Reviewer #3

RC#3.1. The manuscript "Calibrated ensemble forecasts of the height of new snow using quantile regression forests and Ensemble Model Output Statistics" demonstrates the advantages of quantile regression forests (QRF) for postprocessing of the height of new snow. The authors perform an in-depth comparison with ensemble model output statistics (EMOS) in terms of forecast skill and discuss important aspects of QRF with regard to operational implementations. As the paper is well written and its usefulness is clearly motivated, I only have very few comments.

We thank the reviewer for this positive feedback, for these constructive comments and for raising the numerous technical corrections that are required.

RC#3.2. The results suggest that rainfall related predictors improve forecast skill for QRF, probably also because of poor prediction of the snow/rain limit. As the implemented version of EMOS does not use any rainfall (and temperature) related predictors, QRF outperforms EMOS considerably in situations of rainfall, but zero snow, forecasts. For a fairer comparison, I would suggest constructing a second EMOS reference that uses also rainfall (and temperature) based predictors, possibly including interactions. This would help to evaluate, if the increase in skill by QRF really comes from advantageous properties of QRF compared to EMOS or just by the fact that QRF considers more predictors.

Thank you for this comment. We agree that the EMOS version evaluated in Nousu et al. (2019) on HS data and used as a benchmark in our paper is not able to provide correct post-processed forecasts when the snow/rain limit is not predicted accurately, since it does exploit the information provided by other predictors. In our knowledge, there is just one extended version of EMOS-CSGD using more than one ensemble of forecasts (i.e. two or more variables are used as predictors), as proposed by Scheuerer and Hamill (2015). Scheuerer and Hamill (2015) propose to use the precipitable water in addition to the precipitation forecast. This extension could be adapted to our case to add more terms in Eq. A2 and A3 of the current paper. Precisely, these linear combinations can integrate more predictors such as rainfall and temperature forecasts in order to obtain a mean and standard deviation of CSGD that integrate the information about these predictors. We implemented and tested this version, the results in terms of CRPS being shown in Figure 2 below. This version does not lead to an improvement of the CRPS scores.

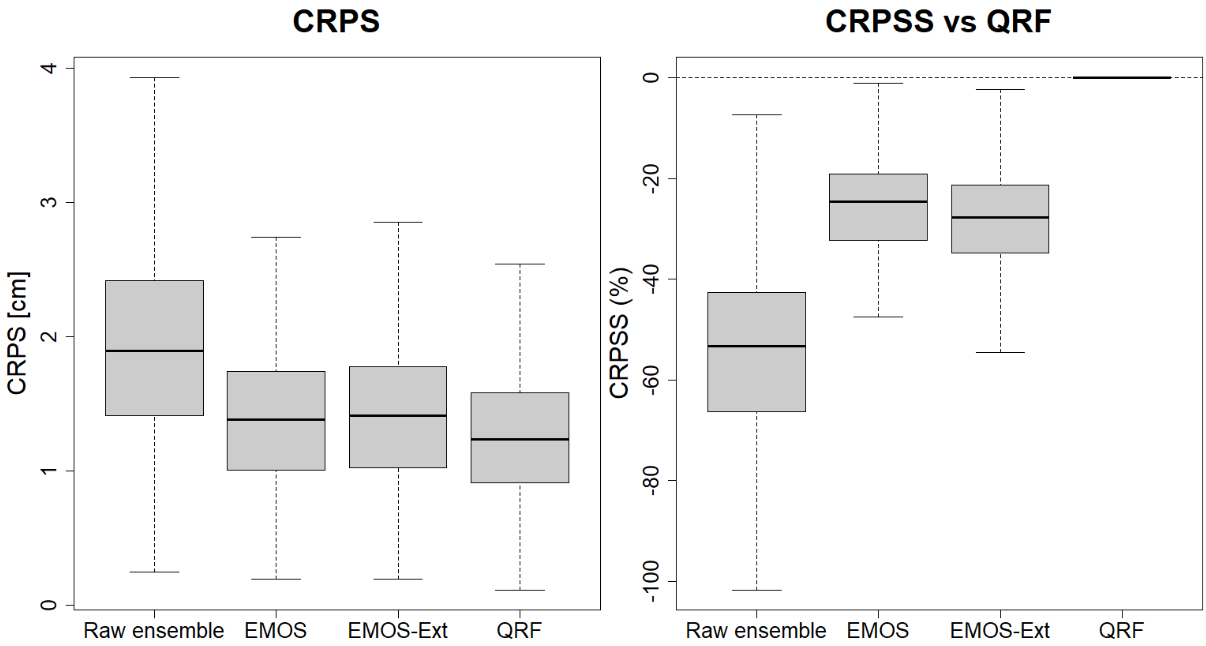

Figure 2: Boxplots of CRPS (left plots) and relative CPRS with QRF as a reference (right plots) with the different methods (Raw, EMOS, EMOS-Ext and QRF) for all locations, for a 1-day lead time. EMOS-Ext corresponds to the extension of the EMOS-CSGD method proposed by Scheuerer and Hamill (2015) where forecasts of rainfall rate and temperature are added to the predictors in addition to the statistics derived from ensembles of HN forecasts.

In this study, the EMOS version proposed by Nousu et al. (2019) was preferred because it has been extensively tested in our region, and we know well its limitations / advantages. It is used operationally in Météo-France. The main purpose of the study presented in the submitted manuscript is to prove that QRF methods could be a valid alternative against this method. However, we will acknowledge in the revised manuscript (in Section 7.1) that different extensions of this EMOS method could also be developed, implemented and tested in order to include more predictors (in particular the boosting extension of EMOS as discussed in the comment RC#1.8).

RC#3.3. An evaluation of forecast sharpness is missing. I appreciate that the authors show not only CRPS values, but also rank histograms as a measure of calibration. However, as forecast quality is determined by both calibration and sharpness, I would suggest adding a figure that compares the sharpness of the raw ensemble, EMOS, and QRF.

Indeed, the current manuscript does not present any criteria that assess the sharpness of the raw and post-processed forecasts. Following Gneiting et al. (2007), we will add a Table presenting the mean width of the predictive intervals (50% and 90% nominal coverages) and boxplots showing the dispersion of these widths for all approaches.

RC#3.3. Figures 3 and 4 are difficult to read because of overlapping intervals. I would suggest modifying the figures such that the limits of the prediction intervals of all forecasts are visible.

Thank you for this comment that was also raised by the two other reviewers (see comment RC#2.7). We will add lines of different types and colors to highlight the different intervals.

RC#3.4. p3/l59: useless question mark

This question mark was actually a missing reference, this problem will be fixed.

RC#3.5. p3/l66: "…statistics of other…"

Thank you, this will be corrected.

RC#3.6. p4/l72: small leading 'e' in ensemble

Ok.

RC#3.7. p4/l75: same leading 'r' in regression

Ok.

RC#3.8. p4/l76: maybe "at zero" instead of "in zero"?

Thank you, "at zero" is better indeed.

RC#3.9. p4/l77: small letters " zero-censored censored shifted-gamma distribution"

Ok.

RC#3.10. p5/l94 to 99: Don't we minimize the within group variance when maximizing homogeneity.

Thank you for this comment. We agree that this paragraph can be confusing since we indicate that we want to maximize the homogeneity (which is correct) and then we define the homogeneity as a sum of variances that we want to be the smallest possible. We will rephrase this part in the revised manuscript.

RC#3.11. p6/l121: "…22 seasons, one…"? as "the" sounds odd to me here

We agree, this will be modified.

RC#3.12. p6/l126-127: only if you apply score decomposition. Or, is the word "simultaneously" missing here.

Yes, we meant simultaneously. This will be added.

RC#3.13. p7/l132: "…variable and equals…"

Thanks, “is” will be removed.

RC#3.14. p7/l136: What do you mean by “technical”?

We mean difficult, not straightforward. “Difficult” might be more adequate here.

RC#3.15. p8/l165: CRPS

Thanks, this will be corrected.

RC#3.16. p8/l172: “…variable, i.e. node…”

Ok.

RC#3.17. Figure 2: I would make clearer that you are analysing the results for 24h accumulations of new snow depths here. Maybe, 1-24h, 25-48h, etc. in the subpanels’ titles.

Ok, this will be done.

RC#3.18. p10/l190: From Figure 4 I guess that the station ID of Saint-Paul-sur-Ubaye should be 4193400.

Thank you for noticing this error, this will be corrected.

RC#3.19. p13/l199: Please do not mention significance without having performed any statistical hypothesis test. Either write, e.g., “considerable” or apply a statistical hypothesis test.

We agree, “significance” will be replaced by “considerable”.

RC#3.20. p14/l203: “…leading to U-shape….”

Thanks, this will be corrected.

RC#3.21. p16/l216: Probably blue instead of read

Thanks, that was a mistake.

RC#3.22. caption of Figure 8: 2nd word: CRPS

Ok.

RC#3.23. caption of figure 8: Definitions in the caption and the titles of the subpanels for subpanels (b) and (c) are interchanged.

Thank you for noticing this error, it will be corrected.

RC#3.24. p18/l261 “adiabatic” instead of “adiabiatic”

Ok.

RC#3.25. p18/l265: Do you mean “statistical post-processing” instead of “statistical processing”

Yes, that what we mean, this will be modified.

RC#3.26. p19/l266: What is "iso-\theta_{w}^{'}"? The 1-degree Celsius isothermal level in terms of pseudo-adiabatic wet-bulb temperature?

The pseudo-adiabatic wet-bulb temperature is the temperature that would have an air particle after an adiabatic cooling until water saturation level and then an adiabatic compression until the 1000 hPa level. The definition is available in *Compendium of Meteorology - for use by class I and II Meteorological Personnel: Volume I, part 2 - Physical meteorology, WMO, 1973, page 122, available at* https://library.wmo.int/index.php?lvl=notice_display&id=7078#.YPVtZKaxVhE
The reference will be added in the revised manuscript.

RC#3.27. p19/l273: "…with statistical post-processing…" w/o the "a"?

The "a" will be removed.

References

Gneiting, T., F. Balabdaoui, and A. E. Raftery. 2007. "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69 (2): 243–68. https://doi.org/10.1111/j.1467-9868.2007.00587.x.

Nousu, Jari-Pekka, Matthieu Lafaysse, Matthieu Vernay, Joseph Bellier, Guillaume Evin, et Bruno Joly. « Statistical Post-Processing of Ensemble Forecasts of the Height of New Snow ». *Nonlinear Processes in Geophysics* 26, nº 3 (26 septembre 2019): 339‑57. https://doi.org/10.5194/npg-26-339-2019.

Scheuerer, Michael, et Thomas M. Hamill. « Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions ». *Monthly Weather Review* 143, nº 11 (1 septembre 2015): 4578‑96. https://doi.org/10.1175/MWR-D-15-0061.1.

WMO. *Compendium of meteorology - for use by class I and II Meteorological Personnel: Volume I, part 1 - Dynamic meteorology*. Publications of blue series, volume 1 (1955-1984) - Education and Training Programme (2004), 1973.