Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2018-74-RC1, 2018

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Spatial consistency and bias in avalanche forecasts – a case study in the European Alps" *by* Frank Techel et al.

K. Müller (Referee)

kmu@nve.no

Received and published: 24 April 2018

The presented study analyses the forecasting goodness of avalanche forecasts from 23 different forecasting centers in the European Alps over a period of four years. The authors use the agreement in danger level between neighboring regions (within and between different forecasting centers) as a measure of forecast consistency and bias. They present a method to explore and quantify spatial consistency of forecast regional avalanche danger levels. Bias between neighboring regions could to some extend be attributed to operational constraints of the involved forecast centers.

The paper gives a good overview of the different practices and concepts for production and communication of avalanche forecasts in the European Alps. The presented statis-

tics give insight into the different approaches and can provide valuable input for future improvements in avalanche forecasting and communication. The dataset presented is extensive and novel and can certainly help to understand and harmonize avalanche forecasting in the European Alps and worldwide.

The text itself is often complicated with long sentences. Simpler and more to the point language throughout the whole paper would be beneficial for the readability and understanding of the paper. Especially for the more technical chapters 3 to 5. Try to avoid repetition. Sometimes terms are defined two or three times throughout the text.

Figures and tables are generally good and informative.

The study is of value to the avalanche community issuing or using regional avalanche forecasts and suited for publication in NHESS after addressing the following general and specific comments.

——————— General comments: ———————

The authors follow Murphy (1993) to assess forecast goodness based on three factors (quality, consistency and value). While they exclude quality since it is nearly impossible to measure, consistency and value are considered. The authors use P_agree as a measure for the consistency of the avalanche forecast. They state that disagreement can be attributed to either climatological or topographical differences or differences in the production of the forecasts between different forecasting centers. I question the value of P_agree as a measure of consistency and miss a discussion on the expected agreement rate or consistency. Aside from political borders, the reason for the delineation of individual forecasting regions is that different avalanche conditions are to be expected. An agreement of close to 100% between two neighboring regions indicates that the boundary between them is superfluous? This point is not addressed. On the other hand, there are only five danger levels. A certain agreement is therefore expected considering that danger levels 2 and 3 are well overrepresented (being issued up 80% of the time) over the course of a forecasting season.

Across political boundaries, avalanche conditions could be expected to be more similar and a disagrrement between danger levels could indicate a substantial difference in assessing avalanche danger or interpreting the avalanche danger scale. The study could be strengthened by filtering regions and considering only those that border to regions of different forecasting centers and exclude those that only border with "internal" forecasting regions. Thus, potential conceptual differences between individual forecasting centers might be easier to identify.

"Value" is presented as being both connected to "quality" and "consistency" in the introduction. The authors should be more precise on if and how they evaluate "value". Section 6.4 presents some general reflections around the value of avalanche forecasts to the users, but an assessment of "value" with regard to the presented statistics is lacking in the methods and conclusions.

The research questions from p3 should be answered in the conclusion. While questions 1 and 2 are addressed answers to questions 3 and 4 should also be given.

Based on the author's analysis, region size seems to be an important parameter for the consistency of a forecast. Region size can be adequately analyzed based on the presented data and should be emphasized in the discussion and conclusions.

——————— Specific comments: ———————

p1 l7: Can we actually expect consistency between neighboring regions wrt danger level. In many cases the situation might actually be different and require different danger levels.

p1 l10: Same as for L7 - could be geographical or meteorological reasons for this.

p3 l5-7: Can you state that more clearly? I think what you mean is that you compare a single categorical value (given for an area and a certain time span) to a complex and dynamic situation (often over a subset of the valid area and time). This will even be more pronounced when comparing regions of rather different size.

p3 l22: a requirement for this would be that forecasters within each center work consistently, at least with respect to other forecasting centers they are compared to. I assume this is an assumption which is difficult to verify.

p3 l24ff: Please be more clear about your use of the terms quality, consistency and especially value. On p3 l19 you state that quality is not measurable. In the abstract and here you state that you focus on consistency which has implications on quality and therefore value. You assume quality to be consistent in your data. On p3 l3 you introduce value as "the benefits or costs incurred by a user as a result of a forecast". Here you state that "implication for the value" are a "result of potential differences in consistency". To me this is somewhat confusing and it is not obvious to me if and how you consider value in your study at all.

p5 l11: difference between forecast center and AWS not clear.

p10 l9: with most of the forecasts during the winter having DL2 or DL3 chances are very high that avalanche danger levels agree between neighboring regions despite differences in size or validity period. Could you present some numbers and discuss this "issue"?

p16 l30: in larger regions the distance to the neighboring region can be larger, which makes it more likely to have different danger ratings due to varying parts of each region influencing the danger level. Please discuss.

p17 l5: the term maximum elevation needs to be introduced and explained earlier; same for the comparison of region sizes. Please explain what you are analyzing and how you calculate rho_elevation and rho_area in the methods section, e.g. 4.2.2.

p19 l19: what is the reason for remove single years? Please state. Later you argue that the chosen four years are a representative excerpt which would imply no need to remove or filter data by individual years.

p21 l1: why not an analysis for moderate avalanche danger?

p23 l32: Is there a difference between forecasting centers? Do some issue the highest while others issue the most representative? If yes, was this considered in the analysis other than for the regions in SWI and VDA?

Sec 5.5: Aggregation of smaller regions to larger forecasting regions will necessarily lead to the same danger rating and it is likely that warning regions within the same larger snow-climate region will more often aggregated together. Therefore it is expected that the (rather small) regions in SWI and VDA have a higher agreement rate than in other parts of the Alps where regions are larger and not aggregated. Please discuss.

p26 l20ff: If you consider your data as sufficiently robust the exercise of removing one of the years does not add value to the study and could be moved to the appendix/supplements.

p27 l11ff: I agree and it would have been interesting to filter the warning regions accordingly and make a separate analysis of regions of neighboring forecasting centers (ideally with an presumably similar snow climate if this information had been available).

p27 l27: It seems like BRI is somewhat special wrt P_v.crit. Have you looked into potential reasons for that? Special climate/topography/size/location or conceptual differences in producing or communicating avalanche forecasts?

p28 l23ff: It is expected that the smaller regions will less often have higher danger levels than larger regions since the chance to have a critical situation increases with size. It would have been interesting to see if and/or how large the differences were if equally large regions from different forecasting centers had been compared. E.g. picking or aggregating a 2000 km2 region from each forecasting center and comparing the frequency of higher danger levels.

p30 l20ff: Please try to answer your research question from p3 in your conclusion, especially questions 3 and 4. Emphasize the impact of the size of a forecast region for

the consistency.

——————— Technical comments: ———————

p2 l20: ...where avalanches may occur...

p2 l21: preferably use trigger/release instead of "initiate"

p2 l22: this sentence does not make sense. What about: "The categorical description of each danger level aims to inform users on the nature of avalanche hazard at hand. However, individual danger levels capture a wide range of differing avalanche conditions (e.g. EAWS, 2005; Lazar et al., 2016; EAWS, 2017a; Statham et al., 2017), and therefore, in isolation, are too basic to be used as a stand-alone decision making tool (e.g. Météo France, 2012) ."

p2 l26: remove "to forecast users"

p2 l29-31: Rephrase this section

p3 l17: remove "crucially"

p3 l17: Could you define what you mean by situation compared to a physical state.

p8 l3: "expected snow and..."

p8 l7: exposed instead of critical terrain?

p9 l3: when conditions change ...

p9 l4: find a better section title

p9 l13: "...takes to internally assess, and externally communicate avalanche danger."

p11 fig4: Change figure text: "Schematic presentation of the spatial arrangement of hypothetical warning regions (bold rectangles) and their role in internal (left column) and externally (right column) communication of the regional danger level, with varying danger levels (D1 and D2).

p14 l15: Make the subdivision of section 4.2.1 clearer. Consider a table instead of a section.

p15 l22: Note, that Dmax...

p15 l30: Consider writing indices i and j in italic throughout the text to increase readability.

p16 l19: What tests? Please be more specific.

p20 fig6: Why not separate in a) map and b) inset?

p25 l7: remove "..., and sometimes unaccounted for, ..."

sec6.3: Use subsections, e.g. "6.3.1 Use of EADS in forecasts"

p28 l5: "...for users..."

p28 l7: "..., improved definitions of the terms and factors found in the EADS and the "matrix", a tool..."

p28 l13: "... but increase consistency, too."

p28 l21: Remove "...the finest spatially delineated units underlying the regional forecasts,..."

p28 l24: Remove "forecast" in "...to more forecasts with higher danger levels,..."

p29 l33: "...the forecast best known..."

p30 l7: Please rephrase the sentence "As discussed...". It is not clear to me what you want to say here.

p31 l10ff. This might be a nice, but very general statement. Maybe adequate to finish a scientific presentation on the topic, but not necessarily to conclude a scientific paper.

p31 l15: www.envidat.ch ?

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-74, 2018.