

Nonlin. Processes Geophys. Discuss., referee comment RC3
<https://doi.org/10.5194/npg-2021-18-RC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on npg-2021-18

Anonymous Referee #3

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-RC3>, 2021

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.

Specific comments:

- 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.
- 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.
- 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.

Technical corrections and typing errors:

- p3/l159: useless question mark
- p3/l166: "...statistics of other..."
- p4/l172: small leading 'e' in ensemble
- p4/l175: same leading 'r' in regression
- p4/l176: maybe "at zero" instead of "in zero"?
- p4/l177: small letters " zero-censored censored shifted-gamma distribution"
- p5/l194 to 99: Don't we minimize the within group variance when maximizing homogeneity.
- p6/l121: "...22 seasons, one..."? as "the" sounds odd to me here
- p6/l126-127: only if you apply score decomposition. Or, is the word "simultaneously" missing here.
- p7/l132: "...variable and equals..."
- p7/l136: What do you mean by "technical"?
- p8/l165: CRPS
- p8/l172: "...variable, i.e. node..."
- 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.
- p10/l190: From Figure 4 I guess that the station ID of Saint-Paul-sur-Ubaye should be 4193400.
- 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.
- p14/l203: "...leading to U-shape..."
- p16/l216: Probably blue instead of read
- caption of Figure 8: 2nd word: CRPS
- caption of figure 8: Definitions in the caption and the titles of the subpanels for subpanels (b) and (c) are interchanged.
- p18/l261 "adiabatic" instead of "adiabiatic"
- p18/l265: Do you mean "statistical post-processing" instead of "statistical processing"
- p19/l266: What is "iso-\theta_{w}^{\wedge}"? The 1-degree Celsius isothermal level in terms of pseudo-adiabatic wet-bulb temperature?
- p19/l273: "...with statistical post-processing..." w/o the "a"?