Reviewer #2

#RC2.1 This is an interesting, well-written and well-thought manuscript.

We thank the reviewer for this overall positive opinion on this study and his constructives comments.

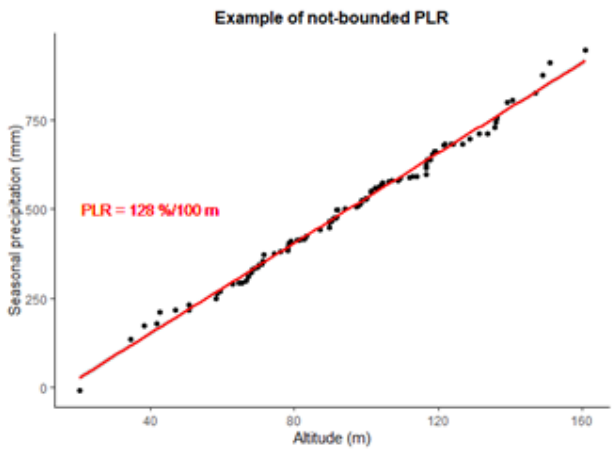
#RC2.2 All or some of the compared precipitation products include altitudinal effects or corrections for such effects. does this create precipitation lapse rates artificially? A comment on this would be welcome.

The interpolator SPAZM applies local linear regressions between altitude and precipitation, conditioned by weather types. This approach aims at adding a physical constraint, with the idea that precipitation lapse rates vary through the territory, depending on the type of precipitation event. Most of the time, the linear regressions reflect a local altitudinal effect and the PLRs are not artificial, that is why SPAZM has been largely used in hydrological modeling (Rouhier 2017). However, we acknowledge that poor regression fits (low R^2 values) can create uncertain altitudinal effects in some situations.

Concerning AROME, the first pre-treatment does not rely on altitude. It aims at correcting the precipitation at problematic pixels, often in high-altitude areas, and to remove artificial precipitation lapse rates. The altitudinal effect on the other gridded precipitation products is not artificial but comes directly from physical models or spatialized observations (radar, satellite).

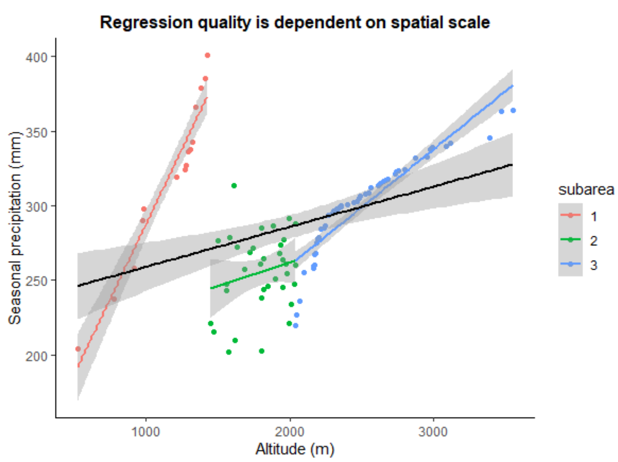
#RC2.3 In eq. 6: the conversion from beta to PLR is unclear to me. A slope is bounded between 0 and infinity, so how can it be converted to a bounded interval between 0 and 100? Also, the denominator P^bar is not defined. Is it the average or median precipitation, and over which area?.

Thank you for this comment. Indeed, we did not indicate that P^bar corresponds to the mean precipitation over the area where the linear regression is performed (either a massif or a catchment). This will be added to the article. The slopes of the linear regressions are not bounded and take values between minus and plus infinity. The altitudinal effect can be negative for hourly/daily precipitation (Formetta 2021). PLRs are not bounded. PLRs are expressed as the increase or decrease of % of the mean precipitation (P^bar) by 100 m. The figure below illustrates these bounds with simulated and nonrealistic data.



#RC2.4 389-390 and other places: I can see how there can be more variability in the PLR when one focuses on smaller scales, but not how the average R2 can change depending on scale. I may be wrong, but in principle the R2 value of a large area should be the average of its component subareas. Showing numerical values might clear any doubt.

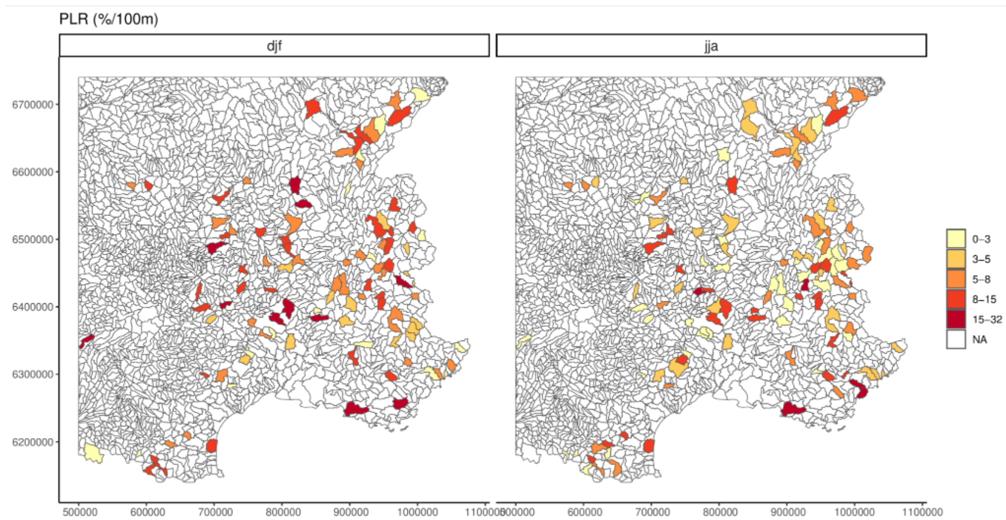
The R2 value of a large area does not necessarily correspond to the average of its component subareas. The following figure illustrates an example.



The R2 calculated on the entire area is 0.14. The R2 values calculated on the three sub-areas are 0.93, 0.05, and 0.84. The mean R^2 is (0.93 + 0.05 + 0.84) /3 = 1.82/3 = 0.607 > 0.14. This often happens when several clouds of dots are present.

#RC2.5 Results in figure 7 (and maybe others): These results are only based on AROME, which is not forced on rain gauge data. While AROME does not perform badly in general, it is not perfect either, and is significantly different than other data-based products. Therefore, I am uncertain of the value of conclusions based only on AROME. Would it be possible to draw a similar figure as figure 7 based only on stations data, even if it means that many more catchments will have NA values?

For most of the catchments, we do not have enough rain gauges to derive PLR using rain gauges (see figure 1c), resulting in a lot of NA values. Furthermore, the PLR values obtained for some catchments where there are only a few gauges may be very uncertain. The figure below these PLR values, and there is no clear spatial patterns.



#RC2.6 The fact that AROME does not use rain gauges is mentioned as an advantage (e.g. l. 461-462), but this poses the following question: if AROME poorly matches the high-elevation gauges, how to know whether it is because of undercatch or because of some deficiency in AROME? If the comparisons were done against stations data that are corrected for a possible undercatch, would AROME still perform as well? I am not asking to do such an exercise, but it could be discussed.

Thank you for this suggestion. We agree that this point should be discussed. We propose to add the following paragraph in the subsection 5.5:

"AROME poorly matches the high-elevation gauges, which could be the combined effect of precipitation undercatch (groisman_accuracy_1994,sevruk_regional_1997, pollock_quantifying_2018} and some deficiencies in AROME (Monteiro, 2022). In Figure 3b, AROME shows a strong agreement with the rain gauges in summer, even the high altitude ones. The differences are larger in winter, where precipitations mostly fall as snow. Precipitation undercatch is limited in summer and more important in winter with solid precipitation. Precipitation undercatch could partly explain the differences among AROME and rain gauges."

#RC2.7 Many of the results are given as figures, but in general it would be good to have tables with actual numerical values (for example in figure 3b the RMSE between stations and models, and also in figure 4)

We decided not to include numerical values for figure 3b and figure 4. For figure 3b, the Pearson correlation coefficient can be calculated for each region and are shown in the tables below. However, the number of altitudinal bands per region is small (five or six observations depending on the region) and limits the statistical interpretation of these values. Moreover, the coefficients obtained from the reanalysis that assimilate rain gauges are close to 1, as expected. We fear that these high values, especially compared to those obtained with AROME, could lead to misinterpretations.

Figure 3b winter

	Foothills Alps	Northern Alps	Pyrenees	Southern Alps
AROME	0.91	0.13	0.97	0.85
CERRA-Land	0.98	0.94	0.97	0.51
COMEPHORE	0.99	0.98	0.84	0.98
SPAZM	0.94	0.99	0.87	0.99

Figure 3b summer

	Foothills Alps	Northern Alps	Pyrenees	Southern Alps
AROME	0.99	0.78	0.39	-0.06
CERRA-Land	0.88	0.59	-0.48	0.69
COMEPHORE	1	0.98	0.82	0.97
SPAZM	0.99	0.96	0.99	0.95

Figure 4

	DJF	JJA
AROME	0.55	0.41
CERRA-Land	0.59	0.42
COMEPHORE	0.85	0.56
SPAZM	0.80	0.71

#RC2.8 l.61: word missing. higher reliability?

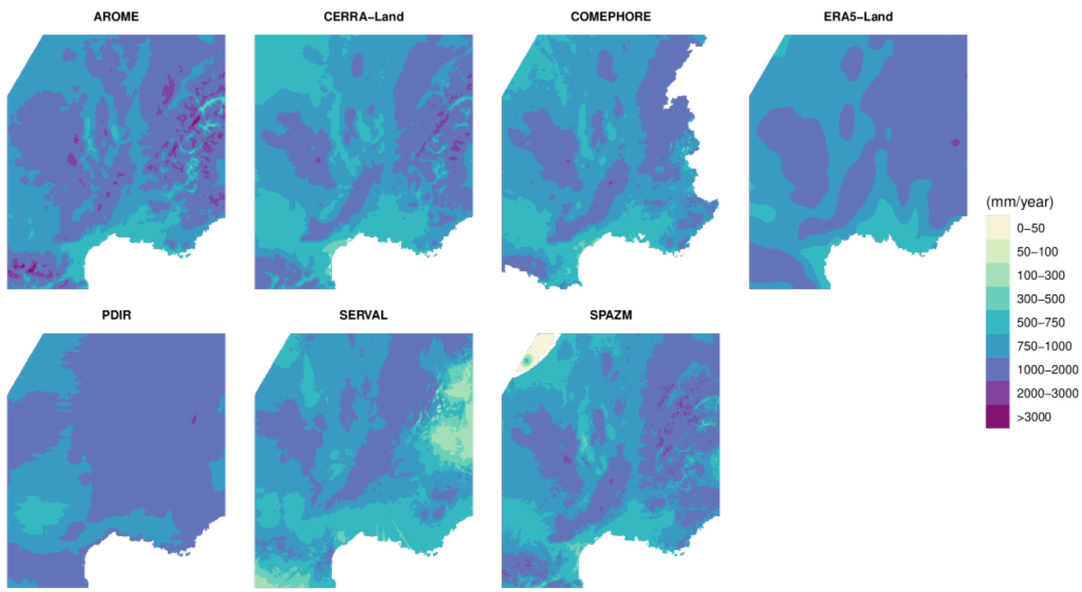
Thank you for pointing out this unclear wording. “higher” will be replaced by “larger”.

#RC2.9 l.180: incomplete sentence, reformulate.

Thank you for noticing this unclear sentence. We propose to change it with the following sentence: “Ground echoes are removed using pixel precipitation probabilities, which are derived from a cloud classification filter.”

#RC2.10 Figure 2: large areas seem to have a mean precipitation close to 0 (e.g. in SERVAL). is this a feature of the data or an artifact of the color scale?

Thank you very much for this comment. We have modified this figure using a discrete color scale (instead of a continuous color gradient):



It was both an artifact of the color scale and a feature of the data. In a large part of the Alps, SERVAL is subject to beam blocking, resulting in a large area of suspicious 50-100 mm of annual precipitation.

#RC2.11 l.275: 50mm is a very small altitude difference. should it be 50m?

Thank you for noticing this typo. It will be corrected by 50 m.

#RC2.12 Figure 4: would be nice to also have numerical correlation values rather than only visual scatters

See our response to the comment #RC2.7.

#RC2.13 l.477: resume -> summarize

Thank you for noticing this wrong choice of word, it will be corrected.

#RC2.14 section 5.4: New methods and results should not be introduced in the discussion section. I recommend moving the part on non-linear results higher in the manuscript.

Thank you for this recommendation. Section 5.4 will be split into two subsections (3.3 method + 4.3 result).

#RC2.15 Figure 9: This figure is very interesting. It shows that larger catchments have more non-linear PLRs. The question that begs for an answer is whether there exists a small subdivision of the catchments where all relations are linear, or if some portions of catchments have irremediably non-linear PLRs.

Thank you for this question. Orographic precipitations are influenced by the leeward and windward mountain faces. To optimize the linearity of the linear regression, we should perform it on small catchments for each slope aspect and wind exposition in order to reduce spatial heterogeneity. However, performing the subdivision would likely require a finer spatial resolution. In addition,some catchments clearly exhibit nonlinear PLRs, irrespective of their size (Jiang 2023) (correlation of precipitation and altitude increase with a decrease of the spatial scale, until a given spatial scale where no improvement is noticed).