

Nonlin. Processes Geophys. Discuss., referee comment RC2
<https://doi.org/10.5194/npg-2021-18-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on npg-2021-18

Anonymous Referee #2

Referee comment on "Calibrated ensemble forecasts of the height of new snow using quantile regression forests and ensemble model output statistics" by Guillaume Evin et al., Nonlin. Processes Geophys. Discuss., <https://doi.org/10.5194/npg-2021-18-RC2>, 2021

Overall:

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.)

Specific Comments:

- 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.

- Line 34: Delete the word “from”: “...This prevents an appropriate correction...”
- Table 1: Abbreviation IQR is not defined – assume it is Inter-quartile Range – but should be defined.
- 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.
- 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.
- 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.)
- Line 200: “QRF leads to an improvement ...” – technically the plots show that EMOS and RAW are degraded relative to QRF.
- 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?
- Line 216: You are describing the blue curve here, not the red one.
- Line 316: “he” should be “the”