Reviewer #2

RC#2.1. This is a useful paper which shows very clear benefits from calibration of ensemble forecasts of snow depth. Two calibration methods are compared and the quantile regression method shows clear advantages over the more standard EMOS approach (although it should perhaps be noted that there are many ways of implementing EMOS and other approaches could perform better than the one used here). It is good also that the authors have included the section 7.3 on the limitations for operational use. This is an important factor, that many calibration methods are mathematically skilful but not practical to apply for real-world forecasting, often due to lack of suitable training data, so it is good to discuss this openly in the paper. I would recommend publication with only minor amendments.

I have suggested to Editor that the title is not understandable to a wide audience (see comment 1 below). I have also suggested that not all figures are of appropriate quality (see note 6 below.)

We thank the reviewer for these positive comments and suggestions. We answer below.

RC#2.2. For me the term "height of new snow" is confusing. I suspect this is simply a slight mistranslation from the authors' native French, but causes confusion to a native English speaker. The normal English term would be "depth of snow", whereas "height" would be used more for the altitude (height up the mountains for example) where snow would occur. (A Google search for "height of new snow" returns many references to depth of snow.) I would recommend changing the word height to depth throughout the paper, including the title, and hence also HN to DN. However, I do note from the references that the authors have published a previous paper on the topic using the same term "height of new snow", so I would understand if they want to keep it for consistency. In this case, it would be worth defining what they mean clearly in the Introduction to avoid confusion.

The use of the term "height of new snow" was asked by a reviewer of a previous publication (Vernay et al., 2015) because this is the official name for this variable in the International Classification for Seasonal Snow on the Ground of the International Association of Cryospheric Sciences (IACS) (Fierz et al., 2009). Therefore we now apply this recommendation in all our publications for homogeneity. We will add the reference at the first occurence of the term in the introduction.

RC#2.3. Line 34: Delete the word "from": "…This prevents an appropriate correction…"

Thank you, this will be corrected in the revised manuscript.

RC#2.4. Table 1: Abbreviation IQR is not defined – assume it is Inter-quartile Range – but should be defined.

This will be added.

RC#2.5. It is interesting that all the predictors used are univariate ensemble summary statistics which means that correlations between variables present in the ensemble members are lost. This might be worth some mention – it is very encouraging that the

methods are successful, but it might be expected that some higher skill might be achievable if correlations between for example precipitation and Near-surface temperature were retained. Might be worth comment.

Thank you for this interesting comment. What we understand is that the correlations between the different variables for the same ensemble could be exploited to improve our prediction. Computing the cross-correlations between some variables could indeed be an option. Note that a closely related perspective could consist in computing additional predictors based on several variables present in the forecasts (as done in, e.g., Zamo et al., 2014; Whan and Schmeits , 2018). The choice of the most relevant combinations in our case remains an open question though. If the reviewer has a specific reference in mind, we will be pleased to add it to the discussion.

RC#2.6. Line 158: The notation used for the intervals looks odd, with opening square brackets at both ends. In figure 6 the closing at the end of the interval uses a closing round bracket, which looks better.

Thank you for this suggestion, this will be modified.

RC#2.7. Figures 3 and 4: I found the colours difficult to interpret when they are overlaid. It does become easier in combination with the text description, but I would suggest some alternative which clarifies the ranges for each colour. Perhaps you could mark the upper and lower bounds (10th and 90th) of each shading with overlaid lines in strong colours. (Also, this would be much worse for someone who is colour blind and cannot distinguish red and green – a different set of colours would be better but if you add lines as suggested then they could also use different line patterns.)

Thank you for this suggestion. We will change the colors to suit colorblind safe colors and add lines of different types (e.g. plain, dotted, dashed) as suggested. It is also asked by the other reviewer.

RC#2.8. Line 200: "QRF leads to an improvement …" – technically the plots show that EMOS and RAW are degraded relative to QRF.

Ok, we will change this comment to "*For most of the stations, EMOS shows a degradation of the performances between 20% and 30%, up to 40% compared to QRF*".

RC#2.9. Line 210: The term ROC has not been defined "Relative Operating Characteristic" (or alternative versions of the name). Also, you do describe ROC here briefly in lines 210-212, but why did you not define it in section 5 where all the other evaluation scores are defined?

Thank you for this comment. As suggested, for the sake of consistency, we will move this paragraph to the evaluation section. We will also add the definition of the term ROC "Relative Operating Characteristic".

RC#2.10. Line 216: You are describing the blue curve here, not the red one.

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

RC#2.11. Line 316: "he" should be "the"

Thank you very much for noticing this typo. "he climatology" will be changed to "the climatology".

References

Fierz, C., R. Armstrong, Y. Durand, P. Etchevers, E. Greene, D. Mcclung, K. Nishimura, P. Satyawali, and S. Sokratov. 2009. “The International Classification for Seasonal Snow on the Ground (UNESCO, IHP (International Hydrological Programme)–VII, Technical Documents in Hydrology, No 83; IACS (International Association of Cryospheric Sciences) Contribution No 1).”

Vernay, M., M.Lafaysse, L. Mérindol, G. Giraud, and S. Morin. 2015. “Ensemble Forecasting of Snowpack Conditions and Avalanche Hazard.” *Cold Regions Science and Technology* 120 (December): 251–62. https://doi.org/10.1016/j.coldregions.2015.04.010.

Whan, K., and M. Schmeits. 2018. “Comparing Area Probability Forecasts of (Extreme) Local Precipitation Using Parametric and Machine Learning Statistical Postprocessing Methods.” *Monthly Weather Review* 146 (11): 3651–73. https://doi.org/10.1175/MWR-D-17-0290.1.

Zamo, M., O. Mestre, P. Arbogast, and O. Pannekoucke. 2014. “A Benchmark of Statistical Regression Methods for Short-Term Forecasting of Photovoltaic Electricity Production, Part I: Deterministic Forecast of Hourly Production.” *Solar Energy* 105 (July): 792–803. https://doi.org/10.1016/j.solener.2013.12.006.