Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-64-RC2, 2020

Nonlinear Processes in Geophysics
Discussions

Interactive comment on "Statistical Postprocessing of Ensemble Forecasts for Severe Weather at Deutscher Wetterdienst" *by* Reinhold Hess

Anonymous Referee #2

Received and published: 27 February 2020

General comments

The manuscript describes the so-called Ensemble-MOS as operated at Deutscher Wetterdienst. The statistical methods applied in Ensemble-MOS are quite standard. Ensemble-MOS is based on univariate multiple linear and logistic regression. The interesting aspects of this manuscript is the operational implementation, which uses a large amount of predictors and predictants and a stepwise variable selection. The manuscript contains interesting aspects of how to apply postprocessing in an operational setting. Nevertheless, an extensive revision of the manuscript is necessary, especially with regard to structure and language. Further, several aspects should be

examined more closely.

Specific comments

1. The general readability of the manuscript should be greatly improved. A clear separation in sections of (i) general aspects in postprocessing with literature, (ii) observations and forecasts, (iii) statistical methodology, (iv) DWD specific implementation and (v) results would largely increase the readability of the manuscript.
2. A thorough revision of the English language is necessary. Language editing should also pay attention to the clarity in terminology.
3. Since the manuscript employs standard statistical methods, the description of the statistics may be shortened. On the other hand, the description of the non-standard approach (why?) for wind gusts (section 3.3) is very confusing.
4. The results and particularly the figures should be discussed more closely. The order the figures are mentioned in the article must be consistent with the figure number.
5. The title emphasizes on severe weather. The manuscript describes both, and severe weather is restricted to wind gusts. So I suggest investigate postprocessing for severe weather more closely. No results are given with respect to the warning issue. No scores are presented, and no comparison is provided of the performance of the not-postprocessed ensemble and the benefit obtained with postprocessing.
6. No results a given with respect to the selected predictors, particularly for severe weather it might be interesting, which variables are most informative.

7. More details could be provide on the effect of model changes. In line 354ff you discuss the inclusion of an indicator. Do you detect any sudden changes? More details may also be provided, how the change from COSMO-DE-EPS to COSMO-D2-EPS affects the postprocessing (you only mention small improvements in COSMO-D2-EPS in lines 363-365).

Minor comments and typos

L. 2 "support warning decision management" –> "support the warning decision management"

L. 8 "capture sufficient number" –> "capture a sufficient number"

L. 8 Unclear terminology "significant training data"?

L. 9 Unclear terminology "threshold probabilities"

L. 24 " ... the estimated forecast errors meet the observed errors of the ensemble mean." This is confusing to me. You refer to the consistency of the spread of an ensemble and the ensemble mean forecast error in a perfect ensemble.

L. 26 "... high atmospheric variables ..." –> "... atmospheric variables in higher vertical layers ..."

L. 30 "This calibration should be done in respect to maximize forecast sharpness" Apart from linguistic errors, it is exactly the opposite, statistical postprocessing maximizes sharpness on the condition of calibration.

L. 85 "..., since predictors also need to be detected that are not correlated to extreme events in order to ..." This sentence is confusing. You may refer here to the 'Forecaster's dilemma: Extreme events and forecast evaluation' by Lerch et al. in 2017.

L. 88 Ben Bouallégue et al. (2015) shows how to derive probabilistic forecasts for specific cost-loss scenarios.

L. 91 "... warning issuance" –> "... warning insurance" ?

L. 135 "underdispersivity" –> "underdispersion" ?

L. 171ff Not sure to understand this paragraph.

L. 195ff Not clear what and why you are doing this.

L. 259 The cost function is indeed the likelihood of the binomial distribution.

Eq. 2 It is unfortunate to use p as the number of predictors, since p is generally associated to probability. Later on this lead to some confusion (line 313).

Eq. 2 Clarify that $\hat{y}_p$ is a probability, and the observation y is a binary variable taking values 0 and 1 (Bernoulli trials).

L. 263 "logarithm" –> "logarithm of the likelihood"

L. 336 You say its effective, but you do not show anything in this respect. I may also be interesting to see, where Ensemble-MOS in not effective (e.g. what about fog).

L. 385 You may refer here to approaches to restore dependence (ensemble copula coupling, Shaake shuffle (e.g. Schefzik, 2016 and others).

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2019-64, 2020.