Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/nhess-2021-215-RC1, 2021

Comment on nhess-2021-215

Krister Kristensen (Referee)

Referee comment on "Methodological and conceptual challenges in rare and severe event forecast-verification" by Philip Alexander Ebert and Peter Milne, Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-215-RC1, 2021

General comments:

The paper addresses the methods for assessing the predictive quality of forecast of rare and severe events (RSE), with special consideration of snow avalanche forecasting. The authors argue that the Peirce Skill Score is the only score that meets all three adequacy constraints that is required for a proper skill score for RSE forecast-verification.

The paper has s comprehensive overview of forecast verification and measures that can be used to assess the quality of forecasts. It raises very important issues for forecasting practice and has valuable discussions and suggestions.

The paper is easy to read, but somewhat long.

Specific comments:

The discussion on the use of the verbal descriptions of the EAWS danger scale in chapter 5 (509) is interesting. Good examples of forecasting problems requiring more probabilistic statements are mentioned later in the chapter. It could maybe have been pointed out that the use by many forecasting services of verbiage instead of probabilities makes both verification and improvement very difficult, and that this is a major weakness for comprehensive risk management.

The paper relates to both binary RSE forecast-verification and multi-categorical skill measures for i.e. regional avalanche forecasts. The focus of the suggestions for quality assessment however is on the categorical, binary forecasts. Multi-categorical verification is discussed, but not dealt with in depth in this paper (proposed for future work). This focus could have been made more explicit from the beginning. Parts of chapter 5 could also be in the introduction chapter.

I have concerns about the statements in the following paragraph (401):

"Hence, if all that mattered was accuracy—Heierli et al.'s hit rate—then the lessons from this study for forecasting in Switzerland 400 is to set the decision-boundary k to ∞, making it impossible to issue any positive predictions and in doing so increase accuracy. Hence accuracy really isn't a good measure to assess a professional avalanche forecaster's

performance. We hope they agree not merely due to concerns about job security."

It seems that this paragraph is not fully representative of what Heiereli et. al. states in their paper. Firstly, I don't think that they suggest that "all that matters is accuracy". Further, I would also suggest reformulating the statement that the "lessons from this study for forecasting in Switzerland is to set the decision-boundary k to ∞". The wording could be interpreted in a way that suggests that the authors of the Heierli et al. paper were not aware of these problems. This does not seem justified, as they for example are very explicit that the Swiss forecasters are to evaluate the quality ("value of information") of "the NN description" of each forecast day.

Technical corrections:

224: non-occurrence
368, 438: decision boundary
389: i.e.
412: Score
429: neighbour
541: our the