Response to review by Benjamin Reuter

Frank Techel, and co-authors

We thank the reviewer for the detailed and constructive review. We sincerely appreciate the time and expertise invested in the review, which helped identify points requiring clarification and improvement.

Below, each reviewer comment is reproduced in **gray**, followed by our response in **blue**. Planned revisions to the manuscript are indicated in **red**.

Where helpful, we refer to our companion paper, Part A (Müller et al., 2025), which provides the conceptual and methodological background for the EAWS Matrix. We would like to note that the published version of Part A has undergone substantial revisions compared to the initial pre-print, which was available to the reviewer at the time of review. As a number of the reviewer's comments touch on aspects addressed there, we point out explicitly where these issues are treated in Part A.

2 Major Comments

2.1 Independent data

Building on expert opinion is an obvious way forward to create a decision aid. The EAWS matrix or look-up table is being developed this way. While one may argue that this approach may also lead to better acceptance in the forecasting community, cultural differences between forecasters due to empirical local knowledge in the Alpine countries (UNESCO intangible cultural heritage), errors from data interpretation (during temporal or spatial extrapolation from point data or when treating uncertainty) or discrepancy in concept application (e.g. stability classes for dry- or wet-snow situations) introduce uncertainty along the way (of building the desired decision tool). Some of the limitations are mentioned in Section 6.3, but readers may wonder how the present study deals with or even mitigates the issues. - It is not the objective of this study to mitigate the sources of uncertainty described by the reviewer. Rather, our aim is to use available data on Matrix usage from many European warning services to identify potential inconsistencies in how the Matrix is applied. As we emphasize in Section 6.1 (Interpretation) and Section 6.3 (Limitations), the presented approach can highlight inconsistencies but it does not allow us to determine their underlying causes, which may range from cultural differences to interpretation practices or conceptual discrepancies (L373-381, L404-414). Recommendations supported by the data and our findings are summarized in Section 6.2 (L415-437). These fed directly into Part A (Müller et al., 2025), where they were taken up. For instance, limitations of the Matrix approach and possible ways forward are discussed in detail in Part A, Section 6.3.

The authors state in the Introduction (L49) and in the Discussion (L337) that the assessed parameters are not measurable. This lack of an "independent reference" (L337) is somewhat true for the danger level, but less so for the input parameters of the look-up table. Frequency distributions were derived from avalanche observations, stability tests or snow pits (including tests). A classification for dry-snow stability classes is available in Schweizer et al. (2021). Hence, examples of typical avalanche situations and their respective stability/frequency classes – and possibly, independent data to verify the danger level, are available. Such data will not do the job of choosing the final danger level (which remains the goal of the look-up table), but uncertainty in data interpretation, and difficulty in concept application are mitigated; possibly making for useful complement data. These data should not go unmentioned, and even be considered for improving the descriptions of the frequency classes and stability classes. - We thank the reviewer for highlighting the distinction between (i) the measurability of the danger level and (ii) the availability of observational data relevant to the Matrix input parameters. To clarify our position, we structure our response around three points:

- The stability classes presented in Schweizer et al. (2021), originating from Schweizer and Wiesinger (2001), and related work (e.g., Techel et al. (2020b)), are derived from the interpretation of stability tests and snowpack stratigraphy. Such data is valuable for defining frequency classes and was therefore used to define the four stability classes, when developing the Matrix in 2022 (see Figure 2). Based on large data sets of stability tests, Techel et al. (2020a) derived data-driven frequency class thresholds (EAWS, 2025, p.13), which were also provided as examples in the document describing the determination of a regional avalanche danger level using the Matrix (EAWS, 2025). However, in a forecast setting, stability must be anticipated from other data. To link the rather abstract stability classes to typical field observations, Figure 2 provides a range of examples. We believe that forecasters apply the range of examples given there.

- We agree that observational datasets as explored in the studies by Schweizer et al. (2021, 2020b); Techel et al. (2020a, 2022)) are valuable. However, even if such data were abundant and available for every day and every warning region - which they are not - assigning frequency classes such as *a few*, *some*, or *many* would still require an element of expert judgment. For this reason, the absence of a fully independent reference for the Matrix inputs persists, especially in a forecast setting where no field data may be available on the day of issuing the forecast.

- We agree that data-driven approaches can help refine class descriptions. Looking ahead, data-driven approaches may become increasingly useful. In particular, stability and frequency distributions derived from distributed snow-cover simulations - as demonstrated in Herla et al. (2024b, a) - offer promising avenues for providing spatially consistent, independent estimates of Matrix input classes.

Please note that existing observational and test-based datasets have already been integrated to contextualize the stability classes. They may also be used to further refine frequency class definitions. - We will add a remark regarding the potential use of distributed snow-cover simulations for providing the necessary real-time data for stability and frequency estimation in a forecast setting.

The authors explain their choice of methodology, which is appreciable. Nevertheless, the article can benefit from clearer arguments and some adjustments regarding the methodology. Hence, the authors are encouraged to include some data from research publications describing the triplet stability/frequency/size and the (forecast/verified) avalanche danger level.

- Data from avalanche observations (Schweizer et al., 2020b), stability tests (Techel et al., 2020a), and possibly other data such as (stratified) numerous field measurements, may help to corroborate / discard danger level choices in some fields.
- Well-documented situations of the past can illustrate this article. Presenting these situations (synthetically or in supplementary material) will improve the readers' understanding of forecasting challenges.

Such situations should cover a wide range of classic cases (dry, wet, glide snow; skier-triggered and natural). They should provide the key for a user of the look-up table to what a classic "poor–some–size 2 = moderate" looks like.

We agree that well-documented situations and a comparison with additional data sources can be valuable for illustrating Matrix usage and related danger-level choices. However, incorporating such case material in detail is beyond the scope of this article, which focuses on detecting deviations and inconsistencies in Matrix usage across European warning services based on the available data. We like to mention that an example of the kind requested is provided in Part A, Section 6.2 in Müller et al. (2025), where an example of applying the Matrix is shown. Part A (Appendix A: Figures A1–A3) links typical field observations to dry-snow, wet-snow, and glide-snow stability classes. These examples serve the illustrative purpose suggested by the reviewer and offer users concrete reference cases spanning different avalanche types.

2.2 Two levels in one field

Readers may wonder why the final table still shows up to two levels per field with one of them not corresponding to the color of the field. Wasn't the goal of the desired decision aid to promote consistency in the choice of the danger rating? Don't optional choices invite for deviation from the concept? Could the authors provide a final version of the look-up table highlighting the suggested changes? (the avalanches.org webpage seems to show an updated version).

When undertaking this study, it was hoped that clarity would emerge regarding the cells with two danger ratings. However, as discussed in Section 6.1 (L335–343), there are limits to how much the Matrix can be refined now that it is in operational use. Based on the data, we provide several recommendations (L415–424), which fed directly into Part A, where the Matrix is revised. However, for many cells, we had no evidence to propose changes to the indicated danger levels. The affected cells and the resulting updated version are shown there (Figure 3b, Part A) and on https://www.avalanches.org/standards/eaws-matrix/. The continuing presence of up to two levels in some cells reflects uncertainty in the survey outcome and this study's findings, rather than an intention to promote optionality.

Moreover, why are secondary danger levels provided in white/no-shaded fields that are considered rare/implausible situations? If data are sparse, shouldn't those fields simply not show a rating or only one rating in parentheses?

As outlined in Section 5, Part A, the danger levels shown in the white cells had limited support in the survey. Due to a lack of data, these cells could also not be validated using the data in this study. We therefore cannot confirm or reject the indicative ratings obtained when developing the initial version of the Matrix (survey described in Section 4 in Part A). Instead, and this is the approach taken in Part A, additional cells are coloured white precisely to highlight the uncertainty attached to the proposed danger level(s) in those Matrix cells (see Figures 3a and 3b, Part A).

Discrete levels are inherent to classifications. As the authors state, the benefit of the look-up table lies in improving consistency by promoting concepts in the forecasting community. This is to a large part due to the classification's simplicity and not due to the ultimate level of detail the classification allows for. Two supporting arguments:

1. The danger level sets the stage for risk-management strategies but alone will never make up the decision. Users need transparent reasoning rather than a single number "that does it all".

2. The possible resolution of factor estimates varies across forecasting services and regions. A small number of well-defined classes seems to be key.

If sub-classes are introduced or fields further divided (Fig. 8), the classification concept is being eroded and the ultimate goal may move out of sight. The authors are encouraged to reflect on the required level of detail. Suggestions to refine the matrix (L434) should be reconsidered with possible disadvantages mentioned. - There are two aspects to this: simplicity is one, which calls for few, well-defined classes. This is supported by the fact that humans can generally estimate only a small number of categories reliably (e.g., Miller, 1956; Kahneman et al., 2021), and it aligns with the conceptual clarity and reproducibility of the Matrix approach. However, forcing forecasters to commit early to a single discrete class is analogous to rounding intermediate results, which can discard relevant information and create discontinuities in the outcome.[1] At the same time, humans are comparatively good at assigning relative rankings, i.e. using sub-classes within absolute classes (e.g., Kahneman et al., 2021). Therefore, combining absolute and relative judgments (as shown in Figure 2b) is a promising path that should be explored. - We'll expand L373-382 with the explanation given before.
To further elaborate on this: Data from the 2023/2024 forecasting season (Techel et al., 2024) showed that forecasters in Switzerland agreed on the exact same factor class only 60–65% of the time, but in more than 90% of cases the discrepancy was less than one full class. More importantly, they were undecided between two adjacent classes 20–25% of the time. Incorporating this structured uncertainty by considering two neighbouring cells in the Matrix when between two classes, avoids premature "rounding" (or forcing to select one class) while keeping the final classification simple. This approach, described in Section 6.1, Part A, preserves the benefits of the EAWS Matrix while keeping unavoidable uncertainty explicit until the final step, where choosing a single danger level is necessary for communication. Whether splitting Matrix cells, as we show in Figure 8, provides a way forward remains open for discussion. For instance, when revising the Matrix in Part A, this approach was not taken up.

2.3 Definition of stability and frequency classes used in the study

[1] e.g., Guidance for laboratory analysis: «Do not round intermediate calculations; rounding intermediate values can cause rounding errors in the final results and should only take place after the final expanded uncertainty has been determined» https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.6969-2019.pdf

Table 2. Snowpack stability classes referring to the point scale and the type of triggering typically associated with these classes. For further details, including typical observations related to each class, see Sect. A1 or EAWS (2025). Values in parentheses indicate that the trigger or evidence is not typical but may occur.

Snowpack stability	Description	Sensitivity (CMAH)	Natural avalanches	Human triggers	Explosive/ Cornice fall	Other indicators of instability
Very poor	Very easy to trigger	Touchy	yes	yes	yes	Shooting cracks, whumpf sounds
Poor	Easy to trigger	Reactive	no	yes	yes	(Shooting cracks, whumpf sounds)
Fair	Difficult to trigger	Stubborn	no	(yes)	yes	
Good	Stable conditions	Unreactive	no	no	no	

Figure 1. Observations related to stability classes used in the EAWS Matrix. Screenshot of table from (Müller et al., 2025).

The presented results for wet and glide-snow situations show a large spread, in particular with respect to stability classes (see use of "very poor"). As current definitions, especially for wet-snow situations, lack tangible elements, it is no surprise that results are somewhat inconsistent. In Europe, traditions to deal with avalanche hazard vary between countries/cultures. Hence, unambiguous definitions are paramount and will condition any multi-cultural evaluation in Europe. Table A1, describing point-scale snow-stability classes, provides little conclusive information (definition of "difficult", "easy", "very easy"?), is misleading ("natural" cannot be a special case of "very easy to trigger") and lacks information to assess stability classes in wet and glide snow situations. - Additional practical observations and a direct link to the terminology used in the CMAH (Statham et al., 2018) have been added to the corresponding table in Part A. As illustrated in Figure 2 (screenshot taken from revised Table in Part A), the revised version clarifies how *very poor* stability relates to natural release without excluding human triggering. - We will replace the table in the Appendix with the updated version from Müller et al. (2025).

Along the lines of the definitions of snow instability on the EAWS webpage, Table 1 should:

– Separate dry, wet and glide-snow problems,

– Explicitly refer to avalanche types (including point releases),

– Clearly distinguish natural release and artificial triggering,

– Provide tangible snowpack descriptions (see Schweizer et al., 2021).

Regarding the request to add specific examples to dry-snow, wet-snow, and glide-snow conditions, such examples were provided for dry-, wet-, and glide-snow avalanches together with the introduction of the Matrix in 2022. These concrete examples complement the stability description. These examples are also shown in Appendix A (Part A, Figures A1–A3; see screenshot of these figures in Figure 2). In addition, the revised Table 1, Part A already provides more practitioner-oriented examples for each of the stability classes.

We acknowledge that the descriptors for the frequency classes are broad and therefore a potential source of uncertainty, as also recognized by the EAWS working group who originally defined them. While this clearly warrants further discussion and future refinement, ideally supported by more data-driven approaches, we do not revise these definitions within this paper.
Table A3 presents the officially accepted EAWS avalanche size classification. These definitions are well established opera-

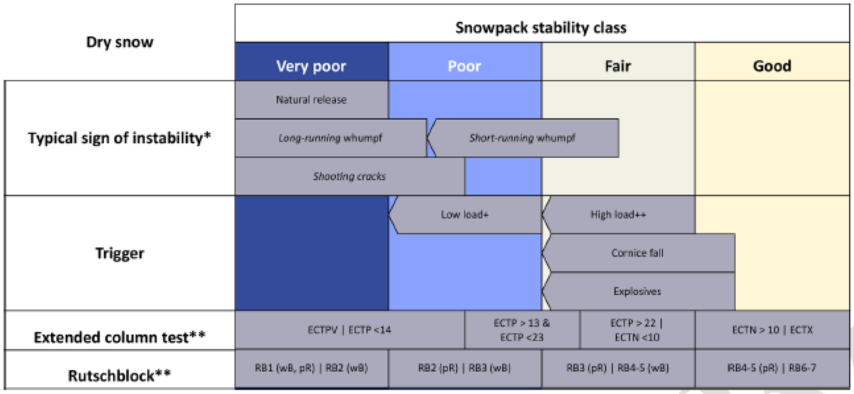

Figure A1. Dry snow conditions: Common evidence or indications for snowpack stability classes focusing on dry-snow slab avalanches. Arrows indicate that existence towards lower stability classes is imperative. Natural avalanches are a clear indication for the class very poor, while a low and a high additional load are considered approximately equivalent to poor and fair stability. Observations and stability test results should be regarded as indicative only. Abbreviations: Extended Column Test (ECT), Rutschblock (RB), whole block (wB), partial release (pR). Schweizer et al. (2020a), Techel et al. (2020), +single skier not falling, ski-cut, ++single skier falling, group of skiers, person on foot.

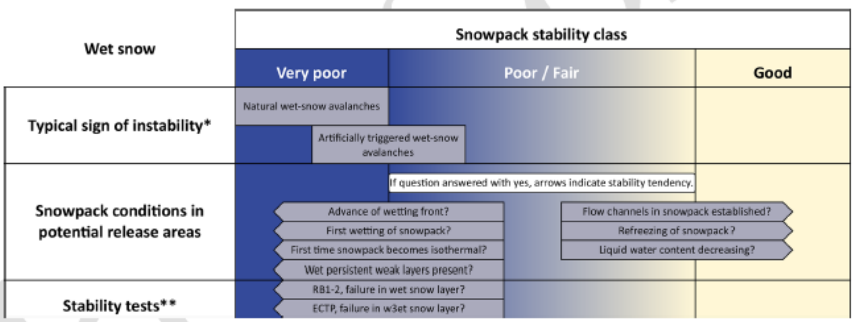

Figure A2. Wet snow conditions: Common evidence or indications related to wet-snow stability. If no liquid water is present in the snowpack, wet-snow avalanches are not possible.

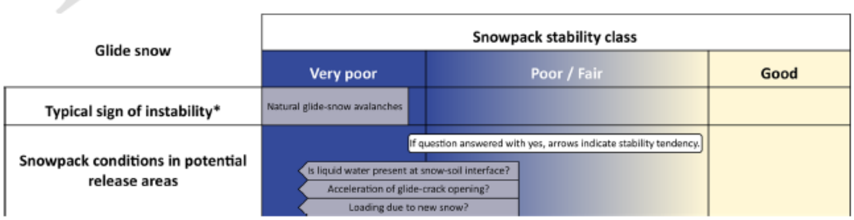

Figure A3. Glide snow conditions: Common evidence or indications related to glide-snow stability. Glide-snow avalanches are not possible if there is no liquid water present at the snow-spoil interface.

Figure 2. Observations related to stability classes used in the EAWS Matrix. Screenshot of figure shown in (Müller et al., 2025).

tionally, and we therefore retain them unchanged here. For expert use, EAWS already provides additional, forecasting-relevant descriptions on https://www.avalanches.org/standards/avalanche-size/.

2.4 Compliance with the suggested danger level

A strong point is that the authors shed light on how forecasting services comply with the matrix. Monitoring matrix compliance seems an interesting path for forecasting services to identify diverging situations. If they manage to identify and train, they can increase forecasting quality and consistency. This could be included in the conclusions (currently touched on at L430). - We'll briefly take this up in the conclusions.

2.5 Avalanche size and methodological clarity

Avalanche size has been identified as a relevant but secondary element in danger-level assessment. In the current matrix it seems to play a more important role. - We are aware of this research (Schweizer et al., 2020b; Techel et al., 2020a). Comparing the data-driven matrix shown in Techel et al. (2020a) with the EAWS Matrix shown in a slightly different layout (Figure 6b in Part A) shows many similarities.

Analysis.

Several specific clarifications are requested:

– L176: clarify computation of disagreement between forecast and matrix-derived levels. - We'll expand this description to clarify the individual calculations.

– L198: justify assumption that Scottish data represent mostly dry-snow conditions. - As we have no data on this linked to Matrix use, we'll refer to the proportions of dry-snow and wet- or glide-snow problems provided by the director of the Scottish Avalanche Information Service. In the six forecast regions, and excluding the cornice problem, which is frequently given in Scotland, dry snow problems were used between 67% and 90% of the time during the 2024/2025 season (data provided by Mark Diggins, head of the Scottish Avalanche Information Service).

– L222–226: reconsider interpretation of "very poor" stability and "natural activity rare". - In addition to natural avalanches, we will mention "very easy to trigger".

– L243–259: re-word statements where "single danger level" is used; check "size 1 vs size 2". - We'll check whether rewording these two sentences can improve clarity.

Interpretation.

Comparisons with benchmark situations are appreciated, but some statements should be double-checked (e.g. L225, L255, L268). Clarify meaning of "neighboring cells"; relate observed patterns to field studies (Techel et al., 2020a; Schweizer et al., 2020a). - The Results section reports solely on factor estimates in a *forecast* setting. We don't analyse field observations in this study; therefore, we do not refer to such studies, except when linking very poor stability and typical release mechanisms for wet-snow and glide-snow avalanches. (L304-306, L409-412). - We will refer to the revised Table providing the stability definitions and examples (see screenshot of this revised Table in Figure 1). We will either clarify the meaning of "neighbouring cells" or replace this with: «cells that differ in one of the factor inputs».

Further notes:

- Glide- and wet-snow avalanches are natural releases; clarify wording. - We will review the manuscript to ensure that glide- and wet-snow avalanches are natural releases, in most cases.
- L351–353: improve clarity of described discrepancies. - We will describe the discrepancies more clearly by stating the specific proportions and referring to the specific figure where these discrepancies can be observed.
- L375–394: avoid circular statements when defining compliance; check interpretation. - We will review and streamline these lines to reduce repetition.
- L408–458: ensure consistent terminology for avalanche types vs problems; reconsider statements on "effective" and "supporting consistent danger-level assessment". - We'll check for consistent terminology throughout.
- L456–458: reflect whether finer granularity actually benefits risk management. - As outlined earlier, there are arguments both for and against using finer granularity in factor assessments. Whether such refinement ultimately improves consistency in danger-level assessment remains an open question. We also note that the Matrix is designed to support *danger* assessment, not *risk* analysis.

3 Minor Comments

- Use one consistent term for the danger level chosen by forecasters ("issued", "forecast", or "assigned"). - We'll change throughout to ...
- L47: clarify meaning of "quality" vs "consistency"; perhaps use "accuracy". - Quality is the term used by Murphy (1993), who we cite for this concept. We'll consider adjusting to accuracy, which is what we mean and describe by quality.
- L56 + L59: choose between "practical" or "operational" implementation. - We'll change to "operational" implementation.

– L75: choose between “factors” or “components”. - We’ll change to input "factors" throughout the manuscript

– L173 + L176: choose either “matrix-derived” or “matrix-suggested”. - We’ll change to "matrix-suggested" throughout the manuscript

– L174: clarify colon usage (“the disagreement was computed as . . . ”). - We’ll split into two sentences.

– L218: specify “dry- or wet-snow conditions”. - For clarity, we’ll repeat the definition from Table 2 and L164-165.

– L235: reword to “. . . stability was as often described by ‘very poor’ as by ‘poor’ . . . ”. - We believe that our wording is correct. But we leave this decision to copy-editor.

– Figure 3 and 5: improve captions and axis labels; highlight matrix-suggested danger levels. - We’ll highlight the Matrix-suggested danger levels. We’ll increase font size of the figure axis.

– L257: clarify whether “avalanche problems” or “dry-snow situations” are meant. - As outlined before dry-snow avalanche problems refers to the group of avalanche problems relating to dry-snow conditions (L164-165). We’ll check throughout manuscript that we use same terminology as described there.

– L261 + L266 + L328 ff: improve transitions and clarity in discussion. - We’ll review phrasing of the transition between these paragraphs.

– L417–424: adjust adjectives and phrasing for clarity (“often”, “frequently”, “under-supported”, etc.). - We’ll review whether other adjectives may be more suitable.

References

EAWS: Determination of the avalanche danger level in regional avalanche forecasting, https://www.avalanches.org/wp-content/uploads/2022/12/EAWS_matrix_definitions_EN.pdf, EAWS working group Matrix and Scale (working group members: Müller, K.; Bellido, G.; Bertrando, L.; Dufour, A.; Feistl, T.; Mitterer, C.; Palmgren, P.; Roux, N.; Sofia, S.; Techel, F.). last access: 10 Jul 2025, 2025.

Herla, F., Binder, M., Lanzanasto, N., Perfler, M., Widforss, A., Reisecker, M., Müller, K., and Mitterer, C.: Synthesizing regional snowpack stability and avalanche problems in the operational AWSOME framework, in: Proceedings, International Snow Science Workshop Tromsø, Norway, pp. 272–277, 2024a.

Herla, F., Haegeli, P., Horton, S., and Mair, P.: A large-scale validation of snowpack simulations in support of avalanche forecasting focusing on critical layers, Nat. Hazards Earth Syst. Sci., 24, 2727—2756, https://doi.org/10.5194/nhess-24-2727-2024, 2024b.

Kahneman, D., Sibony, O., and Sunstein, C.: Noise: A flaw in human judgment, William Collins, London, U.K., 2021.

Miller, G.: The magical number seven, plus or minus two: Some limits on our capacity for processing information, Psychological Review, 63, 81–97, https://doi.org/10.1037/h0043158, 1956.

Müller, K., Techel, F., and Mitterer, C.: The EAWS matrix, a decision support tool to determine the regional avalanche danger level (Part A): conceptual development, Natural Hazards and Earth System Sciences, 25, 4503–4525, https://doi.org/10.5194/nhess-25-4503-2025, 2025.

Murphy, A. H.: What is a good forecast? An essay on the nature of goodness in weather forecasting, Weather and Forecasting, 8, 281–293, https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2, 1993.

Schweizer, J. and Wiesinger, T.: Snow profile interpretation for stability evaluation, Cold Reg. Sci. Technol., 33, 179–188, https://doi.org/10.1016/S0165-232X(01)00036-2, 2001.

Schweizer, J., Mitterer, C., Reuter, B., and Techel, F.: Avalanche danger level characteristics from field observations of snow instability, The Cryosphere, 15, 3293–3315, https://doi.org/10.5194/tc-15-3293-2021, 2020a.

Schweizer, J., Mitterer, C., Techel, F., Stoffel, A., and Reuter, B.: On the relation between avalanche occurrence and avalanche danger level, The Cryosphere, https://doi.org/10.5194/tc-2019-218, 2020b.

Schweizer, J., Mitterer, C., Reuter, B., and Techel, F.: Avalanche danger level characteristics from field observations of snow instability, The Cryosphere, 15, 3293–3315, https://doi.org/10.5194/tc-15-3293-2021, 2021.

Statham, G., Haegeli, P., Greene, E., Birkeland, K., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., and Kelly, J.: A conceptual model of avalanche hazard, Natural Hazards, 90, 663 – 691, https://doi.org/10.1007/s11069-017-3070-5, 2018.

Techel, F., Müller, K., and Schweizer, J.: On the importance of snowpack stability, the frequency distribution of snowpack stability and avalanche size in assessing the avalanche danger level, The Cryosphere, 14, 3503 – 3521, https://doi.org/10.5194/tc-2020-42, 2020a.

Techel, F., Winkler, K., Walcher, M., van Herwijnen, A., and Schweizer, J.: On snow stability interpretation of extended column test results, Natural Hazards Earth System Sciences, 20, 1941–1953, https://doi.org/10.5194/nhess-2020-50, 2020b.

Techel, F., Mayer, S., Pérez-Guillén, C., Schmudlach, G., and Winkler, K.: On the correlation between a sub-level qualifier refining the danger level with observations and models relating to the contributing factors of avalanche danger, Natural Hazards and Earth System Sciences, 22, 1911–1930, https://doi.org/10.5194/nhess-22-1911-2022, 2022.

Techel, F., Lucas, C., Pielmeier, C., Müller, K., and Morreau, M.: Unreliability in expert estimates of factors determining avalanche danger and impact on danger level estimation with the Matrix, in: Proceedings International Snow Science Workshop, Tromsø, Norway, 23-29 Sep 2024, pp. 264 – 271, https://arc.lib.montana.edu/snow-science/item.php?id=3144, 2024.