# ~~An updated~~ The EAWS matrix, a look-up table to determine the regional avalanche danger level (Part A): ~~derivation, usage, and consistency~~Conceptual development

Karsten Müller[1,*], Frank Techel[2,*], and Christoph Mitterer[3]

[1]Norwegian Water Resources and Energy Directorate, Oslo, Norway
[2]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[3]Avalanche Warnings Service Tirol, Innsbruck, Austria
[*]Joint first authors. These authors contributed equally to this work.

**Correspondence:** Karsten Müller (kmu@nve.no)

**Abstract.** Avalanche forecasting plays a crucial role in mitigating risks associated with snow avalanches in mountainous regions. Standards for regional avalanche forecasting were initially developed at national levels. Therefore, the introduction of the European Avalanche Danger Scale (EADS) in 1993, still in use today, represented a milestone in harmonizing the assessment and communication of avalanche danger. ~~However~~Since then, standards, concepts and definitions have evolved~~since then. Here, we reflect on the current standards and definitions used in regional avalanche forecasting, with a focus on~~ . In this study, we present the updated European Avalanche Warning Services (EAWS) Matrix ~~, a~~ – a consensus-based look-up table ~~intended to promote consistency among avalanche forecasters when assigning a danger level . The EAWS Matrix links the factors determining avalanche danger -~~ designed to support consistent and transparent danger level assessments by linking snowpack stability, the frequency of snowpack stability, and avalanche size ~~- to avalanche~~ to the five danger levels. ~~Here, we describe the methodology to obtain a consensus-based EAWS Matrix. By analyzing the operational use of the EAWS Matrix following its introduction, we gain insights into its implementation across European avalanche warning services and obtain an understanding on challenges and short-comings related to its operational use. As a reliable estimation of the factors determining avalanche danger is a prerequisite for consistency in assigning the danger levels using the Matrix, we also explored the consistency of estimating the factors by comparing forecasts prepared by individual forecasters . Noting considerable variations in the assignment of factor classes, we provide recommendations for practice and ways forward, such as refining the definitions of the classes describing the factors, implementing training sessions,~~ We describe the collaborative revision process involving avalanche forecasters from across Europe, including expert surveys and integrating the findings following the operational test phase spanning three winters. The updated Matrix reflects current best practices in regional avalanche forecasting. Its design aligns with the operational forecasting workflow and explicitly addresses inherent uncertainties. We highlight key findings, such as persistent ambiguities, challenges in defining frequency classes, and ~~exploring different matrix layouts. Additionally, the discrepancies between the EADS and current standards and definitions underscore the need for an updated avalanche danger scale. In conclusion, the updated EAWS Matrix represents a next step towards harmonizing avalanche forecasting practices in Europe even though the analysis revealed areas for improvement. Clearly, further efforts are required to~~

~~develop and implement regional avalanche forecasting standards~~ strategies to improve the reliability ~~, credibility and timeliness of avalanche forecasts, regardless of the forecaster or warning service behind the product~~of input factors. Finally, we position the Matrix within the broader framework of the Conceptual Model of Avalanche Hazard (CMAH) and argue that the EADS should be updated accordingly.

## 1 Introduction

Snow avalanches represent a natural hazard in snow-covered, mountainous regions. Avalanches may lead to injury or loss of life, and can cause damage or destroy property and infrastructure. For instance, in Europe in the 50 years between 1974 and 2023, more than 5900 (annual mean: 118) people have died in avalanches (EAWS, 2023a). To reduce adverse effects resulting from avalanches, avalanche warning services disseminate regional avalanche forecasts to inform and warn the general public as well as responsible decision-makers .e.g., in local authorities, on current and expected avalanche conditions.

The assessment of current and future avalanche conditions involves the analysis of a wide variety of heterogeneous data, including field observations, measurements, models, and weather forecasts. Although the interpretation of snow and weather parameters follows a deterministic cause-and-effect approach, actual forecasting decisions are reached using inductive logic (LaChapelle, 1980). Thus, the quality of avalanche forecasts is influenced by a combination of factors, including the forecaster's experience and reliability (Stewart and Lusk, 1994; McClung, 2002), as well as the dynamic nature of the snowpack, which varies spatially and temporally (Schweizer et al., 2008). Due to the inherent uncertainty in predicting the exact timing and location of avalanche events and due to a lack of relevant data, the assessment of avalanche danger maintains a qualitative character. Unlike weather forecasting, which often ~~involve~~ involves precise numerical predictions for variables like air temperature or precipitation, in avalanche forecasts the complex and multifaceted nature of avalanche conditions is assessed and communicated using symbolic representations, encompassing danger levels, classes, terms, and text (Hutter et al., 2021). In regional avalanche forecasting, the focus of this study, the severity of expected avalanche conditions is summarized using the concept of danger levels. Despite advances in model-driven predictions of avalanche danger levels (e.g., Giraud, 1992; Pérez-Guillén et al., 2022), assessing avalanche danger levels has so far remained primarily a subjective decision-making process. While complete consensus between individual forecasters is unattainable, random variations inherent to human judgment should be minimized. Consistency between a forecaster's best judgment and the forecasts they produce is as important as consistency between forecasters, as ~~these~~ both directly impact the quality of avalanche forecasts (Murphy, 1993; Stewart, 2001). ~~High-quality forecasts, however~~High values for consistency lead to high-quality forecasts, which in turn, enhance the potential value of the forecast to decision-makers using them (Murphy, 1993)~~, and~~ . Consequently, high-quality forecasts can therefore increase safety when recreating in terrain exposed to avalanche hazard and contribute to reducing avalanche-related damage and loss.

The information provided in avalanche forecasts is structured following an information pyramid, with the most relevant information, a danger level ($D$), at the top (EAWS, 2023c). The importance of $D$ for decision-making in avalanche terrain has been shown in numerous studies, including, for instance, during the trip planning stage (e.g., Morgan et al., 2023), impacting the decision whether to ski a slope or not (e.g., Furman et al., 2010), or the correlation between the forecast danger level and avalanche risk during back-country skiing (e.g., Techel et al., 2015; Winkler et al., 2021).

Given the importance of ~~the~~ avalanche danger levels to support ~~decision making for users of avalanche forecasts, ensuring a~~ decision-making for avalanche forecast users, ensuring consistent assignment of these levels is paramount. However, several studies have shown considerable variation in the use of danger levels. These variations are greater between forecasters from different or neighboring warning services (Lazar et al., 2016; Techel et al., 2018) than within a single warning service ~~(Techel et al., 2018, 2024a)~~(Techel et al., 2018). Additionally, inconsistencies persist when describing dry- and wet-snow avalanche conditions in terms of the likelihood and size of natural avalanches (Clark, 2019; Hutter et al., 2021).

~~With the aim to increase consistency between~~ Several standards and tools have been developed within EAWS and related organizations to structure the process of public regional avalanche forecasting. However, some of these frameworks date back more than 30 years, during which both knowledge and terminology have significantly evolved. To enhance consistency in the assessment of regional avalanche danger levels across forecasters and warning services~~when deciding on an avalanche danger level for a region, a working group of~~ , the European Avalanche Warning Services (EAWS) ~~revised the definitions of~~ launched a coordinated initiative to revise the core components guiding such assessments. As part of this effort, a dedicated EAWS working group redefined the factors determining the ~~regional~~ avalanche danger level ~~and developed a common workflow for assessing~~ ($D$~~(EAWS, 2022c; Müller et al., 2023). Moreover, the look-up table assisting forecasters in assigning a danger level, referred to as EAWS Matrix , was revised , to be in line with the~~ ) and established a shared operational workflow to support its assessment (EAWS, 2022c; Müller et al., 2023). In parallel, the EAWS Matrix – a central decision aid for assigning danger levels in many European warning services for many years – was revised to align with the updated definitions and terminology~~used to describe avalanche danger by the working group. Definitions, workflowand matrix were accepted~~ . This process unfolded in three main steps:

1. **Conceptual and methodological development:** Definition of key assessment factors and initial revision of the Matrix and workflow. This package of definitions, workflow, and Matrix was formally adopted by the EAWS General Assembly in 2022 (EAWS, 2022c).

2. **Operational testing:** Evaluation of the revised Matrix, workflow, and assessment factors under real operational forecasting conditions, presented in two separate analyses (Techel et al., 2024a, 2025).

3. **Refinement and integration:** Targeted adjustments based on findings from the operational testing phase, resulting in a consolidated Matrix. The proposed changes were formally adopted by the EAWS General Assembly in 2025.

~~Our objective is to reflect on current standards and definitions used in regional avalanche forecasting , with a particular focus on the updated EAWS Matrix. We approach this in three steps: First, we summarize the development and definitions of the major standards used in avalanche forecasting~~ This iterative exchange between methodological development and practical implementation was instrumental in enhancing the clarity, usability, and consistency of the Matrix across forecasters and warning services. It also emphasizes the importance of structured evaluation in the development of operational tools.

The aim of this contribution is to document the iterative revision of the EAWS Matrix and its accompanying workflow, to present the final revised Matrix, and to provide a critical reflection that paves the way for further refinements and contributes to the broader discussion on the future evolution of the European avalanche danger scale. The revision process has the following components outlined in two papers, this one and Techel et al. (2025):

– **Background (Sec. 2):** We provide a structured overview of existing standards for assessing regional avalanche danger in Europe and North America, ~~highlighting their benefits and shortcomings (Sect. 2). Next, we describe~~ with particular focus on the avalanche danger scale and the evolution of various look-up tables supporting danger level assignment. This section outlines the rationale for revising the Matrix.

– **Revision process – Step 1: Definitions and forecaster survey (Sec. 3):** We describe how the EAWS Matrix was revised, including the methodology and ~~outcomes of the revised EAWS Matrix (Sect. 4). Finally, we analyze its use during the first winter following its introduction (Sect. ??) and assess forecaster consistency in~~ key outcomes, and highlight areas of uncertainty and open questions.

– **Matrix and recommended workflow (Sec. 4):** We present the revised Matrix and the accompanying workflow, as adopted by EAWS, and explain how it was intended to be tested by forecasters.

– **Companion analyses (external):** To support potential refinements to the Matrix and workflow, their operational use was evaluated in two parallel studies: one examined the reliability of estimating the input factors ~~required to the EAWS Matrix (Sect. ??). This analysis provides insights into potential challenges and shortcomings of the EAWS Matrix, offering opportunities to refine regional avalanche forecasting standards and fostering the discussion towards an updated European avalanche danger scale.~~ to the Matrix (Techel et al., 2024a), and the other focused on how the Matrix was used in day-to-day forecasting operations (Techel et al., 2025). Together, these studies provided empirical feedback based on real-world application of the Matrix.

– **Revision process – Step 2: Operational testing (Sec. 5):** We summarize the key lessons learned from the two companion analyses and describe how they informed further refinements to the Matrix.

– **Discussion ( 6):** We critically reflect on the consolidated Matrix, discuss known limitations, and explore how the Matrix could be integrated into the existing conceptual framework for regional avalanche forecasting in North America.

## 2 Background

### 2.1 The European Avalanche Danger Scale

Avalanche bulletins have been published since the winter 1945/1946 in Switzerland. Although neither standardized nor defined nor used in a consistent manner, avalanche danger was already described in winter 1951/1952 in Switzerland as being *low*, *moderate*, *considerable*, *high* and *very high*, sometimes in connection with modifiers like *general* and *local* (e.g., SLF, 1953, p.68 ff). A first description of the danger levels used in Switzerland was published in 1985 (Föhn, 1985), allowing consistent use by forecasters and transparent communication to users. Similarly, in France, eight «typical» avalanche situations were used to assess and communicate avalanche conditions (Giraud et al., 1987). These were later on also used in Italy. Despite the formation of a European Avalanche Warning Services (EAWS) working group in 1983, which aimed to promote cooperation across national borders, the Alpine countries France, Italy, Switzerland, Germany and Austria continued to use their own danger scales with a varying number of six to eight danger levels (Mitterer and Mitterer, 2018). In 1993, the EAWS introduced the five-level European Avalanche Danger Scale (EADS, SLF, 1993; Meister, 1995), which was largely based on the wording and definitions used in Switzerland (Föhn, 1985). This adoption of a standardized danger scale marked a pivotal moment for international avalanche warning services, simplifying procedures for all parties involved, and facilitating communication of avalanche danger particularly for forecast users when traveling to different countries (Meister, 1995). Except for minor changes in 1994, the EADS has been unchanged as of today, not only providing a common way of expressing the avalanche danger level across institutions and borders, but impacting «the forecasting process itself, as all forecasters are working to an agreed, common, and at least nominally binding definition of avalanche ~~hazard~~danger.» (Techel et al., 2018, p. 2698).

The EADS uses two columns to describe each danger level (Table 1). The first column describes snowpack stability and includes a qualitative indication of the frequency of the respective locations. The second column describes the likelihood of triggering an avalanche by indicating the typical avalanche size and their distributions, the likelihood of natural avalanches occurring or the typical load required to trigger an avalanche. Frequency of avalanches and potential triggering locations or the likelihood of avalanche release are again described qualitatively.

The EADS has several shortcomings as a tool to summarize avalanche conditions in a region:

– The terminology in the EADS is vague, leaving ample room for interpretation. For instance, clear definitions for classes describing snowpack stability and the frequency of triggering locations are lacking.

– Qualitative terms expressing probability or uncertainty are not defined, which according to Morgan (2017) is inadequate as the same term can have different meaning to different people, but also to the same person in a different context. Not surprisingly, even among avalanche professionals large differences in numeric estimates of probability were observed (Thumlert et al., 2020).

– The load necessary to trigger an avalanche is correlated to snowpack stability (Schweizer and Camponovo, 2001). Thus, both columns in the EADS contain similar and redundant information on *snowpack stability* and triggering.

– The short descriptions of each danger level do not cover the range of all possible combinations. For instance, snowpack stability decreases from *moderately well bonded* to *moderately to poorly bonded* from ~~2-Moderate to 3-Considerable~~ 2 (moderate) to 3 (considerable) while its frequency increases from *some* to *many* steep slopes. But the EADS does not provide guidance when the situation is best described by a snowpack that is *moderately to poorly bonded* on *some* steep slopes.

**Table 1.** European avalanche danger scale (EAWS, 2023b).

| Danger level | Snowpack stability | Likelihood of triggering |
|---|---|---|
| ~~1-Low~~ 1 (low) | The snowpack is well bonded and stable in general. | Triggering is generally possible only from high additional loads** in isolat areas of very steep, extreme terrain**. Only small and medium-sized natu avalanches are possible. |
| ~~2-Moderate~~ 2 (moderate) | The snowpack is only moderately well bonded on some steep slopes*; otherwise well bonded in general. | Triggering is possible primarily from high additional loads**, particula on the indicated steep slopes*. Very large natural avalanches are unlikely |
| ~~3-Considerable~~ 3 (considerable) | The snowpack is moderately to poorly bonded on many steep slopes*. | Triggering is possible even from low additional loads** particularly on t indicated steep slopes*. In certain situations some large, in isolated cas very large natural avalanches are possible. |
| ~~4-High~~ 4 (high) | The snowpack is poorly bonded on most steep slopes*. | Triggering is likely even by low additional loads** on many steep slope In some cases, numerous large and often very large natural avalanches c be expected. |
| ~~5-Very High~~ 5 (very high) | The snowpack is poorly bonded and largely unstable in general. | Numerous very large and often extremely large natural avalanches can expected, even in moderately steep terrain*. |

* The avalanche-prone locations are described in greater detail in the avalanche forecast (elevation, slope aspect, type of terrain): moderately steep terrain: slopes shallower than about 30 degrees; steep slopes: slopes steeper than about 30 degrees; very steep, extreme terrain: particularly adverse terrain related to slope angle (more than about 40 degrees), terrain profile, proximity to ridge, smoothness of underlying ground surface.

** Additional loads: low: individual skier / snowboarder, riding softly, not falling; snowshoer; group with good spacing (minimum 10 m) keeping distances. high: two or more skiers / snowboarders etc. without good spacing (or without intervals); snowmachine; explosives. natural: without human influence.

– When the EADS was translated into other languages, sometimes deviations from the original (German) text were introduced. Moreover, it is possible that individual warning services have developed their own guidelines on how to interpret the danger levels over the years, which may be one source for the observed differences in the use of the danger levels in the European Alps (Techel et al., 2018).

Due to these ~~short-comings, a revision of the EADS is required. Such a revisionshould not only address these points, but must be congruent~~ shortcomings, the European Avalanche Danger Scale (EADS) is currently undergoing revision. The updated version will address the identified issues and align the scale with the terminology and definitions currently ~~used by EAWS to describe~~ adopted by EAWS for describing avalanche danger. ~~In addition, a revised EADS must be connected to the forecasting workflow,~~ Moreover, the revised EADS will need to be integrated with the operational forecasting workflow and the EAWS Matrix presented in this paper.

## 2.2 ~~Historic development of~~ North American Public Avalanche Danger Scale and the ~~EAWS Matrix~~Conceptual model of avalanche hazard

~~In order to harmonize the use of the danger levels among European avalanche forecasters, the EAWS adopted a look-up table developed by Bavarian forecasters in 2005, the so-called *Bavarian Matrix* (BM; shown in the Appendix Figure A1). The BM was split into two sub-matrices: one relating to the potential for human-triggered avalanches and the other to natural avalanche occurrence. Relying on the terminology of the EADS , a danger level was indicated for each possible combination describing the *probability of avalanche release* and the *distribution of hazardous sites* within the two sub-matrices. The main benefit of the BM wasthat it provided a suggestion for scenarios for which the EADS provided no guidance. However, the BM inherited the short-comings noted for the EADS as the factors determining avalanche danger, like spatial distribution, avalanche size and probability, were still not clearly separated nor defined.~~

~~With some variations, the EADS was~~ The EADS was, with minor adjustments, adopted in North America in 1994 (Dennis and Moore, 1996). It was used until 2007, when a revised danger scale, the *North American Public Avalanche Danger Scale* (~~NADS~~NAPADS), was introduced (Statham et al., 2010). In contrast to the EADS, the NAPADS, was intended and designed for public communication only. This revision also triggered work on a general concept for avalanche hazard assessment resulting in the *Conceptual Model of Avalanche Hazard* (CMAH, Statham et al., 2018). The CMAH identifies the key components of avalanche hazard and structures them into a systematic, consistent workflow for hazard assessments. The method is applicable to all types of avalanche forecasting operations, and the underlying principles can be applied at any scale in space or time (Statham et al., 2018). The workflow sequentially addresses the four questions: «What type of avalanche problem(s) exists? Where are these problems located in the terrain? How likely is it that an avalanche will occur? and How big will the avalanche be?» (Statham et al., 2018, p. 671). While the CMAH has become the standard workflow for avalanche forecasting in North America, it was comparably slowly adopted in regional avalanche forecasting in Europe despite there being a general agreement with the concept. Potential reasons for this slow uptake likely include: (i) The CMAH does (deliberately) not conclude with a danger level (Statham et al., 2018). (ii) The CMAH described the locations and spatial distribution of the avalanche problem rather than solely assessing snowpack stability. Analyses in Europe clearly distinguished between the frequency of points with a certain snowpack stability (potential triggering spots) and their actual location (e.g., close to ridge lines, in bowls, ...) (Schweizer et al., 2020; Techel et al., 2020a; Hutter et al., 2021) stating that only the frequency component is relevant for determining the danger level. And lastly, (iii) while the terminology used in the CMAH worked well in the English language, it worked poorly in many European languages (Müller et al., 2016).

Avalanche problems, such as persistent weak layers or wind slabs, describe typical avalanche scenarios and are integral to the CMAH and the avalanche danger assessment (EAWS, 2022a). They are defined as a ‘set of factors that describe the avalanche hazard’ (Statham et al., 2018). These factors include the sensitivity to triggers, spatial distribution and avalanche size in the terminology of the CMAH. The naming convention for avalanche problems differs slightly between North America and Europe, but are generally congruent in operational applications.

### 2.3 Historic development of the EAWS Matrix

The avalanche danger scales offer brief descriptions of the five danger levels, including typical values for the key factors that define each level. However, they lack clarity on the subtleties and do not specify exactly when a transition from one danger level to the next should occur. To address this limitation, look-up tables were introduced with the goal of explicitly linking danger levels to all possible combinations of the determining factors. Compared to the scale alone, these tables provide clearer and more comprehensive guidance for assigning danger levels. The intention is to achieve more consistent application of the scale and aims to harmonize danger level assessments among European avalanche forecasters.

The first look-up table formally adopted by the EAWS was the so-called *Bavarian Matrix* (BM), developed by Bavarian forecasters in 2003 (see Appendix, Figure A1). The BM was split into two sub-matrices: one relating to the potential for human-triggered avalanches and the other to natural avalanche occurrence. Relying on the terminology of the EADS, a danger level was indicated for each possible combination describing the *probability of avalanche release* and the *distribution of hazardous sites* within the two sub-matrices. The main benefit of the BM was that it provided a suggestion for scenarios for which the EADS provided no guidance. However, the BM inherited the short-comings noted for the EADS as the factors determining avalanche danger, like spatial distribution, avalanche size and probability, were still not clearly separated nor defined.

In 2016, Müller et al. (2016) attempted to bridge the gap between the concepts introduced in the CMAH and the structure of the Bavarian Matrix leading to the proposition of the *Avalanche Danger Assessment Matrix* (ADAM; see also Figure A2 in Appendix). This was the first attempt to tailor the CMAH to the specific needs of regional avalanche forecasters - an approach that laid the foundation for subsequent iterations of the EAWS Matrix. ADAM provided a workflow similar to the one suggested by the CMAH and integrated the concept of the spatial distribution in the assessment process. ADAM avoided the issue of the poorly defined probability terms used in the EADS by first evaluating snowpack stability against its spatial distribution separately, resulting in a likelihood-score ranging from *unlikely* to *very likely* when merging them. In a further step, likelihood is combined with avalanche size resulting in a danger level. ADAM was presented in two versions, one using the terminology in line with EADS and another one using the terminology from the CMAH. Thus, ADAM also provided a first translation between the terminologies of EADS and CMAH.

At about the time when Müller et al. (2016) developed ADAM, a working group of EAWS presented an updated version of the BM in 2017, which we refer to as *EAWS-Matrix-v2017*. This matrix introduced *avalanche size* as a separate dimension, and, thus, allowed forecasters to adjust the danger level described by the *distribution of hazardous sites* and the *probability of avalanche release*. However, most identified shortcoming of the EADS and BM were still present ~~. In the following years, avalanche forecasters in Europe did not use a common matrix when assigning a danger level; instead each warning service had a preference for one of the three matriœs (BM, EAWS-Matrix-v2017, ADAM)~~in the EAWS Matrix.

Covering the dimensions of snowpack stability (sensitivity to triggers), frequency (spatial distribution), and avalanche size, ADAM, the EAWS Matrix, and their later iterations can be seen as a specialized adaptation of the CMAH hazard chart (Fig. B1), tailored specifically to the needs of regional avalanche forecasters.

In North America, Thumlert et al. (2020) proposed numerical values to five likelihood terms, which were related to the frequency of natural avalanches releasing in 100 avalanche paths. The five likelihood terms differed compared to any of the other scales in use. Based on the concept presented in ADAM, Thumlert et al. combined these likelihood terms with avalanche size, introducing a first North American version of an avalanche danger assessment matrix (see also Figure A3 in Appendix).

Common to all ~~these~~ the above mentioned matrices was that they were exclusively based on expert judgments and had been designed by small groups of forecasters (sometimes from only one or two warning services). As a result, avalanche forecasters in Europe did not use a common matrix when assigning a danger level; instead each warning service had a preference for one of the three matrices (BM, EAWS-Matrix-v2017, ADAM) or none. What was lacking was either data or a consensus within the European avalanche forecaster community on how to resolve the current issues. Consequently, Techel et al. (2020a) tackled this issue and derived a first data-based characterization of the factors determining avalanche danger, which they termed *snowpack stability*, the *frequency distribution of snowpack stability*, and *avalanche size*. Analyzing a large data set of stability tests and avalanche observations from Switzerland and Norway, Techel et al. showed that the frequency of the locations with the ~~poorest~~ lowest snowpack stability increased with increasing danger level. However, a similarly clear correlation between avalanche size and danger level was not evident. It was observed that the size of the largest avalanche per day and warning region increased only for the higher danger levels. Building upon these insights and drawing inspiration from the matrix layout employed in ADAM, Techel et al. introduced a data-driven matrix. This new matrix utilized simulated stability distributions along with information on the largest avalanche size (refer to Figure A4 in the Appendix).

~~Following these developments~~

## 3 Revision process (step 1): Definitions and forecaster survey

The terminology used in the European Avalanche Danger Scale (EADS) and the EAWS Matrix lacked clear and consistent definitions. As a first step in the revision process, we clarified these terms and established common definitions to ensure a shared understanding among avalanche forecasters. With these definitions in place, forecasters from across Europe were asked to assign a danger level to all possible combinations of the defined factors and their respective classes.

### 3.1 Definition of factors determining avalanche danger (levels)

Following the developments described in Section 2.3, a working group of the EAWS adopted the concept and terminology used in Techel et al. (2020a) for the factors determining avalanche danger, namely *snowpack stability*, the *frequency of snowpack stability*, and *avalanche size*, and provided definitions for these factors and their respective classes (EAWS, 2022c):

– The *avalanche danger level* is a function of snowpack stability, the frequency distribution of snowpack stability and avalanche size for a given unit (area and time). There are five avalanche danger levels: 5 (*Very high*), 4 (*High*), 3 (*Considerable*), 2 (*Moderate*), 1 (*Low*).

**Table 2.** Snowpack stability classes referring to the point scale, and the type of triggering typically associated with these classes. For the full table, including typical observations related to each class, see EAWS (2022c, Table 1)EAWS (2022c, Figures A1-A3).

| Stability class | Description |
|---|---|
| Very poor | very easy to trigger (e.g., natural) |
| Poor | easy to trigger (e.g., a single skier) |
| Fair | difficult to trigger (e.g., explosives) |
| Good | stable conditions |

**Table 3.** Frequency classes of snowpack stability, taken from EAWS (2022c, Table 2).

| Frequency class | Description | Evidence (e.g., observations) |
|---|---|---|
| Many | Points with this stability class are abundant. | Evidence for instability is often easy to find. |
| Some | Points with this stability class are neither many nor a few, but these points typically exist in terrain features with common characteristics (i.e., close to ridgelines, in gullies). | |
| A few | Points with this stability class are rare. While rare, their number is considered relevant for stability assessment. | Evidence for instability is hard to find. |
| None or nearly none | Points with this stability class do not exist, or they are so rare that they are not considered relevant for stability assessment. | |

- *Snowpack stability* is a local property of the snowpack describing the propensity of a snow-covered slope to avalanche (Reuter and Schweizer, 2018). Snowpack stability is described using four classes (Table 2).
- The *frequency distribution of snowpack stability* describes the percentages of points for each stability class relative to all points in avalanche terrain. Thus, the frequency $f$ for all points with stability class $i$ ($n_i$) compared to all points ($n$) is $f(i) = n_i/n$. The frequency distribution of snowpack stability is described in four classes (Table 3).
- *Avalanche size* describes the destructive potential of avalanches (Table 4).

In theory, the EAWS workflow requires forecasters to estimate the frequency distribution of snowpack stability classes across all points in avalanche terrain within a warning region. Independent of the spatial scale of the forecasting problem, assessing snowpack stability has traditionally relied heavily on observations of avalanche activity, signs of instability, and stability test results (Reuter and Schweizer, 2018). More recently, this has been complemented by stability information extracted from one-dimensional physical snowpack models (e.g., Mayer et al., 2022; Herla et al., 2022)(e.g., Mayer et al., 2022; Herla et al., 2022; Binder and . In practice, however, estimating snowpack stability at every point in a large region remains impossible. Forecasters therefore

**Table 4.** Avalanche size classes, taken from EAWS (2022c, Table 3).

| Size class | Label | Destructive potential |
|---|---|---|
| 1 | Small | Unlikely to bury a person, except in run out zones with unfavorable terrain features (e.g., terrain traps). |
| 2 | Medium | May bury, injure, or kill a person. |
| 3 | Large | May bury and destroy cars, damage trucks, destroy small buildings and break a few trees. |
| 4 | Very large | May bury and destroy trucks and trains. May destroy fairly large buildings and small areas of forest. |
| 5 | ~~Extreme~~ Extremely large | May devastate the landscape and has catastrophic destructive potential. |

infer the distribution of stability classes across a region by combining sparse point observations and model data (when available), and their expertise and ~~intuition~~experiences. The estimated proportion of potentially unstable points, relative to a specific triggering level, reflects the average likelihood of triggering an avalanche ~~at a~~ in avalanche terrain within a region. This likelihood, combined with the potential avalanche size, determines the regional danger level. This approach aligns with the hazard chart in the CMAH, which categorizes avalanche danger based on the likelihood and size of avalanches (Statham et al., 2018).

~~Avalanche problems, such as persistent weak layers or wind slabs, describe typical avalanche scenarios and are integral to avalanche danger assessment. However, these problems do not directly correspond to specific snowpack stability classes, which vary spatially and temporally. For example, a persistent weak layer may be widespread, but snowpack stability could range from very poor in a few locations to poor or even fair elsewhere. Thus, the presence of an avalanche problem does not necessarily equate to a specific snowpack stability class at any given location. While the EAWS Matrix focuses on the frequency distribution of snowpack stability classes associated with specific avalanche problems, the CMAH emphasizes the spatial distribution of avalanche problems. This distinction highlights the key difference between these two approaches currently in use.~~

### 3.2 ~~Decomposing the avalanche forecasting task~~EAWS Matrix survey

~~The workflow introduced by the CMAH and adopted by EAWS decomposes regional avalanche forecasting into smaller, more manageable components. This decomposition is generally expected to improve the accuracy of forecasts, as breaking down complex tasks can lead to more precise estimates (MacGregor, 2001). However, as noted by MacGregor (2001, p. 107), Decomposition should be used only when the estimator can make component estimates more accurately or more confidently than the target estimate.~~

~~The accuracy of human estimates depends on various factors, including the quality and quantity of relevant data, the forecaster's expertise in interpreting the data, and their understanding and consistent application of the predefined categories. In a regional avalanche forecasting context, these categories describe snowpack stability, the frequency of the lowest stability~~

~~class, and the largest expected avalanche size (Tables 2 - 4), the key inputs to the EAWS Matrix. Inconsistencies in category assignments among forecasters can reduce the quality of the resulting forecast, particularly the accuracy of the avalanche danger level (Murphy, 1993; Techel et al., 2024a).~~

## 4 ~~Updating the EAWS Matrix~~

The revision of the factors determining avalanche danger by the EAWS in 2022 (~~Sect. 2) lead~~ Sec. 2.3, Tables 2-4) led to a mismatch compared with the terminology used in the *EAWS-Matrix-v2017* and EADS (Table 1). Therefore, ~~an updated matrix~~ as a first step, a matrix with the updated terminology was needed.

Most of the previous matrices (EAWS, 2005, 2017) were developed relying on the joint experience of a small group of forecasters consisting, for instance, for the *Bavarian Matrix* of one forecaster from Austria, Germany, France, Italy, Spain, and Switzerland. Unfortunately, the process on how the avalanche danger levels for individual cells within the matrices were assigned, was not documented. Beside the data-driven matrix developed by Techel et al. (2020a), which relied on Swiss data and the Swiss perspective of interpreting danger levels, there ~~was~~ is a general lack of data allowing a quantitative characterization of the danger levels. Moreover, even if relevant data were available in time and space, assigning a danger level to available evidence remains an expert judgment as avalanche danger cannot be measured or calculated (by algorithms) in a strict sense (e.g., Elder and Armstrong, 1987; Schweizer and Föhn, 1996).

~~Expert elicitation is a particularly suitable approach in cases when suitable data is lacking (e.g., Rowe and Wright, 2001). Therefore, EAWS chose to follow~~

### 3.0.1 Data and Methods

Given the subjective nature of avalanche danger assessments and the lack of relevant objective data for revising the Matrix, expert elicitation offers a structured and transparent method to harness expert judgment in complex, data-sparse contexts where human perception is central (Rowe and Wright, 2001). We therefore followed a similar path as for previous ~~matrices~~ matrix versions by combining multiple expert opinions and drawing on the collective knowledge of avalanche forecasters and their perception of the factors and danger levels. ~~However, unlike prior versions, where a small, though likely representative work group made decisions through group discussions, the~~ It is well known that the meaning of terms can vary across individuals, cultures, and languages, and even within the same individual depending on the context (e.g., Ogden and Richards, 1925; Morgan, 2017) . Therefore, some variability in how combinations of factors were linked to danger levels was expected. To minimize linguistic and cultural bias in the final matrix, we deliberately sought input from a broad cross-section of forecasters across EAWS.

Compared to earlier revisions, the current survey engaged a larger and more diverse ~~pool of domain experts. Experienced~~ group of experts. All EAWS forecasters were considered ~~possessing the appropriate domain knowledge and therefore considered equally capable of performing~~ to possess the necessary expertise and were therefore regarded as equally competent to contribute to this task. This approach ~~was~~ is grounded in the principle that the aggregated ~~judgment of several~~ judgments of multiple independent experts tends to be more ~~precise than any single individual 's opinion, provided judgments are made independently~~

(e.g., Stewart, 2001). Moreover, by inviting accurate than those of a single individual (e.g., Stewart, 2001). Additionally, by actively involving EAWS forecasters to contribute their versions interpretation of the matrix using with the updated terminology and definitions, a higher we aimed to gain broader engagement and greater acceptance of the new matrix was anticipatedrevised version.

### 3.1 Matrix survey

#### 3.0.1 Survey

The matrix was distributed as a survey during the autumn of 2022 with We invited avalanche forecasters to participate in a survey via the EAWS mailing list and/or the heads of warning services during the spring of 2022. We provided the following instructions:

1. Assign a danger level for each combination of classes describing snowpack stability, the frequency of snowpack stability, and largest expected avalanche size (Tables 2-4). For instance, assign a danger level to a scenario that could be described as «*Many* locations exist, where *poor* snowpack stability prevails. Avalanches can reach up to *size 3*.» , where italicized words describe the classes determining avalanche danger.
   (a) Begin with the most unfavorable stability class (*very poor*), which is typically associated with natural avalanches (Table 2), and assign a danger level to every frequency – avalanche size – combination.
   (b) Next, consider *poor* as the determining stability class. Assume that the frequency of locations with stability class *very poor* is *none or nearly none*, or at most *a few* (Table 3).
   (c) Repeat the process for *fair* stability. When *good* is assessed as the lowest stability class, avalanche danger is low.
2. Indicate a primary (more weight) and secondary danger level (less weight) if uncertain between two danger levels.
3. Leave the cell empty if a combination of factors is implausible or if unsure about the appropriate danger level.

Participants were encouraged to fill in all cells for which they felt confident assigning a danger level, leaving the stability category *fair* as optional with the aim to increase participation rates. The experts answered the survey typically in a 'cold state', meaning outside an operational setting with a specific situation at hand.

Following best practices for expert elicitation, we instructed forecasters to complete this task independently of other forecasters. Most importantly, the danger levels determined for specific combinations of stability, frequency, and avalanche size should not be discussed among forecasters until after they had submitted their responses.

We received 60 responses to the survey. To derive the updated matrix, we additionally considered the following sources:

– Working group members provided their version of the matrix at a meeting in 2019, and again in 2022 (N = 5 and 9, respectively). We employed the test-retest reliability methodology (Ashton, 2000) to evaluate the consistency of their responses and to obtain more reliable estimates. Additionally, the second round served as a pilot study to test the survey distributed to EAWS forecasters outside our the working group.

**Table 5.** Distribution of the 76 matrix responses received by country~~, taken from EAWS (2022c, Table D-1)~~. Forecasters in the Czech Republic, Finland, Iceland, Poland and Slovakia were approached, but did not respond.

| Country | N |
|---|---|
| Andorra | 3 |
| Austria | 4 |
| France | 7 |
| Germany | 5 |
| Great Britain | 7 |
| Italy | 18 |
| Norway | 15 |
| Romania | 1 |
| Slovenia | 1 |
| Spain | 5 |
| Switzerland | 8 |
| Sweden | 2 |

- We ~~invited avalanche forecasters to participate in our survey via the EAWS mailing list and/or the heads of warning services, receiving 60 responses in total.~~

- ~~Whenever available, we incorporated~~ incorporated two quantitative studies into our analysis (N = 2; Swiss data: Techel et al., 2020a; Hutter et al., 2021).

In total, we ~~received~~ had 76 responses from 12 different European countries (Table 5)~~. By combining these sources, we aimed to generate a comprehensive and robust~~ , which we consider a comprehensive pool of opinions reflecting the current state of avalanche danger assessment practices in Europe.

### 3.1 ~~Analysis of survey responses~~

#### 3.0.1 Survey analysis

In line with best-practice approaches when combining judgments from experts (e.g., Dietrich and Spiekermann, 2023), and not favoring any one opinion, we opted to calculate the median danger level for each combination of stability, frequency, and avalanche size. In addition, we checked whether the median danger level was also the danger level proposed by the majority of respondents. Since respondents could provide both a first and second danger level, we weighted their answers accordingly:

- If a forecaster provided a single danger level, this danger level was weighted with 100.

- If a forecaster provided two danger levels, the first danger level was weighted with 67 and the second with 33.

### 3.1 Survey results

#### 3.0.1 Survey results

Distribution of survey responses for each danger level (a: $D = 1$ to e: $D = 5$). Shown are the proportions for each combination of stability, frequency, and avalanche size summing up to 1 for each $D$. Values are displayed if they received $\geq 0.01$ of the votes. Stronger color saturation indicates a larger proportion of responses favoring a specific combination.

Figure 1 shows the distribution of responses for each factor combination and danger level. As can be seen, a range of factor combinations was used for each danger level. While the survey provides insights into the most typical combinations for each danger level, there were also some combinations, which were rarely or never selected (blank cells). Our examination of the most frequently selected combinations confirms In line with the definition of the avalanche danger levels, Figure 1 shows that, as danger levels increase, snowpack stability decreases while frequency and avalanche size are expected to increasewhile snowpack stability is considered to decreaseincrease. Notably, the combinations with the highest response rate for each danger level often have secondary choices diagonally above or below that value. This suggests that two factors can offset each other to qualify for the same danger level. For instance, a higher probability of triggering (frequency of snowpack stability) might be balanced by lower consequences (smaller avalanches).

Based on the rank-ordered Rank-ordering the danger level responses for each cellcombination of stability, frequency, and avalanche size, we derived the median $D$ danger level, referred to as $D^1$, and any second $D$ (shown in brackets) danger level, $D^2$, falling within the interquartile range for each combination of stability, frequency, and avalanche size (Figure 2a). We refer to these two danger levels as $D^1$ and $D^2$, respectively. Analyzing the responses across the 45 cells, we find that 27 cells contain a $D^2$, indicating considerable variability in opinions.

Examining the proportion of responses aligning with A clear majority vote existed for only 18 of the median $D$ (45 possible factor combinations (Fig. 2b). Not surprisingly, the cells with highest agreement define the limits of the danger scale (*fair-a few-1*, $D^1 = 1$: 97%; *very poor-many-5*, $D^1 = 5$: 99%; Figure 2b)reveals strong agreement for cells in the upper left and lower right corners (proportion $\geq 0.97$), corresponding to the extreme ends of the stability, frequency and avalanche size spectrum. However, seven combinations, primarily associated with size 1 or size . The other two cells that stand out with regard to a high agreement of responses are *very poor-many-3* ($D^1 = 4$: 85%) and *poor-some-3* ($D^1 = 3$: 84%). These two combinations align well with the description of danger levels 3 (considerable) and 4 avalanches, show lower (high) in the EADS (Tab. 1), which likely explains the clear preference for one danger level in the survey. Otherwise, cells of comparably high agreement are scattered across the matrix, with no obvious pattern connected to one of the factors or to $D$. On average across all cells, one danger level was supported by 67% of the votes, the remaining votes generally went to a second danger level. Seven combinations showed particularly low agreement rates (proportion $\leq 0.55$), emphasizing the uncertainty in describing these cells indicating that the median danger level may be less representative for these cases (e.g., *fair-a few-4*).

Figure 2c illustrates the support, or the percentage of responses, for each specific combination. On average, respondents provided danger level values for 85% of the possible 45 combinations. Notably, cells with *very poor* and *poor* stability received responses from 72 of the 76 respondents ($\geq 95\%$) for 17 of the 30 combinations. Although stability *fair* was optional in our

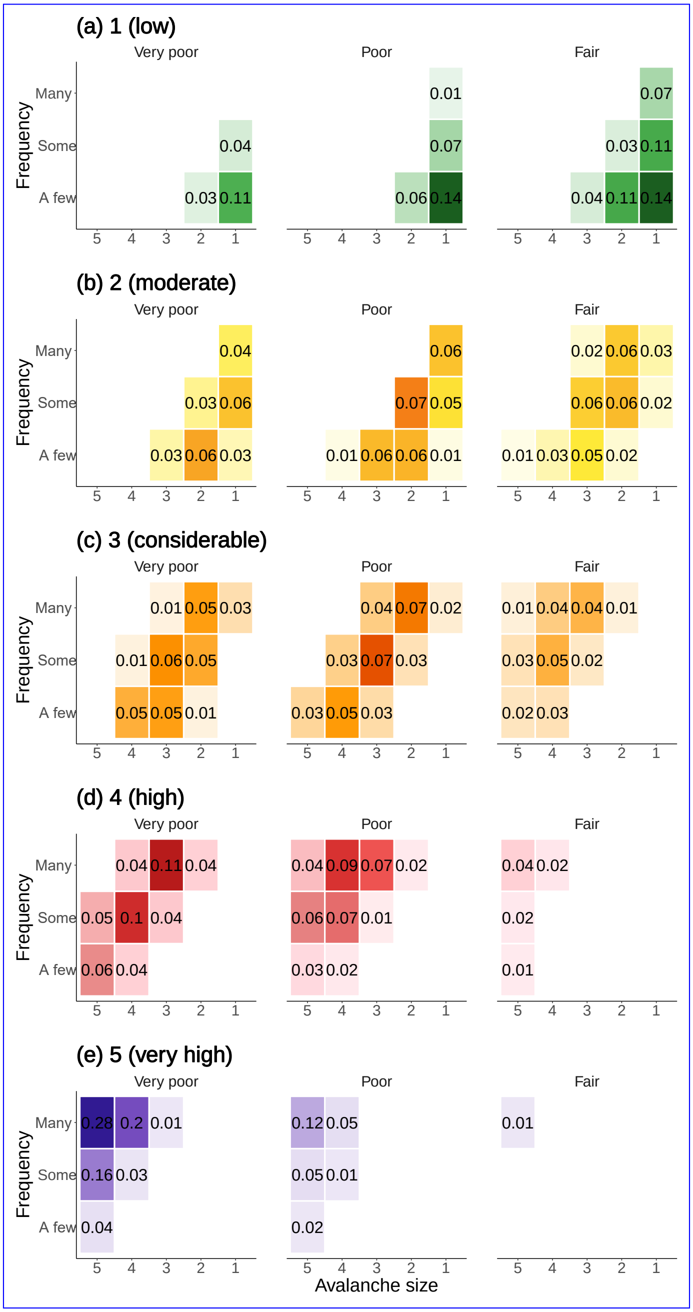


**Figure 1.** Distribution of survey responses for each danger level (a: $D = 1$ to e: $D = 5$). Shown are the proportions for each combination of stability, frequency, and avalanche size summing up to 1 for each $D$. Values are displayed if they received $\geq 0.01$ of the votes. Stronger color saturation indicates a larger proportion of responses favoring a specific combination.

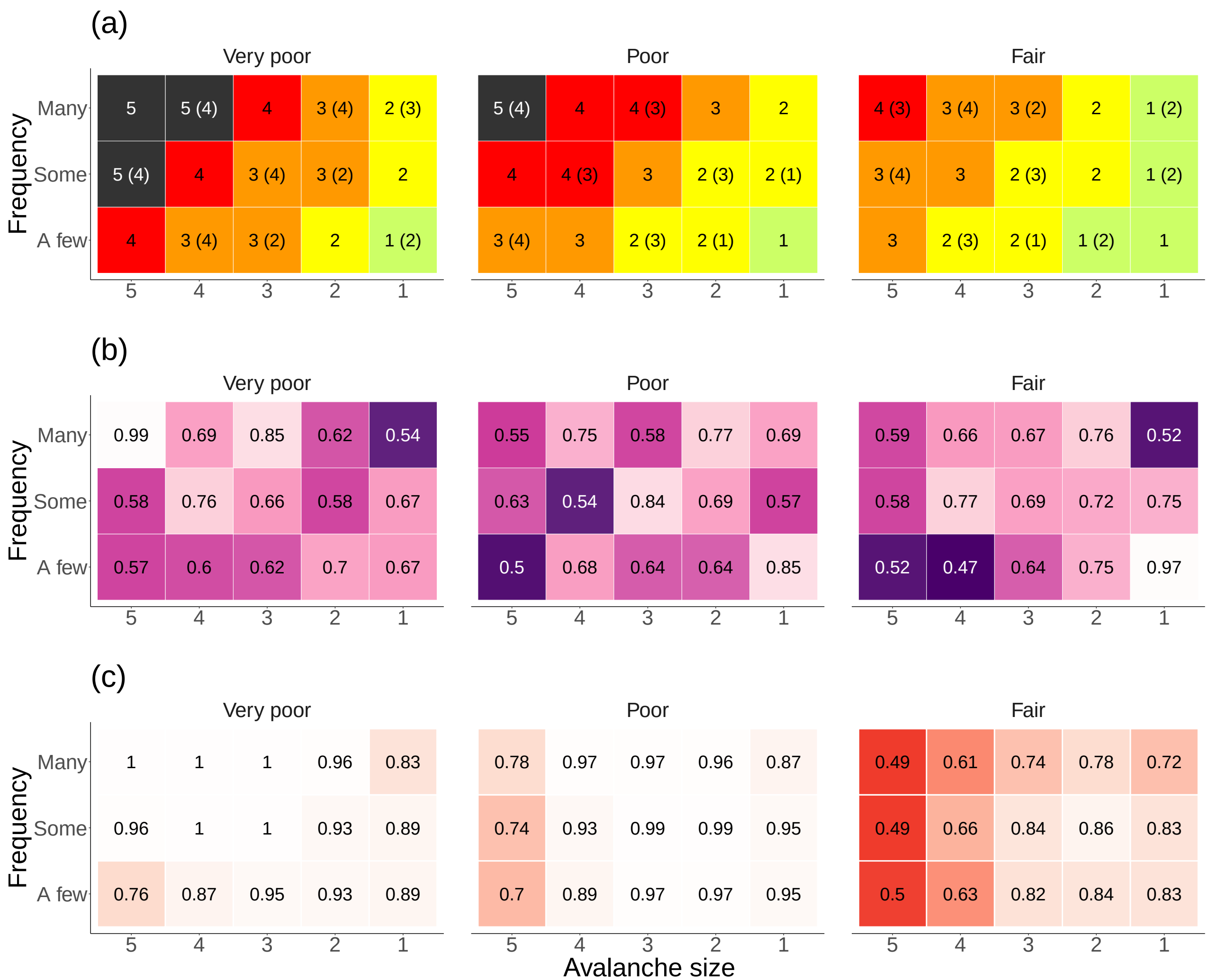


**Figure 2.** Survey responses for each combination of stability (panels), frequency (y-axis), and avalanche size (x-axis). ~~The median~~ (a) Median danger level ($D^1$)~~is displayed in a)~~, ~~while the proportion of responses agreeing~~ with ~~$D^1$ is~~ the second most frequent level ($D^2$) shown in brackets if within the interquartile range (see text for details). (b) Proportion of responses that agree with $D^1$. ~~The proportion~~ (c) Proportion of responses providing a danger level estimate~~is depicted i c)~~. ~~Cells with stronger~~ Stronger color saturation ~~indicate cells with~~ indicates lower agreement (b) or fewer responses (c)~~to emphasize considerable variability in opinions~~.

survey, it received responses in over 82% of cases when combined with frequency classes *a few* and *some* and avalanche *sizes* ~~*1-3*~~*1, 2, and 3*. *Fair* stability had lower response rates when paired with avalanche size 4 ($\leq 66\%$) or size 5 ($\leq 50\%$). Possibly, this indicates that a considerable share of forecasters rated these combinations as less plausible.

It is not surprising that the 76 responses from various European countries and warning services revealed considerable variability in the assignment of $D$ across most factor combinations. Moreover, as shown in EAWS (2022b), 'cultural' differences can be observed when comparing responses by country. For instance, the mean response by Scottish forecasters resulted in five matrix cells with $D^1 = 5$, whereas only two such cells were assigned $D^1 = 5$ by Norwegian or Swiss forecasters. Similar 'cultural' differences have been documented before, such as when assigning danger levels (Lazar et al., 2016; Techel et al., 2018) or estimating avalanche size (Hafner et al., 2023). Despite the EADS being in use for three decades, the absence of unambiguous, standardized guidelines – and a shared understanding of definitions – likely contributes to these variations across European Avalanche Warning Services (Techel et al., 2018).

Given these divergent perceptions of danger levels, it was all the more important to involve a large number of forecasters with varied operational backgrounds to ensure that the updated Matrix would reflect a broad and representative understanding across services.

## 4 ~~The updated~~ EAWS matrix and associated workflow

The findings presented ~~in Section 4~~ above led to the development of an updated matrix (Figure 3), hereafter referred to as ~~EAWS Matrix, which was officially accepted by the EAWS General Assembly in June 2022. The updated matrix provides a comprehensive framework for assessing avalanche danger ratings, taking into account three factors to determine avalanche danger: snowpack stability (shown as panels in Figure 3), frequency (along the y-axis), and avalanche size (along the x-axis).~~ the *EAWS Matrix* or simply the *Matrix*. The design of the ~~EAWS Matrix is based~~ Matrix builds on the recognition that the frequency of locations with the weakest snowpack stability is often ~~decisive~~ the most decisive factor for determining the avalanche danger level (Techel et al., 2020a). ~~Therefore, the matrix is structured to address the three lowest~~ This concept is reflected by displaying three separate panels for the stability classes *very poor*, *poor*, and *fair*, which are connected by arrows from left to right (Figure 3). For each stability class, combinations of frequency (y-axis) and avalanche size (x-axis) are summarized in ~~separate panels (Figure 3). When using this matrix to assess danger levels, forecasters follow a systematic approach, starting from left to right. They begin by considering the lowest stabilityclass. If this class corresponds to a frequency of~~ a separate panel. The layout supports a step-by-step assessment: the forecaster starts in the upper left corner, where the most severe conditions are represented, and works through the Matrix by sequentially ruling out less likely combinations. This progression helps ensure that more serious scenarios are considered first, before settling on the cell that best reflects the expected conditions.

To assign a danger level using the Matrix, forecasters begin by evaluating the frequency of locations with *very poor* stability. If such locations are absent or deemed irrelevant for danger assessment (i.e., *none or nearly none*, ~~they proceed~~ see Table 3), the assessment proceeds to the next stability class~~on the right, and so forth, indicated by the arrows in Figure 3. When the snowpack stability is evaluated~~ , following the directional flow of the Matrix. This stepwise evaluation may continue through the *poor* and, if necessary, *fair* classes. If stability is assessed as *good*, the danger level is ~~automatically assigned as *1-Low*, regardless of the values of the other two factors. It is important to note that while stability and frequency are closely linked, the assessment of avalanche size is independent. Forecasters select the largest size class to~~ set to *1 (low)* by default. In situations where both the lowest and next-lowest stability classes are relevant – e.g., when the latter is significantly more frequent – forecasters may

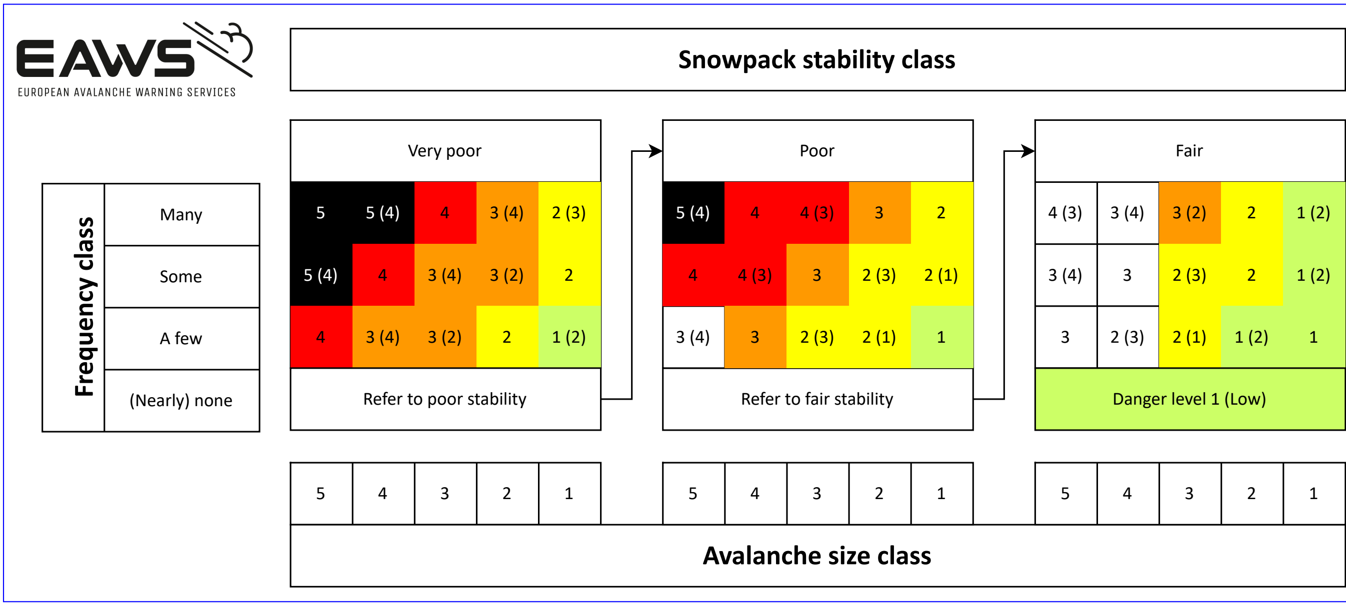


**Figure 3.** EAWS Matrix, as accepted by the EAWS General Assembly in 2022 (taken from EAWS, 2022c). For a detailed explanation refer to the text.

consider more than one panel. The final step involves estimating the largest avalanche size that can be reckoned with ~~given~~ under the observed or anticipated ~~avalanche~~ conditions.

~~To accommodate cases when forecasters assigned different danger levels for the same combination of~~

The combination of selected stability, frequency, and ~~size (Figure~~ avalanche size classes results in one matrix cell indicating the danger level that best represents the situation within a region. However, as the survey results did not always yield a clear danger level consensus for a given factor combination (Figure 2b), the matrix displays either one or two danger levels per cell. Displayed are the respective integer values of ~~a danger level (i.e.~~ the danger levels (e.g., 1 for ~~*1-Low*~~ *1 (low)*). The median danger level, ~~referred to as~~ $D^1$, ~~represents the danger level suggested by the majority of~~ reflects the most common and average response among forecasters and determines the cell~~'~~'s color. ~~In addition, a second danger level ($D^2$) is shown in brackets, if~~ If the interquartile range of ~~the danger level responses included a second~~ responses includes a second, distinct danger level, ~~which was different from $D^1$. By displaying~~ this level is displayed in brackets as $D^2$. By including a second danger level, the ~~variation in forecaster opinions regarding the danger level is intentionally maintained. Taking the factor~~ matrix intentionally retains the variation in expert opinion. For example, for the combination *very poor* ~~-~~ – *some* ~~-~~ – *size 3* ~~as an example (Figure~~ (Figure 3), the ~~resulting danger levels are~~ matrix shows $D^1 = 3$ and $D^2 = 4$. ~~Figure 2b shows that this combination was one where some variation existed with~~ As illustrated in Figure 2b, 34% of forecasters ~~favoring~~ favored a danger level ~~different than~~ other than ~~*3-Considerable. Cells without coloring represent instances where*~~ *3 (considerable)* for this combination. Matrix cells are left uncolored if fewer than 70% of respondents provided a danger level estimate (Figure 2c).

EAWS Matrix, as accepted by the EAWS General Assembly in 2022 (taken from EAWS, 2022c), is organized into three distinct panels, each corresponding to one of the three lowest snowpack stability classes. Within each panel, the frequency (on the y-axis) is plotted against avalanche size (on the x-axis). To navigate through the matrix, arrows interlinking the panels indicate to transition to the next stability class, when the frequency is evaluated as *none or nearly none*. According to the workflow (Müller et al., 2023), forecasters progress from left to right across the panels, assessing the frequency of the lowest stability class first. Upon identifying a relevant stability-frequency combination, forecasters then evaluate the corresponding avalanche size, leading to the recommended danger level(s).

To facilitate the To support the operational application of the EAWS Matrix, the EAWS working group has developed a workflow that outlines outlining the necessary steps for determining the avalanche danger level within a warning region (Müller et al., 2023). The workflowis specifically designed for regional avalanche forecastersThis workflow, like the Matrix itself, is largely aligned with the CMAH, but is explicitly tailored to the context of public regional avalanche forecasting. It assumes that the forecast area is large enough to encompass include multiple mountains, elevation zones, all bands, aspects, and varied terrain features , such as ridges, gullies, and open slopes. ConsequentlyAs a result, terrain is not treated as an independent factor . The workflow involves influencing the danger level. However, specific terrain features might be mentioned in the corresponding avalanche forecast (e.g., Hutter et al., 2021).

The workflow entails assessing all relevant avalanche problems within the given a region, evaluating their snowpack stability, frequency, and avalanche size for each of them, and then using the EAWS Matrix to assign a danger levelto each problem. The highest resulting danger level among the considered avalanche problems is communicated for the given warning region. This structured approach ensures that all relevant factors are considered , aiming for a more consistent evaluation of avalanche danger.

Use of matrix during winter 2022/2023. Warning services are aggregated to groups of services using the same workflow and forecasting software or when spatially continuous. country N cases Factors (regions) (n days) use published Factors $\rightarrow D$ $D$ in forecast software deviations from $D^1$ ATB **Aus**tria (except Tyrol) Germany (**B**avaria) 2347 (155) operat'l no direct link $D^1$ suggested, $D^2$ shown allowed (5.7%) CAT Spain (**Cat**alunya) 289 (56) operat'l yes direct link only $D^1$ indicated (0%)NOR **Nor**way 7014 (181) operat'l yes direct link only $D^1$ indicated allowed (16%) SWE **Swe**den 671 (146) operat'l yes direct link $D^1$ and $D^2$ shown allowed (9.7%) SWI **Swi**tzerland 595 (81)* test (live) no no link matrix not used (47%) TST Austria, Italy (**T**yrol, **S**outh Tyrol, **T**rento) 1020 (142) operat'l yes direct link only $D^1$ is shown allowed (2.0%) VAR Spain (**V**al d'**Ar**an) 216 (129) operat'l yes direct link only $D^1$ indicated allowed (6.5%)

### 4.1 Use of EAWS Matrix in winter 2022-2023

EAWS members were encouraged to apply the definitions, workflow and updated EAWS Matrix in their operations during the winter season 2022-2023. In total, 15 EAWS avalanche warning services logged their choices for the three factors during operational forecasting, either for the entire season or for extended periods. For the purpose of this analysis, we aggregated adjacent warning services to groups of services when using the same workflow and forecasting software. Table **??** provides an overview. In most of these warning services, workflow and matrix were implemented in the operational workflow and forecasting software. An exception was the Swiss warning service (SWI), where the objective was to first gain insights on the reliability of estimating the factors describing avalanche danger according to the partly revised definitions (Tables 2-4). In SWI, forecasters were advised not to use the EAWS Matrix.

Analyzing each of the three factors determining avalanche danger separately reveals several interesting findings. Firstly, snow stability was most commonly rated as *poor* (P), with a median proportion across the warning services of 0.56 (Figure **??**a). The terms *a few* and *some* were the predominant choices for frequency, with proportions of 0.53 and 0.42, respectively (Figure **??**b). Avalanche size also displayed a clear dominance of *size-2* with a median of 0.59 of all responses (Figure **??**c). As can be noted in the range of values displayed in the boxplots in Figure **??**, the use of the classes varied quite strongly between warning services. However, we refrain from interpreting these variations, as a multitude of factors, including seasonal and snow-climatological differences, data collection methodologies, and potential differences in the interpretation of class definitions, may all contribute to the observed differences.

Use of classes during winter 2022-2023 describing (a) snowpack stability, (b) frequency, and (c) avalanche size. The boxplots represent the distribution of the class labels summarized for all warning services in Table **??**. Abbreviations for i) stability are VP-very poor, P-poor, F-fair, G-good, ii) frequency are Ma-many, So-some, Fe-a few, Nn-none or nearly none, iii) and the avalanche size classes 1-5 (size-5 never used).

Use of individual matrix cells. The values represent the unweighted means of the respective proportions for all groups of warning services as in Table **??**. Stronger color saturation indicates a larger proportion of responses.

Examining the use of specific combinations of these factors (Fig. **??**), we observe that the two single most frequent combinations were *poor-some-2* and *poor-a few-2*, representing 17% and 16% of the responses. Combinations featuring *very poor* or *fair* stability were only used about half of the time compared to *poor*. Combinations denoting *many* for frequency were seldom used, with proportions falling below or equal to 1% . Furthermore, combinations featuring avalanches of *size 4* were exceptionally rare, while *size 5* was virtually never selected, underscoring their association with extremely rare and unusual avalanche conditions.

Analyzing differences between $D$ given by a forecaster for a specific combination of stability, frequency, and size, and $D^1$, as indicated in the Matrix, shows that almost all warning services exhibited some level of deviation from the Matrix during the winter occasionally (Tab. **??**). With 47% of the cases, deviations were most frequent in SWI, where the Matrix was not consulted. In NOR, where the Matrix was used operationally, deviations were also observed comparably often (16%). In contrast, CAT never deviated from $D^1$, and in TST the proportion of deviations was also small (2%).

Figure **??** shows the agreement between $D$ and $D^1$ for all combinations of factors and the seven groups of warning services, as defined in Table **??**. The left column shows the proportions of cases when $D = D^1$, and the right column where $D \neq D^1$ for each group of warning services. Of interest are primarily the combinations for which high proportions of $D \neq D^1$ were observed. These deviations were most frequent for SWI and NOR, followed by SWE and VAR. $D \neq D^1$ was $> 0.5$ in eight cases (SWI 6, NOR 2), being most pronounced for *poor-many-3* with 100% deviation for SWI and 75% for NOR. NOR had comparably low proportions of $D = D^1$ for *fair*, while in SWI avariety of cells showed deviations, with no obvious pattern. For the most frequently used combination, *poor-some-2*, all services (except CAT)deviated from the Matrix in at least 4% of cases.

For this cell, disagreement with $D^1$ was particularly large in SWI, where $D^2 = 3$ was chosen in 53% of cases. The pattern of optioning for $D^2 = 3$ was also comparably evident in SWE, NOR, and VAR (14-19% of cases). As several forecasters assessed each situation in SWI, a majority opinion can be derived for each case. Considering the 40 occasions when the factor combination using majority voting resulted in *poor-some-2*, the tendency to deviate from $D^1$ was even more pronounced: the majority-voted $D$ was $D^2 = 3$ in 25 cases (63%) and $D^1 = 2$ in 12 cases; in three cases it was undecided between the two danger level options.

Proportion of cases where forecasters agreed (left panel) or disagreed (right panel) with $D^1$ as were used by groups of warning services (center column). Combinations involving stability class *fair* and frequency class *many* were not recorded during 2022-2023. Abbreviations according to Table **??**. Cells are displayed if the number of cases was $\geq 5$.

Proportion of cases that a specific danger level (a: $D = 1$ to d: $D = 4$) was used for a specific matrix combination. Proportions are derived by summing up all cases $n$ across warning region groups. Values are shown for all cells with a proportion $\geq 0.01$. Colour shading correspond to the proportion of cases.

The spread of cells used for each $D$ is shown in Figure **??**. $D = 1$ ($n = 2203$, 18% of all cases) is closely linked to avalanche size 1 and *a few* locations with predominantly *fair* or *poor* stability. In addition, $D = 1$ is associated with *good* stability in 10% cases (*good* stability is not shown in Fig. **??**). $D = 2$ ($n = 6806$, 56%) is primarily clustered around *poor* stability, with *a few* or *some* locations and an avalanche size of 2. $D = 3$ ($n = 3025$, 25%) is scattered around the combinations *very poor - some - size 2* and *poor - some - size 3*, while $D = 4$ ($n = 118$, 1%) is clearly centered at *very poor - many - size 3*.

In Figure **??** we show the typical factor combinations for each group of warning services. Overall, the most commonly used descriptions of $D$ exhibited strong similarities, whether we examined the factors independently or in combination. Nevertheless, discernible distinctions emerge. For instance, in NOR and SWE, $D = 1$ were most frequently associated with *poor* stability, whereas elsewhere *fair* stability dominated. Similarly, at $D = 3$ in NOR, CAT, and to some extent in ATB, there was a greater prevalence of *very poor* stability compared to other warning services. Frequency terms employed to describe danger levels generally exhibited similarities, with the exception being NOR, where frequency for $D = 4$ was more frequently described as *some* rather than *many*, often in combination with larger avalanches. For example, at $D = 4$, avalanche size in NOR ranged between size 3 and 4, while in SWI, it typically varied between size 2 and 3. However, it's worth noting that the days of $D = 4$ were relatively rare (118 cases), suggesting that these results may not be fully representative. Avalanche size exhibited greater variation at higher danger levels but demonstrated remarkable uniformity at $D = 1$ and $D = 2$. And lastly, in SWI the most frequently used combination for $D = 3$ (*poor-some-2*) was the combination, which was the most frequent for $D = 2$ in many other warning services (CAT, SWE, TST, VAR). Nevertheless, when assigning equal weight to the various datasets, distinct patterns emerge regarding the most prevalent combinations (see Figure **??**).

Typical descriptions of $D$ based on the European Avalanche Danger Scale (EADS), the updated matrix (Figure **??**), and their frequency of use during the winter season of 2022/2023. In the "individual" rows, we have tallied the classes describing the factors independently, irrespective of their combination with others. These rows display the most commonly occurring class, and when a second class is present in at least 30% of cases, it is enclosed in brackets. Conversely, in the "combined" rows, we showcase the classes from the most prevalent combination of stability, frequency, and size. For reference, "n.d." denotes instances where no specific class was defined.

### 4.1 Consistency in assessing factors determining avalanche danger

To investigate potential inconsistencies in assigning classes to the three determining factors, two warning services, NOR and SWI, conducted tests during the winter season 2022-2023. In SWI, three forecasters made independent assessments of the classes on a daily basis, but only for a subset of the Swiss forecasting domain, resulting in 564 pairwise comparisons (219 unique cases on 73 forecast days). In NOR, on three separate occasions, between 6 and 12 forecasters independently conducted comprehensive avalanche danger assessments for one selected region for the following day, resulting in 117 pairwise comparisons. As mentioned before, forecasters in SWI were explicitly instructed not to consider the Matrix for assigning $D$ in order to mitigate any influence the Matrix might have on their class choices. Conversely, in NOR, incorporating the Matrix is a standard part of the forecaster workflow when assigning a danger level .

We consider the mean agreement between any two forecasters assessing the same forecasting scenario as a measure of consistency among forecasters. Here, agreement means the selection of the same class for a specific factor. With this approach , we obtain an estimate of how reliably two randomly chosen forecasters would obtain the same class. In SWI, the pairwise agreement rate for selecting the same class describing the factors was 59% for stability, 63% for frequency, and 74% for avalanche size (Table **??**). The agreement rate for the combination of all three factors was 30%, while in 4% of cases, the classes for all three factors differed. In NOR, the classes are set for each avalanche problem where the main problem is said to be decisive for the danger level. Therefore, we conducted the analysis for the avalanche problem considered decisive but also for cases where they were given in a different order (values in brackets). The respective pairwise agreement rates were 49% (54%) for stability, 50% (51%) for frequency, and 77% (73%) for avalanche size (Table **??**). All three factors were the same in 21% (21%) of cases, while in 3% (7%) of cases, the values on all three axes differed. It is of note, however, that the most frequently chosen classes were *poor* for stability (67%), *some* for frequency (54%), and *size-2* for avalanche size (73%). Therefore, simply using the most typical value would already yield a relatively high agreement rate.

Pairwise agreement when estimating the factors describing avalanche danger. $n$ indicates the number of pairwise comparisons, $N$ Each avalanche problem is assessed separately. If two Matrix cells are considered relevant for a single problem, the number of unique cases (unique day and warning regionone resulting in the higher danger level is chosen (Figure 4a). Source Avalanche problem Stability Frequency Size All the same All different n (N) #Forecasters NOR decisive 49% 50% 77% 21% 3% 117 (3) 6, 8, 12 same 54% 51% 73% 21% 7% 169 (3) 6, 8, 12 SWI decisive 59% 63% 74% 30% 4% 564 (219) 2 or 3

Discussion

### 4.1 Matrix survey and updated EAWS Matrix

A new EAWS matrix was needed to be congruent with the revised definitions of the factors determining avalanche danger and the workflow for assessing regional avalanche danger (Müller et al., 2023, summarized in Sect. 2). Due to a lack of objective data, expert elicitation was conducted by asking European avalanche forecasters in a survey to assign a danger level ($D$) to each combination of factors (Section 4). This task required survey participants to understand, interpret and apply the partly revised, purely descriptive definitions of the classes for snowpack stability, frequency, and avalanche size (Tables 2-4), and link them to the danger levels.

It is well known that the same word can have different meaning to different people depending on their background, culture or language, and that this meaning may also differ to the same person in a different context (e.g., Ogden and Richards, 1925; Morgan, 2017) . Thus, variations in the combination of factors and corresponding danger level had to be expected. However, to reduce the influence of a specific language or cultural background on the final matrix, forecasters from throughout EAWS were approached. It is, therefore, not surprising that the 76 responses revealed considerable variability in the assignment of $D$ across most factor combinations (refer to Section 4). Moreover, as shown in EAWS (2022b), 'cultural' differences can be noted when comparing responses by country, with, for instance, the mean response by Scottish forecasters resulting in five matrix cells with $D^1 = 5$, while only two of the matrix cells were assigned $D^1 = 5$ by Norwegian or Swiss forecasters. Such 'cultural' differences have been noted previously, as, for instance, when assigning danger levels (Lazar et al., 2016; Techel et al., 2018) or when estimating avalanche size (Hafner et al., 2023). Despite the EADS being in use for three decades, the absence of unambiguous, standardized guidelines and a common understanding and interpretation of the definitions across European Avalanche Warning Services likely contributes to these variations (Techel et al., 2018).

The revised EAWS Matrix presented in Section 3.0.1 (Figure 3) is based on 'consensus', which we defined as the most frequently proposed danger level for each factor combination ($D^1$). In addition, the second-most frequently suggested danger level ($D^2$) is displayed in brackets in the Matrix if proposed by $\geq$25% of respondents. A clear majority vote existed for only 18 of the 45 possible factor combinations (Fig. 2b). These cells of comparably high agreement are scattered across the matrix, with no obvious pattern connected to one of the factors or to $D$. In the remaining 27 cases, up to 50% of respondents suggested the same alternative danger level. Not surprisingly, the cells with highest agreement define the limits of the danger scale (*fair-a few-1*, $D^1 = 1$: 97%; *very poor-many-5*, $D^1 = 5$: 99%; Figure 2b). The other two cells that stand out with regard to a high agreement of responses are *very poor-many-3* ($D^1 = 4$: 85%) and *poor-some-3* ($D^1 = 3$: 84%). These two combinations align well with the description of danger levels 3 (considerable) and 4 (high) in the EADS (Fig. 1), which likely explains the clear preference for one danger level in the survey.

### 4.1 Matrix usage

The majority of the analyzed warning services integrated the EAWS Matrix into their forecasting software, typically defaulting to $D^1$ following selection of input parameters, and, thus, nudging forecasters towards this option. While forecasters could override $D^1$, indicating a disagreement with the EAWS Matrix, there was also the undesired option to reach a desired danger level by adjusting one or more of the input factors. This practice could mask a forecaster's true factor assessment. Although the extent of the nudging effect and the frequency of factor adjustments are unclear, these likely influenced decisions to some extent. While the nudging effect is deliberate to promote consistent danger level assignment, care is needed to avoid compromising forecasters' choices in parameter selection.

Moreover, it is noteworthy that feedback obtained from various seminars with forecasters following the winter season indicated that less-experienced forecasters highly appreciate the guidance offered by the EAWS Matrix, potentially also leading to a greater tendency to agree with $D^1$. In Switzerland, however, forecasters assessed factors and $D$ independently, without directly referencing the EAWS Matrix. This approach is similar to the North American system, where factors (e.g., likelihood and avalanche size) are evaluated for each avalanche problem on the hazard chart, but where the final decision on $D$ remains with the forecaster as no clear assessment rules are defined (Clark and Haegeli, 2018). Comparing Swiss forecasters choice of $D$ given the selected factor combination, $D$ frequently deviated from $D^1$ in the EAWS Matrix . However, absolute differences were generally minor as forecasters most frequently selected the sub-level closest to $D^1$ (Techel et al., 2024a, for a description of sub-levels . Given the different setups and intentions, it is not surprising that some warning services adhered to $D^1$ most of the time (i.e., ATB, CAT, TST chose $D^1 \geq 93\%$), while other services deviated more frequently from $D^1$, with a particularly large share of deviations observed in SWI (46%).

Our analysis suggests that forecasters largely concurred with the avalanche danger level proposed by the EAWS Matrix(Figure **??**). This observation is supported by the fact that for most Matrix cells there was a tendency to agree with $D^1$ (Fig. **??**) and that for most danger levels a similar set of typical factor combinations resulted (Fig. **??**), regardless whether the Matrix was used or not (SWI). This supports the assumption that, overall, the approach of deriving a consensus-based matrix by eliciting expert knowledge resulted in generally applicable factor combinations.

Despite the overall usage patterns aligning well with the danger level descriptions in the EADS (Tab. 1), two specific cases require particular attention, namely (1) the shift in stability for $D = 2$ from *fair* (EADS) to *poor* (usage), and (2) the closeness of the typical factor combinations describing $D = 2$ and $D = 3$, which are the same for two of the three factors (Figure **??**).

All warning services described stability most often as *poor* when giving $D = 2$, and , therefore, deviated from the description of stability in the EADS (which translates to *fair* stability). However, *poor* stability ratings were commonly used in combination with a lower frequency class *a few* (to *some*) by some warning services (ATB, NOR, SWE, SWI), while CAT and TST generally used *poor* stability in combination with *some* locations. In other words, the latter two warning service groups kept the frequency class from EADS, but shifted stability to a lower class, which can be interpreted as a true shift from EADS.

Danger levels $D = 2$ and $D = 3$ account for 81% of the data, aligning with previous analyses for the European Alps (about 80%; Techel et al., 2018). In the EADS and in the EAWS Matrix, these levels primarily differ by stability ($D = 2$: *fair* vs. $D = 3$: *poor*), and - in the Matrix - also by avalanche size ($D = 2$: *size 2* vs. $D = 3$: *size 3*), but less by frequency (Figure **??**). The shift to *poor* stability as seen in the usage data for $D = 2$ reduces the variations between these two danger levels. Moreover, it must be expected that the distribution of cases within $D = 3$, that is the number of cases at the lower and higher end of the level, are unbalanced towards the lower end. While we have no direct evidence for that within our data-set, the distribution of sub-levels in Switzerland, which attempt to capture exactly these variations within a level, showed that a 3- (avalanche conditions considered to be low in the level) is typically between two to four times more frequent compared to a 3+ (high in the level, (Techel et al., 2020b, 2024a). Assuming a similar distribution of avalanche conditions in other regions in the Alps (ATB, TST) suggests that there are many more situations with $D = 3$ close to $D = 2$ compared to cases close to $D = 4$, which would partly explain the comparably similar factor combinations for these two danger levels. In contrast, in Norway, there was a preference for *very poor* stability for $D = 3$. We can only speculate whether this is related to a more balanced distribution of cases within $D = 3$, whether Norwegian forecasters have a slightly different representation of *very poor* stability, or whether this is related to the much larger regions in Norway. However, the typical combinations obtained for the five danger levels fall well within the responses obtained in the survey (Fig. 1). In other words, the underlying distribution must be taken into account when considering the most frequently used combination as an approach to obtain a typical description for each danger level. Seen from that perspective, the two approaches - survey and usage - provide valuable different views to obtain a typical characterization of each danger level.

When examining the EADS, it becomes apparent that the combination *poor-some-2* and *poor-a few-2* encompass elements of the description of both $D = 2$ and $D = 3$. Consequently, it represents one of these intermediate states, which are frequently encountered and utilized by avalanche forecasters, but which are not adequately captured by the EADS. Moreover, the EADS does not assign a specific avalanche size to $D = 2$ and only states that very large avalanches (*size 4*) are unlikely (Table 1), thus only excluding *size 4 and size 5*. In the case of these combinations, the agreement rates in the survey do not significantly differ from the agreement rates observed during operational usage in 2022-2023. Figure**??** illustrates that SWI, in particular, tends to favor $D = 3$ for the combination *poor-some-2*. While NOR, SWE, and VAR also occasionally assigned $D = 3$ for this specific scenario, the other services and the EAWS Matrix currently favor $D = 2$. As mentioned before, SWI most often deviated from $D^1$, which is likely related to the fact that SWI forecasters generally assign $D$ without consulting the Matrix.

It is potentially problematic that the two most common danger levels strongly intersect and that there is no consensus among forecasters on when to assign them given the current definitions. Obviously, a clear distinction is necessary to improve consistency between forecasters and simplify communication with the public.

### 4.1 Estimating the factors determining avalanche danger

The current EAWS Matrix , along with its predecessors, rely on the concept of decomposing the complex task of assigning a danger level by breaking this task into smaller components. Such decomposition is expected to increase the accuracy of the final estimate if the estimator is able to make more accurate estimates of the individual components (MacGregor, 2001). Moreover, it is also a logical step to consider all relevant input parameters, which can reliably be estimated in the judgment process. However, there are numerous factors potentially influencing the ability to make accurate estimates, whether concerning individual input factors or the danger level as a whole. These include the environmental predictability of the avalanche conditions, the correspondence between the available data (input) and the avalanche conditions (output), the match between the environment and the forecaster, the reliability of the forecaster acquiring the relevant information, or the skill of the forecaster processing this information (e.g., Stewart and Lusk, 1994; Stewart, 2001).

In the matrix survey, highest danger level across all considered avalanche problems is then communicated for the task was detached from a real forecasting situation as the participating forecaster had to assign combinations of the three factors to danger levels without relating to a specific avalanche situation. Therefore, variations likely resulted primarily from the fact that each forecaster has a slightly different representation of what the terms describing snowpack stability, frequency distribution, and avalanche size mean, and how these relate to the danger level. In contrast, during operational use, all the before-mentioned points come into play, each of which may potentially increase variations and may thus reduce consistency between forecasters. However, in addition to these points, rather vague and imprecise definitions, some of which were only introduced prior to this winter season (i.e., the frequency classes; Tab. 3) , likely also impacted consistency when estimating the axis labels.

The comparably low agreement rate between any two forecasters estimating the factors for exactly the same situation (Sect. **??**)certainly raises questions regarding the reliable estimation of the factors determining avalanche danger, and, consequently, how this unreliability impacts the assignment of the danger level. To make this clearer, imagine that two forecasters assess exactly the same avalanche conditions and that their assessments agree for two of the three factors, and differ for one of the factors by one neighboring class. Supposing this kind of disagreement is the normal case, and testing this for all possible combinations in the Matrix, $D$ would differ between the two forecasters in 42% of cases if $D^1$ would be selected. Assuming that forecasters become more consistent with continued use, say two forecasters differ on one of the factors 50% of the time while agreeing in all other situations, the resulting $D^1$ would still differ 21% of the time. As only one of the two forecaster's estimate can be correct and hypothesizing equal competence for both forecasters, such inconsistency in the assessment of factor classes inevitably reduces accuracy (Techel et al., 2024a). In the given example, accuracy would be at most 0.89 - the factor required to achieve an agreement rate between two forecasters of 0.79, solely due to this inconsistency (Techel et al., 2024a).

In Switzerland, forecasters historically did not use the Matrix nor were they constrained to select from a predefined set of classes to describe the factors determining avalanche danger. Instead, a diverse range of words describing these factors (see also Hutter et al., 2021) were employed in daily forecaster discussions, where the team collectively determined the danger level. Therefore, the introduction of a short list of classes for stability and frequency distribution, and forcing forecasters to use these classes, likely contributed to the low agreement rates (stability: 59%, frequency: 63%). This stands in contrast to Swiss forecasters achieving significantly higher pairwise agreement rates when selecting an avalanche problem ($\geq 74\%$, $N = 15000$, unpublished data) or the danger level (87%, Techel et al., 2024a). Interestingly, in Norway, where forecasters have long used this system (Müller et al., 2016) and where the factors are publicly communicated in avalanche forecasts (Engeset et al., 2018) , the agreement rates were similarly low ($\leq 54\%$) . As observed in Switzerland, the agreement on the danger level is generally higher (76%, Techel et al., 2024a). However, the Norwegian numbers are only based on three situations with comparatively dynamic weather and avalanche conditions, which may have heightened the decision-making complexity for forecasters. Additionally, Swiss forecasters routinely work in pairs, fostering collaboration and shared decision-making, whereas in most other services, including Norway, forecasters typically work independently, with fewer opportunities for coordination and peer feedback.

Another potential cause impacting a reliable estimation of the factors is linked to the workflow and design of the Matrix, which require a forecaster to choose exactly one single cell in the matrix. However, there may be situations when multiple combinations describe the prevailing or anticipated avalanche conditions equally well, or when a forecaster is uncertain with regard to selecting a specific class. In both these situations, a forecaster has to settle on one option only, which is a much less flexible approach compared to the approach used in North America (Statham et al., 2018), where forecasters can mark several cells . A more informative approach would be to allow forecasters to mark all combinations in the matrix, which are considered relevant for the given situation, rather than mandating a single choice. For example, imagine that *poor* stability

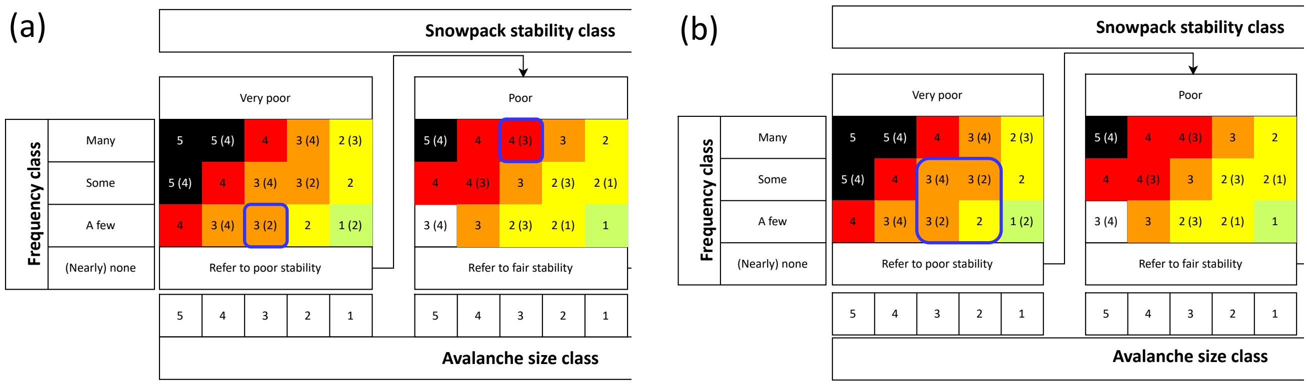


**Figure 4.** EAWS Matrix (extract) with several cells marked as relevant (blue border) to accommodate (a) a range of possible combinations or (b) uncertainty. In (a), two combinations of stability and frequency are considered (*very poor - a few*, *poor - some*), in (b) uncertainty relates to frequency (*a few* or *some*) and avalanche size (*size 2 or 3*). For explanations refer to text.

is expected on *many* slopes with avalanches up to *size 3*. In addition, in *a few* places stability is considered *very poor*, as natural avalanches are expected (Fig. 4a). In this case, warning region. In situations with overlapping avalanche problems, the forecaster would highlight these two cells. A similar approach could also be employed to convey uncertainty (Figure 4b). For example, a forecaster may be confident that stability is *very poor* but may be uncertain whether the frequency of these locations should be considered *a few* or *some*, and whether avalanche size will be size 2 or size 3. In such cases, the respective cells in the matrix could be marked, expressing the uncertainty of these classes. Ideally, more information relevant to assess these factors should be collected to reduce the uncertainty associated with these labels. In both cases, rules for selecting $D$ need to be established, such as selecting the cell resulting in the highest danger level in case of the first scenario (poor-many - 3with $D = 4(3)$, Fig. 4a), or by choosing the median or majority of the indicated $D$ in the second case ($D = 3$, Fig. 4b). In addition, communication with forecast users should convey this information in a clear manner, for instance, by stating that both natural and human-triggered avalanches are possible in scenario 1. issued danger level may exceed that of the individual problems. This structured approach ensures that all relevant factors are systematically evaluated, promoting consistency in the assessment of avalanche danger levels.

By imposing a single discrete choice on forecasters, uncertainty inherent in the assessment process may inadvertently be masked. Moreover, forcing forecasters to express their detailed judgments by forcing them to choose from a small number of discrete choices may violate the basic maxim of forecasting postulated by Murphy (1993, p. 282), namely that a forecast should always correspond to a forecasters best judgement. Assume, for example, that a forecaster expects a range of avalanche conditions within the forecast domain, but that spatial variations cannot be expressed with this level of detail in the forecast. This results in a mismatch between the forecasters best judgement and the forecast, requiring the forecaster to simplify and deviate from the best-possible estimate, inevitably impacting the accuracy of the forecast (e.g., Murphy (1993), Techel (2020, p. 72-74) ). Thus, to reduce a potential mismatch between the forecasters best judgment and the forecast, an alternative way could be to express the tendency within a class or the uncertainty between choosing one of two options by either using a finer-grained scale (more classes, more spatial units) or by allowing for sub-categories within existing classes. For instance, a forecaster might choose a class, such as *a few*, but indicate a tendency toward the next class, such as *some* by using *a few→some*. Using more classes competes with the human capability of reliably being able to choose at most from seven classes (e.g., Miller, 1956; Kahneman et al., . Moreover, each class would need a clear definition. However, more promising is combining absolute and relative judgments (e.g., Kahneman et al., 2021), as this capitalizes on the fact that humans are generally good at making relative rankings within classes established before-hand. Such concepts are already used in avalanche forecasting. For instance, in North America, it is common that avalanche practitioners assess avalanche size using intermediate classes. While there are definitions for avalanches of size

## 5 Revision process – Step 2and size 3 (i.e. Table 4), a size 2.5 is simply in between these classes (e.g., Hafner et al., 2023). Similarly, in Switzerland, sub-levels indicate whether avalanche danger is expected low, in the middle, or high within a level, requiring the forecaster to first make an absolute judgment (: Operational testing

The EAWS Matrix reflects the collective judgment of many European avalanche forecasters, but was initially developed as a desktop exercise—a so-called *cold-state* assessment (Roiser and Sahakian, 2013), as the experts answered the survey outside an operational context and without the emotional or situational pressures of a real forecast setting. In psychology, this distinction refers to cold cognition – reasoning in a calm setting without emotionally charged consequences – versus hot cognition, which involves decision-making under time pressure or stress (e.g., Roiser and Sahakian, 2013; Loewenstein, 2005) . To transition from the conceptual foundation to a reliable tool for operational danger level assessment, we evaluated how the Matrix performed in day-to-day forecasting practice.

We focused on two critical aspects of working with the Matrix, presented in two separate studies: one assessed the reliability with which forecasters could estimate the Matrix input factors (Techel et al., 2024a), as reliable input estimation is a prerequisite for meaningful use of the Matrix; the other examined how the Matrix was applied during real-time operational forecasting across Europe (Techel et al., 2025). Below, we summarize the key findings most relevant for improving the Matrix and its associated workflow, based primarily on data collected from 26 European warning services over one to three forecasting seasons, following the Matrix's initial release in 2022 (EAWS, 2022c).

The operational analysis presented in Techel et al. (2025) revealed two main insights, both of which contributed directly to the revised Matrix presented in Figure 5. First, several Matrix cells were rarely or never used in operational practice ($< 1\%$ of cases), meaning the danger levels assigned to these cells could not be empirically validated. Notably, these same cells corresponded to cells with low support ($\leq 80\%$) in the expert survey (Figure 2b). To reflect the resulting uncertainty, these six cells (labeled *e–g* in Figure 5) are now white in the revised Matrix. Second, the analysis showed that some Matrix cells were used almost exclusively for a single danger level, even though the Matrix had displayed two levels (with one in brackets). To reflect this stronger operational consensus, we removed the bracketed danger level from the four affected cells (marked *a–d* in Figure 5).

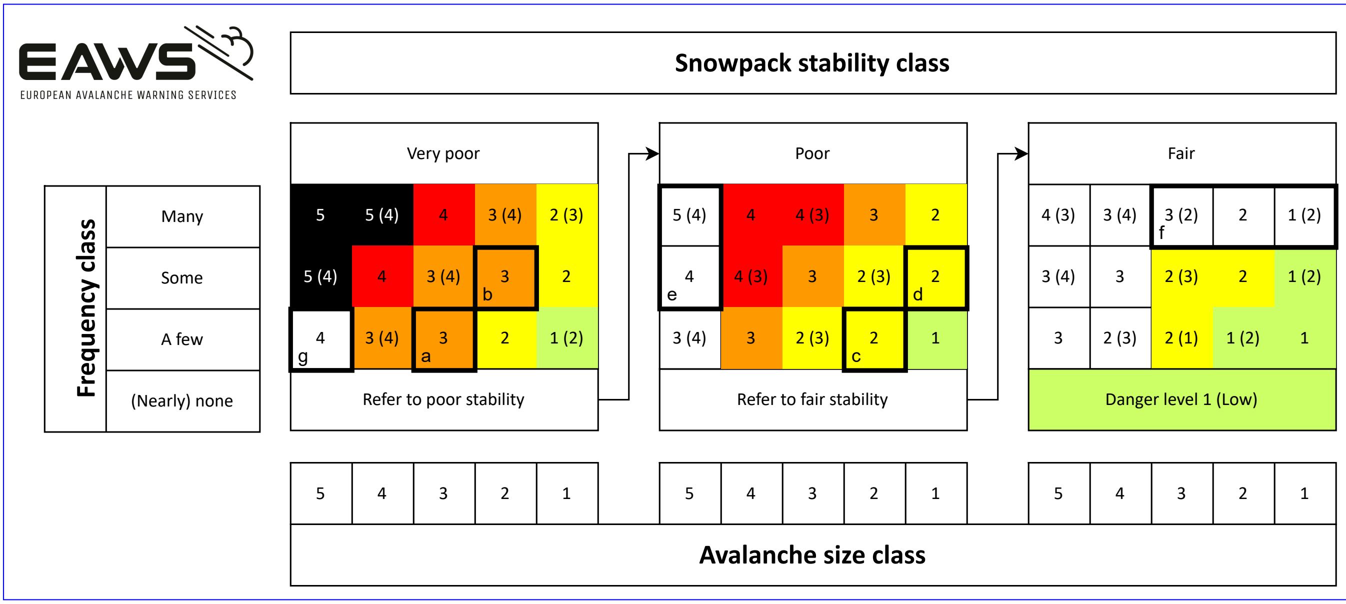


**Figure 5.** Revised EAWS Matrix integrating insights from Techel et al. (2025). Cells *a–d* consistently associated with a single danger level ($D$~~), and then a relative ranking (Techel et al., 2022; Lucas et al., 2023).~~) during operational use and therefore no longer display an alternative level ($D^2$). Cells *e–g*, which received limited support in the expert survey and were not or rarely selected during operational forecasting, are shaded white to highlight elevated uncertainty and the lack of consensus among European avalanche forecasters.

## 6 Discussion

The main purpose of the EAWS Matrix is to enhance consistency between individual forecasters and warning services. To ensure a consistent and practical application of the EAWS Matrix, it is crucial that the definitions for the factors are clear and easily applicable in an operational setting. For instance, avalanche size is defined based on physical measurements, such as volume or mass, or by the destructive potential of the avalanche. Similarly, the definition of snowpack stability is closely linked to observable triggering mechanisms (e.g., explosives or a skier). Thus, both avalanche size and snowpack stability can be assessed using observational data. These definitions, grounded in physical evidence, also support clearer mental imagery – that is, forecasters can more readily form vivid internal representations of what a size 3 avalanche or *very poor* stability might look like, even in the absence of direct observation. In contrast, it is difficult to unambiguously define frequency classes, particularly when considering that the frequency of locations with *very poor* or *poor* stability is generally low. Moreover, the number of potentially unstable locations must often be inferred from sparse observations~~and, increasingly, from models~~, though, models may increasingly provide this information (Herla et al., 2024; Techel et al., 2024b). When relevant data is limited or unevenly distributed within a region, the uncertainty of the assessment increases. Given clear definitions, it would be crucial that all forecasters possess a sound understanding of and consistently adhere to these. However, even if definitions for all factors and their classes were clear, inconsistencies cannot be completely eliminated. The interpretation of current conditions, often based on limited observations, combined with the inherent uncertainties in numerical weather prediction models, will inevitably lead to slight variations in interpretation among forecasters. Thus, inconsistency also becomes a function of data availability and reliability.

### 6.1 Towards an updated avalanche danger scale

The EAWS Matrix and related look-up tables have undergone continuous development over the past two decades. The European avalanche danger scale, in contrast, has remained unchanged since 1993. As a result, EADS and EAWS Matrix, initially closely linked, have gradually diverged. To ensure a unified and coherent framework for avalanche danger assessment and communication, it is essential to reestablish a robust connection between them.

The survey on the EAWS Matrix and data on its operational usage provide valuable insights on how avalanche danger levels are typically described in forecast products. Drawing on this information, the connection between the EADS and the EAWS Matrix can be reestablished by updating the EADS using the terms typically chosen in the Matrix to describe the typical character of every avalanche danger levelImproving the reliability of factor assessments is essential for consistent and accurate use of the EAWS Matrix (Techel et al., 2024a). This can be supported by: (1) increasing the availability of relevant data, such as short-term snowpack simulations to reduce uncertainty; (2) strengthening forecasters' skills in information retrieval and interpretation through targeted training and regular operational exchange; and (3) aligning forecast resolution – spatially, temporally, and categorically – with the resolution of forecasters' assessments given the available data. It is also important to regularly compare how different forecasters assess the factors for the same conditions. If their assessments vary widely, consistency can be improved through focused training and by combining input from several forecasters, making use of the "wisdom of crowds" to increase overall accuracy (Techel et al., 2024a).

Figure ?? showed that there is a high level of agreement among warning services regarding the typical description of the danger levels using the factors. Considering the median value as a representative and consensus-oriented approximation to describe a factor or factor combination for a specific danger level , we summarized the warning service specific descriptions of avalanche danger (Fig. ??)in Table ??. For comparison, we also show the respective combinations for the EAWS Matrix and EADS. As can be seen, there are some differences in what is the typical class for each factor for each danger level across all three sources. $D = 4$ shows the least variation. For $D = 1$ Currently, about half of the Matrix cells do not show full consensus and still include a secondary danger level in brackets ($D^2$). While many of these cells are rarely used, two stand out as being frequently selected for two adjacent danger levels—both in the survey and in operational practice: *poor–some–size 2* and $D = 2$ we see the previously mentioned tendency to emphasize *poor* stability, combined with either an avalanche *very poor–some–size 3* (Techel et al., 2025). Showing two danger levels thus reflects a persistent ambiguity. Operational testing, however, did not clarify the reasons behind this ambiguity or offer clear strategies for resolving it. Particularly in warning services where forecasters more frequently chose $D^2$ over $D^1$, *poor–some–size 2* was often also linked to 3 (considerable). In contrast, other services assigned this danger level primarily when conditions were assessed as having either lower stability (*size 1very poor*or the frequency class a few, instead of ), higher frequency (*fairmany*stability as in the EADS. $D = 3$ has a tendency to *size 2* avalanches in the usage data compared to the EADS and the Matrix. The differentiation between $D = 2$ and $D = 3$ is substantially less pronounced in the usage data compared to the Matrix and especially to the EADS.

If factors were used similarly by European forecasters and if the distribution of cases within a danger level were approximately balanced in the usage data, it would be possible to derive an updated danger scale based on usage patterns. However, as seen in data from Switzerland (Sect.**??**), cases tended to be more frequent at the respective lower end of danger levels 3-considerable and 4-high, biasing the usage data towards the lower end of these danger levels. Moreover, it is important to recognize that winter seasons vary in terms of snowfall, the frequency and duration of cold spells, and other meteorological factors. Consequently, these variations lead to differences in avalanche conditions, the severity of avalanche problems, and ultimately the assessment of avalanche danger. Our usage data is based on a single winter season, which may have been influenced by the unique characteristics of that particular season. Thus, data from more seasons including information on sub-classes of the factors and sub-levels of $D$ need to be analyzed before an updated danger scale can be derived based on the usage data.

Characterization of danger levels as described in the EADS and EAWS Matrix, and the most frequent combinations used during the season 2022/2023. The latter summarizes the results shown in Figure **??** for the seven groups of warning services, with the most frequent factor shown in the column *individual* and the most frequent combination of factors in the column *combined*. The first value indicates the most frequently used class or combination, values in brackets indicate if a second class or combination was associated with a danger level more than 30% of the time. $D$ stab freq size stab freq size stab freq size stab freq ), or larger avalanche size (*size* *1 (low) F (or P) Fe 1 F (or P) Fe 1 F Fe 1 F Fe 1 or 2 2 (moderate) P So (or Fe) 2 P Fe (or So) 2 P or F So (or Fe) 2 F So n.d. , <4 3 (considerable) P So 3 P (or VP) So 2 (or 3) P So 3 P So or Ma ≤ 3 or 4 4 (high) VP Ma 3 VP Ma 3 VP Ma 3 or 4 VP (or P) Ma* *3or 4 5 (very high) – – –* ) (Techel et al., 2025). This highlights a core challenge: the Matrix's structure itself influences how it is applied – particularly among forecasters who tend to stick with the primary danger level. As such, further targeted discussion and investigation of these transition zones is critical to refine guidance and reduce interpretive variability in Matrix use.

It is essential that forecasters can express uncertainty inherent in the forecasting task when working with the Matrix. As suggested in Sec. 4, this may include the option to select multiple cells. To facilitate discussion in case of variations, forecasters could indicate where within a class their assessment lies – – VP Ma 5 VP Ma 4 or 5 toward the lower or higher end, or near the center (e.g., as in Switzerland Lucas et al., 2023). Indicating the relative trends within a factor class could help identify patterns, inform consensus-building, and ultimately highlight whether further subdivision of broad classes like *some* is warranted.

## 6.1 Recommendations for practice and ways forward

## 6.1 Relation to Conceptual Model of Avalanche Hazard (CMAH)

The CMAH provides a systematic framework for avalanche hazard assessment by addressing avalanche problems, their location, the associated likelihood of avalanches, and their size. It has a broad scope and was designed to serve various avalanche operations from back-country guiding to road safety to regional public forecasting. In contrast, the EAWS Matrix was developed specifically to standardize regional public avalanche forecasts across Europe.

~~We identified areas for improvement in the current standards and definitions used in avalanche forecasting (Sect. 2) but also with regard to~~ While both the CMAH and the ~~operational use of the Matrix. The latter relate primarily to~~ EAWS Matrix aim to support consistent and reliable hazard assessments, they handle the spatial component differently. The CMAH focuses on the ~~inconsistent assignment of factor classes but also to the design of the Matrix. From these,~~ spatial distribution of the avalanche problem. The EAWS Matrix, however, centers on the ~~following recommendations and potential areas for further development emerge with the goal to reduce the variations due to an individual forecaster or warning service preparing a forecast, and, thus, to ultimately lead to better forecast products for forecast users:~~

~~**Increase reliability of estimation of factors**: Particularly in warning services where forecasters tend to work primarily by themselves, regular training sessions are essential to foster a common understanding of factor categories. This will permit calibration in the use of~~ frequency of locations with the weakest snowpack stability – an approach that reflects empirical findings that the frequency of the lowest stability class is often the most decisive factor when determining the danger level (Techel et al., 2020a). Importantly, these two concepts – spatial distribution of avalanche problem and frequency of snowpack stability – do not always align. Presumably, it is generally easier to assess where the avalanche problem exists in the terrain than it is to assess the frequency of the ~~categories between forecasters. Ideally, such exchanges would also occur across warning service boundaries, to develop and maintain a common understanding of categories. **Adjust size of warning regions**: Spatial units for regional forecasting should be of comparable size and fall within a standardized range of square kilometers, ensuring consistency. Their boundaries should be determined based on the availability and density of relevant data. **Evaluate methodologies to estimate factors**: It should be evaluated whether choosing multiple cells in the matrix (as discussed in Sect. **??** or as used , or whether continuous sliders or sub-categories are alternative and more effective ways to factor estimation compared to choosing a single category from a small number of classes (concept currently used in the EAWS Matrix). **Evaluate alternative Matrix layouts**: Investigate different matrix layouts, such as used in ADAM (Müller et al., 2016, see also Fig. A2) or the data-driven matrix by Techel et al. (2020a, see also Fig. A4), to determine if they offer advantages over the current design of the EAWS Matrix . **Revise Matrix based on data**:~~ different classes of snowpack stability. Thus, simply translating the spatial distribution of an avalanche problem to the frequency of potential triggering locations given a specific trigger, can lead to a mismatch likely overestimating the frequency term and, consequently, leading towards higher danger ratings. Whether the reliability of assessing the spatial distribution or the frequency class is higher has not yet been investigated. However, in forecasts, often a combination is used (Hutter et al., 2021): frequency descriptors are often used to describe the number of triggering locations in a region and can be linked to specific danger levels. However, they are commonly paired with location-specific narrative, helping forecasters convey both *how often* avalanches may occur and *where* within the terrain they are most likely.

Despite these differences, the Matrix and CMAH are complementary. Several European services where forecasters are trained primarily using the CMAH – such as in Sweden and Scotland – have successfully integrated the Matrix into their operations. However, it is important to emphasize that neither CMAH nor the Matrix in its current state with a $D^2$ should be applied rigidly. Avalanche forecasting is an iterative process carried out under significant uncertainty, often with sparse or ambiguous data. The Matrix is based on expert estimates. However, data-driven analyses describing danger levels using a variety of relevant data sources should be leveraged, as for instance the comprehensive study by Techel et al. (2022), which provides insights into how observations and model predictions relate to forecast danger levels. Additionally, data sources such as avalanche detection data from satellites or terrestrial systems (e.g., Eckerstorfer et al., 2017; Mayer et al., 2020) or snowpack modelling combined with machine-learning approaches (e.g., Techel et al., 2022; Mayer et al., 2023) may be suitable to obtain intended to support structured thinking and reduce inconsistency in similar conditions. Its design has been shaped by the CMAH's structured approach, adapted to the operational needs of regional forecasting in Europe.

As discussed in Section 4, the Matrix guides forecasters from left (very poor stability) to right (fair stability), reflecting the central role of the weakest stability class in determining $D$ (Techel et al., 2020a). Conceptually, the Matrix can be seen as a variant of the CMAH hazard chart. In this interpretation, snowpack stability and its frequency jointly represent the CMAH dimension of the *likelihood of avalanches*, which is then combined with *avalanche size*. This approach follows the alternative layout proposed as ADAM (avalanche danger assessment matrix) by Müller et al. (2016, , see also FigureA2) and further developed in the data-driven versions of the Matrix. **Reduce the subjective component of avalanche forecasting**: Currently, the quality of avalanche forecasts relies strongly on the competence and experience of a human forecaster interpreting a variety of heterogeneous data. However, with the increasing availability of reliable, highly-resolved models in combination with state-of-the art data analysis techniques (e.g., Herla et al., 2022), forecasting should become more data- Matrix by Techel et al. (2020a, .

Figure 6 shows the EAWS Matrix rearranged according to the CMAH workflow. This layout facilitates comparisons across stability–frequency combinations and highlights opportunities for simplification, such as merging rows with near-identical danger ratings. A compacted version is presented in Figure 7.

Rearranging the Matrix according to the ADAM design aligns it with the classic structure of hazard matrices, where hazard increases from the lower left to the upper right corner (Duijm, 2015). This makes the connection to CMAH more visible and intuitive. The workflow also becomes closer to CMAH logic: first determine the likelihood of avalanches (through stability and frequency), then assess consequences (avalanche size). The layout also reveals structural improvements – such as merging similar rows – and highlights clusters of rarely used or unsupported combinations. Notably, problematic combinations like *very poor–some–size 3* and model-driven. This may be achieved by combining model predictions and the forecaster's best judgment using hybrid intelligence methods (Dellermann et al., 2019). Moreover, using historic data and machine-learning approaches, forecaster decisions may be directly modeled (Pérez-Guillén et al., 2022; Maissen et al., 2024), providing the forecaster with a *poor–some–size 2* appear adjacent in this layout and span danger levels 2, 3, and 4, highlighting the continuous nature of avalanche danger and emphasizing the need for refinement. Compared to earlier matrices, specifically the ADAM matrix (Fig.A2) and the data-driven second opinion when choosing a danger level (Winkler et al., 2024). **Validate the correctness of forecast factor categories and danger levels**: Although difficult to achieve due to a general lack of relevant data, the evaluation of the performance of avalanche forecasts should become standard practice. Such evaluations should go beyond the evaluation of the correctness of forecast danger levels (as done in several previous studies, e.g., Techel and Schweizer, 2017; Logan and Greene, 2023

**(a)** Frequency of snowpack stability

| Stability Matrix | | Frequency class | | |
|---|---|---|---|---|
| | | Few | Some | Many |
| Snowpack stability class | Very poor | VpFe | VpSo | VpMa |
| | Poor | PFe | PSo | PMa |
| | Fair | FFe | PSo | FMa |

**(b)** Sorted by lowest stability class

| Danger Matrix | | Avalanche size class | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| From stability matrix | VpMa | 2 (3) | 3 (4) | 4 | 5 (4) | 5 |
| | VpSo | 2 | 3 | 3 (4) | 4 | 5 (4) |
| | VpFe | 1 (2) | 2 | 3 | 3 (4) | 4 |
| | PMa | 2 | 3 | 4 (3) | 4 | 5 (4) |
| | PSo | 2 | 2 (3) | 3 | 4 (3) | 4 |
| | PFe | 1 | 2 | 2 (3) | 3 | 3 (4) |
| | FMa | 1 (2) | 2 | 3 (2) | 3 (4) | 4 (3) |
| | FSo | 1 (2) | 2 | 2 (3) | 3 | 3 (4) |
| | FFe | 1 | 1 (2) | 2 (1) | 2 (3) | 3 |

**(c)** Sorted by resulting danger level

| Danger Matrix | | Avalanche size class | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | |
| From stability martix | VpMa | 2 (3) | 3 (4) | 4 | 5 (4) | 5 | |
| | PMa | 2 | 3 | 4 (3) | 4 | 5 (4) | Very similar |
| | VpSo | 2 | 3 | 3 (4) | 4 | 5 (4) | |
| | PSo | 2 | 2 (3) | 3 | 4 (3) | 4 | |
| | VpFe | 1 (2) | 2 | 3 | 3 (4) | 4 | Very similar |
| | FMa | 1 (2) | 2 | 3 (2) | 3 (4) | 4 (3) | |
| | FSo | 1 (2) | 2 | 2 (3) | 3 | 3 (4) | Very similar |
| | PFe | 1 | 2 | 2 (3) | 3 | 3 (4) | |
| | FFe | 1 | 1 (2) | 2 (1) | 2 (3) | 3 | |

**Figure 6.** EAWS Matrix (Fig. 5) rearranged to follow the CMAH workflow. Panel (a) shows a matrix of snowpack stability and frequency combinations, representing the *likelihood of avalanches* as defined in the CMAH. In panel (b), these combinations are paired with avalanche size and reordered by increasing snowpack stability. Panel (c) then sorts the same rows by decreasing resulting danger level (*D*), allowing visual identification of rows that produce very similar danger ratings. Rows where the maximum difference across all avalanche sizes is less than one danger level are marked with grey triangles and labeled "very similar." These rows can be considered for merging without significant loss of information, as demonstrated in the compact Matrix shown in Figure 7.

**(a)**

| Stability Matrix | | Frequency class | | |
|---|---|---|---|---|
| | | Few | Some | Many |
| Snowpack stability class | Very poor | D | B | A |
| | Poor | E | C | B |
| | Fair | F | E | D |

**(b)**

| Danger Matrix | | Avalanche size class | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| From stability matrix | A | 2 (3) | 3 (4) | 4 | 5 (4) | 5 |
| | B | 2 | 3 | 4 (3) | 4 | 5 (4) |
| | C | 2 | 2 (3) | 3 | 4 (3) | 4 |
| | D | 1 (2) | 2 | 3 | 3 (4) | 4 |
| | E | 1 | 2 | 2 (3) | 3 | 3 (4) |
| | F | 1 | 1 (2) | 2 (1) | 2 (3) | 3 |

**Figure 7.** Compact version of the EAWS Matrix layout shown in Fig. 6 panels (a) and (c). Rows highlighted in Figure 6c as being similar have been merged and are given the same letter, which can be transferred to (b) when combining with avalanche size.

, but should consider all published forecast parameters as emphasized by Lucas et al. (2023). As validation data is scarce, data sources like automated avalanche detection data or snowpack modelling should be explored. Such evaluations should be ongoing to maintain forecast quality, but, importantly, these should be made before introducing new parameters in avalanche forecasts. **Update avalanche danger scale**: Clearly, congruence between definitions, workflow, EAWS Matrix and the EADS is of utmost importance.Therefore, the EADS should be updated using available data, including usage data as described in Section **??**.**Multiple cells**: we recommend to assign multiple cells where applicable for describing avalanche conditions to account for complexity and uncertainty and avert from the practice of enforcing the assignment of a single combination. matrix (Fig.A4), the rearranged EAWS Matrix (Fig. 7) offers greater detail and is fully compatible with current EAWS standards.

A trade-off of this compact representation is that it obscures which specific stability–frequency combination led to a given danger level on the right hand side of the chart (Fig. 7). However, this information could still be tracked or annotated during operational use.

## 7 Conclusions

Public avalanche forecasting involves both the assessment of the requires both a robust assessment of current and future avalanche conditions as well as their communication to the public. Avalanche forecasting is currently of categorical nature (e.g. factorsconditions and effective communication of the associated hazard. Today, this process is largely categorical in nature—using defined factor classes, avalanche problemsand danger level), and danger levels. The quality of a forecast depends and consistency of these forecasts depend not only on data availability , the skill of the forecaster , and the definition and and forecaster expertise, but also on the clarity, applicability, and shared understanding of the categories used. In this study we investigated the latter by evaluating themselves.

In this paper, we presented the revision process behind the updated EAWS Matrixand its associated factors and ascertain the extent of consistency in their understanding and application.

The updated EAWS Matrix , as presented in Section**??**, represents the latest consensus among European avalanche forecasters . The EAWS Matrix has been favorably received as a valuable tool in operational avalanche forecasting and in forecaster training . It offers a standardized framework for evaluating avalanche danger by linking , involving broad engagement from European avalanche warning services. The updated Matrix (Section 5) reflects the most recent consensus among forecasters across Europe. It has been well received as both an operational forecasting tool and a training aid, offering a structured framework for determining avalanche danger levels ($D$) based on three key factors: snowpack stability, frequency distribution the frequency of snowpack stability, and avalanche size, to the danger level ($D$). .

The introduction of a second secondary danger level ($D^2$) acknowledges the remaining inconsistencies among avalanche forecasters and combinations with limited response rates($\leq$70%) in the matrix survey signify rare or unconventional scenarios, as perceived by most forecasters (Figure 3).

Our analysis of the EAWS Matrix's implementation by 15 European avalanche warning services across eight countries during its inaugural winter unveiled a generally consistent pattern in class assignments to each danger level.However, consistency among forecasters when choosing individual factors is currently relatively low, necessitating refinements in definitions and the workflow for avalanche danger assessment. In addition, training sessions are essential to improve consistency and enhance forecaster skills in applying the EAWS Matrix effectively. Diverging trends between services currently still exist, leading to an overlap of factors used at different danger levels.Notably, distinguishing between danger levels 2 in selected cells recognizes that uncertainty and disagreement still persist, especially in combinations with low survey response rates, different perceptions, or inconsistent operational use. Particularly noteworthy are the cells *very poor–some–size 3* and *poor–some–size 2*, which frequently span two danger levels (i.e., 3posed more challenges compared to levels 1 and 2, and 3 and 4 (no data was available for level 5) . While a revision of the EAWS Matrix after a single winter is challenging due to seasonal variations, our goal is to find a consensus that suggests one danger level per cell to reduce inconsistency to a minimum in the future–4 and 2–3, respectively). These cases highlight the need for better definitions – especially of the frequency class *some* – and call for closer examination of transition zones between danger levels.

Improving the consistency of factor assessments will require progress on several fronts: increasing the availability of relevant data (e.g., high-resolution snowpack simulations), enhancing forecaster training and operational exchange, and aligning forecast resolution with the granularity of expert judgment. Evaluating and improving inter-rater reliability (IRR) is critical for ensuring forecast quality. Where IRR is low, interventions such as targeted training and forecast aggregation can improve consistency.

When applying the EAWS Matrix The Matrix should not be applied rigidly. Instead, we recommend to assign multiple cells where applicable for describing avalanche conditionsto account for complexity and uncertainty and avert from the practice of enforcing the assignment of a single combination.Further, considering categories on a more nuanced scale can contribute to a more consistent assessment of avalanche danger level, provided that classes can be clearly defined and assessed. Especially a better understanding and more accurate measurement of frequency distribution of snowpack stability are crucial for improved consistencyallowing forecasters to select multiple cells when appropriate to reflect complex or uncertain conditions. Additionally, enabling forecasters to indicate where within a class their assessment lies (e.g., near the lower or upper end of "some") may support more nuanced and transparent decision-making.

CurrentlyWhile originally developed as a standalone tool, the EAWS Matrix closely aligns with the Conceptual Model of Avalanche Hazard (CMAH). Its structure and logic can be visualized in an ADAM-style layout – mirroring classic risk matrices – where likelihood and potential impact are considered sequentially. Such a redesign could simplify the Matrix, clarify transition zones, and emphasize its conceptual link to the CMAH.

At present, the updated definitionsMatrix, workflow, and EAWS Matrix presented by Müller et al. (2023) represent a departure from the European avalanche danger scale. Using repeatedly collected data on matrix usage, along with ongoing discussions and training among European avalanche forecasters, can provide a robust foundation for developing an updated avalanche danger scale incorporating the defined termsterminology described here—and formally documented in EAWS guidelines (EAWS, 2022c)—diverge from the current European Avalanche Danger Scale (EADS). A revision of the EADS that incorporates these updated definitions, structures, and principles is overdue. The Matrix and its accompanying tools represent a significant step forward in harmonizing hazard assessment and risk communication across Europe, and ongoing monitoring, data collection, and collaborative refinement will be key to further progress.

## Appendix A: Avalanche Danger Matrices

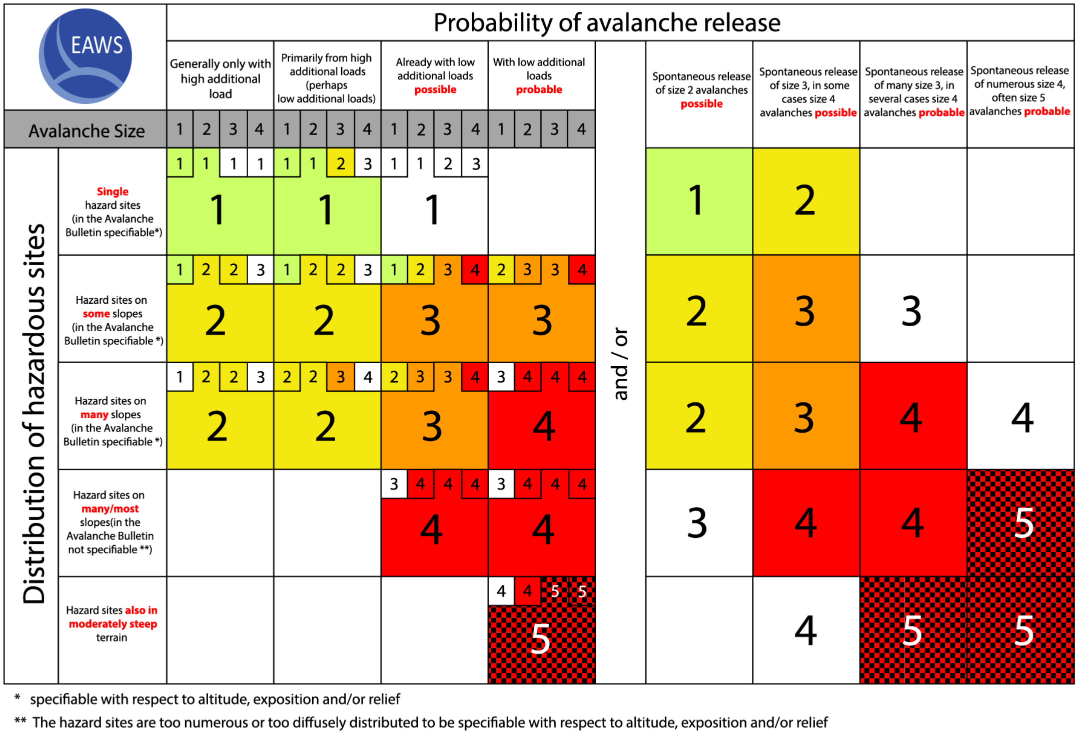


**Figure A1.** The figure depicts the EAWS-Matrix-v2017 (figure taken from EAWS, 2017), which is essentially identical to the Bavarian Matrix (BM EAWS, 2005) when considering the large cells only. The BM illustrates the distribution of hazardous zones against the probability of avalanche release. Initially, the left-hand side, which dealt with artificially triggered avalanches, lacked the factor of avalanche size entirely. Meanwhile, the right-hand side focused on naturally triggered avalanches, providing indications of expected sizes in the column headings. However, an update in 2017 integrated avalanche size (small cells) into the left-hand side as a third dimension, leading to its transformation into the EAWS-Matrix-v2017 (EAWS, 2017).

## Appendix B: CMAH hazard chart

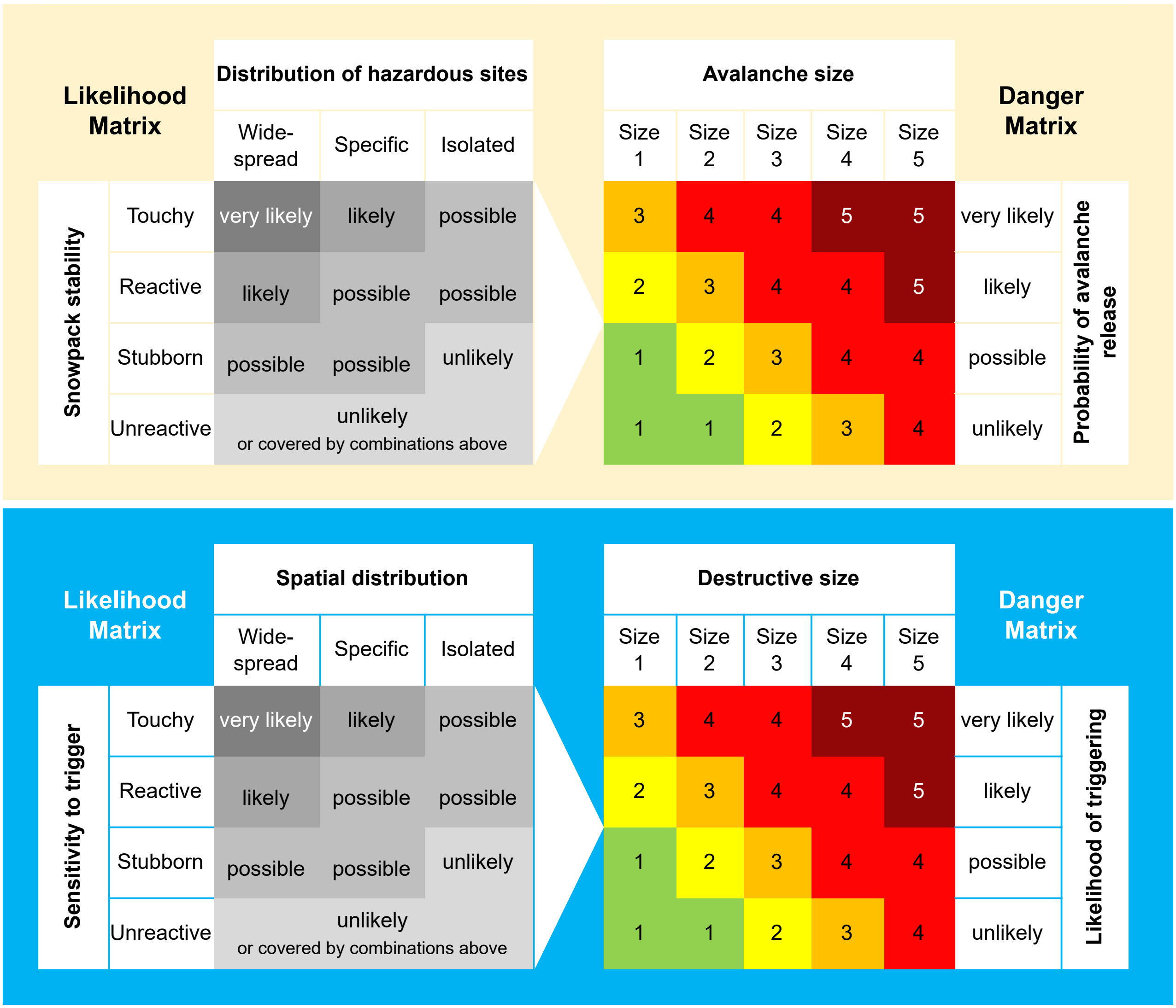


**Figure A2.** The Avalanche Danger Assessment Matrix (ADAM), as published by Müller et al. (2016), presents two versions: one aligning with the terminology of the European Avalanche Danger Scale (EADS, Table 1) at the top and another adhering to the Conceptual Model of Avalanche Hazard (CMAH) at the bottom. ADAM consists of a likelihood matrix (left-hand side), which defines likelihood terms based on the spatial distribution and snowpack stability (sensitivity to trigger), and the Danger Matrix (right-hand side), which provides guidelines for determining the appropriate danger level by combining the likelihood of triggering with avalanche size.

| | Size D1 | Size D2 | Size D3 | Size D4 | Size D5 |
|---|---|---|---|---|---|
| Strong (> 30%) | 1 or 2 | 3 | 4 | 5 | 5 |
| Good (10 to 30%) | 1 | 2 or 3 | 3 or 4 | 4 | 5 |
| Fair (1 to 10%) | 1 | 1 or 2 | 3 | 4 | 4 |
| Small (1 to 3%) | 1 | 1 | 2 | 3 | 4 |
| Slight (< 1%) | 1 | 1 | 1 | 2 | 3 |

FIG. 2: GUIDANCE FOR COMBINING LIKELIHOOD OF AVALANCHES WITH AVALANCHE SIZE TO ASSIGN AVALANCHE HAZARD RATINGS (AFTER MULLER ET AL., 2016A; CLARK AND HAEGELI, 2018).

**Figure A3.** The likelihood matrix proposed by Thumlert et al. (2020) addresses concerns highlighted in their survey, revealing a broad spectrum of interpretations of likelihood terms outlined in the Conceptual Model of Avalanche Hazard (CMAH, Statham et al. (2018)) among avalanche professionals. The aim was to move away from likelihood terms commonly linked with higher probabilities and instead introduce levels of chance paired with a percentage range. The resulting likelihood matrix shown here bears resemblance to the danger matrix on the right-hand side of ADAM Müller et al. (2016). Intermediate levels are recommended for situations with higher chances and smaller avalanches.

*a) stability matrix*

| snowpack stability | frequency: none* | few | several | many |
|---|---|---|---|---|
| very poor | ** | D | B | A |
| poor | ** | E | D | C |
| fair | - | - | E | E |
| good | - | - | - | - |

* *none or nearly none*

** *if none, refer to next higher stability class*

\- *no data*

(hatched) C *cell contains less than 1% of the data*

*b) danger matrix*

| stability matrix | largest avalanche size: 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| A | **3** -4 | **4** (-3) | **4** | **4** |
| B | **3** (-2/-1) | **3** (-2) | **3** (-2) | **4** -3 |
| C | **2** (-3) | **2** -3 | **3** -2 | - |
| D | **1** -2 | **2** -1 | **2** -1 | **3** (-2) |
| E | **1** | **1** (-2) | **1** (-2) | - |

*-3: >30%*

*(-3): 15-30%*

**Figure A4.** Data-driven look-up table for avalanche danger assessment (figure and caption taken from Techel et al., 2020a). The (a) *stability matrix* combines the frequency class of the most unfavorable snowpack stability class (columns) and the snowpack stability class (rows) to obtain a letter describing specific stability situations, the (b) *danger matrix* combines the largest avalanche size (columns) and the specific stability situations (letter) obtained in the stability matrix (rows) to assess the danger level. In (b): The most frequent danger level is shown in bold. If the second most frequent danger level was present more than 30% of the cases, the value is shown with no brackets, if present between 15 and 30% it is placed in brackets. In (a) and (b): Cells containing less than 1% of the data are marked.

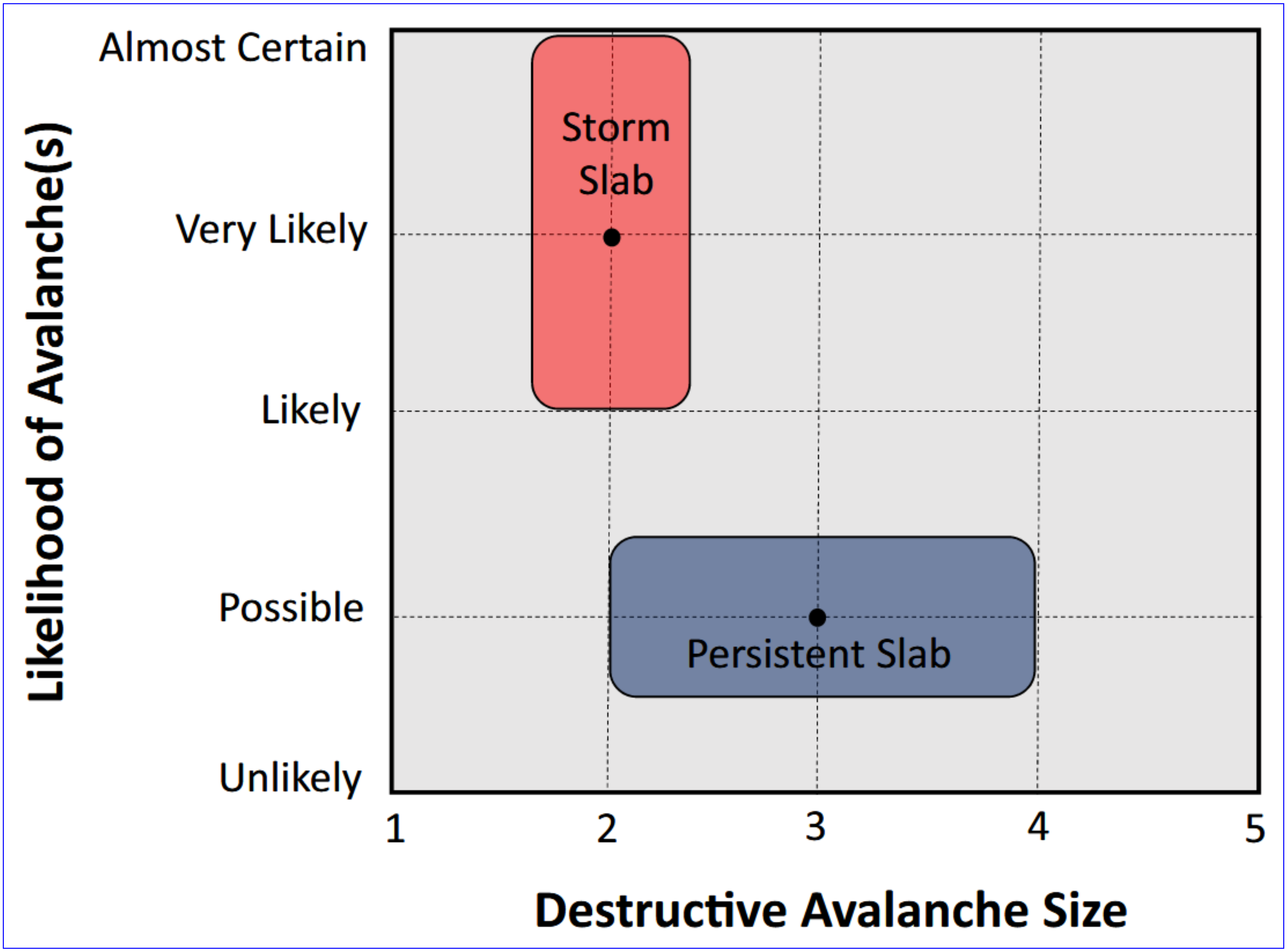


**Figure B1.** Avalanche hazard chart illustrating two avalanche problems, adapted from Statham et al. (2018, Fig. 3). The y-axis represents increasing *likelihood of avalanches*, while the x-axis indicates increasing *avalanche size*. In this example, persistent slab avalanches are considered *possible* with sizes ranging from 2 to 4, whereas storm slab avalanches are assessed as *likely to almost certain*, predominantly around size 2.

*Code and data availability.* Data and code will be published at the repository envidat.org, and will also be indexed at https://opendata.swiss/en/.

*Author contributions.* KM (project lead, study design, data curation, writing, reviewing), FT (study design, data curation, formal analysis, writing, reviewing), CM (study design, data curation, writing, reviewing).

*Competing interests.* We declare that they have no conflict of interest.

*Acknowledgements.* We thank Stefano Sofia, Petter Palmgren, Anne Dufour, Nicolas Roux, Fabiano Monti, Giacomo Villa, Guilem Martin Bellido, and Lorenzo Bertranda for invaluable discussions within the EAWS working group *Matrix & Scale*. We thank the EAWS workgroup coordinators Thomas Stucki and Thomas Feistl for their support over the years we were working on the EAWS Matrix. We greatly appreciate the valuable feedback provided by Rune Engeset, Markus Eckerstorfer, Matthias Walcher, and by the two reviewers Grant Statham and Erich Peitzsch. We further thank Editor Pascal Haegeli for his valuable feedback on the initial manuscript, which ultimately led to the decision to separate the conceptual development and the operational analysis into two distinct publications.

## References


Ashton, R. H.: A review and analysis of research on the test–retest reliability of professional judgment, Journal of Behavioral Decision Making, 13, 277–294, https://doi.org/10.1002/1099-0771(200007/09)13:3<277::AID-BDM350>3.0.CO;2-B, 2000.

Binder, M. and Mitterer, C.: INITIALIZING SNOW COVER SIMULATIONS WITH OBSERVED SNOW PROFILES, in: Proceedings International Snow Science Workshop 2023, BEND, OREGON, 2023.

Clark, T.: Exploring the link between the Conceptual Model of Avalanche Hazard and the North American Public Avalanche Danger Scale, Master's thesis, Simon Fraser University, 115 p., 2019.

Clark, T. and Haegeli, P.: Establishing the link between the Conceptual model of avalanche hazard and the North American public avalanche danger scale: initial explorations from Canada, in: Proceedings International Snow Science Workshop, Innsbruck, Austria, 2018.

Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J.: Hybrid Intelligence, Bus Inf Syst Eng, https://doi.org/10.1007/s12599-019-00595-2, 2019.

Dennis, A. and Moore, M.: Evolution of Public Avalanche Information: The North American Experience with Avalanche Danger Rating Levels, in: Proceedings of the 1996 International Snow Science Workshop, Banff, Canada, pp. 60–66, 1996.

Dietrich, F. and Spiekermann, K.: Jury Theorems, in: The Stanford Encyclopedia of Philosophy, edited by Zalta, E. N. and Nodelman, U., Metaphysics Research Lab, Stanford University, Spring 2023 edn., 2023.

Duijm, N. J.: Recommendations on the use and design of risk matrices, Safety science, 76, 21–31, 2015.

EAWS: Bavarian matrix, European Avalanche Warning Services, 2005.

EAWS: EAWS Matrix (version 2017), European Avalanche Warning Services, https://www.avalanches.org/standards/eaws-matrix/, last access: 21 May 2022, 2017.

EAWS: EAWS standards, https://www.avalanches.org/standards/, 2022a.

EAWS: The revised EAWS-Matrix, https://www.avalanches.org/wp-content/uploads/2022/12/eaws_matrix_poster_v1.pdf, presented at EAWS General Assembly, Davos, June 2022; last access: 12 Feb 2024, 2022b.

EAWS: Determination of the avalanche danger level in regional avalanche forecasting, European Avalanche Warning Services, https://www.avalanches.org/wp-content/uploads/2022/12/EAWS_matrix_definitions_EN.pdf, last access: 21 June 2023, 2022c.

EAWS: Fatalities statistics, https://www.avalanches.org/fatalities/fatalities-statistics/, last access: 2023/11/09, 2023a.

EAWS: Standards: European Avalanche Danger Scale, Tech. rep., European Avalanche Warning Services (EAWS), https://www.avalanches.org/wp-content/uploads/2022/09/European_Avalanche_Danger_Scale-EAWS.pdf, last access: 2023/07/05, 2023b.

EAWS: Standards: Information pyramid, https://www.avalanches.org/standards/information-pyramid/, last access: 2023/11/13, 2023c.

Eckerstorfer, M., Malnes, E., and Müller, K.: A complete snow avalanche activity record from a Norwegian forecasting region using Sentinel-1 satellite-radar data, Cold Regions Science and Technology, 144, 2017.

Elder, K. and Armstrong, B.: A quantitative approach for verifying avalanche hazard ratings, in: Symposium at Davos 1986 on Avalanche Formation, Movement and Effects, vol. 162 of *International Association of Hydrological Sciences Publication*, pp. 593 – 603, 1987.

Engeset, R. V., Pfuhl, G., Landrø, M., Mannberg, A., and Hetland, A.: Communicating public avalanche warnings – what works?, Nat Hazards Earth Syst Sci, 18, 2537–2559, https://doi.org/10.5194/nhess-2018-183, 2018.

Föhn, P.: Das Schweizerische Lawinenbulletin - eine Interpretationshilfe für den Benützer (translation: The Swiss avalanche bulletin - guideline for users), Tech. Rep. 38, Eidgenössisches Institut für Schnee- und Lawinenforschung SLF, Weissfluhjoch/Davos, Switzerland, 1985.

Furman, N., Shooter, W., and Schumann, S.: The roles of heuristics, avalanche forecast, and risk propensity in the decision making of backcountry skiers, Leisure Sciences, 32, 453–469, https://doi.org/10.1080/01490400.2010.510967, 2010.

Giraud, G.: MEPRA: an expert system for avalanche risk forecasting, in: Proccedings International Snow Science Workshop ISSW 1992, Breckenridge, CO., USA, pp. 97–106, 1992.

Giraud, G., Lafeuille, J., and Pahaut, E.: Evaluation de la qualité de la prevision du risque d'avalanche, in: Avalanche Formation, Movement and Effects (Proceedings of the Davos Symposium, September 1986), vol. 162, pp. 583–591, 1987.

Hafner, E. D., Techel, F., Daudt, R. C., Wegner, J. D., Schindler, K., and Bühler, Y.: Avalanche size estimation and avalanche outline determination by experts: reliability and implications for practice, EGUsphere, 2023, 1–31, https://doi.org/10.5194/egusphere-2023-586, 2023.

Herla, F., Haegeli, P., and Mair, P.: A data exploration tool for averaging and accessing large data sets of snow stratigraphy profiles useful for avalanche forecasting, The Cryosphere, 16, 3149–3162, https://doi.org/10.5194/tc-16-3149-2022, 2022.

Herla, F., Binder, M., Lanzanasto, N., Perfller, 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, 2024.

Hutter, V., Techel, F., and Purves, R. S.: How is avalanche danger described in textual descriptions in avalanche forecasts in Switzerland? Consistency between forecasters and avalanche danger, Natural Hazards and Earth System Sciences, 21, 3879–3897, https://doi.org/10.5194/nhess-2021-160, 2021.

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

LaChapelle, E.: The fundamental process in conventional avalanche forecasting, Journal of Glaciology, 26, 75–84, https://doi.org/10.3189/S0022143000010601, 1980.

Lazar, B., Trautmann, S., Cooperstein, M., Greene, E., and Birkeland, K.: North American avalanche danger scale: Do backcountry forecasters apply it consistently?, in: Proceedings International Snow Science Workshop, 2–7 October 2016, Breckenridge, Co., pp. 457 – 465, 2016.

Loewenstein, G.: Hot-cold empathy gaps and medical decision making, Health Psychology, 24, S49–S56, https://doi.org/10.1037/0278-6133.24.4.S49, 2005.

Logan, S. and Greene, E.: Assessing the Colorado avalanche information center's backcountry avalanche forecasts, in: Proceedings International Snow Science Workshop, Bend, OR, USA, 2023.

Lucas, C., Trachsel, J., Eberli, M., Grüter, S., Winkler, K., and Techel, F.: Introducing sublevels in the Swiss avalanche forecast, in: International Snow Science Workshop ISSW 2023, Bend, Oregon, USA, 2023.

MacGregor, D.: Principles of forecasting: a handbook for researchers and practitioners, vol. 30 of *International Series in Operations Research & Management Science*, chap. Decomposition for judgmental forecasting and estimation, pp. 107–123, Springer, Boston, MA, https://doi.org/10.1007/978-0-306-47630-3_6, 2001.

Maissen, A., Techel, F., and Volpi, M.: A three-stage model pipeline predicting regional avalanche danger in Switzerland (RAvaFcast v1.0.0): a decision-support tool for operational avalanche forecasting, EGUsphere, 2024, 1–34, https://doi.org/10.5194/egusphere-2023-2948, 2024.

Mayer, S., van Herwijnen, A., Ulivieri, G., and Schweizer, J.: Evaluating the performance of an operational infrasound avalanche detection system at three locations in the Swiss Alps during two winter seasons, Cold Regions Science and Technology, 173, 102 962, https://doi.org/10.1016/j.coldregions.2019.102962, 2020.

Mayer, S., Herwijnen, A., Techel, F., and Schweizer, J.: A random forest model to assess snow instability from simulated snow stratigraphy, The Cryosphere, pp. 4593–4615, https://doi.org/10.5194/tc-16-4593-2022, 2022.

Mayer, S., Techel, F., Schweizer, J., and van Herwijnen, A.: Prediction of natural dry-snow avalanche activity using physics-based snowpack simulations, Nat. Hazards Earth Syst. Sci., 23, 3445—-3465, https://doi.org/10.5194/nhess-23-3445-2023, 2023.

McClung, D.: The elements of applied avalanche forecasting part I: The human issues, Natural Hazards, 26, 111–129, 2002.

Meister, R.: Country-wide avalanche warning in Switzerland, in: Proceedings ISSW 1994. International Snow Science Workshop, 30 Oct - 3 Nov 1994, Snowbird, UT, pp. 58–71, 1995.

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.

Mitterer, C. and Mitterer, L.: 25 Jahre Europäische Lawinengefahrenstufenskala", Bergundsteigen, pp. 67–76, https://bergundsteigen.com/wp-content/uploads/2021/08/67-7625-jahre-europaeische-lawinengefahrenstufenskala.pdf, 2018.

Morgan, A., Haegeli, P., Finn, H., and Mair, P.: A user perspective on the avalanche danger scale – insights from North America, Natural Hazards and Earth System Sciences, 23, 1719–1742, https://doi.org/10.5194/nhess-23-1719-2023, 2023.

Morgan, M. G.: Characterizing, analyzing, and communicating uncertainty, p. 209–243, Cambridge University Press, https://doi.org/10.1017/9781316882665.009, 2017.

Müller, K., Mitterer, C., Engeset, R., Ekker, R., and Kosberg, S.: Combining the conceptual model of avalanche hazard with the Bavarian matrix, in: Proceedings International Snow Science Workshop, 2–7 October 2016, Breckenridge, Co., USA, pp. 472–479, 2016.

Müller, K., Techel, F., Mitterer, C., Feistl, T., Sofia, S., Roux, N., Palmgren, P., Bellido, G., and Bertranda, L.: The EAWS Matrix, a look-up table for regional avalanche danger level assessments, and its underlying concept, in: Proceedings International Snow Science Workshop, 8 - 13 Oct, Bend, Or., 2023.

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.

Ogden, C. K. and Richards, I. A.: The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism, vol. 29, Harcourt, Brace, 1925.

Pérez-Guillén, C., Techel, F., Hendrick, M., Volpi, M., van Herwijnen, A., Olevski, T., Obozinski, G., Pérez-Cruz, F., and Schweizer, J.: Data-driven automated predictions of the avalanche danger level for dry-snow conditions in Switzerland, Natural Hazards Earth System Sciences, 22, 2031–2056, https://doi.org/10.5194/nhess-22-2031-2022, 2022.

Reuter, B. and Schweizer, J.: Describing snow instability by failure initiation, crack propagation, and slab tensile support, Geophysical Research Letters, 45, 2018.

Roiser, J. P. and Sahakian, B. J.: Hot and cold cognition in depression, CNS Spectrums, 18, 139–149, https://doi.org/10.1017/S1092852913000072, 2013.

Rowe, G. and Wright, G.: Expert Opinions in Forecasting: The Role of the Delphi Technique, pp. 125–144, Springer US, Boston, MA, https://doi.org/10.1007/978-0-306-47630-3_7, 2001.

Schweizer, J. and Camponovo, C.: The skier's zone of influence in triggering slab avalanches, Annals of Glaciology, 32, 314–320, https://doi.org/10.3189/172756401781819300, 2001.

Schweizer, J. and Föhn, P.: Avalanche forecasting - an expert system approach, Journal of Glaciology, 42, 318–332, https://doi.org/10.3189/S0022143000004172, 1996.

Schweizer, J., Kronholm, K., Jamieson, J. B., and Birkeland, K. W.: Review of spatial variability of snowpack properties and its importance for avalanche formation, Cold Regions Science and Technology, 51, 253–272, https://doi.org/10.1016/j.coldregions.2007.04.009, 2008.

Schweizer, J., Mitterer, C., Techel, F., Stoffel, A., and Reuter, B.: On the relation between avalanche occurrence and avalanche danger level, The Cryosphere, 14, 737–750, https://doi.org/10.5194/tc-14-737-2020, 2020.

SLF: Schnee und Lawinen in den Schweizer Alpen. Winter 1951/52, Eidg. Institut für Schnee- und Lawinenforschung Weissfluhjoch Davos: 231 pages (Nr. 15), 1953.

SLF: Interpretationshilfe zum Lawinenbulletin des Eidgenössischen Institutes für Schnee- und Lawinenforschung Weissfluhjoch, Davos, in: Mitteilung des Eidgenössischen Institutes für Schnee- und Lawinenforschung, vol. 49, p. 24, Eidgenössischen Institutes für Schnee- und Lawinenforschung, 1993.

Statham, G., Haegeli, P., Birkeland, K., Greene, E., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., and Kelly, J.: The North American public avalanche danger scale, in: Proceedings International Snow Science Workshop, 17 - 22 Oct, Lake Tahoe, Ca., pp. 117–123, 2010.

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, 2018.

Stewart, T. and Lusk, C.: Seven components of judgmental forecasting skill: implications for research and the improvement of forecasts, Journal of forecasting, 13, 579–599, 1994.

Stewart, T. R.: Principles of forecasting: a handbook for researchers and practitioners, chap. Improving reliability in judgemental forecasting, pp. 81–106, Springer Science + Business Media, LLC, 2001.

Techel, F.: On consistency and quality in public avalanche forecasting: a data-driven approach to forecast verification and to refining definitions of avalanche danger, Ph.D. thesis, Department of Geography, University of Zurich, Zurich Switzerland, 2020.

Techel, F. and Schweizer, J.: On using local avalanche danger level estimates for regional forecast verification, Cold Regions Science and Technology, 144, 52 – 62, https://doi.org/10.1016/j.coldregions.2017.07.012, 2017.

Techel, F., Zweifel, B., and Winkler, K.: Analysis of avalanche risk factors in backcountry terrain based on usage frequency and accident data in Switzerland, Nat. Hazards Earth Syst. Sci., 15, 1985–1997, https://doi.org/10.5194/nhess-15-1985-2015, 2015.

Techel, F., Mitterer, C., Ceaglio, E., Coléou, C., Morin, S., Rastelli, F., and Purves, R. S.: Spatial consistency and bias in avalanche forecasts – a case study in the European Alps, Natural Hazards and Earth System Sciences, 18, 2697–2716, https://doi.org/10.5194/nhess-18-2697-2018, 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., Pielmeier, C., and Winkler, K.: Refined dry-snow avalanche danger ratings in regional avalanche forecasts: consistent? And better than random?, Cold Regions Science and Technology, 180, 103 162, https://doi.org/10.1016/j.coldregions.2020.103162, 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, pp. 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, 2024a.

Techel, F., Mayer, S., Purves, R. S., Schmudlach, G., and Winkler, K.: Forecasting avalanche danger: human-made forecasts vs. fully automated model-driven predictions, Natural Hazards and Earth System Sciences Discussions, 2024, 1–31, https://doi.org/10.5194/nhess-2024-158, 2024b.

Techel, F., Müller, K., and Mitterer, C.: The EAWS matrix, a look-up table to determine the regional avalanche danger level (Part B): operational testing, Natural Hazards and Earth System Sciences Discussions, in discussion, 2025.

Thumlert, S., Statham, G., and Jamieson, B.: The likelihood scale in avalanche forecasting, The Avalanche Review, 38, 31–33, 2020.

Winkler, K., Schmudlach, G., Degraeuwe, B., and Techel, F.: On the correlation between the forecast avalanche danger and avalanche risk taken by backcountry skiers in Switzerland, Cold Regions Science and Technology, 188, 103 299, https://doi.org/10.1016/j.coldregions.2021.103299, 2021.

Winkler, K., Trachsel, J., Knerr, J., Niederer, U., Weiss, G., Ruesch, M., and Techel, F.: SAFE - a layer-based avalanche forecast editor for better integration of machine learning models, in: Proceedings International Snow Science Workshop, Tromsø, Norway, 23-29 Sep 2024, 2024.