to perform such a comparison as part of this paper, yet, the above consideration would make a useful complement to section 4.5.

Reply: We thank the reviewer for this comment. When planning the experiments we carefully discussed what type of snow model was best for our purpose. We considered a range of models, from simple degree-day models to the most sophisticated ones such as SNOWPACK. Simple models have the advantage of requiring few input parameters, so they allow to avoid uncertainties associated with other forcing fields, as said by the reviewer. The disadvantage of these models is that they need to be calibrated over each study site, so sufficiently long time series of forcing and validation data are necessary to calibrate and validate the model over independent time periods. However, retrospective seasonal forecasts cover a period of 20-25 years, which is already quite short, and further dividing it into two parts would lead to unrepresentative results. On the other hand, sophisticated snow models have higher input requirements and higher computation-load compared to simple snow models, but they have the advantage that they can be directly used without calibration and their snow snow estimates have high accuracy (Terzago et al., 2020) so one can make the hypothesis that the model uncertainty is neglectable with respect to the uncertainty on the forcing. After considering all these points, we agreed that it is better to employ a sophisticated snow model, which guarantees higher accuracy and lower model uncertainty compared to a degree-day model. We better explained this in Section 4.5

Terzago, S., Andreoli, V., Arduini, G., Balsamo, G., Campo, L., Cassardo, C., Cremonese, E., Dolia, D., Gabellani, S., von Hardenberg, J., Morra di Cella, U., Palazzi, E., Piazzi, G., Pogliotti, P., and Provenzale, A.: Sensitivity of snow models to the accuracy of meteorological forcings in mountain environments, Hydrol. Earth Syst. Sci., 24, 4061–4090, https://doi.org/10.5194/hess-24-4061-2020, 2020.

Line 627: Being able to predict that snow will melt at a certain location in, say, May is fairly easy even if the forecasted weather data is uncertain. Having said that, the increase of some of the skill scores towards the end of the season is no real surprise.

Reply: Snow depth seasonal forecasts do not provide only the information mentioned by the Reviewer, i.e. that snow will melt at the end of the snow season, which would be a quite obvious result. They also provide information on the expected snow depth anomaly with respect to the average conditions for the period. In other words, the forecast answers the questions: what is the deviation (excess or shortage) of the forecasted snow depth from the average snow depth value in May? What is the probability that in May snow depth will be below/near/above normal conditions for that period? Seasonal forecasts predict the expected distance of a given variable from the normal ("average") conditions. So the skill in correctly predicting the anomaly at long lead times is indeed quite surprising and a non-obvious result. The snowpack predictability that we find in late spring, i.e. April, can be probably explained by the fact that April is the snow depth accumulation is maximum (see Figure 3) and snowpack is an integrator of the weather conditions and snowfalls over the previous snow season up to April, so even if seasonal forecast models are not able to capture the correct timing/amount of snowfalls, they seem to be able to predict the overall conditions in the winter/early spring period. This has been better clarified in the discussion, thank you for the comment.