

Nonlin. Processes Geophys. Discuss., referee comment RC1
<https://doi.org/10.5194/npg-2021-18-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on npg-2021-18

Anonymous Referee #1

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-RC1>, 2021

The authors propose a quantile regression forest (QRF)-based postprocessing method for the height of new snow (HN). The results are compared to a recently proposed ensemble model output statistics (EMOS) approach for postprocessing HN forecasts. QRF shows clear improvements over the EMOS model, in particular the inclusion of additional predictor variables seems to be beneficial.

Overall, I found the paper to be interesting, well-written and easy to follow throughout. I only have some minor and technical comments that are outlined below.

Minor comments

- I found the description of the forecast distribution in Section 4.3 a bit confusing. Perhaps it would help to specifically clarify here that the resulting forecast is a set of quantiles derived from the observations from the final nodes, and not the empirical distribution in equation (1).
- line 145 ff: Perhaps a few more details should be provided on why the CRPS is computed in this form. The motivation to create more quantile forecasts than raw ensemble members is not entirely clear to me. In the end, you compute an ensemble of quantiles that has many more members than the raw ensemble. While this makes it easier to account for the necessarily finite size of the sample from the forecast distribution and makes the comparison to EMOS (with a continuous forecast distribution) more "fair", doesn't this represent an "unfair" advantage when comparing to the raw ensemble?
- line 169 ff: Compared to the description of the QRF model, the description of feature importance is rather short and probably difficult to understand for readers unfamiliar with QRF. Perhaps a few more details and formulas here might help make this more clear.
- Section 6, first paragraph: Are the CRPS values computed for the test set, the training

set, or another validation period?

- Figure 2: Given that the show forecasts seem to be of particular importance, wouldn't include more summary statistics from that variable further improve results?
- Figure 3: I find the confidence intervals difficult to distinguish due to the overlap and would suggest to split up the plot into three panels for all of the 3 models.
- Overall, the paper in particular demonstrates that the inclusion of additional predictor variables improves performance. This is very much in line with the previous work on QRF and also several other machine learning-based postprocessing methods (for example proposed in Messner et al (2017, <https://doi.org/10.1175/MWR-D-16-0088.1>), Rasp and Lerch (2018, <https://doi.org/10.1175/MWR-D-18-0187.1>), Bremnes (2020, <https://doi.org/10.1175/MWR-D-19-0227.1>), and others). Since EMOS only uses forecasts of the variable of interest as predictor, it would have been more "fair" to compare the the boosting extension of EMOS proposed in the paper by Messner et al. While this is beyond the scope of the paper in the current form, I'd suggest to include this aspect in the discussion as an avenue for future work.

Technical comments

- line 59: Missing reference?
- Section 4.2: I'd suggest to consider moving this Section to the beginning of Section 6. In the current form, the meaning of the hyperparameters `mtry` and `nodesize` is not yet explained and will be difficult to understand for readers not familiar with QRF.
- Code and data availability: The limitations on availability of the EMOS and NWP model code are clearly explained, but it is unclear to me whether (or where) the QRF code is available.