Combining physical modeling and machine learning to improve avalanche hazard forecasting

Léo Viallon-galinier

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *13-12-2022* **par :**

Léo Viallon Galinier

Apport croisé de la modélisation physique et de l'apprentissage automatique pour la prévision du risque d'avalanches

Combining physical modeling and machine learning to improve avalanche hazard forecasting

---

JURY

Etienne Berthier	Directeur de recherche	Président du jury
Ingrid Reiweger	Professeure associée	Examinatrice
Jürg Schweizer	Professeur adjoint	Examinateur
Pascal Haegeli	Professeur associé	Rapporteur
Emma Suriñach	Professeure émérite	Rapporteuse
Pascal Hagenmuller	Ingénieur des Ponts, Eaux et Forêts	Directeur de thèse
Nicolas Eckert	Ingénieur en chef des Ponts, Eaux et Forêts	Directeur de thèse

---

École doctorale et spécialité :
SDU2E : Océan, Atmosphère, Climat

Unité de Recherche :
CNRM - Centre National de Recherches Météorologiques, Météo-France - CNRS
INRAE - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement

Directeur(s) de Thèse :
Pascal Hagenmuller et *Nicolas Eckert*

Rapporteurs :
Pascal Haegeli et *Emma Suriñach*

Abstract

Snow avalanches are a major concern in mountainous areas as they threaten recreationists and also infrastructures or urbanized areas. One action to reduce the associated risk is to warn the public and the authorities efficiently by forecasting the avalanche hazard. Avalanche forecasting relies on both snowpack and avalanche observations and representation of the snowpack evolution with snow cover models. However, there is currently no avalanche hazard forecasting model that combines avalanche observations and the knowledge of physical processes leading to avalanches. Moreover, models using observations and based on statistical methods, so-called machine-learning, are still mainly limited to meteorological data or simple snowpack variables as predictors and do not exploit our current understanding on snow evolution and avalanche release. The objective of this thesis is to combine the pros of snowpack and avalanche observation, physical and statistical methods to propose new avalanche hazard indicators for short-term avalanche activity forecasting. The avalanche-related information is extracted from meteorological and snow models with the help of our knowledge of mechanical processes involved in avalanche release and then statistically related to observed avalanche activity with a machine learning model. We thus develop a statistical-physical model that predicts avalanche activity based on snow and meteorological conditions and subsequent stability indicators. The proposed model aims to be numerically efficient while keeping a comprehension of physical phenomena occurring in the snowpack. This overall objective is addressed in different tasks summed up as research articles. We first identify and evaluate the different mechanically-based stability models existing in the literature and adapt them to the Crocus snow cover model. We highlight the need to combine different stability indicators to represent the different processes leading to avalanche formation. A selection of mechanically-based stability models then complements meteorological and snow cover information as an input for the random forest statistical model. The model trained on past avalanche observations on the Haute-Maurienne massif of the French Alps is then evaluated in detail. We showed the interest of the statistical model to represent avalanche activity, with a recall of 75% and specificity of 76%. We also point out the interest of using stability indices as inputs for the statistical model. However, avalanche being a rare phenomenon, numerous false positive are observed, with a precision of around 3.3%, which illustrates the difficulty of predicting avalanche occurrences with a high spatio-temporal resolution, even with the current cutting-edge data and modeling tools. Then the statistical model is successfully generalized to other data sources and geographical areas: similar performances are obtained on a selection of four other mountainous areas of French Alps and Pyrenees, with another avalanche observation dataset as learning support, which underlines the robustness of the selected model. A first insight into triggered avalanches, a complementary information to the natural avalanche susceptibility in avalanche warning bulletins, also shows similar performances and thus provides encouraging perspectives to use similar models for this complementary problem. We proposed a new model for avalanche activity estimation from snow cover model representation of the snowpack that takes advantage of both the knowledge of physical processes and past observations through machine learning. Our model would, in the future, replace the MEPRA model currently available to avalanche forecasters. Further work will nevertheless be required to improve this model: the combination of observational data sources and the use of new satellite detection of avalanche deposits would help in improving the accuracy of the model as well as refinements of the statistical model.

Résumé

Les avalanches représentent un défi pour les habitants et les infrastructures des zones de montagnes. Une des actions pour réduire le risque associé est l'information préventive du public et des autorités par la rédaction de bulletins de prévision de l'aléa avalancheux, couramment appelé prévision du risque d'avalanche. La prévision du risque d'avalanche repose à la fois sur des observations de l'état du manteau neigeux et des avalanches ainsi que sur la modélisation numérique. Il n'y a pourtant aujourd'hui aucun modèle qui combine à la fois les observations et la connaissance physique des processus physiques à la source des avalanches. De plus, les modèles statistiques sont principalement limités à l'usage de variables météorologiques ou de variables simples et intégrées sur l'ensemble du manteau neigeux sans utiliser notre connaissance des processus de déclenchement des avalanches. L'objectif de cette thèse est de combiner les avantages respectifs de notre connaissance du passé au travers des observations du manteau neigeux et des avalanches, la connaissance de la physique des phénomènes et des méthodes statistiques pour proposer de nouveaux indicateurs de l'aléa avalancheux pour la prévision du risque d'avalanche. L'information pertinente pour l'activité avalancheuse est d'abord extraite des modèles météorologiques et d'évolution du manteau neigeux à l'aide de la connaissance des processus physiques à l'œuvre dans le déclenchement des avalanches, puis fourni à un modèle d'apprentissage pour relier ces variables d'entrée à l'activité avalancheuse observée. Ce travail est découpé en différentes parties et résumé dans des articles publiés, soumis ou en préparation. Nous identifions et évaluons dans un premier temps les différents modèles de stabilité à base physique existants. Ils sont ensuite adaptés au modèle de manteau neigeux Crocus. Nous montrons la nécessité de combiner plusieurs indicateurs de stabilité pour rendre compte de l'ensemble des phénomènes mis en jeu dans le départ d'une avalanche. Une sélection de modèles de stabilité est ensuite utilisée, conjointement aux informations nivo-météorologiques, comme variable d'entrée pour un modèle statistique de type forêt d'arbres aléatoires. Le modèle statistique est entrainé et évalué sur des observations passées de l'activité avalancheuse dans le massif français de Haute-Maurienne, puis évalué en détail sur cette zone géographique. L'intérêt de la méthode statistique est démontré avec une probabilité d'identification des situations avalancheuses de 75% et une sélectivité de 76%. L'intérêt de l'utilisation des modèles mécaniques de stabilité comme variable d'entrée du modèle statistique est également confirmé. Le phénomène avalancheux reste néanmoins un phénomène rare et difficile à prévoir avec une haute résolution spatio-temporelle, ce qui se traduit par un nombre élevé de faux positifs et une précision autour de 3,3%, y compris avec un jeu de données et des outils de modélisation de pointe. Le modèle statistique est ensuite généralisé à d'autres sources de données et d'autres zones géographiques avec succès : des performances similaires sont obtenues sur une sélection de quatre massifs français des Alpes et des Pyrénées et avec un nouveau jeu de données. Ces résultats confirment la robustesse du modèle statistique choisi. Des résultats préliminaires sur l'activité avalancheuse provoquée, complémentaire de l'activité naturelle pour les bulletins d'estimation du risque d'avalanche, donne des perspectives encourageantes. Le modèle que nous proposons pourra dans le futur remplacer le modèle MEPRA actuellement mis à disposition des prévisionnistes. Des travaux supplémentaires seront néanmoins nécessaires pour continuellement améliorer ce modèle. Cela inclut notamment la combinaison des différentes sources d'observation disponibles pour l'activité avalancheuse, ainsi que l'introduction de nouvelles sources de données et des améliorations des modèles statistiques utilisés.

Résumé étendu

The following summary is dedicated to non-English speakers. English speakers are suggested to read directly the expanded introduction in Chapter 1.

Les régions de montagne sont soumises à divers aléas naturels, comme les glissements de terrain, chutes de pierres, crues, laves torrentielles ou avalanches. Ces aléas menacent les pratiquants de la montagne, les infrastructures ainsi que les zones urbanisées et leurs habitants. L'exposition associée à ces aléas a augmenté au cours des dernières décennies suite au développement des activités de montagne et au développement en conséquence des infrastructures et de l'urbanisation de montagne [Pörtner et al., 2019]. Parmi ces aléas, les avalanches sont le résultat d'un déséquilibre dans le manteau neigeux qui va ensuite s'écouler le long d'une pente. Le manteau neigeux est formé durant l'hiver suite à diverses chutes de neige dans différentes conditions. Il évolue ensuite en fonction des conditions météorologiques. A un certain moment, le manteau neigeux peut se révéler instable et une avalanche peut se déclencher, soit sous le poids propre du manteau neigeux (avalanche naturelle) ou suite à une perturbation extérieure (skieur, chute de pierre ou de glace, déclenchement préventif par explosifs, etc., formant une avalanche provoquée). Durant la période hivernale, les avalanches sont un aléa généralisé dans les zones de montagne, pour peu que la pente soit suffisante. Il s'agit d'une préoccupation pour un grand nombre de parties prenantes, incluant les stations de sports d'hiver, les skieurs ou raquettistes mais aussi les gestionnaires de routes de zone de montagne, les populations de montagne ou de vallée et les autorités locales [Bründl and Margreth, 2021].

Pour réduire le risque associé à l'aléa avalancheux, les pays les plus exposés ont introduit des règlementations pour protéger les populations dans les zones les plus menacées. Des réseaux d'observations ont été développés pour améliorer la connaissance du phénomène à long terme et fournir de l'information à court terme sur l'état du manteau neigeux. Des infrastructures de protection ont également été mises en place depuis plusieurs siècles pour protéger les infrastructures principales et les zones habitées. Il reste néanmoins des zones soumises à un risque élevé et des populations comme des infrastructures restent menacées [Glass et al., 2000]. Bien que les plans de prévention des risques naturels limitent l'extension de la vulnérabilité sur les zones les plus à risque, il est impossible de protéger chaque espace de montagne. Ainsi, 110 personnes en moyennes sont tuées chaque année en Europe par avalanche [EAWS, 2022]. L'information sur les situations pour lesquelles l'aléa est le plus élevé reste nécessaire, afin de limiter l'exposition, par l'évacuation pour les zones habitées, la limitation ou la réorientation de la fréquentation dans les régions plus reculées.

Pour prévenir des situations critiques pour le risque d'avalanche, les pays concernés ont mis en place des services de prévision du risque d'avalanche [LaChapelle, 1977; Morin et al., 2020]. Même s'il n'est pas possible de prédire l'occurrence d'une avalanche en particulier ni son étendue précise, l'intensité de l'activité avalancheuse attendue dans des zones géographiques adaptées peut être estimée. Des bulletins d'estimation du risque d'avalanche sont ainsi produits quotidiennement par les services en charge de cette prévision. Dans cette thèse, nous proposons un nouvel outil d'aide à la décision pour l'estimation à court terme de l'aléa avalancheux.

L'alerte nécessite des outils de prédiction de l'activité avalancheuse attendue. L'activité avalancheuse est généralement estimée à partir de deux sources d'informations : l'observation de la stratigraphie du manteau neigeux et de l'activité avalancheuse ainsi que la modélisation de l'évolution future à l'aide de modèles numériques. La plupart des services en charge de la prévision du risque d'avalanche ont organisé des réseaux d'observations du manteau neigeux et de l'activité avalancheuse [p. ex. Giard et al., 2018]. L'observation de terrain est précieuse mais elle ne donne qu'une information ponctuelle sur l'état présent ou passé du manteau neigeux et de l'activité avalancheuse. Pour compléter ces observations, la connaissance de la physique des phénomènes à l'œuvre dans le manteau neigeux permet de modéliser son évolution future [Brun et al., 1992; Vionnet et al., 2012; Bartelt and Lehning, 2002]. Les modèles

détaillés de manteau neigeux représentent la stratigraphie du manteau et son évolution dans le temps à l'aide de lois physiques ou phénoménologiques. Ils peuvent aider à imaginer l'évolution future du manteau neigeux ainsi que modéliser le manteau neigeux là où aucune observation n'est disponible. En revanche, ils ne donnent pas directement une analyse de la stabilité du manteau neigeux. Une connaissance supplémentaire des processus physiques menant aux avalanches est nécessaire pour estimer la stabilité d'un manteau neigeux et traduire l'information des modèles en risque d'avalanche. Dans cette thèse nous proposons un outil d'analyse de la stabilité de manteaux neigeux modélisés afin de prédire l'activité avalancheuse attendue dans le contexte de prévision à court terme du risque d'avalanche.

Plusieurs méthodes utilisant notre connaissance de la physique des avalanches ont été conçues pour analyser la stabilité du manteau neigeux à partir de profils stratigraphiques. Ces méthode comprennent des règles expertes, des analyses mécanique du comportement du manteau neigeux ou une combinaison des deux. L'application de ces méthodes au manteau neigeux modélisés est un premier défi, car les modèles de stabilité ont pour la plupart d'abord été développés pour l'analyse de profils simplifiés ou nécessitent la connaissance de propriétés mécaniques des couches de neige qui ne sont pas modélisées directement par les modèles de manteau neigeux. De plus, chaque indice de stabilité a pour but de représenter un aspect spécifique de l'activité avalancheuse et des processus qui y mènent. Il n'y a donc pas un indice de stabilité qui fournirait l'ensemble de l'information pertinente pour évaluer l'activité avalancheuse attendue.

Une première partie de la thèse s'attache donc à répondre aux questions suivantes : quels indices de stabilité à base physique sont pertinents pour déterminer les situations avalancheuses à partir de modèles numérique de manteau neigeux ? Et comment adapter ces modèles de stabilité au modèle numérique de manteau neigeux Crocus, que nous utilisons ? A l'aide d'une revue de la littérature, nous avons identifié les modèles mécaniques de stabilité pertinents pour un usage avec des modèles détaillés de manteau neigeux (en terme de résolution verticale, processus représentés, échelle spatiale et temporelle, coût numérique, etc.). Nous avons sélectionné un ensemble de modèles de stabilité pour représenter au mieux l'ensemble des processus en jeu dans le déclenchement d'une avalanche, en l'état de la littérature scientifique. Enfin, nous illustrons leur comportement et leur complémentarité dans des situations typique de risque d'avalanche. En conclusion, cette première partie a permis de sélectionner et d'implémenter un ensemble d'indices de stabilité permettant de réduire l'information du modèle de manteau neigeux Crocus en extrayant les indicateurs pertinents pour le risque d'avalanche.

Une approche alternative, ou complémentaire, aux indices de stabilité à base physique, vient des statistiques et de l'utilisation des observations de l'activité avalancheuse passée. Plusieurs réseaux d'observations ont été développés pour observer l'activité avalancheuse, notamment par l'ONF et Inrae pour identifier et suivre les zones à risques, ou par Météo-France en partenariat avec les stations de ski pour suivre l'évolution au jour le jour des conditions nivo-météorologiques et de l'activité avalancheuse. Ces réseaux produisent ainsi chaque année, et depuis plusieurs dizaines d'années, une quantité importante d'informations sur l'activité avalancheuse passée. Il est donc possible d'envisager des approches d'apprentissage automatique pour la prévision du risque d'avalanche. Dans la communauté neige et avalanches, la méthode des plus proches voisins (identification de situations passées proches de la situation à prédire) a été utilisée dans des stations de ski suisses comme françaises dans les années 1980 [Navarre et al., 1987; Buser, 1989]. Depuis ces travaux pionniers, peu de nouvelles améliorations ont été apportées aux utilisateurs finaux malgré quelques études sur des méthodes alternatives d'apprentissage ou l'usage de nouveaux prédicteurs de l'activité avalancheuse. De plus, l'approche statistique et l'approche à base physique ne sont pas incompatibles. L'analyse mécanique de la stabilité permet de résumer l'information pertinente pour la stabilité du manteau neigeux parmi la grande quantité de données produites par les modèles numériques de manteau neigeux. L'approche statistique permet ensuite de tirer parti des observations passées de l'activité avalancheuse. Pourtant, l'analyse de stabilité est peu utilisée dans les modèles d'apprentissage actuels.

Une deuxième partie de la thèse propose de combiner la connaissance des processus physiques impliqués dans la formation des avalanches et l'observation passée de l'activité avalancheuse à l'aide de méthode d'apprentissage automatique pour développer un nouvel outil d'aide à la décision pour prédire l'activité avalancheuse attendue à partir de la modélisation numérique du manteau neigeux. Les questions principales sont comment les indices de stabilité à base physique peuvent se montrer complémentaires des méthodes d'apprentissage pour la prédiction de l'aléa avalancheux ? Que peut apporter l'apprentissage à un prévisionniste qui doit quantifier l'activité avalancheuse attendue ? Les indices de stabilité sont combinés à l'apprentissage pour produire un indicateur intégré de l'activité avalancheuse attendue. Cette adjonction systématique de l'analyse mécanique en entrée des modèles d'apprentissage pour réduire l'information produite par les modèles de neige avant d'utiliser les méthodes statistiques est totalement nouveau. L'intérêt de l'usage de ces indices de stabilité en combinaison avec l'apprentissage automatique est démontré. Il est également démontré que l'ensemble choisi des indices de stabilité

contient quasiment à eux seuls l'ensemble de l'information pertinente, ce qui valide le choix des indices de stabilité utilisés.

Étant adressé à la prévision du risque d'avalanche, par définition un secteur critique pour la sécurité, une évaluation approfondie des résultats du modèle statistique est ensuite menée pour évaluer les possibilités d'usage opérationnel de ce modèle. Cela comprend notamment une caractérisation précise des résultats en fonction de l'activité avalancheuse observée, au delà des scores globaux de la classification binaire entre jours avalancheux et non avalancheux. Il est notamment démontré que le modèle statistique est significativement plus performant que le modèle MEPRA aujourd'hui mis à disposition des prévisionnistes.

Enfin, la méthode est généralisée à d'autres sources de données et d'autres zones géographiques, pour répondre à la question : comment ces modèles statistiques se comportent et se généralisent en dehors du bac à sable du chercheur ? La méthode est ainsi généralisée à une sélection de quatre massifs des Alpes et des Pyrénées français et avec un autre jeu de donnée d'observation. L'évaluation montre des résultats similaires ce qui prouve la robustesse de la méthode. Des résultats préliminaires sur les avalanches provoquées permettent aussi d'imaginer l'usage d'une méthode similaire pour cette problématique connexe. Ces évaluations confortent l'utilité de la combinaison de l'approche mécanique et statistique pour la prévision du risque d'avalanche à plus large échelle.

Remerciements

Ce manuscrit de thèse est le résultat de trois années de thèse mais surtout de nombreuses rencontres et interactions. Tout d'abord, ce travail n'aurait pas été possible sans la rencontre avec Pascal, qui m'a permis d'imaginer travailler sur ce matériau magique qu'est la neige. Pascal m'a ensuite présenté Nicolas et ils ont ensemble accepté de m'encadrer, et de me supporter, sur cette thèse pendant trois ans. La thèse a également été possible grâce à un financement du ministère de la transition écologique et solidaire et l'accompagnement de nombreuses personnes pour la constitution du dossier de financement (Françoise Prêteux, Alexandre Péry, Marc Pontaud, Samuel Morin, Mohamed Naaim, Patrick Flammarion, Dominique Marbouty, Patrick Le Buhan, Bernard Commère, Éric Brun-Barrière, Laure Tourjansky, Catherine Calmet, Olivier Marco, Jean-Marc Allain et Luc Dormieux).

Je suis donc arrivé en Septembre 2019 au Centre d'Études de la Neige (CEN). J'y ai été accueilli avec beaucoup de bienveillance par Jean-Louis et Delphine, pour la partie administrative, pas toujours simple. Je tiens également à remercier l'ensemble des agents du bâtiment pour leur accueil chaleureux, du rez de chaussée au deuxième étage ainsi que le service informatique, qui a eu a subir mes très nombreuses demandes.

Le résultat du travail de thèse que je présente dans ce manuscrit est le résultat de nombreuses interactions qui ont nourri mon travail. Je souhaite en particulier remercier Sophie Giffard-Roisin, Alec van Herwinjen, Guillaume Chambon et Philippe Naveau d'avoir accepté de se pencher sur mon projet de thèse et participer aux comités de suivi. Je tiens aussi à remercier Matthieu et son équipe pour les échanges qui ont alimenté ma thèse et les réunions d'équipe stimulantes, tout particulièrement pendant la période du Covid.

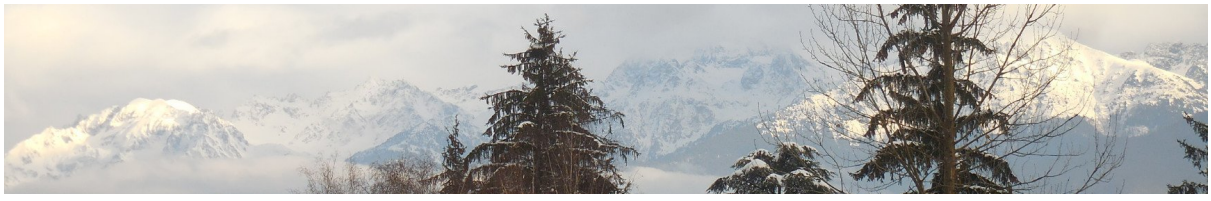
Je remercie également l'ensemble de mon Jury qui a permis l'aboutissement de ces trois années, Étienne Berthier, Ingrid Reiweger, Jürg Schweizer ainsi que Pascal Haegeli et Emma Suriñach qui ont accepté de rapporter ce travail et m'ont fourni des retours constructifs.

La thèse est un travail scientifique mais aussi des hauts et des bas au jour le jour. J'ai eu la chance de travailler dans un bureau avec une très bonne ambiance, d'abord avec Josué et César puis avec Lisa et Carlo (et ses correspondants visio). J'ai aussi été accueilli par les anciens doctorants, qui m'ont guidé et conseillé, Bertrand et Iheb. J'ai aussi beaucoup échangé avec les doctorants de ma promotion, Laure, Anna et Antoine, puis des suivantes, Lisa, Matthieu, Ange, Alvaro et Diego. Je remercie également Yves et Yannick qui m'ont fait sortir un peu du bureau pour aller sur le terrain et s'aérer les idées et être ensuite plus productif. J'ai aussi appris avec eux plein d'autres choses sur le contreventement des structures ou les travaux en terrains caillouteux.

Je tiens à remercier tout particulièrement Samuel et Marie qui ont toujours veillé à préserver une bonne ambiance, de la bienveillance et de l'écoute dans le laboratoire, ce qui fait dire à certains que nous vivons au pays des bisounours, et c'est très appréciable !

Enfin, je voudrais remercier ma famille (alors, tu trouves ?) de leur soutien, bien évidemment Chloé qui m'a accompagné dans l'aventure ainsi que mes amis pour toutes les soirées autour d'une tisane (et les journées en montagne) à discuter de choses et d'autres et à partager nos expériences de thèses, pour la plupart.

Enfin, merci aux prévis qui ont concocté un paysage magnifique pour le jour de ma soutenance.



Contents

Chapter 1

Introduction

Contents

1.1 Context

Mountainous regions are prone to gravity-driven natural hazards, including landslides, rockfalls, floods, debris flows and snow avalanches. These hazards threaten recreationists in backcountry areas, human infrastructures (e.g. roads, communication networks), and even urbanized areas and their inhabitants. Moreover, the associated exposure has increased in the past decades due to the development of mountain activities, subsequent development of infrastructures and population increase [e.g. Pörtner et al., 2019; Zgheib et al., 2020]. Among these hazards, snow avalanches result from an instability in the snowpack which then flows down the slope. The snowpack is formed during the winter through snowfall events with different characteristics (temperature, wind, temperature and humidity during snow formation in the clouds, etc.). It then evolves depending on meteorological conditions (incoming radiations from the sun, clouds or its environment; heat and mass exchanges at the surface; internal evolution under temperature and humidity gradients or melting). At some point of this evolution, the snowpack can become unstable, and snow avalanches may release, under the load of the snowpack itself (natural release) or triggered by an external perturbation (skier, rockfall, icefall, explosives for preventive triggering, etc., referred to as artificially triggered avalanches). During winter, snow avalanches are a widespread hazard in mountain areas (as long as the slope is sufficient). They are a significant concern for a wide range of stakeholders, including ski resorts, ski mountaineers and snowshoers, but also road managers in high elevation roads or local authorities and populations in high valleys [Bründl and Margreth, 2021].

1.1.1 Risk mitigation

To reduce the risk associated with avalanche hazard, the most exposed countries have introduced regulations to protect populations in threatened areas. For instance, in France, reforestation campaigns have been conducted since the middle of the 19th century with the creation of the RTM (Restauration des Terrains en Montagne, part of the French forestry service) [Jeudy, 2006], to reconstruct protective forests (Figure 1.1a). The need for knowledge on this specific hazard led to the creation of the first observation networks reporting avalanche occurrences systematically in defined areas. The first avalanches report started in 1899 in France [De Crécy, 1965] (Figure 1.1b), and constitutes the first steps toward the still active Enquête Permanente Avalanche (EPA). Nowadays, this survey comprises the systematic observation of more than 3 000 avalanche paths in the French Alps [Bourova et al., 2016]. This historical knowledge allows identifying the most hazardous places and restricting development in these places through urbanism regulations as the French PPRN (Plan de Prévention des Risques Naturels, including avalanches since 1982, Figure 1.1c). In complement, remaining threatened areas may be protected through protective structures. The goal can be to stabilize the snowpack in release areas (e.g. supporting structures, Figure 1.1) or to deviate or slow down the avalanche flow near urbanized areas or infrastructures (e.g. dissipative structures, dams, Figures 1.1d, e, f). These actions significantly reduce the risk where they are applied, but they do not ensure perfect protection, especially for extreme scenarios. Moreover, these defensive techniques are bounded to limited and well-identified areas. These defences can neither fully protect the whole distance of roads or railways crossing mountains, nor protect isolated stakes or people in backcountry areas. In fact, some population and infrastructures remain threaten by avalanches [Glass et al., 2000] and around 110 people are killed each year in avalanches in Europe [EAWS, 2022] (Figure 1.3). These fatalities mainly occur in uncontrolled area, that is to say in backcountry, where no systematic hazard reduction methods are possible at reasonable economic and environmental costs [Techel et al., 2016]. The information on critical situations, with high avalanche activity, in which only the reduction of exposure by limitation of frequentation of most exposed slopes, can significantly reduce this risk. In more controlled areas, the identification of hazardous areas and the evaluation of protective measures are challenging, and the residual risk may remain important in specific situations. It is reminded by several tragic accidents, such as those of Val d'Isère (39 young people killed) in 1970 in France or Montroc (12 fatalities) in France, Evolene in Switzerland (12 fatalities), Galtür (39 fatalities) in Austria during winter 1998-1999 or Rigopiano in 2017 in Italy (29 fatalities) (Figure 1.2). Hence, warnings on critical situations are required to complement permanent protective infrastructures and regulations.

Most countries facing avalanche hazard rely on operational services for avalanche hazard forecasting [e.g. LaChapelle, 1977; Morin et al., 2020]. Even though the timing of the release and the extent of avalanches cannot be predicted precisely, the intensity of avalanche activity in defined areas can be estimated. Warning services provide avalanche bulletins daily with an estimation of avalanche danger. This danger is often reported on the most commonly used avalanche danger scale, ranging from 1 (low) to 5 (very high in Europe or extreme in North America) and accompanied by a description of the typical avalanche situation expected, most impacted areas, for both natural and triggered avalanches [EAWS,

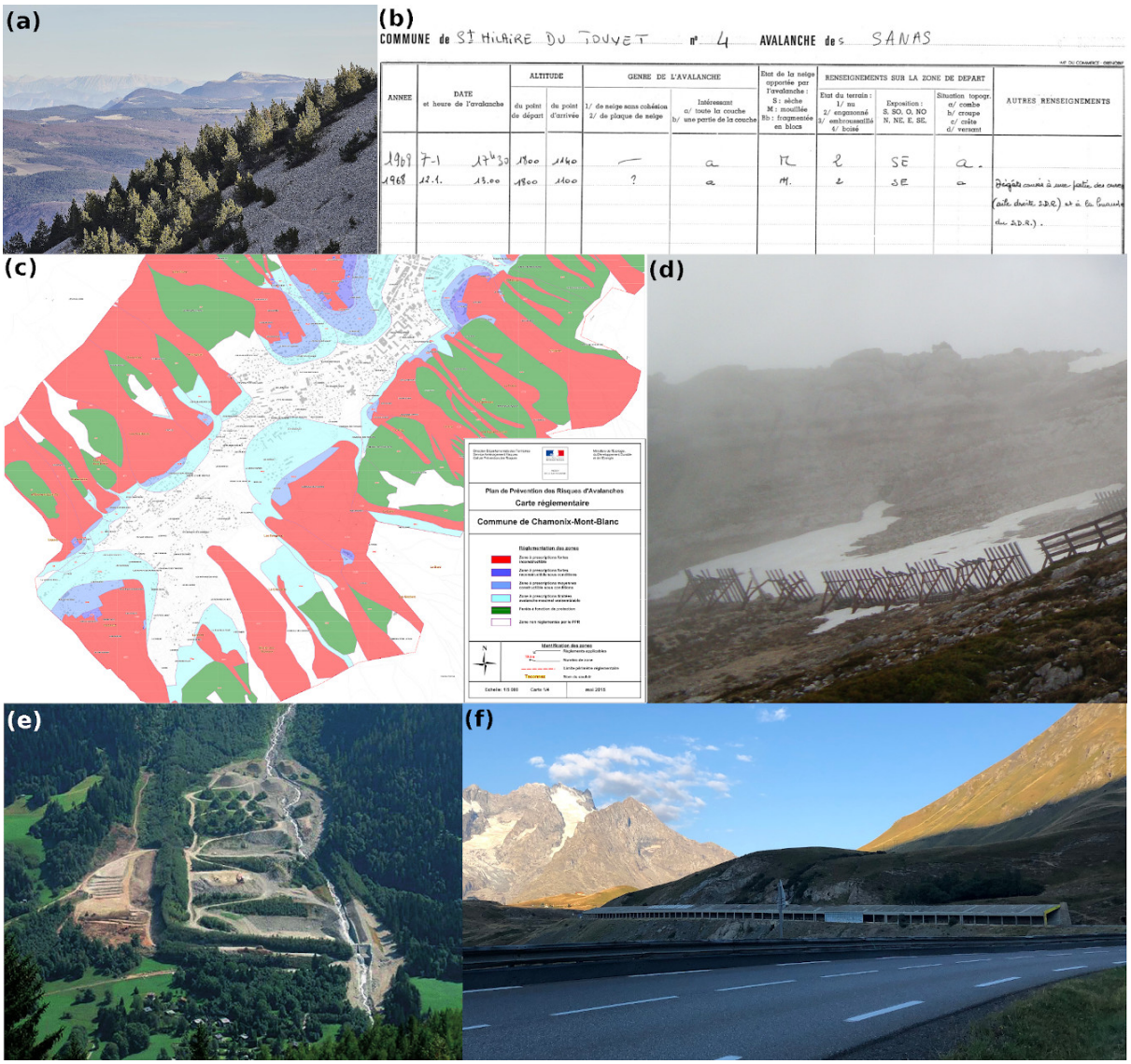


Figure 1.1: Different tools and infrastructures to protect against avalanches: (a) reforestation, example of the slopes of Mont Ventoux, France, image RTM. (b) Avalanche report from book of ONF, example of one avalanche path of Chartreuse massif. (c) urbanization regulation, example of PPRN of Chamonix, France. (d) Structure for snow retention and limitation of avalanche triggering in release areas, example of Chamechaude, Le Sappey en Chartreuse, France. (e) Structure for snow retention at the bottom of avalanche path of Taconnaz, Chamonix, France. Image INRAE. (f) Structure to protect a road, example of Galerie de la Marionnaise, Le Monêtier les Bains, France. Image from Marie Dumont.

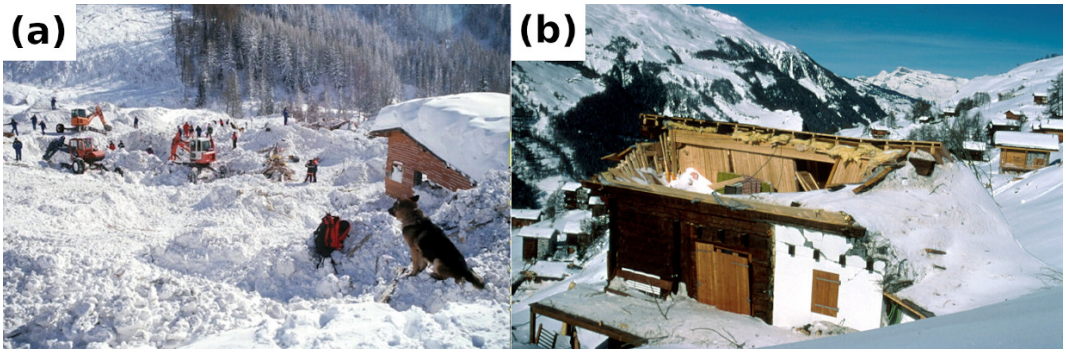


Figure 1.2: Destructive avalanches in 1999: (a) Montroc, France, the 9th of February and (b) Evolene, Switzerland, the 21st.

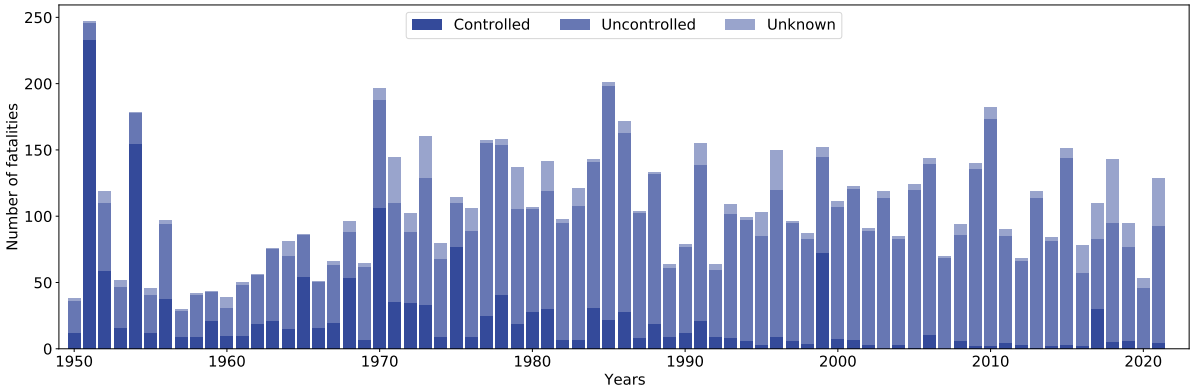


Figure 1.3: Avalanche deaths in Europe in controlled and uncontrolled areas, from 1950 to 2022. Data from [EAWS, 2022].

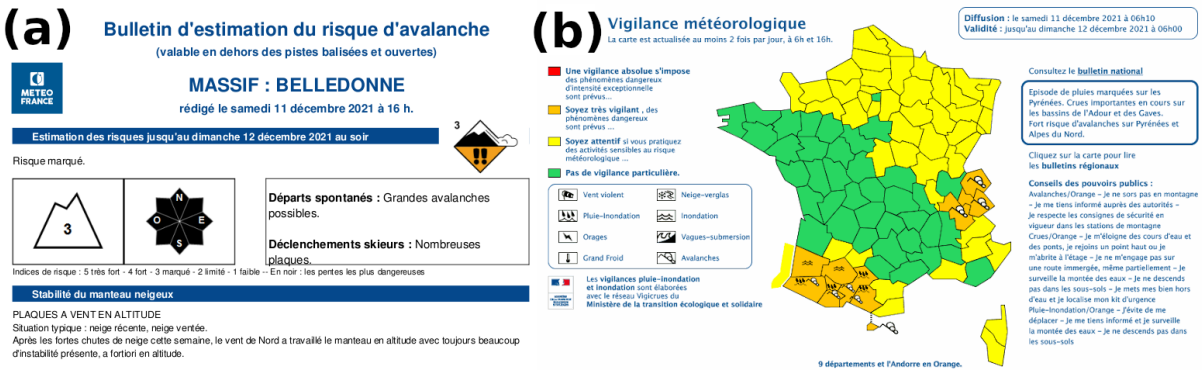


Figure 1.4: Bulletins warning populations of avalanche hazard, the 11th of september, 2021: (a) BRA bulletin focused on avalanche hazard in mountainous areas and (b) Vigilance map at the national scale, dedicated to the general public and covering 10 meteorological phenomena.

2022; Statham et al., 2010]. In Europe, these bulletins are designed for ski resorts, recreationists, and local authorities in charge of infrastructure or public safety. In France, the warning is divided into two productions. The BRA (Bulletin d'estimation du Risque d'Avalanche) with a large scope and a detailed description of expected avalanche activity, and a focus on backcountry skiing (Figure 1.4a). Firstly produced in 1971, it was subject to several modifications up to its current form. In complement to the BRA, the vigilance bulletins describe the weather-related hazards for infrastructures or urbanized areas on a 4-colour scale (green, yellow, orange, red) designed explicitly for populations and public safety authorities since 2001 [Calmet, 2018] (Figure 1.4). Avalanche is one of the hazard covered by vigilance system [Coléou and Morin, 2018], spread over a large audience (television, radio, etc.). For avalanches, orange or red levels mainly concerns major avalanche events (corresponding to danger levels 5 or 4). This forecasting is of crucial interest to protect populations and infrastructures in places where no protective structures are in place and for extreme avalanche events. It guides public safety authorities to take preventive actions to reduce the risk: evacuating populations, closing roads or train lines, restricting access to backcountry areas, or preventively triggering avalanches.

Warning requires short-term forecast of the avalanche activity. The avalanche forecaster prepares his forecast by combining two main sources of information: the observation of current snowpack and avalanche activity and modeling of its future evolution. Most avalanche warning services have organized observation networks of the snowpack and avalanche activity [e.g. Giard et al., 2018]. In addition to meteorological parameters and snow depth, full snowpack profiles are reported regularly through a standardized observation method [Fierz et al., 2009]. Snowpack profiles contain, for each identified layer in the snowpack, measurements of penetration resistance, temperature, density, water liquid content, estimations of hardness and identification of the shape of grains composing the layer. In complement, observers often report the observed avalanche activity, whether it is a natural activity or avalanches triggered preventively or accidentally. Snow profiles are one of the bases of the stability analysis of the snowpack. However, it provides only spatially limited information (limited number of measurement points) on past snowpack stratigraphy and avalanche activity. To overcome these limitations, obser-

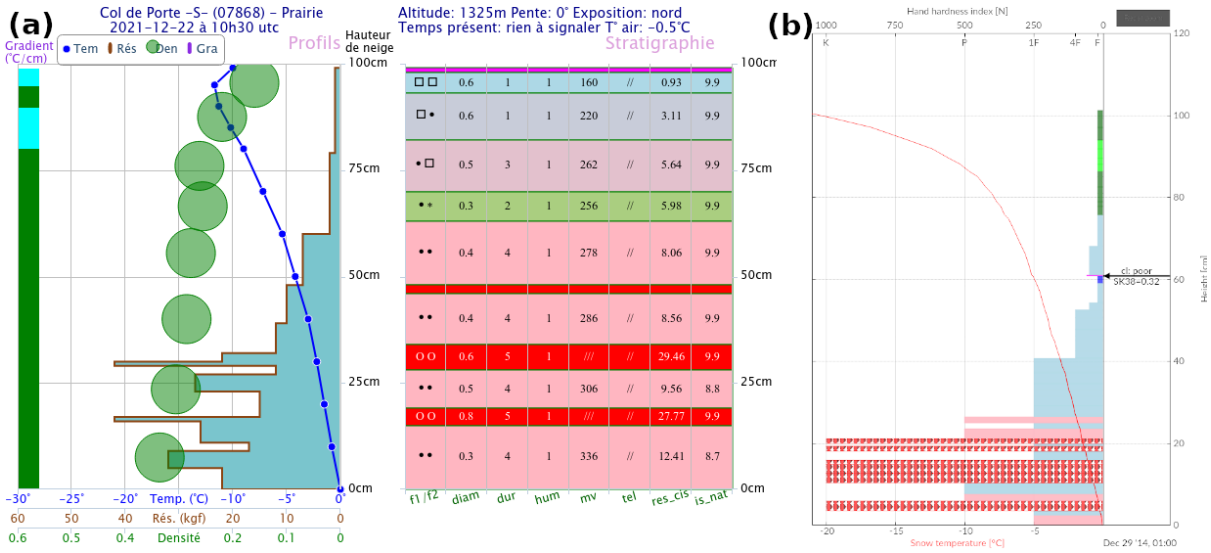


Figure 1.5: Observed and modelled stratigraphies. (a) Observation on 22th December 2021 at col de Porte, Chartreuse, France. (b) Snowpack simulated profile for Weissfluhjoch station, Switzerland on 14th December 2014.

vations are combined with physical knowledge of snowpack evolution, which is used to model future snowpack evolution [Brun et al., 1992; Vionnet et al., 2012; Bartelt and Lehning, 2002]. Detailed snow cover models, such as SURFEX/ISBA-Crocus [Brun et al., 1992; Vionnet et al., 2012] in France, Snowpack in Switzerland [Bartelt and Lehning, 2002; Lehning et al., 2002b] or Sntherm in the US [Jordan, 1991] were developed to complement and extend snow profile information available. All models aim at modeling the complete stratigraphy of the snowpack, including, for each layer, thickness, density, temperature, liquid water content, some mechanical resistance parameters or grain shape and size. Such models can be used to forecast the future state of the snow cover from meteorological forecasting models [e.g. Vernay et al., 2022; Bartelt and Lehning, 2002]. Both approaches provide stratigraphy of the snowpack as presented in Figure 1.5. However, none provide a snowpack stability analysis directly. Therefore, additional knowledge of avalanche physics is required to estimate the corresponding stability and interpret it to deduce avalanche hazards.

1.1.2 Risk and hazard

According to the IPCC, risk is the potential for adverse consequences for human or ecological systems [e.g. Pörtner et al., 2019]. It results from interaction between a hazard with the exposure and vulnerability of the affected human or ecological systems. The hazard is the potential occurrence of a natural or human-induced physical event that may cause an impact. In our case, the hazard is avalanche release. Indeed, an avalanche in uninhabited regions without any human people involved may not have any consequence for human societies. A hazard thus require an exposure, that is to say the presence of humans or systems of interest (infrastructures or ecosystems) to become produce consequences. Moreover, even in case of the presence of elements at risk, they can be protected (naturally or by protective infrastructures), which defines its vulnerability [Reisinger et al., 2020]. In this work, we focus on the avalanche hazard. The output of the developed models have then to be combined with exposition and vulnerability to estimate the avalanche risk in given areas. The Vigilance avalanche takes into account these different parameters. However, the avalanche bulletins (BRA, in France), even though referring to a level of risk, only reports the hazard because exposition is not known and depends on the behavior of each one, especially in remote areas.

1.1.3 Avalanche formation

Determining snowpack stability from snowpack stratigraphy requires specific knowledge of processes involved in avalanche formation. A snow avalanche results from an evolution of the snowpack up to an unstable snowpack structure on a sufficiently steep slope so that the involved snow starts moving down. Several interpretative frameworks exist to classify avalanches: by type of release, sliding mechanism, processes at starting point, type of flow, or on characteristics of the final deposit for instance [Interna-

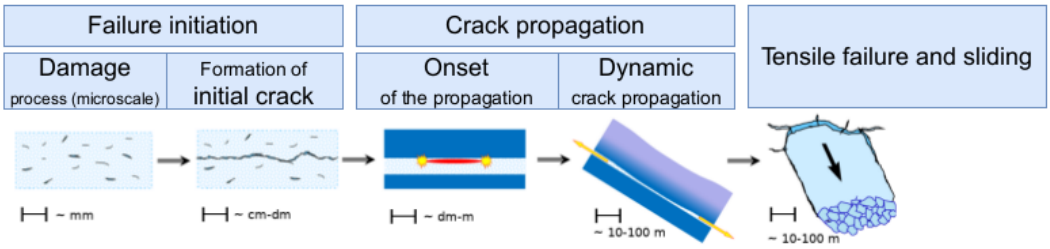


Figure 1.6: Processes involved in dry slab avalanches, according to Schweizer et al. [2003].

tional Association of Hydrological Sciences, 1981]. It is usual to first classify avalanches into two main types when considering the snowpack in the release area: dry snow avalanches and wet snow avalanches depending if the initiation of the avalanche occurs in dry or wet snow. It is also possible to classify avalanches depending on the initiation mechanism. Loose snow avalanches start at a single point near the surface and then grow by entraining additional snow while going down the slope. Slab avalanches involve the almost simultaneous release of a cohesive slab over a weak layer extending from meters to one kilometer [Schweizer et al., 2003].

In this context, these classifications lead to two categories for dry snow avalanches (when no liquid water is present in the snowpack), experiencing different processes of formation.

- **Dry loose snow avalanches** start at a single point near the surface due to a lack of cohesion between snow grains. The avalanche then grows progressively by basal and lateral entrainment leading to a typical triangular shape in the starting area (Figure 1.7a). This process generally occurs during snowfall.

- **Dry slab avalanches** are due to the presence of a slab structure inside the snowpack: relatively cohesive layers on the top of a weaker layer. Processes involved in the avalanche release include failure initiation, with the progressive damaging under loading of the weak layer or rapid formation of an initial crack under additional load (e.g. skier). If the crack reach a sufficient size, crack propagation can occur: it will start propagating in the weak layer, destroying the basal slab support in all directions. Finally, tensile failure occurs and the slab will finally release if the slope is sufficient for gravity to overtake the remaining friction (Figure 1.6) [Schweizer et al., 2003]. These avalanches are responsible for most fatalities by avalanche among recreationists [Schweizer and Lütschg, 2001]. The slab structure may remain in the snowpack for long periods until an additional loading initiates a crack in the weak layer. The basal crack in the weak layer can propagate on long distances (up to kilometer scale) and therefore involve substantial snow volumes (Figure 1.7b).

Wet snow avalanches occur when the snow is isothermal at 0 °C and contains liquid water:

- **Wet loose snow avalanches** are the consequence of the reduction of the cohesion between snow grains due to an excess of liquid water. The typical shape of dry loose snow avalanches are present. However, the flow dynamic is different as the snow involved generally has higher densities and contains liquid water (Figure 1.7c).

- **Wet slab avalanches** occur with similar processes as dry slab avalanches, except the snowpack contains liquid water. Moreover, the initiation process may involve an additional phenomenon: when liquid water reaches for the first time the weak layer, it could reduce its mechanical resistance and therefore initiate the failure of the weak layer [Bellaire et al., 2017].

1.1.4 Scope of this thesis

Avalanches are still a significant issue in mountain areas by threatening outdoor recreationists and infrastructure. The forecasting of avalanche activity is of critical interest to reduce the exposure both in uncontrolled areas by limiting visits in unstable conditions as well as in controlled areas by evacuating populations or closing infrastructures in critical situations. Knowledge on the future snowpack state could be obtained by current observations of the snowpack and expert analysis of future evolution. However, processes involved are complex and observations are rare especially in exposed areas. Hence, the numerical modeling of snowpack evolution have been developed to provide information on the snowpack at a tunable spatial and temporal resolution [e.g. Brun et al., 1989; Morin et al., 2020]. It is also able to

Figure 1.7: Different types of avalanches: (a) loose dry snow avalanche, experimental avalanche from INRAE test site, Lautaret, France, (b) slab avalanche and (c) wet snow avalanche, images from Météo-France.

provide snowpack evolution in the few coming days based on meteorological forecast. These snow cover models, such as Crocus [Brun et al., 1989; Vionnet et al., 2012] or SNOWPACK[Bartelt and Lehning, 2002], provide snow profiles representing the structure of the snowpack, but not directly information on stability or avalanche activity. An additional step is therefore required to analyze snowpack stratigraphy in terms of stability and expected avalanche activity. In this thesis we aim at providing tools to analyze the stability of modelled snow profiles in order to predict the expected avalanche activity in the context of short term forecasting.

A first approach in the community was to analyze the stability of snowpack with snow profile analysis based on our knowledge of avalanche release, on the form of stability indicators. It includes sets of expert rules, mechanical analysis of the snowpack behavior or a combination of both [e.g. Föhn, 1987b; Heierli et al., 2008; Schweizer and Jamieson, 2007]. The application of all these methods to a snow cover model is a first challenge as most of them were firstly developed on simplified profiles or require information on mechanical behavior of each snow layer which is not directly represented by snow cover models. Moreover, most of the indicators focus on a specific facet of avalanche activity and none provide a synthetic information on the expected avalanche activity.

An alternative approach comes from statistics and observations of past avalanche activity. Several networks report avalanche activity regularly, providing a significant amount of data on the past decades [e.g. Bourova et al., 2016; Coléou and Morin, 2018]. It is a situation in which machine learning have shown to be relevant. In the snow community, nearest neighbor (identification of past situations close to the current one) were used in French and Swiss ski resorts [Navarre et al., 1987; Buser, 1989]. However, since these pioneering works, no significant improvements reached the final user until very recent results [Pérez-Guillén et al., 2022] despite a few studies on alternative machine learning methods or different combinations of predictors.

These two approaches are complementary, as the mechanical analysis of stability summarizes the huge amount of data produced by the snow cover models and the statistical approach allows for taking advantage of the past avalanche observations. However, stability analyses are not yet used in combination with machine learning methods despite the possibility it provides to improve models with the knowledge gathered on past events. Therefore, this thesis aims to combine knowledge of physical processes leading to avalanche release and records of past avalanche activity with machine learning methods to develop new tools to predict expected avalanche activity based on modelled snowpack profiles.

The research of this thesis is conducted in the context of short-term forecast. It has several implications. It means that we focus on modelled snow profiles rather than observed profiles. It may not be exactly the profiles observed in the field but it has the advantage of being available operationally in forecasting mode in the coming days. It also means that we pay attention to the computational efficiency of the models used, which leads us to dismiss some of existing mechanical models that are useful for processes study but are not applicable in real time. Moreover, we do not consider the climate scale. We examine the relation between snowpack profile and avalanche activity and therefore do not consider the climate evolution of observed snowpacks.

1.2 State of the art

This thesis relies on snowpack simulations and mechanical knowledge to estimate stability from snow profiles. These two inputs are then combined with observations of past avalanche conditions through statistical methods. This section gives an overview of the state of the art of each of the four main research topics combined in this work: snowpack simulations, snowpack and avalanche observations, estimation of

snowpack stability from snow profiles and use of statistical tools for short-term evaluation of avalanche hazard estimation based on snow and weather information.

1.2.1 Snowpack simulation

Simulation of the snow cover evolution has been developed for many scientific and socio-economic applications, including hydrological concerns (floods, hydroelectricity), weather and climate applications or avalanche hazard forecasting [Vionnet et al., 2012]. Different models aim to represent the snowpack and its evolution through the season, depending on the application and expected resolution. They ranges from single-layer snow schemes [e.g. Douville et al., 1995; Bazile et al., 2001] to detailed snow cover models [e.g. Vionnet et al., 2012; Bartelt and Lehning, 2002]. Models mainly differ by the vertical resolution, physical properties represented, and physical processes taken into account [Vionnet et al., 2012].

Single-layer snow schemes represent the bulk snow mass and height as well as properties necessary to compute interaction with atmosphere such as albedo (ratio between received and re-emitted light, for radiative energy balance) [e.g. Douville et al., 1995; Bazile et al., 2001; Saloranta, 2012]. Several layers are required to represent better snow evolution, ageing and liquid water penetration in the snowpack. Some intermediate models aim to better represent snow properties by including the interaction between a limited number of snow layers [e.g. Loth and Graf, 1998; Boone and Etchevers, 2001]. However, if this may bring significant improvements with a limited computation overhead, this is not sufficient for avalanche hazard applications [Brun et al., 1989].

For avalanche hazard applications, the knowledge of detailed snow stratigraphy is crucial as most processes involved are linked to changes in snow properties [Schweizer et al., 2003]. For instance, a prerequisite for slab avalanche is the presence of a weak layer below a more cohesive slab. This structure can be formed during a snowfall due to different conditions of snow deposition (different temperatures or effect of wind) or between two layers that experience different evolutions (a pre-existing snow layer remaining at surface during days, experiencing a high temperature gradient then covered by a new layer). Detailed snow cover models include Crocus originally developed in France [Brun et al., 1989; Vionnet et al., 2012], SNOWPACK in Switzerland [Bartelt and Lehning, 2002; Lehning et al., 2002b,a] and SNTHERM in the United States [Jordan, 1991]. These models aim to represent the complete stratigraphy of the snowpack, with an unlimited (or sufficiently high) number of layers to accurately represent the vertical profile of snow properties required for both precise estimations of physical processes and avalanche hazard estimation. This level of detail allows for comparison with manual stratigraphy [Fierz et al., 2009] and for mechanical analysis of the profile to estimate snowpack stability, for instance.

In this work, we focus on SAFRAN-SURFEX/ISBA-Crocus model, hereafter referred to as Crocus, which representation of snowpack is reproduced on Figure 1.8. The SAFRAN weather analysis combines vertical atmospheric profiles estimates from the ERA-40 and ARPEGE (from 2002) reanalysis with observed data to provide atmospheric forcing data on an hourly basis with a vertical resolution of 300 m [Durand et al., 2009; Vernay et al., 2022]. The analysis, performed on pre-defined areas supposed to be climatologically homogeneous (the so-called massifs), is available from 1958 to present [Vernay et al., 2020]. Then the Crocus snow model [Brun et al., 1989; Vionnet et al., 2012], integrated to the soil surface scheme ISBA [Decharme et al., 2011] of the surface modeling platform SURFEX [Moigne, 2012] is used for snowpack modeling.

Snow cover modeling allows for providing detailed snow profiles nearly in any condition. However, it inherits uncertainties from meteorological data and physical processes represented in snow models. In fact, models are commonly driven by meteorological data without the use of snowpack observations during a whole season. Therefore, errors can accumulate through winter. Moreover, evaluations on the detailed stratigraphy are seldom [Viallon-Galinier et al., 2021]. Assimilation of snow observations in snow simulations are an active research field [e.g. Viallon-Galinier et al., 2021; Cluzet et al., 2020]. However, relying only on physical parameterizations, for instance, for stability assessment, can be limited due to uncertainties in previous simulation steps. To overcome these limitations, a first approach is to use an ensemble of meteorological forcings that are representative of the uncertainty on meteorological conditions [Charrois et al., 2016] and an ensemble of parameterizations for snow physics [Lafaysse et al., 2017] or a combination of both as, e.g., Cluzet et al. [2021]. The other main approach is to correct models a posteriori with statistical tools [e.g. Obled et al., 2002; Evin et al., 2021a].

1.2.2 Snowpack, stability and avalanche observations

Snow profiles and stability field measurements

Avalanche warning services have developed networks of local observers regularly reporting snow profiles [Pahaut and Giraud, 1995; Fierz et al., 2009]. Observers are generally aware of the interest of paying

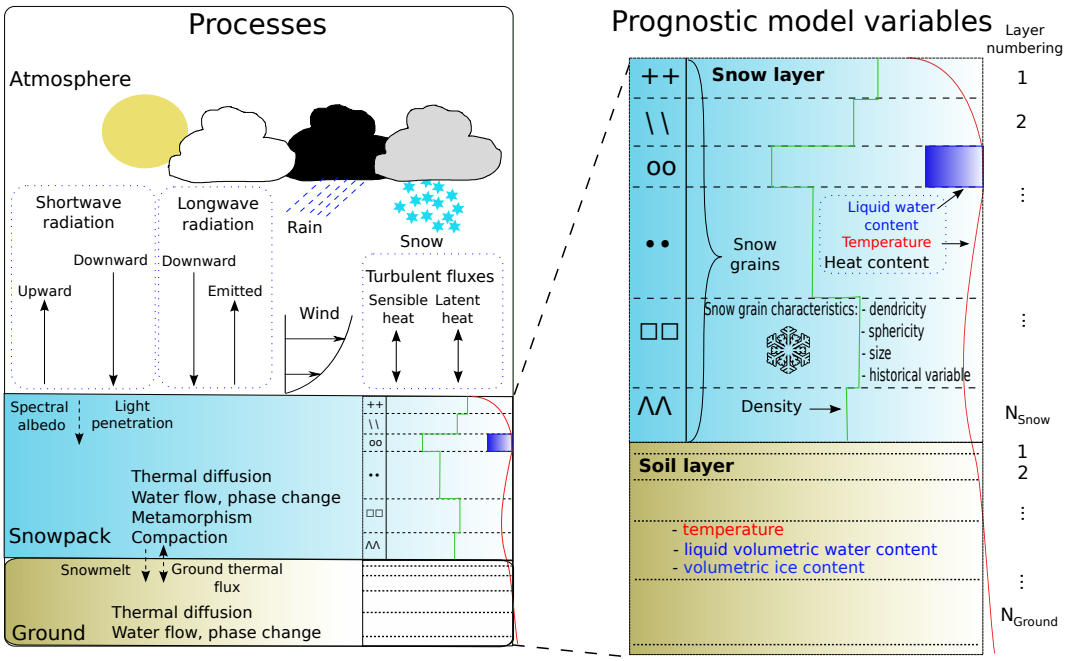


Figure 1.8: Diagram of the Crocus detalied snowpack numeric model and interaction with ground and atmoshpere: main variables and physical processes taken into account, from Vionnet et al. [2012].

attention to weak layers that can be involved in snowpack instability, leading to accurate reports on these layers. Observed snowpack profiles can then be analyzed with expert rules to infer snowpack stability [Schweizer and Wiesinger, 2001; Jamieson and Schweizer, 2005; Coléou and Morin, 2018]. However, measurements on weak layers are generally difficult to conduct without destroying the layer. Hence, valuable parameters which are complex to measure with a correct reproducibility are rarely reported, such as shear strength or density of weak layers.

Hence, an alternative approach consists in simulating controlled stress tests on the snowpack to evaluate its stability. Point stability of the snowpack can be estimated through different standardized stability tests such as the compression test [van Herwijnen and Jamieson, 2007], the extended column test [Simenhois and Birkeland, 2009], the rutschblock [Föhn, 1987b; Schweizer and Jamieson, 2010] or the propagation saw test [Gauthier and Jamieson, 2008]. All tests aim to create an initial perturbation of the snowpack (compression stress or carving of weak layers) on an isolated column of snow and observe an eventual failure of the tested snowpack. These tests allow for the identification of the critical weak layers in the snowpack (except for the propagation saw test that requires a pre-identified weak layer to evaluate) and an estimation of the load required to break these weak layers, as well as information on the rupture characteristics.

Snow profile and stability tests are highly complementary as snowpack profiles allow for understanding the potential processes involved in instabilities, evaluating volumes that can be mobilized in avalanches and estimating the evolution of the snowpack in the future, whereas tests help in identifying active weak layers and estimating the load required for avalanche triggering. The relation between snow profile and observed stability or the interpretation of snow profile in terms of stability were the first insights into tools for estimating stability from snow profiles. However, both of these measurements are time-consuming, leading to a limited set of data combining both snow stratigraphy and stability tests [e.g. Schweizer et al., 2021].

Avalanche observation networks

The knowledge of past avalanche activity is of crucial importance in avalanche risk management. It is used for two main goals: identifying sites at risk by determining the reachable area downstream based on observation of previous large avalanches and identifying meteorological and snow situations prone to avalanches. For these different goals, different observation networks were developed. In France, two main networks report avalanche activity: the Enquête Permanente Avalanche (EPA) and the network of snow observers (mainly in ski resorts).

The EPA consists of forest rangers reporting all avalanches that reach a defined threshold [Bourova et al., 2016]. It is maintained for the French Ministry in charge of the environment by ONF (Office

National des Forêts) and INRAE (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement). Starting at the beginning of the 20th century, it currently covers around 3000 avalanche paths in the French Alps and Pyrenees. Initially developed to capture large avalanche events in exposed areas for hazard mapping, it provides valuable information on avalanche activity over a long period. Each avalanche record indicates the location and date of the avalanche (with its uncertainty) with additional information such as elevation and aspect of the starting zone, volume and type of avalanche deposit. This unique dataset suffers from two main drawbacks. The first one is the uncertainty on dates of avalanches and the absence of non-event reports: during periods of low visibility or for remote sites, the uncertainty on the date could be substantial. As there is no report when the observation was possible, but no avalanches were detected, it is easy to identify days prone to avalanches but uneasy to be sure that no avalanche released on other days. The second limitation is linked to its goal: as it focuses on large avalanches reaching valley floors or infrastructures, small avalanches or high-elevation avalanche activity are weakly reported.

The alternative network is the network of snow observers that regularly report meteorological conditions, bulk or surface snow information, snow profiles and also information on snowpack stability [Coléou and Morin, 2018; Giard et al., 2018]. This network is mainly composed of ski resorts. It reports both natural and artificially triggered avalanche activity: the number, size, type, preferential aspect and altitude of natural avalanche release are reported as well as the same information on triggered avalanches (triggered by explosives or by skiers). Compared to EPA, this is a more recent source of data, but it covers more than 100 observation stations in the French Alps and Pyrenees, with the first reports in 1970. The uncertainty on dates is significantly lower than in EPA, as observers are asked to report only avalanches that occurred during the last 24 hours. Moreover, the observation contains a field to inform that observation was not possible and a different one to inform that there is no avalanche observed, which is a piece of additional information compared to EPA. However, it does not cover well-defined areas as the observer reports activity that the observer can view or is aware of around its ski resort, which highly depends on communication between ski resort patrollers, time available for observation and local topography.

Other methods for recording avalanches were experimented, such as seismic detection [e.g. van Herwijnen and Schweizer, 2011] or infrasound detection [e.g. Mayer et al., 2020]. None of them proposes the same spatial coverage for long periods. A wider spatial coverage now becomes possible with the analysis of satellite images to detect avalanche deposits [e.g. Karas et al., 2022]. This approach provides a potentially broad spatial coverage, but mainly on recent periods and with an uncertainty on the date linked to the periodicity of acquisition of satellites used and cloud coverage for visible satellite methods. The detection is also mainly limited to sufficiently large avalanches and remain an active field of research to improve the overall precision of the methods [e.g. Leinss et al., 2020; Hafner et al., 2021].

1.2.3 Estimation of snowpack stability from snow profiles

Snow profiles are reported by snow observers or produced by snow models. They provide valuable snowpack information, but an additional step is still required to translate this into stability information which is one of the keys to avalanche hazard assessment [Statham et al., 2018]. Since the pioneering work of Roch [1966a], several mechanical models were developed to assess snowpack stability from snow profiles. These models can be used with snow cover models. In parallel, alternative approaches based on expert knowledge are used by operational avalanche services. We detail below existing stability indices and the way they apply to snow cover model output.

Stability indices: models to analyze the stability from snow profiles

Snow profiles contain information on a succession of snow layers. From this data, it is possible to mechanically model the behavior of the snowpack and deduce some information on the stability, on the form of continuous indicators, called stability indices. Different models have been developed to represent the different stages of avalanche formation, from failure initiation to propagation and release.

Failure initiation models Failure initiation models mainly comprise the shear strength-stress ratio between the shear stress in each layer compared to the shear strength of the layer [Roch, 1966a]. Several indices exist, depending on the contribution to the stress considered: Föhn [1987b] first separates between natural contribution considering the stress due to weight of overlying slabs and skier ratio that add the contribution of an external load (typically a skier) applied at the top of the snowpack. Reuter et al. [2015] also uses a shear strength-stress ratio with stress only consisting of the additional contribution of an external load. To represent the early stages of natural avalanche processes, with progressive

breakage of individual bonds between grains of the weak layer, Nadreau and Michel [1986] introduced the deformation rate stability index or deformation rate ratio. This index compares the rate of bond cracking and the typical rate of bond healing through sintering. These four stability indices are static ones, whereas the chronology of events is known to have crucial importance in natural release [Schweizer et al., 2003]. To represent the relative effect of snowfall and sintering, Conway and Wilbour [1999] developed a stability index taking into account the time dimension, primarily through the snowfall rate: the expected time to failure.

Crack propagation models After the formation of an initial crack, it can grow up to a critical size from which it starts to propagate. The corresponding stability index is the critical crack length and can be estimated from snow stratigraphy [Sigrist, 2006; Heierli et al., 2008; Gaume et al., 2017]. During the crack growth, the propagation process can be slowed down or stopped by a vertical fracture of the slab. To measure this effect, the slab tensile criterion compares tensile strength and stress in layers above the weak layer at the onset of crack propagation [Reuter and Schweizer, 2018]. New models are developed to represent the dynamic propagation of the crack, which impacts the size of the avalanche to be expected [e.g. Gaume et al., 2018]. However, it remains for now limited to idealized situations and not full snow profiles [e.g. Bobillier et al., 2021] or measured in experimental setups [e.g. Bergfeld et al., 2021b] and no summarizing indicator emerged yet.

Expert approaches In complement to mechanically-based indices, expert rules are also used, sometimes combined with previously presented stability indices. The easiest and most common is the so-called lemons method [Jamieson and Schweizer, 2005] which uses six simple rules (yellow flags or twist of lemon) on grain size, hardness, grain shape, depth and difference in properties (grain size and hardness) between two consecutive layers. A total of 5 or 6 lemons indicate that triggering is likely. With similar rules, the SSI stability index combines the shear strength-stress ratio with two simple rules on hardness and grain size differences between layers to identify potential weak layers in SNOWPACK simulated profiles [Schweizer et al., 2006]. Based on the Crocus modelled profiles, the expert module MEPRA (Module Expert pour la Prévision du Risque d'Avalanches) [Giraud et al., 2002] was created. MEPRA uses the shear-strength stress ratios as a starting point and combines it with many expert rules to represent other phenomena not taken into account by the simple shear strength-stress ratio. MEPRA produces two indices, one for natural release and one for triggered avalanches (representation of an overloading of the snowpack).

Time-evolution based indices In addition to values at a given time of stability indices, the time-evolution of stability indices is also of interest. The most obvious example is the cumulative precipitations (or snowfall amount) [e.g. Navarre et al., 1987; Schweizer et al., 2008b]. Other meteorological parameters are also used in the form of cumulative values or differences on several days [e.g. Bovis, 1977; Davis et al., 1999; Kronholm et al., 2006]. However, using time-evolution derived variables can also be of interest for stability indices to identify the most critical period. Indeed, avalanches are less probable in a stabilization phase rather than during the weakening phase of the snowpack. This is one of the ideas behind the work of Conway and Wilbour [1999] on strength-stress ratio evolution under snowfall but remains not much considered.

Wet snow indices For wet snow avalanche activity, liquid water content has emerged as the primary driver of instability [Wever et al., 2016a]. Hence Mitterer et al. [2013] introduced a threshold on liquid water content per layer to identify instabilities. Naaim et al. [2016] aggregated the information at the snowpack scale through a humid snowpack depth. Details on wet snow indices and how they can be transposed to the Crocus snow cover model are provided in Appendix C.

Summary of dry-snow stability indices Chapter 2 of this thesis summarizes all the mechanical models that can be used to infer stability indices from simulated snow profiles. The reader is referred to this chapter for further details. For wet snow indices, the reader is referred to Appendix C.

Application of stability indices from outputs of snow cover models

Stability indices were presented above as a way to estimate snowpack stability from the physical and mechanical properties of each layer reported in a snow profile. However, traditional snow profiles were not designed for mechanical analysis and do not report all mechanical parameters necessary for computing these indices. Modelled snow profiles do not represent the evolution of mechanical properties either.

They rather represent the snow layers with basic properties (density, mass and heat content) and snow microstructure is represented by ad hoc variables that are oriented to describe snow metamorphism but not mechanical properties (e.g. grain size, shape, sphericity). An additional step is therefore required to infer mechanical parameters required by stability indicators. Some diagnostics are already integrated into snow cover models, such as cone penetration resistance because it was useful to compare to field measurements or shear strength because it is a fundamental parameter to estimate stability [Giraud et al., 2002; Richter et al., 2019]. However, parameters such as elastic modulus, Poisson ratio, tensile strength or weak layer fracture energy are not represented in common snow cover models.

Different parameterizations have been proposed in the literature for all parameters, mainly from density and grain shape. These parameterizations often derive from linear, power or exponential fits on measurements [e.g. Smith, 1965; Gerling et al., 2017; Scapozza, 2004, for elastic modulus]. An overview of the variety of parameterizations of mechanical properties is presented in Chapter 2, Figure 2.2. Most of them are based on relatively old measurements. This important variety illustrates the uncertainty on these parameterizations. Indeed, snow is a very fragile material which makes measurements complex, especially in weak layers that are of crucial interest for stability assessment.

A wide variety of stability indices started to be used in combination with snow cover models [e.g. Schweizer et al., 2016b; Reuter and Bellaire, 2018; Richter et al., 2019]. However, there is no systematic evaluation of the added values of the stability indices. Validations on limited datasets have been conducted for almost all indices (See Table 2.1). Some of them are used in operational avalanche hazard forecasting [Morin et al., 2020]. However, more systematic evaluation and relative interest in the stability indices are hardly reported.

Operational use of stability indices

Avalanche warning services rely on both observed and simulated profiles to assess the snowpack stability [Morin et al., 2020]. To help in analyzing these profiles, stability indices are generally provided. The lemons flags are common on field observation profiles, whereas shear strength-stress ratio (with skier for triggered avalanches and without for natural activity) are more common on simulated profiles (Figure 1.9). Relevant weak layers are pointed out with SSI index in SNOWPACK model and through the MEPRA algorithm in the Crocus model.

Using basic stability indices allows for a first reduction of the information and a first estimation of the stability to identify avalanche-prone situations to study more in detail. However, the strength-stress ratio usually employed only covers a limited part of the processes of avalanche formation. Expert rules, as extensively used in MEPRA, can cover the remaining processes. However, expert rules should be adapted or re-evaluated for each further model evolution, whether it is evolutions of the inputs, such as weather models, SAFRAN pre-computing of forcings, or snow cover models. Aiming at representing all processes, MEPRA is a very complex set of rules (see Appendix B) with empirical parameterizations validated on old versions of the Crocus model. Finally, MEPRA is based on a succession of thresholds used to classify avalanche situations and hazard levels. This separation of classes with successive decisions, as in decision trees, is known to have a large variance [Hastie et al., 2009]: slight variations of parameters can pass a threshold and lead to entirely different results for two close snow profiles. It results in a high instability of the results. Hence, MEPRA remains difficult to interpret on a single simulation which limits its practical use.

1.2.4 Statistical approaches to relate avalanche activity to snow and weather conditions at short time scale

Avalanches are a complex phenomenon to model. Moreover, models used to represent the snowpack inevitably contain some bias [e.g. Vernay et al., 2022]. As presented above, we have, on the one hand, snow cover models that provide a large amount of data (Section 1.2.1) from which it would be interesting to extract snowpack stability information. On the other hand, networks of avalanche observations have reported information over a long period and in numerous areas of the alpine regions (Section 1.2.2). This is a context (complex non-linear phenomena, with an observation dataset available) in which statistical, or so-called machine learning tools, have often shown to provide interesting results [e.g. Obled and Good, 1980; Navarre et al., 1987; Sielenou et al., 2021]. Combining the raw modelled data with past observations can help in identifying avalanche-prone situations.

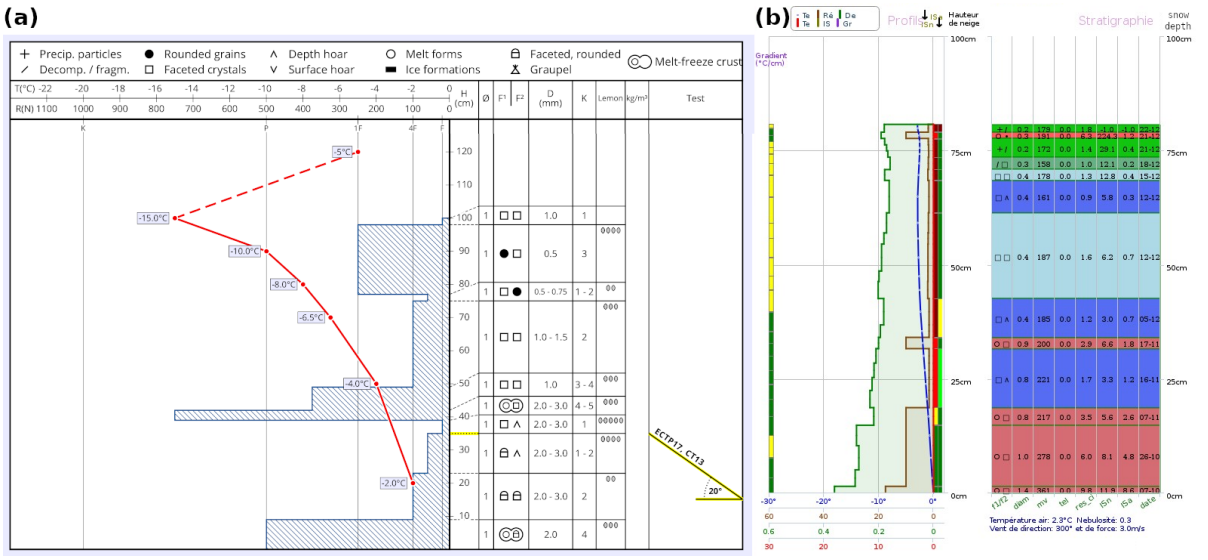


Figure 1.9: (a) Observed and (b) simuated profiles with stability information. Stability tests and lemons are reported on observed profiles while strength-stress ratios S_n and S_r are reported with colors for each layer on simulated profile. Example of (a) profile observed by Benjamin Reuter near Mont de Lans Glacier, Les Deux Alpes, France on 16th November 2021 and (b) Crocus simulation for Grandes Rousses massif, 2700 m, NE aspect the 22nd December 2020.

Methods

Different statistical methods have been tested in the literature to predict avalanche hazard from snow cover model outputs. First, linear models were intensively used, mainly linear discriminant analysis (LDA) [e.g. Föhn et al., 1977; McGregor, 1989; Bois and Obled, 1976; Bovis, 1977; Obled and Good, 1980]. It consists in drawing optimal linear hyperplanes to separate between classes. It has the advantage of high numerical efficiency and the possibility of graphical analysis of the results, at least on a limited number of dimensions.

The second highly used method is nearest neighbors (NN), which started to be intensively used around 1980 [e.g. Obled and Good, 1980; Buser, 1983; Navarre et al., 1987; Gassner and Brabec, 2002; Purves et al., 2003]. The principle is very intuitive: the class is chosen based on the class of the nearest elements of the training set. The only hyperparameter to choose is the number of considered neighbors. However, this method requires defining a distance metric between observations. For this purpose, several norms and variable normalizations are possible, which enlarge the possibilities and require some additional tuning. Moreover, with a large number of predictors (input variables), the problem can come across the curse of dimensionality [Bellman, 1957]: with a high number of dimensions, all observations tend to be isolated, separated by a similar distance.

In a third time, classification (and regression, to a minor extent) trees [Breiman et al., 1984] were introduced in snow community [e.g. Davis et al., 1999; Hendrikx et al., 2005; Kronholm et al., 2006; Hendrikx et al., 2014; Schweizer et al., 2009; Dreier et al., 2016]. Statistical trees are very similar to expert decision trees. They can be easily analyzed and plotted and then appear very understandable. However, it is optimized based on statistical correlations, which are not necessarily directly linked to physical relations (describing causality relations). Moreover, as presented for MEPRA, which is an expert tree, this method experiences large variance, and high instability [Hastie et al., 2009] as splits with thresholds can lead to large changes in the result with small variations on the inputs around the thresholds.

Random forests [Breiman, 2001] were developed to overcome these limitations of classification and regression trees, reducing the variance by averaging a set of classification trees constructed with different subsets of data to make them somehow independent. It started to be used in avalanche research fields in years 2010 [e.g. Mitterer and Schweizer, 2013; Möhle et al., 2014; Marienthal et al., 2015; Dreier et al., 2016; Chawla and Singh, 2021; Mayer et al., 2022]. It is a popular method as it allows for non-linear separations between classes, with largely better stability than classification trees, and remains simple to understand as it remains a family of decision trees.

As an alternative, generalization of linear methods such as support vector machine (SVM) provides linear separations with a different optimization method than LDA, but not directly in the space of input

variables. The problem is first transposed in an alternative space, defined by the kernel of the scalar product. This trick (the so-called kernel trick) allows for working in a space with more dimensions than the original input parameter space, potentially infinite. In this new space, groups of data may be linearly separated, whereas they are not linearly separable directly in the original space. Working in a different space which is generally not describable, this method does not allow for a straightforward interpretation of its internal mechanics. It is less present in the snow community but was still tested [e.g. Pozdnoukhov et al., 2008; Choubin et al., 2019; Sielenou et al., 2021]. Generalization of the linear regression through logistic regressions was also experimented in the avalanche community [e.g. Mosavi et al., 2020].

Finally, convolutional neural networks [LeCun et al., 1998] have become more and more popular in the last years, and snow avalanche research does not entirely withstand this trend, with some prospective studies [e.g. Singh and Ganju, 2008; Dekanova et al., 2018].

In this work, we will focus on Random forest models, as they have been shown to be a relevant method for avalanche activity forecasting [Sielenou et al., 2021] when compared to other methods.

Target of the prediction

This research focuses on the context of avalanche forecasting, from nowcasting to forecasting in 1 to 4 days. The most common use of statistical methods is the relationship between avalanche activity and meteorological and snow conditions at the daily or sub-daily time scale [e.g. Navarre et al., 1987; Buser, 1983; Mayer et al., 2022].

However, what is called avalanche activity remains to define. It can be represented as a binary classification [Floyer and McClung, 2003; Navarre et al., 1987; Hendrikx et al., 2005, 2014]. The two classes of binary classification can be adjacent or rather closer to two more extreme classes [e.g. Sielenou et al., 2021]. In cases when the goal is to separate extreme events from more common avalanche activity, several classes can also be used [e.g. Sielenou et al., 2021]. The several classes can also describe different types of avalanche activity (e.g. dry and wet) [Bovis, 1977]. A step further, to overcome the concept of classes completely, is to use a continuous indicator [e.g. Davis et al., 1999; Dreier et al., 2016]. In the snow avalanche community, continuous values are often interpreted as a probability of avalanche occurrence [e.g. Föhn et al., 1977; McGregor, 1989]. We will focus in a first time on binary classification but also use continuous value in the present work.

Variables used as input for the forecast

Depending on the goal and available information, different data can be used as an input for statistical models. Inputs are usually classified into three groups: meteorological, snowpack and stability variables [McClung, 2023]. The first one is the easiest to get (whether it is measured or computed), but meteorological variables are not directly linked to processes of avalanche formation: complex models representing numerous processes are required to represent the snowpack evolution from only meteorological forcing [e.g. Vionnet et al., 2012]. Snowpack variables mainly gather bulk variables that are a first measurement on the snowpack that is not focused on the stability markers, whereas stability variables are designed to represent more directly processes involved in avalanche formation. Stability variables include both expert combinations of other variables that are expected to be of interest, such as wind drift indices [e.g. Hendrikx et al., 2014] or results of mechanical models.

The first models mainly gather meteorological or easy to measure bulk snowpack variables such as snow depth, the temperature of snow surface or penetration in specified conditions (foot, cone penetration) [e.g. Föhn et al., 1977; McGregor, 1989; Bois and Obled, 1976; Bovis, 1977]. First improvements of the first methods come with an increase in the number of variables. It includes more information from snowpack [e.g. Obled and Good, 1980] and computation of derivatives of snow conditions (such as snow depth differences) or accumulated meteorological conditions (sum of temperatures or precipitation on several days) [e.g. Bovis, 1977; Navarre et al., 1987; Kronholm et al., 2006]. Only a few studies introduce stability variables in combination with statistical models [e.g. Mayer et al., 2022]. In this work, we focus on the added value of stability variables.

Spatial scale

The target area and time resolution also have large impacts on the chosen method and variables of interest. In forecasting context, the time scale is well defined and ranges from hours to one day. On the contrary the spatial scale ranges from hundreds of squared kilometers [e.g. McGregor, 1989; Hendrikx et al., 2014; Chawla and Singh, 2021] to very local areas (typically one squared kilometer or a single avalanche path) [Mitterer and Schweizer, 2013; Mayer et al., 2022] by way of relatively homogeneous areas of few hundreds squared kilometers [e.g. Bois and Obled, 1976; Kronholm et al., 2006; Sielenou et al.,

2021]. With reduced areas of interest, it is possible to associate specific input variables. Avalanches occur when an unstable or metastable snowpack leaves the stable state to flow down the slope. However, many ingredients of instability may be present while no avalanche release, especially when a small trigger would be required. It is then difficult to identify avalanche-prone situations by observation without endangering the observer. On the contrary, in larger areas, the larger dataset allow for easier statistical analysis. This second approach is more designed for identifying critical snowfalls for large natural avalanche activity. In contrast, the more local approach provides insights for local avalanche forecasting and the possibility of evaluating the spatial distribution of avalanche hazard, which is one of the keys to avalanche hazard [Statham et al., 2018]. We evaluate prediction performance at different spatial scales in this thesis.

Evaluations of machine learning model predictions

In machine learning science, best practices for prediction skills evaluation have emerged. In terms of methods, evaluation processes are now clearly defined, with two independent datasets: train and test set [Hastie et al., 2009]. A third set is also used for hyperparameter optimization. To compare models to each other, classical datasets allow for intercomparison of the results. Several classical datasets exist for each problem type, such as the Iris dataset (classification of Iris varieties given flower parameters [Fisher, 1936]) or MNIST (image classification of handwritten digits [LeCun et al., 1998]). The snow and avalanche community has not yet agreed on shared databases to evaluate snow avalanche related models.

In snow-related problems, a random drawing of observations to define a train and a test set, as usually done in machine learning problems, is insufficient to ensure the independence of observations. Indeed, snow evolves slowly and stores the history of past conditions in its layers. Consequently, using classical techniques to separate train and test sets may result in having one day in the train set and the following one in the test set, which is not independent in general and may result in overestimating the algorithm skills. However, it is still sometimes used because it is easier to use as the methods were intensively developed by machine learning researchers. It is, for instance, the case for the out of bag method (OOB) in random forest classifiers that allow for a straightforward evaluation on an independent set in the general case but that mix non-independent observations in the train and the test set when applied to avalanche related problems [Marienthal et al., 2015; Sielenou et al., 2021].

Moreover, there is not yet a classical dataset to compare scores when dealing with snow and avalanche related questions. As we see above, there is a large variety of questions that may be answered by machine learning models (in terms of scale, variables of interest, and precise goal). Each one requires specific datasets. This is not a problem by itself, as each study will aim at using the best data available for its precise goal. However, it limits the ability to compare related studies. For instance, Mayer et al. [2022] use a test set with well-defined situations (completely stable or highly unstable, based on rutschblock experiments), whereas other studies use the whole chronicle of winter [e.g. Sielenou et al., 2021], including many days of intermediate stability. Both approaches have advantages: the first one allows for identifying critical days for which misclassification should be avoided, whereas the second one is more representative of the operational use. However, it precludes comparing the results directly as the first option will likely present higher scores due to the absence of intermediate situations that are more difficult to classify.

Depending on the goal and tolerance to the different misclassification, different scores are of interest. The overall score or accuracy, which is the part of correctly classified situations, is the most obvious for a first impression. However, it can also be misleading when dealing with relatively rare phenomena such as avalanches. In that case, a classifier answering invariably "no avalanche" would have a high score but is of no interest. That is why the full confusion matrix or other scores are of interest to judge the interest of the method. To this end, classical scores include true positive rate or recall (TPR, part of positive events that are correctly classified), false positive rate (FPR, part of negative events that are misclassified, also reported as true negative rate or TNR that is the contrary), precision (part of correct positive predictions) or balanced accuracy (the mean TPR and TNR). However, here also, there is not yet a consensus on a minimal set to report. Older studies focus on overall score [e.g. Föhn et al., 1977; McGregor, 1989; Bois and Obled, 1976] but more recent ones do not systematically enlarge the panel of reported scores [e.g. Floyer and McClung, 2003; Sielenou et al., 2021] whereas others give the whole confusion matrix or compare different scores [e.g. Hendrikx et al., 2014; Dreier et al., 2016; Choubin et al., 2019; Mayer et al., 2022]. Ebert and Milne [2022] identified a set of scores relevant for rare events such as avalanches that may provide a basis for future studies. However, the current wide variety of reported scores makes them once again difficult to compare.

Further comparison of existing literature

To facilitate the comparison of the different models together, we present in Appendix A a synthetic overview of literature related to machine-learning methods for short-term avalanche hazard forecasting.

1.3 Scientific questions

The main objective of this thesis is to predict avalanche activity with snow cover models by combining two approaches: mechanical analysis of snow profiles and machine learning on avalanche observations. The final goal is to provide decision support system for avalanche forecasters, adapted to the snow cover models, and with better performances than existing tools.

Several mechanically-based indicators of instability have been developed since the pioneering works of Roch [1966a]. These stability indices were firstly developed to analyze field measurements [e.g. Föhn et al., 1977] or to study mechanical processes in the snowpack during avalanche release [e.g. Gaume et al., 2017]. They were therefore used in combination with snow cover models to analyze snow profiles. However, this application is not straightforward as snow cover models do not represent the wide variety of mechanical parameters that can be measured in the field [e.g. Roch, 1966b] or modelled by mechanical models [e.g. Wautier et al., 2015]. The scientific question here is to choose the appropriate indices to represent avalanche activity in combination with a snow cover model (Crocus in this work) and evaluate and adapt the behavior of the selected indices to the snow model Crocus which initially contains only natural and skier strength-stress ratios.

In a second step, selected stability indicators will be used in combination with machine learning methods to provide an integrated indicator of expected snow avalanche activity. We showed in previous section that numerous machine learning methods were used in the context of prediction of avalanche activity. However, stability indicators were not much used to summarize the information of snow cover models coherently with known processes of avalanche formation before applying statistical models. The second main objective of this work is thus to fill this gap and evaluate the interest of using mechanically-based stability indices in machine learning models for avalanche activity forecasting.

Finally, a careful evaluation is conducted on the results of the statistical model in the context of usability for operational avalanche forecasting services. It includes in-depth evaluation with a focus on operational use, that is to say a precise characterization of the results behind the overall score, especially in critical situations. It also includes the extension to different sources of data and different geographic areas. These evaluations contribute to give an idea of the model usefulness outside of the sandbox of the well defined test conditions.

To summarize, this thesis focuses on the following scientific questions:

- *Which mechanically-based stability indices are of interest for identifying avalanche-prone situations when using with snow cover models in the current state of the art?* Most of the mechanical analyses of snowpack stability were not initially developed for use with snow cover models. We first reviewed the literature for identifying the relevant mechanical theories that are well suited to be used with snow cover models (in terms of vertical resolution of the processes represented, scale, numerical cost, etc.).

- *How to adapt selected mechanically-based indices to the snow cover model Crocus?* Snow cover models do not yet represent the evolution of the mechanical behavior of the snow layers and approximations in the representation of physical processes (e.g. liquid water percolation) do not allow for a direct application of mechanically-based stability indices after snow cover models. We propose a set of indices as well as parameterizations adapted to the snow cover model and illustrate the usefulness of these models on typical situations as well as the uncertainty they carry.

- *How physical indices can complement machine learning methods for avalanche activity forecasting?* Machine learning algorithms already have been widely tested but mainly with meteorological and bulk snowpack information. The goal is to use physical indices to complement available information and evaluate the performance of the resulting model and the interest of the use of pre-processed physically-based variables in a statistical approach.

- *What can a machine learning model bring to the forecaster when characterizing the expected avalanche activity?* Except the nearest neighbor developed in ski resorts in the 80s, machine learning methods are not often used in operational services. In France, the current avalanche diagnostic available to forecasters is the MEPRA module. We then compare the constructed model to the existing information available to forecasters and carefully investigate the behavior of the model in critical situations.

- *How do these models extend from the sandbox of the researcher to the large variety of input data and geographical areas?* The models are generally trained and evaluated on restricted areas where they are intercompared. After the development and tuning of such models, and before any operational use, it remains crucial to check the behavior of the model in less controlled environments. We evaluate the performance of the models with other sources of data and in other geographical areas to validate the reproducibility of the results in slightly different situations.

1.4 Structure of the manuscript

Following this introduction, this thesis combines a series of manuscripts that are either articles or results presented at conferences either published, under review or under preparation for publication. As every research work, it includes collaborations and then co-authorship. However, for all the content of the main chapters of the thesis, numerical developments and experiments, analysis and writing were mainly done by myself. All chapters, except Chapter 4, which is closely linked to the third one, can be read independently. As a consequence, the manuscript inevitably contains a few repetitions.

Chapter 2 Snow cover modeling provides information otherwise unavailable on the present and future state of the snow cover and can be used to evaluate snowpack stability. The main goal of this chapter is to summarize the broad spectrum of existing models to assess snowpack stability from simulated snow profiles and analyze their behaviour in specific practical situations through examples. The basic mechanical concepts behind these stability models include the maximum stress criterion, which characterizes the failure initiation propensity, and the critical crack length to evaluate the crack propagation propensity. However, many subtle differences between models, mainly due to additional expert rules or the effective implementation of the concepts, can be confusing. We try to disentangle this diversity. We discuss the differences and present an overview of the mechanical parameterizations of snow material properties, such as strength or stiffness, as they are a key ingredient for stability modeling. In addition, we apply the stability models to typical and simplified snow profiles in order to illustrate the influence of the underlying assumptions and the model's sensitivity to the mechanical input.

This work was published as Viallon-Galinier L., Hagenmuller P., Reuter B. and Eckert N. (2022), *Modeling snowpack stability from simulated snow stratigraphy: Summary and implementation examples*, Cold Regions Science and Technology (201), doi:10.1016/j.coldregions.2022.103596.

Chapter 3 Several numerical and statistical methods were used to forecast avalanche activity. However, it remains unclear how the combination of avalanche processes knowledge, mechanical analysis of snow profiles and observed avalanche data improves avalanche activity prediction. In this chapter, we combine extensive snow cover and snow stability simulations with observed avalanche occurrences within a Random Forest approach to predict avalanche days at a spatial resolution corresponding to elevations and aspects of avalanche paths in a given mountain range. We take advantage of a selection of stability indices presented in the first chapter that we introduce in the machine learning model. We then evaluate the benefit of these mechanical and physical stability indices in the context of statistical learning. We also develop a rigorous leave-one-out evaluation procedure, including an independent test set. In a region of the French Alps (Haute-Maurienne) and over the period 1960-2018, we show the added value within the statistical model of considering advanced snow cover modeling and mechanical stability indices instead of using only simple meteorological and bulk information.

This work is under review in The Cryosphere (doi:10.5194/tc-2022-108).

Chapter 4 In the perspective of operational use of statistical models, chapter 4 is dedicated to in-depth and sensitivity analyses of the model presented in the previous chapter. First of all, the results are compared to the results of the current operational system MEPRA. In addition to the reliability of the system, the spatial scale of the results and the correct identification of days with high avalanche activity are confirmed.

These results were accepted for presentation at the international symposium on snow, International Glaciological Society, Davos, 2022.

Chapter 5 In the two previous chapters, the model was trained and evaluated on avalanche observations from Enquête Permanente Avalanche (EPA) in the specific region of Haute-Maurienne. This last chapter goes one step further by expanding the methodology to other areas and data sources. The interest of avalanche activity reports by the Meteo-France network of observers is highlighted, by demonstrating

similar predictive performances compared to the one obtained when the EPA is used as avalanche observation source. Four different areas are then selected to compare the results with the Haute-Maurienne area. We show that similar results can be obtained for all areas, with some underperformance for areas with less observations. These results allow for a first generalization of the algorithm and give confidence in the possibility of applying it in any area providing data is available for algorithm learning.

This work is planned to be submitted in Natural Hazards and Earth System Sciences.

Chapter 6 The last chapter summarizes the main results of this work and opens perspectives for future research.

In addition to these main chapters, a collection of appendices is assembled. They gather additional details that we refer to from previous chapters or publications in which I am not alone as the main contributor and that give a different or complementary analysis and point of view of this work without directly answering the primary goal of the thesis.

Appendix A gives a detailed overview of the literature on statistical tools applied to forecasting avalanche hazard. It details, for each publication, the statistical method involved, the geographical scale, the variables used and briefly present the results and the methods used to compute them. It complements this introduction and allows for putting in perspective this work.

Appendix B is a complete presentation of MEPRA. MEPRA is the expert tool currently at the disposal of operational forecasters in France. Except for the source code that is public, there is not yet an extensive description of its internal structure. As this work is designed to propose an alternative to this system, it helps to understand the existing tool first.

Appendix C presents indices to characterize humid snow situations. Chapter 2 mainly focuses on dry snow indices. For avalanche activity prediction, considering only dry avalanches is insufficient, especially as it is not obvious how to separate datasets between dry and wet snow avalanches. Hence, characterization of wet snow requires to be added to dry snow indices to cover all the avalanche types. Some indices are published in the literature and reminded here, and their application to the snow cover model Crocus is discussed. Some others were not yet published and are explained in detail in this appendix.

Moreover, I participated during this thesis to other publications, for which I am in the main contributors and which are in close relation with the scientific questions of this thesis:

- Sielenou et al. [2021] evaluate different statistical methods to determine expected avalanche activity corresponding to given snow and meteorological conditions. Avalanche activity from Enquête Permanente Avalanche is classified into three classes and importance of the different input variables are studied on the 23 massifs of French Alps, highlighting coherence with current avalanche literature and different behavior depending on the massifs. Most of the developments were carried out by Pascal Dkengne Sielenou, whereas I actively contributed to the redaction of the paper and the review process with Nicolas Eckert. This article, among other goals, validates the use of the Random Forest method for avalanche forecasting based on EPA dataset.

- Reuter et al. [2022b] develop a new expert model for avalanche forecasting providing both information on avalanche-prone conditions and the expected avalanche problem types, based on the analysis of mechanically-based stability indices. I participated mainly to the application of this methodology to the Crocus snow cover model.

- Viallon-Galinier et al. [2021] evaluate the internal properties of the stratigraphy produced by the snow cover model Crocus. In this thesis we consider that the model provides a reasonable representation of the snow layering. However, such models have generally been evaluated on bulk or surface properties, such as snow depth, water equivalent of snow cover or surface albedo, but not on the detailed stratigraphy. When using the snow stratigraphy extensively in mechanical models, it must be checked carefully and corrected if relevant. This side issue has been addressed in an article introducing a new method to (1) directly compare simulated and observed snow layering and (2) allow for the insertion of observed profiles to initialize a snow cover model. It also allow to

quantify the errors of the snow cover model to therefore be able to estimate the errors on mechanical properties, stability indices and predicted avalanche activity.

- Dick et al. [2022] study the effect of Saharan dust deposition on avalanche activity. On the example of the impact on stability of the dust deposition on snow, we study and illustrate the behavior of MEPRA model and validate the use of wet snow instability indicators with Crocus snow cover model. This is linked to this thesis because it illustrates and tests in practical situation wet snow stability indices which are not covered by Chapter 2.

A list of publications resulting from this thesis is available in Appendix D.

Chapter 2

Modeling snowpack stability from simulated snow stratigraphy: summary and implementation examples

This chapter is published as Viallon-Galinier, L., Hagenmuller, P., Reuter, B. and Eckert, N. (2022). Modelling snowpack stability from simulated snow stratigraphy: Summary and implementation examples. Cold Regions Science and Technology.

Contents

Abstract

Information on snowpack stability, i.e., on the propensity for failure initiation and crack propagation in a weak layer, is essential for forecasting snow avalanches. To complement field observations, snow cover modelling provides information otherwise unavailable on the present and future state of the snow cover, and can be used to evaluate snowpack stability. The main goal of this paper is to summarize the broad spectrum of models to assess snowpack stability from simulated snow profiles. The basic mechanical concepts behind these stability models include: the maximum stress criterion which characterizes the failure initiation propensity and the critical crack length to evaluate the crack propagation propensity. However, many subtle differences between models, mainly due to additional expert rules or the effective implementation of the concepts, can be confusing. We try to disentangle this diversity in this summary. We discuss the differences and also present an overview of the mechanical parameterizations of snow material properties such as strength or stiffness as they are a key ingredient for stability modelling. In addition, we apply the stability models to typical and simplified snow profiles in order to illustrate the influence of the underlying assumptions and the model sensitivity to the mechanical input. As we point out scientific challenges and model limitations, the examples we discussed can provide guidance on the interpretation of similar model results.

2.1 Introduction

Avalanches are a significant issue in mountain areas by threatening outdoor recreationists and infrastructure [Wilhelm et al., 2000; Stethem et al., 2003]. Assessing avalanche hazard is therefore important in these areas and, to this end, several countries rely on operational forecasting services [LaChapelle, 1977; Morin et al., 2020].

Avalanche hazard forecasting requires information about the current state of the snowpack, which is the result of the meteorological history, and its future evolution under meteorological conditions [LaChapelle, 1977]. To fulfill these requirements, two main sources of information have been generally used. On one hand, observation networks have been developed with the goal of regularly reporting meteorological conditions and vertical profiles of snow properties including estimations of grain shape and size, density, humidity or temperature [Pahaut and Giraud, 1995; Fierz et al., 2009]. On the other hand, numerical snow cover models [Morin et al., 2020], such as Crocus [Brun et al., 1989; Vionnet et al., 2012], Snowpack [Bartelt and Lehning, 2002; Lehning et al., 2002b,a] or Sntherm [Jordan, 1991], can describe the evolution of physical properties of the snowpack with time. Both approaches aim at providing detailed snowpack stratigraphy, including vertical profiles of physical and mechanical properties, which is the basis for snowpack stability analysis, but they differ in spatial and temporal resolution.

Slab avalanches, whether they release naturally or are artificially triggered, result from a sequence of processes occurring in the snowpack [e.g. Schweizer, 2017]. First a failure initiates in a weak layer, this can happen progressively and lead to natural release or rapidly when caused by an external trigger (e.g. a skier). The initial failure can grow into a crack which will start self-propagating if the crack reaches a critical size. If not arrested by the tensile failure of the slab, the crack then propagates dynamically [Bergfeld et al., 2021b]. Finally, an avalanche releases if the basal friction within the damaged weak layer is insufficient to prevent sliding of the slab on its substratum. The first processes, namely failure initiation and onset of crack propagation, describe the snowpack stability at the point scale, which is paired with spatial information for avalanche forecasting [Statham et al., 2018]. We here focus on snowpack stability at the point scale. Low snowpack stability means the snow layering is prone for failure initiation and crack propagation.

Point stability can be observed in the field with stability tests, such as the compression test [van Herwijnen and Jamieson, 2007], the extended column test [Simenhois and Birkeland, 2009], the rutschblock [Föhn, 1987a] or the propagation saw test [Gauthier and Jamieson, 2008]. Besides, observed snowpack profiles can be interpreted in terms of stability by applying expert rules [Jamieson and Schweizer, 2005; Coléou and Morin, 2018]. A complementary approach is to use models computing instability indicators that describe the processes leading to avalanche release, which would facilitate interpretation of snow cover model output for avalanche hazard assessment and numerical avalanche forecasting [Morin et al., 2020].

Since the pioneering work of Roch [1966a], several mechanical models were developed to assess snowpack stability from either simulated or measured vertical profiles of snow properties. Schweizer [1999] or Schweizer et al. [2003] reviewed the processes involved in avalanche formation. Podolskiy et al. [2013] compiled the different methods used to model a simplified slab structure with the finite element method. However, to the best of our knowledge, there is no recent review of the models used to assess snowpack

stability from snow profiles even if this field has evolved quickly in the past decade.

The aim of this paper is to provide an overview of existing methods to compute stability indicators from detailed snow stratigraphy based on a mechanical analysis. More precisely, we focus here on mechanical models which provide point stability information from modeled snow cover data. Technically speaking, it means that snow layer properties from the snow cover model are used as input of a mechanical model which provides indicators of stability. We only consider snow cover models which aim at representing the whole layering of a snowpack, such as SNOWPACK or Crocus, and mechanical models with demonstrated applicability to this kind of snow cover models.

We provide a snapshot of currently applied mechanical models in Section 2.2. These models generally rely on the information on the mechanical properties of the snow layers. Commonly used parameterizations of the mechanical properties are thus detailed in Section 2.3. Then the main stability models are applied to different typical situations to point out their strengths and limitations in the light of the underlying assumptions (Section 2.4). Finally, we conclude and discuss some guiding lines for future research.

2.2 Stability models

Stability tests mimic the processes involved in avalanche release and have long been used as snow stability information at the point scale [e.g. Föhn, 1987b]. Extending from those concepts, mechanical models have been developed and have improved our understanding of the processes contributing to snow instability (Figure 2.1a). In particular, fracture mechanics helped formalizing the distinction between failure initiation and crack propagation. In this section, we review different models that have been used to characterize snow stability at the point scale.

Snow stability models can be separated into two groups: purely mechanical models and expert models. The first group consists of models that rely on material properties and a mechanical theory. For instance, some failure initiation models derive from the maximum stress criterion, which assumes that a material fails when the stress in a material element exceeds its strength. The second group comprises the so-called expert models. These models have a mechanical basis, but also include empirical thresholds and adjustments which do not derive from a mechanical theory but expert knowledge. For example, the maximum stress criterion can be adjusted by considering differences between the properties of adjacent layers [e.g. Schweizer et al., 2006].

Stability models of each group are listed in Tables 2.1 and 2.2 according to the main mechanical criterion they rely on. In particular, we present the input variables specifically required to run the model (in addition to basic layer properties such as thickness, density and slope angle). The main goal is to refer to the original study which describes the model and its theoretical basis. Besides, we direct the reader to applications of the model to snow cover simulations. Model evaluations are also cited when available and the model computation complexity is roughly estimated. We try to briefly summarize the important computation steps in equations where possible. The notations used are summed up in Section 2.6.

2.2.1 Purely mechanical models

The purely mechanical models are listed in Table 2.1 with references to theoretical work on each model. They are presented according to the main processes they represent: failure initiation or crack propagation. We focus in this section on the concepts of presented models. Practical implementations details are discussed further in Section 2.3.

Failure initiation models

The natural strength-stress ratio S_n, often called natural stability index, is a mechanical criterion comparing the shear strength τ_c of the weak layer to the shear stress τ due to the weight of the overlying snowpack: $S_n = \tau_c/\tau$. This concept was introduced by Roch [1966a] and further evaluated by Föhn [1987a].

A model also taking into account the skier-induced stress $\Delta\tau$ in addition to the snow load was developed by Föhn [1987a]: $S_a = \tau_c/(\tau + \Delta\tau)$. This ratio is referred to as the skier stability index. For consistency, we call it the skier strength-stress ratio. The additional stress due to a skier is generally calculated based on the analytical solution for load distribution in an elastic half space [Boussinesq, 1885], with some simplifications and empirical adjustments based on grain shape or bond characteristics [e.g. Lehning et al., 2004; Giraud et al., 2002]. Finite element models can also capture the full distribution of the skier induced stress in a layered snowpack [e.g. Habermann et al., 2008; Gaume and Reuter, 2017].

In both cases, the calculation of $\Delta\tau$ assumes that snow is an elastic material and, for instance, plastic skier penetration on top of the snowpack cannot be accounted for.

In artificial triggering, an external load is required to fail the weak layer which may subsequently lead to an avalanche. Combining external load and the associated material property (strength) through dimensional analysis, Reuter et al. [2015] introduced a failure initiation criterion directly relating the skier-induced stress $\Delta\tau$ to the shear strength of the weak layer τ_c: $S_r = \tau_c/\Delta\tau$, omitting the stress due to the load of the slab. Here, we call it the external strength-stress ratio.

The damage process preceding natural avalanche release can be slow with the progressive failure of individual bonds in the weak layer. If the rate of bond cracking overcomes the rate of bond healing through sintering, a failure can initiate in the weak layer [Capelli et al., 2018]. To quantify this damage process potentially leading to failure, Lehning et al. [2004] introduced the so-called deformation rate stability index S_d, here called the deformation rate ratio. This index is based on the theoretical work of Nadreau and Michel [1986] and compares the stress at the bond scale σ_b to the bond strength σ_{b_c} (both computed in the model SNOWPACK): $S_d = \sigma_b/\sigma_{b_c}$.

Natural, skier, external and deformation rate ratios (S_n, S_a, S_r and S_d, respectively) describe failure initiation. The ratio S_d also includes the stage prior to formation of initial crack: the progressive damage of bonds into a macroscopic crack. All are oriented as stability indices, that is to say low values are associated to poor stability whereas high values are associated to stable conditions.

Crack propagation models

Failure initiation is required to release an avalanche but it is not sufficient [van Herwijnen and Jamieson, 2007]. Only if the initial crack reaches a critical size, it may become self-propagating. Models describing the conditions at onset of crack propagation employ the concept of the critical crack length [e.g. Anderson, 2017].

Although propagating cracks were observed in the field and concepts suggested [e.g. McClung, 1979], it was only after specific field tests were introduced such as propagation saw test (PST) [Gauthier and Jamieson, 2006] and analyzed [Sigrist and Schweizer, 2007] that crack propagation models were developed for the onset of crack propagation. The onset of crack propagation corresponds to the state where the specific fracture energy of the weak layer equals the energy release rate of the material around. In other words, it refers to the equilibrium between the fracture energy required to extend the crack in the weak layer and the strain energy released in the material surrounding the weak layer that deforms during this process. The models presented here assume that energy release is due to elastic bending and change in potential energy of the slab. The first estimations of critical crack length assumed a homogeneous slab to compute the strain energy release [Sigrist et al., 2006; Heierli et al., 2008; Schweizer et al., 2011]. Accounting for slab layering [Reuter et al., 2015] and enhancing the formulation of strain energy with finite element simulations [van Herwijnen et al., 2016], these authors derived the critical crack length from measured penetration profiles and showed that it is related to the length directly measured in propagation saw tests [Reuter and Schweizer, 2018]. When we calculate the critical crack length, denoted a_c, we use the equations described by Schweizer et al. [2011].

An alternative approach considering shear induced stresses at the crack tip and their influence on the weak layer has been developed by Gaume et al. [2017]. Based on discrete element simulations, they related the critical crack length to the slab and weak layer properties with an analytical formula [see eq. 9 in Gaume et al., 2017]. Note that this formula depends on the thickness of the specific weak layer used in this study and could somehow be related to the collapse height of the weak layer. More recently, this equation was adjusted with correction factors to fit field measurements to snowpack simulations and fill the gap between the layer thickness corresponding to the snow cover model vertical resolution and the one required by Gaume's parameterization [Richter et al., 2019].

Once a crack starts to propagate, a vertical tensile fracture through the slab may still arrest the crack in the weak layer. In order to quantify the capacity of the slab to support crack propagation in the weak layer, the slab tensile criterion T compares the tensile strength of the slab layers to the tensile stress due to slab bending at the onset of crack propagation, i.e. when the critical crack length is reached [Reuter and Schweizer, 2018]. It represents the portion of the slab where the tensile stress exceeds the tensile strength, at the onset of crack propagation.

In the conceptual representation of avalanche formation (Figure 2.1b) the indices a_c, a_g and T refer to the process of crack propagation. Low values of the critical crack length a_c or a_g indicate high crack propagation propensity. Low values of the slab tensile criterion T indicate sufficient support of the slab for crack propagation. Indices describing the dynamic phase of crack propagation haven't been suggested, yet, as the associated theory and measurements are still under development [Bergfeld et al., 2021b].

The required input to calculate the described snow instability indices varies. Whereas the natural strength-stress ratio just requires weak layer strength, slope angle, slab thickness and average slab density, computation of the critical crack length or the slab tensile criterion can require more input variables and can be computationally more expensive. Moreover, the obtained indices also depend on the mechanical material properties computed at the layer scale. Hence, in section 2.3, we briefly discuss different parameterizations that are commonly used with the described snow stability indices.

2.2.2 Expert models

A list of mechanical models including expert knowledge is given in Table 2.2. The expert models are all somehow based on a shear strength-stress ratio representing failure initiation complemented with expert rules related to other contributing factors of avalanche formation such as propagation propensity and to the temporal evolution of the snow cover (Figure 2.1b).

The Structural Stability Index (SSI) is based on the skier ratio S_a, but also considers snow structural properties of the weak layer and its adjacent layers to refine the estimate of snow instability and to help identify weak layers. The skier ratio is adjusted by hardness and grain size differences between the weak layer and the adjacent layers [Schweizer et al., 2006]. Typically, marked vertical differences of structural properties are related to lower stability.

The MEPRA index called "natural hazard" R_{nat} combines the natural strength-stress ratio S_n with empirical rules to an index describing the danger related to natural avalanche release on a 6-level scale. Expert rules concern the amount of new snow and snow wetness, which represent the two main drivers of natural avalanche activity. Moreover, the hazard level is adjusted according to the temporal evolution of the snow cover to account for the short time persistence of natural instabilities [Giraud et al., 2002; Lafaysse et al., 2020].

The MEPRA index called "accidental hazard" R_{acc} blends the skier strength-stress ratio S_a with complex rules into an index describing the danger related to artificial triggering on a 4-level scale. These rules include the identification of a cohesive slab, sitting on a weak layer with typically composed of depth hoar, faceted crystals and precipitation particles and characterized by low values of S_a. The hazard associated with this weak layer is then estimated with additional expert rules [Giraud et al., 2002; Lafaysse et al., 2020].

Table 2.1: Summary of purely mechanical models, sorted by theoretical stability models. The first column presents the model and the original publications. The second column guides the reader to the references with the practical implementations and also lists the required input parameters (bullet points in italics) in addition to slope angle, the density and the thickness of the layers. The last column guides the reader to the evaluation of each model, when existing. Computational costs and implementation complexity is low except for the models including FEM simulations. The notations used are summarized in Section 2.6.

Model	Implementation and input parameters	Evaluation
Natural strength-stress ratio: $S_n = \tau_c/\tau$ Roch [1966a], Föhn [1987b]	**MEPRA** [Giraud and Navarre, 1995; Lafaysse et al., 2020] • *Shear strength* τ_c computed by snow cover model mainly from density and grain shape, with sophisticated adjustments based on sphericity, dendricity, grain size, liquid water content and history [Giraud and Navarre, 1995]. **SNOWPACK** natural stability index or Sn38 [Lehning et al., 2004] • *Shear strength* τ_c computed by snow cover model from density and grain shape [Richter et al., 2019, Figure 2, therein]. The *shear stress* τ is the shear component of the stress derived from the weight of the overlaying layers.	e.g. Nishimura et al. [2005]
Skier strength-stress ratio: $S_a = \tau_c/(\tau + \Delta\tau)$ Roch [1966a], Föhn [1987b]	**MEPRA** [Giraud and Navarre, 1995; Lafaysse et al., 2020]: • *Shear strength* τ_c computed by snow cover model mainly from density and grain shape, with additional adjustments for MEPRA [Giraud and Navarre, 1995] • *Skier induced stress* $\Delta\tau$ computed with an analytic function adapted to the grain shape of slab layers to represent the reduction of stress by hard layers (bridging effect) [Giraud and Navarre, 1995] **SNOWPACK** skier stability index or Sk38 [Lehning et al., 2004]: • *Shear strength* τ_c computed by the SNOWPACK snow cover model with parameterizations of Jamieson and Johnston [2001] • *Skier induced stress* $\Delta\tau$: pre-defined function, taking into account ski penetration according to Jamieson and Johnston [1998] Some refinements to take into account normal stress and bond size dispersion have also been proposed by Lehning et al. [2004].	- Schweizer et al. [2006]; Nishimura et al. [2005]
External strength-stress ratio: $S_r = \tau_c/\Delta\tau$ Reuter et al. [2015]	**SNOWPACK** Failure initiation criterion [Reuter et al., 2022b] • *Shear strength* τ_c from snow cover model [Richter et al., 2019] • *Skier induced stress* $\Delta\tau$: use of FEM for computing additional shear stress	Reuter and Schweizer [2018] Reuter and Bellaire [2018]

Model	Implementation and input parameters	Evaluation
Deformation rate ratio: $S_d = \sigma_b/\sigma_{b_c}$ Nadreau and Michel [1986]	**SNOWPACK** Deformation rate ratio [Lehning et al., 2004] : • *Critical stress in bonds* σ_{b_c} given by the model [Lehning et al., 2002b] • *Stress in bonds* σ_b derived from output of the snow cover model as: $\sigma_b = -p\tan(abs(\dot{\epsilon_b}))\sqrt{\frac{1-p}{p-\sigma_{0_{ice}}}}$ with pressure in bonds p computed as [Nadreau and Michel, 1986] with modelled temperature	-
Critical crack length a_c Sigrist et al. [2006] Heierli et al. [2008]	Critical crack length based on beam theory where a_c is the solution of a polynomial equation [Schweizer et al., 2011, see eq. 4]. van Herwijnen et al. [2016] further developed this approach with FEM to account for size effects. Adapted by Reuter and Bellaire [2018] to be used after the SNOWPACK model. • *Weak layer fracture energy* w_f estimated from modelled shear strength [Reuter and Bellaire, 2018; Gaume et al., 2014]. • *Slab equivalent modulus* E_{eq} determined by an FEM model [Reuter et al., 2015] where each layer is assumed to be elastic with a modulus computed with an exponential law on density [Scapozza, 2004]	Reuter and Bellaire [2018]
Alternative critical crack length: $a_g = \sqrt{\frac{E'DD_{wl}}{G_{wl}}}\frac{-\tau+\sqrt{\tau^2+2\sigma(\tau_c-\tau)}}{\sigma}$ Gaume et al. [2017]	Applied by Schweizer et al. [2016b] and Richter et al. [2019]. Simplified by [Richter et al., 2019] • *Slab equivalent modulus* E_{eq} determined by FEM modelling [Schweizer et al., 2016b] on elastic modulus determined with an exponential low on density [Scapozza, 2004] • *Shear modulus of the weak layer* G_{wl} set at a constant value, 0.5 MPa in Schweizer et al. [2016b]	Richter et al. [2019] Schweizer et al. [2016b]
Slab tensile criterion: T Reuter and Schweizer [2018]	**SNOWPACK** Tensile criterion [Reuter and Schweizer, 2018, see eq. 1], with implementation by Reuter and Bellaire [2018] • *Tensile strength* σ_c estimated with [Jamieson and Johnston, 1990] • *Tensile stress* based on FEM computation to determine tension at onset of crack propagation	Reuter and Bellaire [2018]

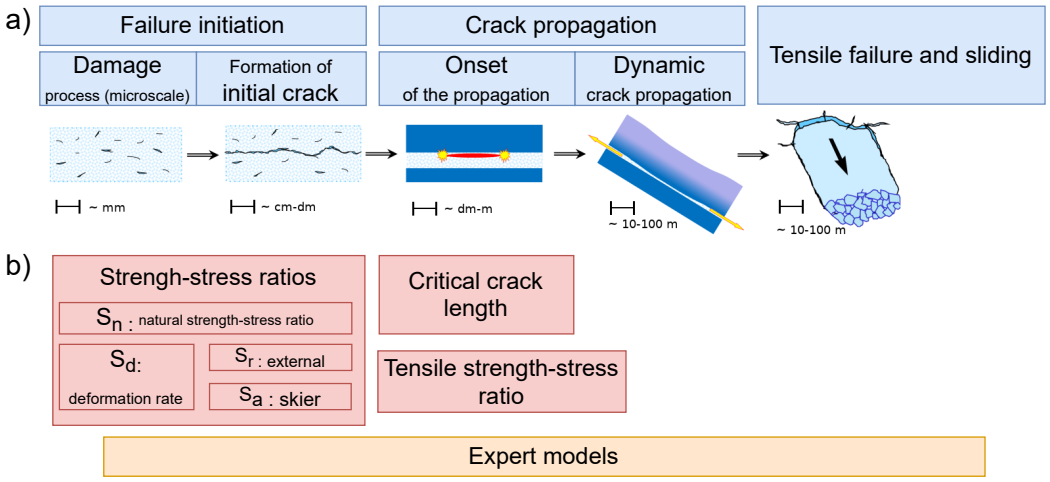


Figure 2.1: (a) Processes involved in avalanche formation according to Schweizer et al. [2016a] and (b) classification of stability models according to the processes they represent.

Table 2.2: Summary of expert stability models

Model	Implementation	Evaluation
MEPRA indices: R_{acc} R_{nat} Giraud et al. [2002]	**MEPRA** hazard indices [Vionnet et al., 2012], based on criteria S and S_a with : • *shear strength*, • *penetration resistance*, • *grain shape*, • *temperature* • *liquid water* all output of the snow cover model for each layer [Vionnet et al., 2012; Lafaysse et al., 2020]	-
SSI stability index Schweizer et al. [2006]	**SNOWPACK** based on Sk38 [Schweizer et al., 2006]. Uses thresholds on variations between adjacent layers of: • *Hardness* • *Grain size* all output of the snow cover model for each layer [Bartelt and Lehning, 2002; Lehning et al., 2002b]	Schweizer et al. [2006]

2.3 Mechanical parameters

Detailed snow cover models such as Crocus or SNOWPACK simulate snow layer properties such as thickness, density, temperature, liquid water content and various grain shape proxies [e.g. Vionnet et al., 2012; Bartelt and Lehning, 2002]. However, the stability models presented in the previous section also require mechanical properties as presented in the second column of Table 2.1. The link between snow cover model properties and mechanical properties such as weak layer strength, or slab equivalent modulus thus rely on additional parameterizations, which are not specific to the considered snow cover model. A variety of parameterizations exists for the same mechanical parameter, as for instance, for the skier-induced stress that represents the bridging effect in Crocus [Giraud et al., 2002] or the skier penetration in SNOWPACK [Lehning et al., 2004] (see Table 2.1). The choice of such parameterizations can have a significant impact on the results of the stability models. Without being exhaustive, we provide an

overview of the diversity of the existing methods to obtain the properties required for computing stability indices (Figure 2.2).

The input to the stability models includes properties simulated by snow cover models (e.g. thickness, density and grain shape), mechanical properties for a particular layer (e.g. layer shear strength, see Figure 2.2a) and mechanical properties depending on many layers of the stratigraphy (e.g. slab equivalent modulus, Figure 2.2b).

2.3.1 Mechanical properties per layer

The mechanical properties required per layer are material properties which do not depend on the surrounding layers. They include elastic and failure properties. They are mainly determined in field or lab measurements that are then used to fit more general parameterizations.

The **elastic modulus** E describes the amount of reversible deformation for a given stress and has been measured with various techniques at different strain rates. Mellor [1974] reported values of modulus measured with different techniques and strain rates from 10^{-6} to $10 \times 10^{-2}\,\mathrm{s}^{-1}$. Camponovo and Schweizer [2001] used a rheological setup to derive elastic shear modulus while Gerling et al. [2017] used wave propagation to measure elastic modulus at high strain rates. Based on these measurements, different parameterizations emerged, mainly as functions of density. They include linear or power relations of density, temperature or strain rate [e.g. Smith, 1965], power laws of density [e.g. van Herwijnen et al., 2016; Gerling et al., 2017] and exponential laws of density [e.g. Scapozza, 2004]. Another way of estimating snow elastic properties is from microtomography images and numerical modelling of the material behaviour from ice and air properties [e.g. Köchle and Schneebeli, 2014; Wautier et al., 2015; Srivastava et al., 2016]. Elastic properties also include the **Poisson's ratio** ν which describes the deformation perpendicular to the loading direction. It is either usually chosen as a constant value in the typical range [0.25, 0.3] [Podolskiy et al., 2013] or determined from a density parameterization [Mellor, 1974; Sigrist et al., 2006].

Strength is a measure of the maximum stress that a material can support before starting to fail. For avalanche formation, considering different loading directions is important: shear or compressive strength of the weak layer for failure initiation and tensile strength of the slab during crack propagation [Roch, 1966a; Bobillier et al., 2021; Reuter and Schweizer, 2018]. Different measurement techniques such as the rotating vane, the shear or tension frame, have been used in the field or in the lab [Mellor, 1974]. Based on these measurements, power laws on density, sometimes specific to some grain shapes, have been determined [Perla et al., 1982; Jamieson and Johnston, 1990, 2001; Chalmers and Jamieson, 2001]. For shear strength, adjustments accounting for normal load have been developed by [Jamieson and Johnston, 1998] and [Zeidler and Jamieson, 2006]. A combination of such parameterizations are currently used in snow cover models [Lehning et al., 2004; Giraud et al., 2002].

Fracture energy corresponds to the energy per unit surface required to grow a crack. It is also related to the fracture toughness which describes the critical stress intensity at the crack tip [Griffith, 1921]. The fracture energy was derived from experimental propagation tests with finite element modelling [Schweizer et al., 2011] or by measuring the deformation of the slab with particle tracking velocimetry [van Herwijnen et al., 2016]. In addition, Reuter and Schweizer [2018] presented an empirical relation relating the snow micropenetrometer force to the fracture energy. In the lab, the fracture toughness was measured on notched beams composed of typical slab snow but not on weak layer snow [Schweizer et al., 2004; Kirchner et al., 2002].

2.3.2 Non local properties

The non-local properties are the **skier induced stress** $\Delta\tau$, the **slab equivalent modulus** E_{eq} and the **shear stress** τ. For their computation, mechanical properties of several layers are important. The skier-induced stress can be computed from an analytical solution [e.g Boussinesq, 1885] or estimated, for instance with a piecewise linear approximation [e.g. McClung and Schweizer, 1999] or with finite elements [e.g. Jones et al., 2006]. The slab equivalent modulus is the result of a homogenization of the mechanical behaviour of the slab, assuming that in the specific loading situation, the slab deforms as a homogeneous material with elastic modulus E_{eq}. An estimate of the slab equivalent modulus can be obtained by an analytic averaging method [Sigrist, 2006; Monti et al., 2016] or by finite element modelling [e.g. Reuter et al., 2015]. Full representation of snow stratigraphy by finite elements is expected to provide more accurate estimates but at the cost of longer computation times compared to bulk average for instance that do not account for order of layers [e.g. Habermann et al., 2008; Monti et al., 2016]. A simple approach is to use the average of properties over layers (taking into account thicknesses) [Sigrist et al., 2006]. The benefit of more complex computations has to be weighted in particular in the light of the

uncertainties carried out by the choice of parameterization of mechanical parameters per layer. The shear stress is derived from the weight of the overlying snowpack, always computed as the sum of the product of thickness and density of the layers above the weak layer and acceleration due to gravity.

2.3.3 Limitations of the parameterizations

The presented parameterizations mainly rely on experimental work and provide straightforward methods to compute mechanical properties from snow cover simulations. However, snow is a very fragile material [Kirchner et al., 2002] which limits sampling and testing, and presents different microstructural patterns due to very active metamorphism [Hagenmuller, 2014]. Therefore, these parameterizations based on experimental work ineluctably have limitations.

Most of the presented parameterizations only rely on density as the main (or unique) descriptor of the snow microstructure [e.g. Keeler and Weeks, 1968; Scapozza, 2004; Schweizer et al., 2011] which cannot fully describe its complexity [Shapiro et al., 1997]. Moreover, the parameterizations often do not take into account temperature, liquid water content or the loading conditions such as the strain rate and the loading direction, which are critical for the material behaviour in some cases [Mellor, 1974; Denoth, 1980; Shapiro et al., 1997].

Some properties are also difficult to measure or even some samples are difficult to carry to the lab, especially for weak snow (precipitation particles, depth hoar or surface hoar, for instance) [Reiweger and Schweizer, 2010; Walters and Adams, 2014]. This limits the number of measurement points on such snow. The measurements used for fitting parameterizations do not cover all snow types. Numerical experiments based on tomographic images can help to characterize fragile snow types [e.g. Hagenmuller et al., 2014, 2015; Mede et al., 2018]. Nevertheless, these tomography-based models are mainly limited to elastic properties and strength values and were simulated for a small number of samples.

Models estimating the crack propagation propensity require the slab equivalent modulus to represent slab deformation. The crack propagation models rely on the idea that, at the onset of crack propagation, the energy for crack extension in the weak layer equals the change of gravitational potential energy and the strain energy corresponding to the deformation of the overlaying slab [e.g. van Herwijnen et al., 2016]. Hence, the commonly made assumption is that the mechanical behavior of the slab can be approximated by linear stress-strain relations with an equivalent modulus [Reuter et al., 2019]. This equivalent modulus must not be related to the elastic modulus of snow as their definitions from a material science point of view differ. Stability indicators may require equivalent modulus representing the whole deformation rather than the well-defined material property. This assumption on the snow deformation regime can also lead to inconsistency in weak layer fracture energies depending on the chosen measurement techniques as discussed in LeBaron and Miller [2016].

2.4 Illustration

We applied the models from Section 2.2 to simplified snow profiles. Comparing their results allows for a short sensitivity analysis to assess how the models may describe the process of avalanche release and to emphasize their underlying assumptions. The goal is neither to provide an exhaustive sensitivity analysis of the different models nor to evaluate the models but we guide the reader to dedicated literature.

2.4.1 Methodology

Selected models

For illustration, we selected a subset of the stability models presented in Section 2.2. We chose models which require simple physical quantities (layer thicknesses, density or hardness...) and do not show expert models. In particular, S_d is based on the grain bond size which is a variable specific to SNOWPACK and was therefore not considered. Overall, we considered three strength-stress ratios namely S_n, S_r, S_a, two critical crack lengths a_c and a_g and the slab tensile criterion T. The onset of crack propagation used for the ratio T is here defined with the critical crack length a_c.

Selected snow profiles

We applied the stability models to a set of snow profiles inspired by Habermann et al. [2008] for a slope angle of 38°. Figure 2.3 presents the basic geometry of these profiles, composed of a slab, a weak layer and a substratum. Each layer is described by its density, hardness and elastic modulus according to Habermann et al. [2008] (see Table 2.3).

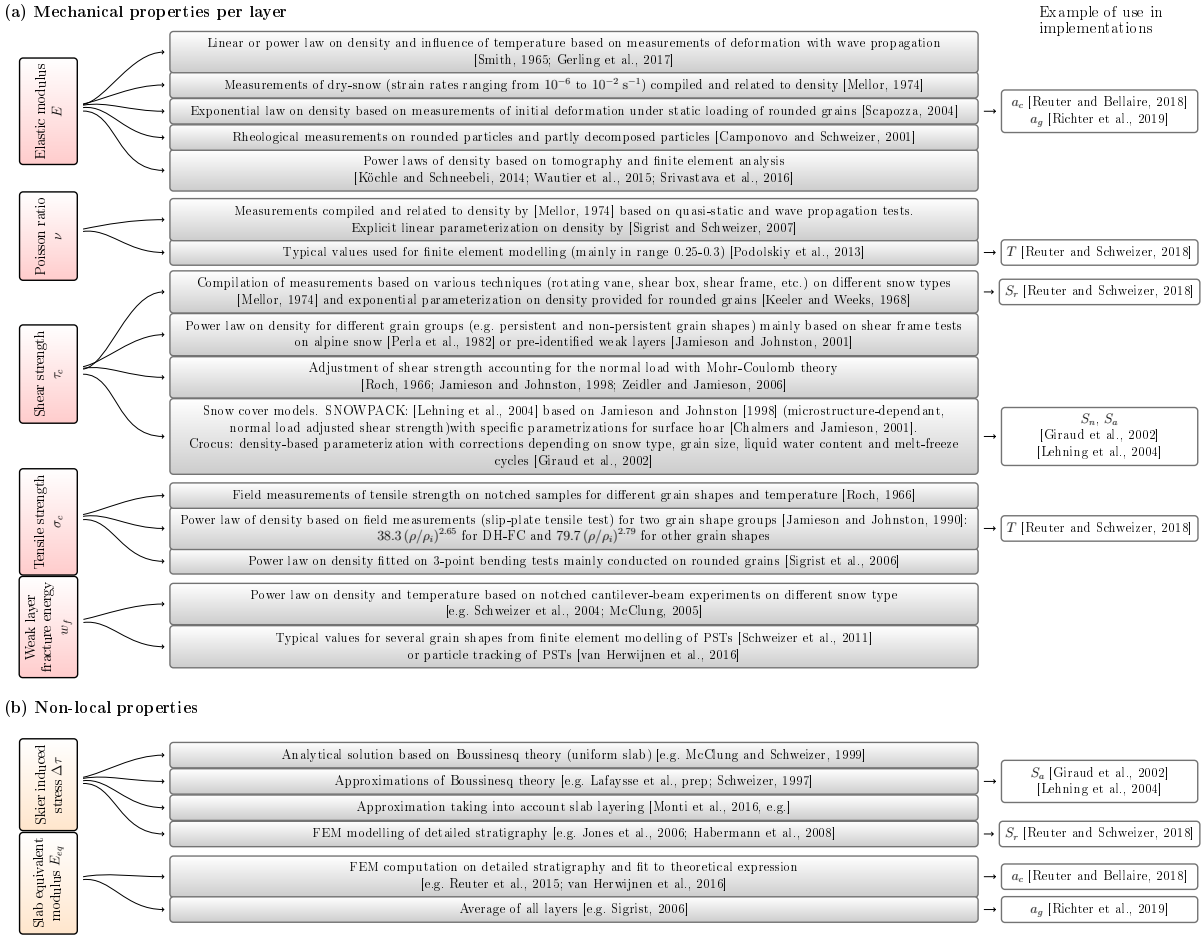


Figure 2.2: Map of the (a) mechanical properties per layer (red) or (b) non-local properties (orange). An overview of associated methods to determine mechanical properties from snow cover model output is given in the second column (grey).The stability models presented in Table 2.1 can be used with any of these mechanical parameterizations. However, the implementations cited in Table 2.1 used specific parameterizations, which are indicated in the last column of this figure.

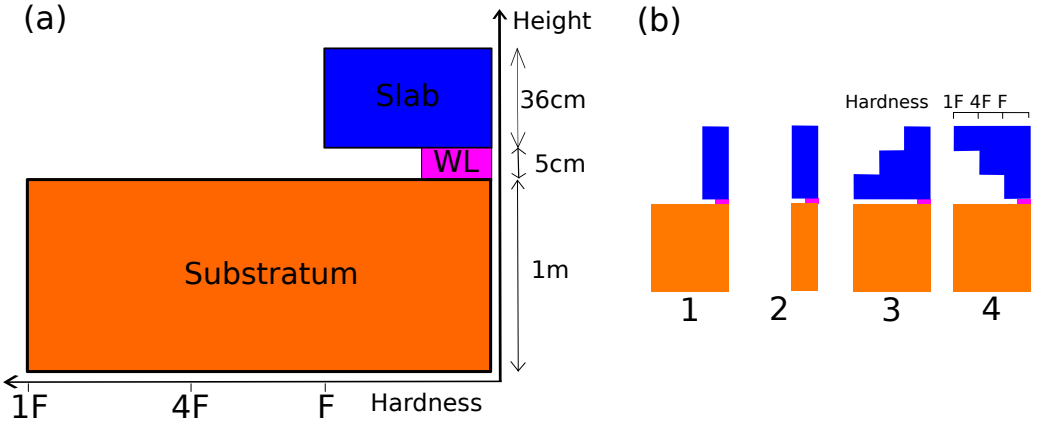


Figure 2.3: (a) Simplified snowpack stratigraphy, including three parts (substratum, weak layer (WL), slab) and (b) four simplified snow profiles from Habermann et al. [2008].

Table 2.3: Table of mechanical properties derived from hardness for each layer.

Layer hardness	Density ($\mathrm{kg\,m^{-3}}$)	Elastic modulus (MPa)
Weak layer (WL)	100	0.15
Soft (F)	120	0.3
Medium (4F)	180	1.5
Hard (1F)	270	7.5

We also considered the temporal evolution of a snow profile after a snowfall. The substratum has the same properties as in Figure 2.3a. On top, a snow layer with a thickness of 5 cm with an initial density of $50\,\mathrm{kg\,m^{-3}}$ acts as a weak layer. Snow falls at a $10\,\mathrm{cm\,h^{-1}}$ rate for 10 hours creating layers with an initial density of $100\,\mathrm{kg\,m^{-3}}$. The snow settles under its own weight. Settlement is described with a viscous law $\frac{\mathrm{d}D}{D} = \frac{-\sigma}{\eta}\mathrm{d}t$ where σ is the normal load on each layer, D the layer thickness and η the viscosity. The viscosity η depends on density: $\eta = \eta_0 \frac{\rho}{c_\eta} e^{b_\eta \rho}$ with $\eta_0 = 7.6\,\mathrm{kg\,s^{-1}\,m}$, $\mathrm{c}_\eta = 250\,\mathrm{kg\,m^{-3}}$ and $\mathrm{b}_\eta = 0.023\,\mathrm{m^3\,/kg}$ as in [Vionnet et al., 2012].

Mechanical properties

To apply stability models other properties are required apart from thicknesses, density and elastic modulus. As shown in Section 2.3, numerous mechanical parameterizations exist. We selected the following ones, based on their simplicity and use in previous implementations:

- The shear strength of the weak layer was set to 500 Pa as this value is in the range measured by Jamieson and Johnston [2001].
- The tensile strength of the slab is derived from Jamieson and Johnston [1990].
- The weak layer fracture energy was set to $0.2\,\mathrm{J\,m^{-2}}$, consistent with the range observed by Schweizer et al. [2009].
- The skier induced stress $\Delta\tau$ was computed according to McClung and Schweizer [1999].
- The equivalent modulus of the slab was estimated by the mean elastic modulus of the slab layers [e.g. Sigrist, 2006].

To illustrate the role of the slab layering (Section 2.4.3), we also computed the skier induced stress $\Delta\tau$ and the slab equivalent modulus E_{eq} with finite element simulations [Reuter and Schweizer, 2018]. The slab equivalent modulus is computed by considering the deformation energy of the slab only and does not include the substrate. The strength-stress ratios S_{r_F} and S_{a_F} were simulated with finite element simulations. The critical crack length using finite element simulations for equivalent modulus is denoted a_{c_F}.

For temporal evolution, all mechanical parameterizations were set as in the previous paragraph except the elastic modulus, shear strength and weak layer fracture energy. The elastic modulus was computed as $E = 1.8 \cdot 10^5 \cdot \exp(\rho/\rho_0)$ (Pa) [Scapozza, 2004] with $\rho_0 = 67\,\mathrm{kg\,m^{-3}}$ and shear strength as $\sigma_c = 14.5 \cdot \frac{\rho}{\rho_{ice}}^{1.73}$ (kPa) [Jamieson and Johnston, 1990]. Given the lack of parametrizations of the weak layer fracture energy, we derived plausible values to describe the temporal evolution. We estimated the weak layer fracture energy as $\alpha \cdot \tau_c^2/E$ as this assumption has provided reasonable results in specific applications (see Birkeland et al. [2019, Equation 4] or Gaume et al. [2014, Equation B7]). The scaling coefficient $\alpha = 0.2\,\mathrm{J\,m^{-2}}$ was chosen so that w_f is in the range observed by Schweizer et al. [2011]. Hence, it can be regarded an order-of-magnitude best guess.

2.4.2 Homogeneous slab

We conducted a sensitivity analysis on the simplified stratigraphy presented in Figure 2.3a to discuss the influence of mechanical parameters in each stability model. This stratigraphy is used as a reference and six parameters (slab thickness, slab equivalent modulus, weak layer shear strength, weak layer fracture energy, weak layer thickness and weak layer elastic modulus) were altered one after the other. By doing so, we did not account for possible correlation between the different parameters. For instance, we do not consider that, in the field, a thick slab is on average denser and stiffer than a thinner slab. However, this is on purpose to discuss the processes that are represented or not by the different models and the influence of mechanical parameters on their results. The sensitivity analysis is first presented for

the shear strength-stress ratios mainly representing the failure initiation processes, then for the critical crack length representing the crack propagation propensity and eventually for the slab tensile criterion (representing potential slab fracture breaking off crack propagation).

Shear strength-stress ratios are, as expected, affected by slab thickness and weak layer shear strength (Figure 2.4), in addition to the density of the slab (not shown). The shear strength-stress ratios S_n (natural), S_a (skier) and S_r (external) are sensitive to the slab thickness. A thicker slab means a heavier slab and consequently, a higher stress is induced in the weak layer, so a lower value of S_n is obtained. In contrast, the skier induced stress in the weak layer is reduced by a thicker slab and consequently leads to a higher value for S_r. The skier ratio S_a combines the two effects and hence presents a non-monotonic evolution with a maximum for a thickness of 0.5 m. The shear strength-stress ratios are proportional to the weak layer shear strength: they increase linearly with this parameter. The other parameters, weak layer thickness and weak layer stiffness, do not affect the shear strength-stress ratios. The natural ratio S_n is by definition only sensitive to the load of the overlying slab (stress) and to the strength of the weak layer. In contrast, the skier induced stress can be affected by the layering of the elastic modulus as shown by Habermann et al. [2008]. However, in this sensitivity analysis we used an analytical expression for skier induced stress which only depends on depth and slope angle [McClung and Schweizer, 1999] and thus cannot capture the effect of layering.

The critical crack lengths, a_c and a_g, decrease with increasing slab thickness as this parameter affects slab bending and the subsequent stress concentration at the crack tip. The other way around, the critical crack length increases with slab equivalent modulus as stiff slabs deform less. Low values of weak layer strength ease failure initiation and facilitates crack propagation: a higher strength yields a larger critical crack length. Indeed, weak layer strength explicitly appears in the expression of a_g. However, a_c is not affected by weak layer shear strength as the mechanical parameter used in this stability model is the weak layer fracture energy. Indeed, a_c increases with weak layer fracture energy.

The weak layer thickness is also taken into account to compute a_g. When using a snow cover model, this thickness does not refer to the simulated layer thickness (vertical resolution of the model) but rather to a "useful thickness" [Richter et al., 2019], probably related to the zone where the strain accumulates [Walters and Adams, 2014]. Moreover, the critical crack lengths decrease with the weak layer elastic modulus, as the modulus affects the stress concentration at the crack tip. It is an explicit parameter in the expression of a_g and this effect is also indirectly captured by a_c with the chosen parameterization of the weak layer fracture energy w_f.

The slab tensile criterion T describes the ratio between tensile stress and strength in the slab at the onset of crack propagation, when critical crack length is strictly positive. It is highly related to the value a_c: the shorter the critical crack length, the smaller the unsupported part of the slab and the smaller the deformation. That means we obtain low values of tensile stress and low values of T.

2.4.3 Layered slab

We applied the different stability models to snow profiles with different slab layering (Figure 2.5). We first present the results for failure initiation, then for crack propagation.

The natural ratio S_n is only sensitive to the load of the slab. Values are the same for the profiles 1 and 2 and for the profiles 3 and 4 but the load is larger for the profiles 3 and 4 compared to the profiles 1 and 2. The external ratio S_r is only dependent on slab thickness since the skier stress distribution is computed with a very simple analytical expression which neglects the potential impact of layering [McClung and Schweizer, 1999]. The value of S_r is thus the same for all presented profiles. In contrast, the external ratio S_{rF} relies on finite element simulations and is able to account for the impact of snow layering, including the substratum [e.g. Habermann et al., 2008; Thumlert and Jamieson, 2014]. The value of S_r is slightly lower for a stiffer substratum (profile 1 compared to profile 2) and significantly lower for density increasing with depth in the slab (profile 3) in contrast to a slab density decreasing with depth (profile 4). In the latter case, note that the model for S_r does not account for any potential skier penetration in the snowpack. Therefore, the effect of a hard layer at the surface might still be under-estimated by this model [Habermann et al., 2008]. The skier ratios S_a and S_{aF} exhibit little variations, with the slab layering or with the elastic modulus of the substratum, with values ranging between 0.5 and 0.6 for the provided examples.

The critical crack length a_g is related to both slab equivalent modulus and load. Layering is not accounted for since the slab modulus is here simply computed as the average of all layers. The values of a_g appear to be almost the same in all profiles. Indeed, the effects of additional load are here counterbalanced by the effects of an increased stiffness. In contrast, a_c exhibits more variations with the different snow profiles. As expected, the values of the critical crack length do not depend on the substratum (profiles 1 and 2) as it is not represented in the models. Moreover, slab properties are

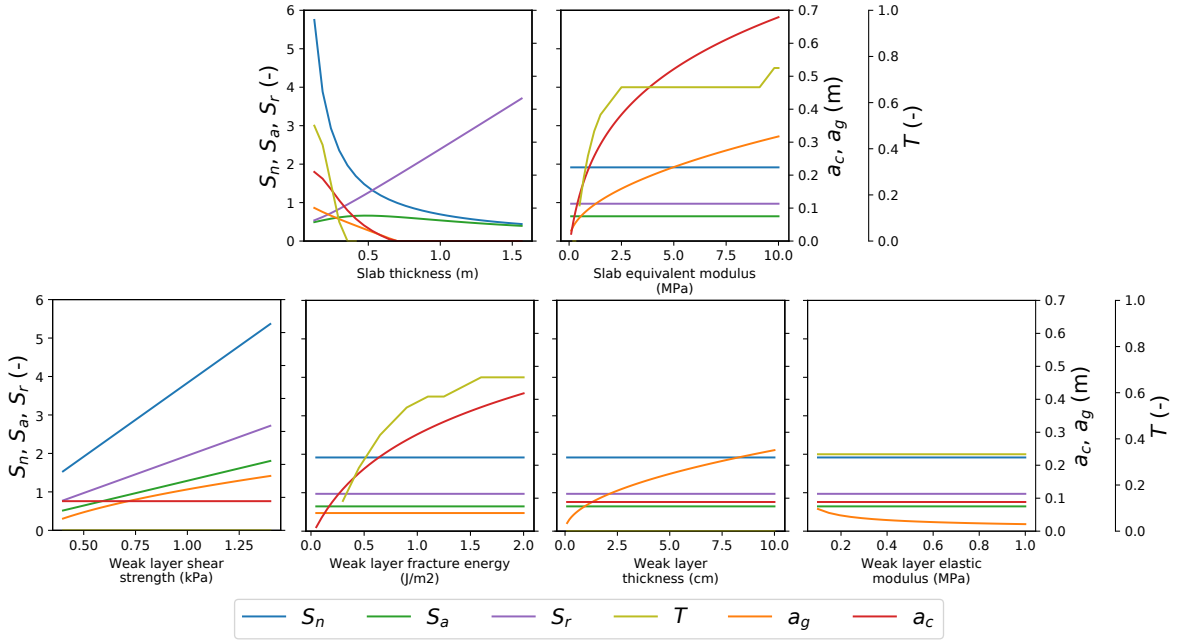


Figure 2.4: Sensitivity analysis of the stability models to their mechanical parameters on the simplified snow profile described on Figure 2.3a. Legend names refer to models described in Table 2.1.

averaged in the computation of a_c and a_g, and thus the models do not represent the layering of the snowpack (profiles 3 and 4). Yet, a finite element simulation accounts for the layering of the slab in the computation of a_{cF}. Noticeably, in profile 4, where density decreases with depth, the critical crack length a_c is larger by a factor of 2.8 compared to a_{cF}, which is in agreement with the findings of Gaume and Reuter [2017]. In profile 3, the critical crack lengths a_c and a_{cF} are very close, which shows that in this case averaging the elastic modulus to define the slab equivalent modulus may be sufficient. The values of a_c and a_g, i.e. the two different models of the crack length, are very different for profiles 3 and 4 ($a_c = 26\,cm$, $a_g = 5\,cm$). Sensitivity studies on a representative set of profiles help to assess how important it is to explicitly account for this layering, as done by e.g. Habermann et al. [2008] or Monti et al. [2016] for skier induced stress and Reuter and Schweizer [2018] for critical crack length. The slab tensile criterion T computed with FEM shows larger variations between the profiles. Being representative of the tensile stress in the slab, which often concentrates near the surface, T has a higher value for profile 3 with lower strength at surface than profile 4, but it is also higher than for a uniform slab (profiles 1 and 2, value of zero) as a higher elastic modulus for one layer will concentrate the stress. The value is also correlated to a_c: higher values are linked to higher values of a_c.

2.4.4 Temporal evolution

Finally, we applied the selected stability models to a typical new snow situation. The slab is initially composed of new snow falling on a non-persistent weak layer composed of very low density snow. The time-evolution of the snowpack is studied up to five days after the snowfall (Figure 2.6).

The stress on the weak layer due to the new snow increases during the snowfall. When the snowfall ends, the stress remains constant, but other snowpack properties continue to evolve with time as snow settles: thickness decreases while density and strength increase, as measured by Roch [1966b]. Consequently, the natural ratio S_n decreases during snowfall and then increases with weak layer hardening. By contrast, S_r and S_a increase during snowfall due to the increasing distance between the additional load (applied at surface) and the weak layer. This distance is then slightly reduced due to the settlement leading to a reduction of the increase rate of the values of S_a and S_r. In a real situation, the skier may penetrate the snowpack and the effective skier load may be higher at beginning due to skier penetration, which is not represented here [Schweizer and Reuter, 2015; Monti et al., 2016; Thumlert and Jamieson, 2014].

Both models of the critical crack length show decreasing values during snowfall due to the increased stress on the weak layer. After the snowfall, the values of the critical crack length increase, representing a stabilization which is mainly a consequence of snow settling leading to the stiffening of the slab and

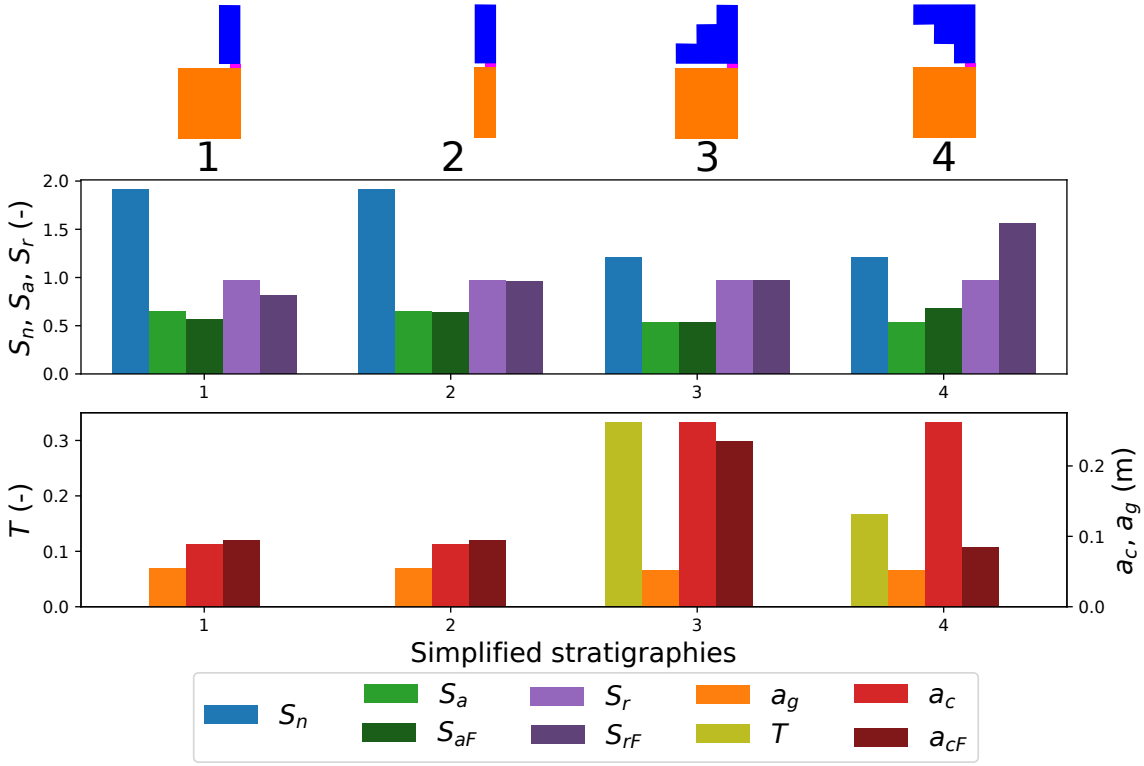


Figure 2.5: Sensitivity analysis of the stability models to different profiles presenting a heterogeneous layering. Legend names refer to models described in Table 2.1.

the strengthening of the weak layer. The stabilization following snow storms, with modelled settling, is faster immediately after the snowfall than later when processes slow down, which is consistent with observations of Birkeland et al. [2019]. The values of T remain very close to zero, which indicates that the tensile stresses due to a crack of length a_c in the slab never exceed the tensile strength of the different layers. In other words, it means that tensile failure in the slab would be unlikely and crack would rather tend to keep propagating than be arrested.

2.5 Concluding remarks

We compiled different mechanical models which were developed to assess snowpack stability from simulated snow profiles. Two main mechanical concepts are behind the stability models: maximum stress criterion characterizes the failure initiation propensity and Griffith's energy criterion based on weak layer fracture energy allows to evaluate crack propagation propensity. However, model implementations present many subtle differences and the results are not all directly comparable, as models often complement each other (Figure 2.1). Moreover, some empirical rules are often included or different implementations of the mechanical properties co-exist. The main goal of this review was to explain the differences in a synthesis including the references to the relevant literature (Tables 2.1 and 2.2) and illustrative examples. We highlight the sensitivity of the stability models to the mechanical input such as strength or stiffness and thus also present an overview of the available mechanical parameterizations (Figure 2.2).

We provided an overview of the current state of research on snowpack stability assessment based on snow cover modelling, but also point out some scientific challenges and draw some guiding lines for future research.

It appears that all stability models share the same assumption that snow behaves as an elastic brittle material which can be represented with continuum and fracture mechanics. However, it has long been known that snow also exhibits visco-plastic behaviour [e.g. Narita, 1980]. In practice, the assumption of brittle elasticity makes sense for many of the processes involved in avalanche formation but may still limit the scope of the presented models both for failure initiation and crack propagation. In particular, an elastic model cannot correctly represent skier penetration and thus may fail to correctly reproduce skier-induced stress. Empirical adjustments based on snow depth and density have been developed [e.g.

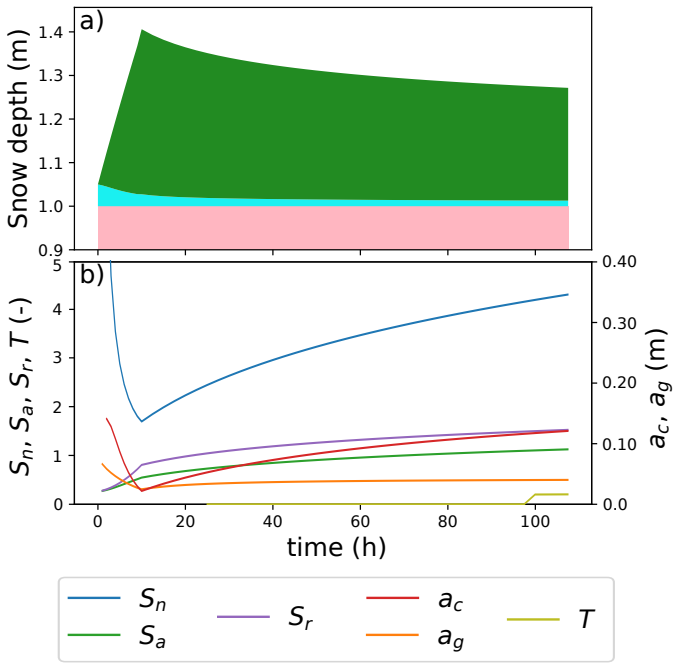


Figure 2.6: Temporal evolution of the snow depth (a) and of the associated stability models (b) during and after a snowfall. The slab (dark green) is composed of new snow with a moderate density (initial value of 100 kg m^{-3}). The weak layer (blue) is composed of very light snow (initial density of 50 kg m^{-3}). The slab and the weak layer evolve with time due to snow settlement. The substratum (pink) is composed of hard snow and do not evolve with time. Legend names refer to models described in Table 2.1.

Schweizer and Reuter, 2015; Reuter et al., 2015] but the associated theoretical framework remains to be evaluated. Besides, all crack propagation models presented above assume that the slab deforms in an elastic manner whereas energy may dissipate notably near the crack tip or in the slab itself. Even if the slab modulus may be adjusted to an effective value to reproduce the observed behavior, the chosen elastic framework can lead to discrepancies between the input parameters of the stability models and the measured values of material properties. For instance, the weak layer fracture energy ranges between 0.2 and 2.2 J m^{-2} when derived from propagation saw tests [van Herwijnen and Heierli, 2010; Schweizer et al., 2011; Bergfeld et al., 2021b] but is estimated around 0.05 J m^{-2} if derived from ice properties and 3D microstructure [LeBaron and Miller, 2016]. This discrepancy is not a problem in itself but limits stability models to benefit from broader studies focusing on snow as a material.

To represent the different processes of avalanche formation that contribute to point stability, several indices have to be combined (Figure 2.1). At least indices for failure initiation and crack propagation have to be considered simultaneously. Several studies proposed possible combination of stability indices, such as Reuter et al. [2015] who combined S_r and a_c or Reuter and Schweizer [2018] who added T. Gaume and Reuter [2017]; Rosendahl and Weißgraeber [2019b] merged a failure stress condition and an energy condition for crack propagation. Besides, the presented stability models focus on failure initiation and on the onset of crack propagation, which corresponds only to the central part of the series of mechanical processes leading to avalanche release (Figure 2.1). So far, less attention was paid to the progressive damaging of the weak layer which can lead to the creation of an initial crack. After Conway and Wilbour [1999] had presented a time derivative based on natural strength-stress ratio, their model has long not been used with snow cover models, but very recently implemented to assess natural release from snow cover modelling [Reuter et al., 2022b]. Recent studies have investigated the competition between bond breaking and the healing due to sintering [Capelli et al., 2018] and their translation into material visco-plasticity [e.g. Puzrin et al., 2019]. Explicitly adding time dependent processes in the stability models can eventually help to pinpoint natural avalanche activity and improve existing models [Brown and Jamieson, 2008]. At the very end of the series of processes, dynamic crack propagation has seen more research more recently [e.g. Gaume et al., 2018; Bergfeld et al., 2021b]. Trottet et al. [2021] suggested that once the crack has reached a "super-critical" size (larger than the size at the onset of crack propagation), the involved failure mechanism or mode might switch and affect the size of the avalanche release area. However this idea was challenged by the results of Bergfeld et al. [2021a]. Research on the processes involved in the dynamic phase of crack propagation that follows the onset is ongoing. At

last, even for point stability assessment, a minimal spatial homogeneity is assumed (at least on several meters), especially for crack length modelling whereas terrain or wind effects induce spatial heterogeneity of the snowpack [Gaume, 2012] and highly influence propagation results [Gaume et al., 2013]. Detailed snow cover models are mainly 1D models and do not represent meter-scale spatial variability.

The most recent studies designed to understand the physics of avalanche formation mainly work with a simple two-layer stratigraphy composed of a homogeneous slab and a weak layer [e.g. Gaume et al., 2015; Bobillier et al., 2021; Rosendahl and Weißgraeber, 2019a]. These models can thus not directly be applied to heterogeneous vertical snow profiles (see influence of layering in Figure 2.4). These models have either to be adapted to more complex snow profiles or an additional step is required to reduce the complex stratigraphy, for instance with the use of equivalent mechanical properties [Monti et al., 2016]. Moreover, the substratum has been shown to influence snowpack stability. The mechanical properties of the substratum can modify the stress distribution especially in the weak layer just above [Habermann et al., 2008] and can also influence the crack propagation propensity since layering affects the stress distribution in the slab [e.g. Gaume and Reuter, 2017; Reuter and Schweizer, 2018]. However, it appears from this review that accounting for the substratum is not systematically taken into account in the stability models.

For understanding the concepts involved in avalanche formation, simplified stratigraphy (for instance with homogeneous slab) is of crucial help to separate between involved processes. However, for practical use in avalanche forecasting, the full stratigraphy is crucial as it can decide whether a weak layer fails for example. To get this stratigraphic information, both observed profiles or simulation [Schweizer et al., 2006] can be analyzed with expert rules to identify critical weak layers. Some stability indices have been developed specifically to identify weak layers prior to further stability analysis such as SSI [Schweizer et al., 2006]. Other expert methods may also be used to identify weak layers such as the so-called lemons [Jamieson and Schweizer, 2005] or the tracking approach developed by Reuter et al. [2022b].

The assessment of the snowpack stability highly depends on the knowledge of the material properties of the layers. There is a large variety of mechanical parameterizations (Section 2.3) and their choice affects the stability indices (Figure 2.4). This large variety is due to the complexity of snow as a material. In particular, mechanical properties are highly sensitive to temperature, liquid water content [e.g. Mellor, 1974] or the microstructure [e.g. Hagenmuller et al., 2015; Jamieson and Johnston, 1990]. Moreover, the snow types which are critical to assess snowpack stability, for instance weak layers or snow near the melting point, are also difficult to measure. Thus, despite the amount of measurements available, there is no best practice for computing mechanical parameters for snow cover models [e.g. Podolskiy et al., 2013; Reuter and Schweizer, 2018]. This limitation might also be linked to the rough representation of microstructure by snow cover models (or in observation reports) which may be insufficient to represent the evolution of mechanical properties.

If snow cover model output is used to run a stability model, also the meteorological forcing represents a source of uncertainty, which adds to the snow cover model uncertainty [e.g. Vernay et al., 2015; Lafaysse et al., 2017] and propagate to the finally computed stability indicators. Improving stability models needs to go with improvements of the snow cover models [e.g. Simson et al., 2021], including a better representation of snow microstructure or the explicit representation of the evolution of some mechanical properties [e.g. Hagenmuller et al., 2015]. In any case, as they are computed from the output of snow cover models, stability indicators are sensitive to previous model parameterizations. A way to overcome these limitations is to use statistical tools to identify critical situations rather than only considering physically based models. These models range from simple statistical adjustments of the threshold on stability indicators [e.g. Reuter and Schweizer, 2018] to more advanced statistical methods such as random forests [e.g. Sielenou et al., 2021; Evin et al., 2021b; Mayer et al., 2021]. Besides, these techniques are versatile: they can be straightforwardly adapted to different versions of meteorological and snow models.

2.6 Notations used

Notations used are reported below:

- σ is the slope-normal stress in a snow layer
- τ is the slope-parallel (shear) stress in a snow layer
- τ_c is the shear strength of a snow layer
- σ_c is the tensile strength of a snow layer

- σ_b is the stress in bonds
- σ_{b_c} is the strength in bonds
- $\Delta\tau$ is the skier-induced stress
- D is the slab thickness
- D_{wl} is the weak layer thickness
- E denotes the elasitc modulus
- E_{eq} denotes the equivalent modulus
- ν denotes the Poisson ratio
- $E' = E/(1-\nu^2)$ denotes the plane stress elastic modulus
- ρ denotes the density and ρ_i the density of ice
- w_f is the weak layer fracture energy
- G is the shear modulus ($G = E/2(1+\nu)$)

Chapter 3

Combining modelled snowpack stability with machine learning to predict avalanche activity

This chapter is under review as Viallon-Galinier, L., Hagenmuller, P. and Eckert, N. (2022). Combining snow physics and machine learning to predict avalanche activity: does it help?. The Cryosphere Discussions.

Contents

Abstract

Predicting avalanche activity from meteorological and snow cover simulations is critical in mountainous areas to support operational forecasting. Several numerical and statistical methods have tried to address this issue. However, it remains unclear how combining snow physics, mechanical analysis of snow profiles and observed avalanche data improves avalanche activity prediction. This study combines extensive snow cover and snow stability simulations with observed avalanche occurrences within a Random Forest approach to predict avalanche situations at a spatial resolution corresponding to elevations and aspects of avalanche paths in a given mountain range. We develop a rigorous leave-one-out evaluation procedure including an independent evaluation set, confusion matrices, and receiver operating characteristic curves. In a region of the French Alps (Haute-Maurienne) and over the period 1960-2018, we show the added value within the machine learning model of considering advanced snow cover modelling and mechanical stability indices instead of using only simple meteorological and bulk information. Specifically, using mechanically-based stability indices and their time-derivatives in addition to simple snow and meteorological variables increases the probability of avalanche situation detection from around 65% to 76%. However, due to the scarcity of avalanche events and the possible misclassification of non-avalanche situations in the training data set, the predicted avalanche situations that are really observed remains low, around 3.4%. These scores illustrate the difficulty of predicting avalanche occurrence with a high spatio-temporal resolution, even with the current data and modelling tools. Yet, our study opens perspectives to improve modelling tools supporting operational avalanche forecasting.

3.1 Introduction

Avalanches are a significant issue in mountain areas where they threaten recreationists and infrastructures [Wilhelm et al., 2000; Stethem et al., 2003]. The mapping [Keylock et al., 1999; Eckert et al., 2010b] and forecasting [Schweizer et al., 2020] of avalanche hazard and related risks are therefore important challenges for local authorities [Bründl and Margreth, 2021; Eckert and Giacona, 2022]. Most of the countries facing such hazards rely on operational services for avalanche hazard forecasting [LaChapelle, 1977; Morin et al., 2020] and hazard mapping [Eckert et al., 2018]. In this work, we focus on the issue of forecasting (estimation of the outcomes of unseen data) of daily avalanche activity from simulated meteorological and snow data. Indeed, inferring the relation between avalanche activity and given weather and snow conditions is one of the essential components of operational avalanche hazard forecasting (prediction in the future based on predicted snow and weather conditions).

Prediction of avalanche activity is mainly based on the knowledge of the snowpack evolution and of the mechanical processes leading to avalanches [e.g. LaChapelle, 1977; Morin et al., 2020]. Information on the snowpack evolution can be collected through field observations and measurements [e.g. Coléou and Morin, 2018], and numerical simulations [e.g. Bartelt and Lehning, 2002; Vionnet et al., 2012]. These data typically include a detailed description of the snowpack stratigraphy with vertical profiles of snow properties [Fierz et al., 2009]. Several methods allow for identifying avalanche-prone situations from these profiles. Detection of weak layers based on mechanical and expert rules, such as the so-called lemons technique [Schweizer and Jamieson, 2007], comprises one qualitative approach. Numerical computation of stability indices based on mechanical theories constitutes an automated method to quantify the snowpack stability [Roch, 1966a; Föhn, 1987b; Lehning et al., 2004; Schweizer et al., 2006; Viallon-Galinier et al., 2021]. These approaches rely on the knowledge of mechanical processes involved in avalanche release [Schweizer, 2017; Viallon-Galinier et al., 2021]. Numerical models, which are currently used as an aid to decision-making for avalanche forecasters, generally combine mechanical stability indices and expert rules to provide information on snowpack stability [Morin et al., 2020; Schweizer et al., 2006; Giraud et al., 2002; Viallon-Galinier et al., 2021].

Machine learning techniques can approach the complex link between simple snow cover variables and avalanche occurrence. These methods allow taking advantage of the knowledge of past avalanche activity to determine objective delimitation of avalanche-prone conditions within the space defined by their potential drivers. The first attempt to use machine learning techniques in the avalanche community was performed using linear methods by Bois et al. [1974]. In the next decade, several attempts were made to use nearest neighbors for local avalanche danger forecasting [e.g. Navarre et al., 1987; Buser, 1989]. Classification trees quickly became another common choice, as it is conceptually close to decision processes used by forecasters [e.g. Kronholm et al., 2006; Hendrikx et al., 2014]. The first use of random forests was performed by Mitterer and Schweizer [2013]. This method became then popular in the community [e.g. Sielenou et al., 2021; Pérez-Guillén et al., 2022; Sielenou et al., 2021]. Other techniques have also been tested, such as support vector machine [e.g. Pozdnoukhov et al., 2011; Choubin et al.,

2019; Sielenou et al., 2021] and more advanced techniques appeared in the last years such as convolutional neural networks [e.g. Singh and Ganju, 2008; Dekanova et al., 2018].

Most existing studies use meteorological variables as input or simple bulk variables such as snow depth to feed the machine learning model. The first machine learning models [Navarre et al., 1987; Buser, 1989] mainly relied on meteorological observations, simple snow observations and avalanche records. The use of modelled snow information was therefore developed to complement or replace observations [e.g. Schirmer et al., 2009; Sielenou et al., 2021] and expert analyses were introduced to provide appropriate variables [Schweizer and Föhn, 1996]. However, most of the commonly used variables are only surrogates for the true drivers of avalanche processes. By contrast, studies using mechanically-based variables closely related to the processes involved in avalanche formation [e.g. Viallon-Galinier et al., 2021] are less frequent in machine learning approaches [e.g. Schweizer and Föhn, 1996; Mayer et al., 2022]. However, these variables could increase the interpretability of the algorithm results and bring complementary non-linear information readily oriented toward the prediction of avalanche activity. Hence, they may reduce the complexity of statistical tools to implement (simpler statistical relations and a smaller number of variables to consider) compared to a model that directly uses the snow model output, and improve the overall predictive power.

Existing statistical prediction approaches are difficult to compare. Different spatial extensions are considered from large mountain ranges [e.g. Kronholm et al., 2006; Sielenou et al., 2021] to avalanche paths [e.g. Choubin et al., 2019]. In the literature, different measures of avalanche activity are also considered from binary classes [e.g. Kronholm et al., 2006; Hendrikx et al., 2014] to ordinal multi-classes [e.g. Mosavi et al., 2020; Sielenou et al., 2021]. Yet, the most important difficulty for the comparison is that existing studies do not share a common evaluation process which includes a relevant segmentation of the training and evaluation datasets and common performance metrics. This absence of a homogeneous methodology for evaluating machine learning approaches within the snow and avalanche community limits the comparison between studies.

On this basis, this paper aims to determine whether combining machine learning on avalanche data and mechanical stability analysis of snow profiles helps predict avalanche activity. In particular, we compare the prediction score of the model trained either only on meteorological and simple snow variables as input or also on variables related to the snowpack stability and derived from the full snowpack stratigraphy. We use random forest techniques to relate meteorological, modelled snowpack information and mechanically-based stability indices to observed avalanche occurrences. We also employ time-derivatives of mechanical indices to account for short-time persistence of avalanche-prone conditions in certain cases. We eventually present a rigorous leave-one-out evaluation procedure of broad interest for evaluating avalanche prediction efficiency that includes an independent evaluation set, confusion matrices, receiver operating characteristic (ROC) curves and additional scores derived from the confusion matrix. The study area is located in Haute-Maurienne in the French Alps where extensive avalanche data and snow cover reanalyses over 58 years (1960-2018) are available.

3.2 Material and methods

3.2.1 Study area

We selected an area belonging to the Haute-Maurienne massif in the Northern French Alps, consisting of the three district municipalities of Bessans, Bonneval-sur-Arc and Lanslevillard (Figure 3.1). This area is frequently studied for avalanche-related issues [e.g. Ancey et al., 2004; Eckert et al., 2009; Favier et al., 2014; Kern et al., 2021; Zgheib et al., 2020] because it is prone to intense avalanche activity. The area is characterized by a relatively high elevation ranging from 1400 to 3700 m, and its avalanche activity does not yet seem to be reduced by adverse climate warming effects [Lavigne et al., 2015; Zgheib et al., 2022]. Located in the eastern French Alps next to the Italian border, the area experiences extreme snowfall events known as "easterly return", which drive most of the avalanche activity [Eckert et al., 2010a; Roux et al., 2021]. We considered data on the winters between 1960 and 2018. When referring to the winter season, we consider days between the 15th of October and the 15th of May. These dates are consistent with the dates of production of avalanche bulletins in France and were already selected as suitable bounds in other studies [e.g. Sielenou et al., 2021].

3.2.2 Avalanche observations

Our proxy of avalanche activity relies on the *Enquête Permanente sur les Avalanches* (EPA). The EPA reports all avalanches in approximately 3,000 pre-defined paths over French mountain ranges [Bourova et al., 2016]. About 110 of these are located in the studied area and are shown in Figure 3.1. Each

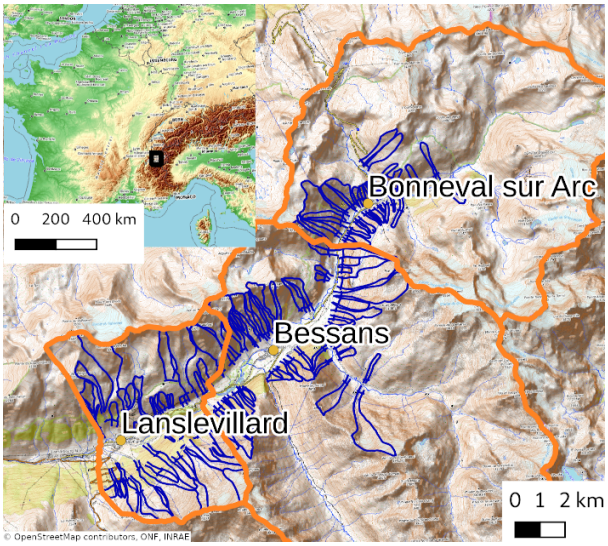


Figure 3.1: Studied area. General situation on the left and contour of avalanche paths surveyed each day (EPA) in blue for the three district municipalities of our studied area (delimited with orange lines). Only the avalanches that flow below a certain threshold (blue line at the bottom of each avalanche path) are systematically reported.

avalanche record indicates the period during which the avalanche is likely to have released and some additional information, such as the elevation and the aspect of the starting zone. EPA was initially designed to capture large natural avalanche events in exposed areas and was extensively used for hazard mapping [Bourova et al., 2016]. Hence, only avalanches whose run-out reaches a certain pre-identified run out threshold (defined for each avalanche path, with a threshold elevation, e.g.: a road, or the valley floor, see Figure 3.1) are systematically recorded. The avalanche activity derived from the EPA depends on this specific sampling procedure. Moreover, it relies on human-based observations and inevitably contains some uncertainties. However, EPA remains one of the longest avalanche activity records. The selected area is characterized by a dense observation network covering a large variety of avalanche paths. Besides, the steep topography of Haute-Maurienne reduces the effect of the observation threshold as most avalanches flow far downslope, close to the valley floor. Further discussion on the EPA strengths and weaknesses is out of the scope of the paper and can be found in Jomelli et al. [2007] or Eckert et al. [2013].

One of the drawbacks of this data for the current study is the uncertainty of the date of some avalanche events, which can be large for remote paths or during low visibility periods (28.6% of the reports have an uncertainty above one day and 23.6% above 3 days, as estimated by the observers). To associate meteorological and snow conditions to each observed avalanche, we remove observations with an uncertainty (length of the period on which the avalanche can have occurred) of more than three days on the release date, from the dataset. When the uncertainty is larger than one day, the last day of the period was defined as the day of the avalanche event. For instance, if an observer reports that an avalanche has occurred between the 21st and 23rd of January in a given path, we consider that the uncertainty of the report is 3 days ($\leq$ 3 days) and we arbitrarily consider that the avalanche occurred on the 23rd of January. Moreover, the aspect and the elevation of the starting zone were not reported in a few cases (representing less than 5% of the total number of events) because the starting point was not visible from the observation point or due to a lack of time for the observation. In these cases, the starting zone was defined by the average elevation and aspect of the typical release area defined for each avalanche path. We applied this definition of release day and zone to the 2779 observed avalanches in the studied domain.

We grouped these observations into eight aspect sectors (from North to North-West) and three elevation bands (centered at 1800, 2400 and 3000 m). This choice defines the spatial resolution of our model. All observations are represented in this geometry in Figure 3.2. When considering all avalanche and non-avalanche situations, the avalanche situations represent 1.1% of the overall dataset. This is called the base rate and acts as a reference for further comparisons.

3.2.3 Simulated snowpack

The SAFRAN-SURFEX/ISBA-Crocus model chain [Durand et al., 1999; Lafaysse et al., 2013] was used to simulate the snow and meteorological conditions in the Haute-Maurienne massif. SAFRAN provides meteorological information adapting numerical weather prediction on a gridded domain to the area of interest and assimilates observed meteorological data [Durand et al., 2009]. We used the publicly available reanalysis [Vernay et al., 2020]. This modelling scheme assumes that meteorological conditions depend only on elevation and aspect. The SURFEX/ISBA-Crocus model is a one-dimensional snowpack model representing snowpack evolution with a multi-layered scheme based on physical evolution laws [Brun et al., 1989; Vionnet et al., 2012]. It uses as an input the meteorological data from SAFRAN model, and it is coupled to the soil scheme ISBA-DIF [Decharme et al., 2011] to represent energy and mass exchange at the bottom of the snowpack. Accordingly to the spatial resolution of the avalanche observations, snow conditions are computed for eight aspects and three elevation levels (1800, 2400 and 3000 m). The temporal resolutions of the meteorological and snow conditions considered here were 1 h and 3 h, respectively.

These simulations retrieve meteorological and bulk snow conditions but also the full snowpack stratigraphy. Hence an additional step is required to take advantage of this information, which is here done through the computation of stability indices as presented right after.

3.2.4 Stability indices

Nine stability indices have been selected based on their applicability with our snow cover model: five for dry snow avalanches and four for wet snow avalanches. In addition, we also computed time-derivatives of these indices.

Dry snow indices

For dry snow, three indices are related to failure initiation, namely natural strength-stress ratio (S_n, [Föhn, 1987b]), skier strength-stress ratio (S_a, [Föhn, 1987b]) and external strength-stress ratio (S_r, [Reuter et al., 2015]). These indices compare shear strength to shear stress, for a given layer interface, where the stress originates from the weight of the overlying layers (S_n and S_a) and/or of an external load (skier, for instance) at the top of the snowpack (S_a and S_r) [Viallon-Galinier et al., 2021]. Moreover, we selected two formulations of critical crack length for representing crack propagation [Viallon-Galinier et al., 2021]: the original formulation by [Heierli et al., 2008; van Herwijnen et al., 2016] and the alternative approach by Gaume et al. [2017]. Both approaches require a slab modulus, determined from density according to Scapozza [2004], and fracture energy estimated from strength. Details on these indices are available in Viallon-Galinier et al. [2022b].

These indices were computed for each layer. For each time step, based on the values of each index, we identified five weak layers (one per index). We defined a weak layer as a layer characterized by a local minimum of the considered stability index (excluding the top and the bottom layers). This approach allows identifying the five weakest layers, with five complementary ways of estimating the weakness (five stability indices). It has the advantage of providing a constant number of variables (25 variables: 5 stability indices on 5 weak layers) for further statistical analysis.

Wet snow indices

To characterize the conditions prone to wet snow avalanches, we used the mean liquid water content in the whole snowpack [Mitterer et al., 2013, 2016] and the thicknesses of humid snow layers. For the latter index, we considered a snow layer as humid as soon as its liquid water content exceeds either 0, 1 and 3% in volume. These three indices are denoted I_{h0}, I_{h1} and I_{h3}. We also introduce the snow depth as an indicator of the amount of snow that can be involved in a potential avalanche.

Time-derivatives

Stability indices values at a given time may not be sufficient to represent the avalanche activity. The time evolution of snow properties is supposed to be represented by snow cover models. However, considering snowpack properties only at a given date and disregarding its past evolution does not indicate whether the snowpack is becoming more prone to avalanches or is in a stabilization phase. For instance, low values of a stability index may indicate an avalanche-prone situation. However, if these values are preceded by even lower ones, the possible avalanche should already have occurred when the stability was minimal, or even before, but not after. Yet, few stability indices include the time dimension in the literature.

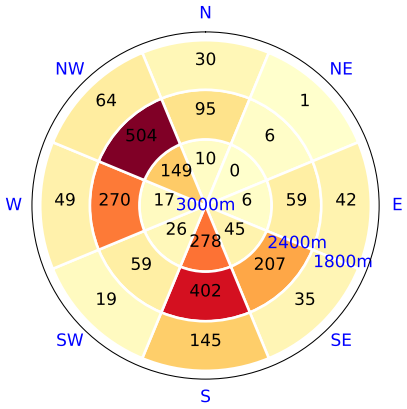


Figure 3.2: Number of avalanche situations recorded in our study area over the full time period at the presented spatial resolution, i.e. per elevation band (1800, 2400 and 3000 m) and aspects (8 from N to NW).

To our knowledge, only Conway and Wilbour [1999] (also used by Reuter et al. [2022b]) and MEPRA natural hazard [Giraud et al., 2002; Viallon-Galinier et al., 2021] include explicit time dependence. Here, we used time derivatives of the previously defined stability indices. We defined the time-derivative of stability index f on a given weak layer as $(f(t) - f(t - \mathrm{d}t))/\mathrm{d}t$ with several time intervals $\mathrm{d}t$ (6, 24 48, 72, 120 and 240 h). The derivatives represent 150 variables for dry snow indices and 24 variables for wet snow indices. Time derivatives on snow depth were used as a straightforward indicator of stability for dry snow conditions (accumulation of new snow) and wet snow conditions (settling and melting).

3.2.5 Learning procedure

Random forests are used to relate snow and meteorological conditions to avalanche activity in the presented spatial resolution.

Avalanche activity

Avalanche activity is based on EPA records in the selected area. For each day, aspect and elevation band, we classify avalanche and non-avalanche situations. A given day on given aspect sector and elevation band is considered as an avalanche situation if at least one avalanche is reported for this day (after filtering of observations and attribution of dates). All other situations are non-avalanche ones. The model resolution defined by classes of elevation and aspect is more demanding than more common approaches applied to whole mountain ranges [e.g. Hendrikx et al., 2014; Kronholm et al., 2006; Sielenou et al., 2021] but provides more detailed information, closer to the spatial resolution used in avalanche operational forecasting [Morin et al., 2020]. The number of avalanche situations observed by elevation and aspect range is shown in Figure 3.2.

General overview of input variables

For each elevation and aspect selected, input variables used are summarized in Table 3.1.

These variables gather information from the meteorological SAFRAN model (Meteo), SURFEX-ISBA/Crocus (Simple snow), stability indices (Stability) computed on the basis of modelled snowpack and derivatives of these variables (Derivatives) as described in Section 3.2.4. Hereafter, if no special mention is added, all these variables (All) are used but for studying variable importance, subsets of this list are also used.

Machine learning algorithm

To relate snow and meteorological conditions to avalanche activity as defined above, we used Random Forest (RF) techniques [Breiman, 2001; Hastie et al., 2009]. Random Forest is an ensemble method used for classification. Each decision tree in the ensemble is built from a random subset of the data. This technique allows going beyond the limitations of single decision trees but without dramatically increasing the algorithm complexity and with similar introspection capabilities. Once trained, each

Table 3.1: Variables used to predict avalanche activity using machine learning

Category	Sub-category	Name	time intervals	Number of variables
Meteo	Snowfall	Snowfall accumulation (mm)	24 and 72h	2
	Rainfall	Rainfall accumulation (mm)	24 and 72h	2
	Temperature	Min, max, mean values (K)	24 and 72h	6
	Wind	Max and mean wind speed (km/h)	24 and 72h	4
		Projected mean direction on N-S axis and E-W axis	24 and 72h	4
Simple snow	Snow depth	Snow depth (m)	—	1
	Depth of new snow	Depth (m) of snow fallen since (see intervals)	24, 72, 120h	3
Stability	Dry snow	Stability indices (S_n, S_a, S_k, a_c, a_g) for the 5 identified weak layers and depths of each weak layer	—	25
		Depth of the corresponding weak layers (m)	—	5
	Wet snow	Maximum mean liquid water content	24h	1
		Maximum height of wet snow with thresholds of 0, 1, 3% of liquid water to consider layer as wet (m)	24h	3
	Snow depth	Snow depth	—	1
Derivatives	Dry snow indices	All dry snow indices	6, 24, 48, 72, 120, 240h	150
	Wet snow indices	All wet snow indices	6, 24, 48, 72, 120, 240h	24
	Snow depth variation	Snow depth variation (m)	24, 72, 120h	3

tree of the Random Forest predicts a class for the input data. Aggregating all trees allow to define a probability for each class as the portion of trees predicting the given class.

Random Forest classifiers require two hyper-parameters: the number of trees and the tree depth. Here, we let the trees fully grow until there is only one element in each leaf, as usually done [Hastie et al., 2009]. An optimization on our full dataset showed that 3000 trees were sufficient (more trees did not improve the results), so that this value was selected for the whole study.

We use two classes, namely avalanche and non-avalanche situations, that are highly unbalanced (mean of 1.1% of avalanche situations in the winter season depending on elevation and aspect, see Figure 3.2). Machine learning techniques, if not handled with care, do not perform well on unbalanced data [e.g. Hastie et al., 2009; Sielenou et al., 2021]. They are designed to optimize the overall classification accuracy or a similar score. Their results thus tend to be biased towards the majority class [Chawla et al., 2004; Sielenou et al., 2021], here the non-avalanche situations. The most common techniques to limit this effect are oversampling of the minority classes, undersampling of the majority classes or dedicated learning algorithms. We here used a combination of these techniques. We only considered situations of the winter season characterized by a simulated snow depth larger than 10 cm. This first selection led to the undersampling of the majority class. Note that we chose this conservative threshold to remove very obvious non-avalanche situations from the dataset (no snow in the starting zone means no avalanche). We do not expect this threshold to be optimal as this is the goal of the training phase of the machine learning algorithm. However, this first step was not sufficient to fully balance the dataset. We therefore used an adaptation of RF classifier to deal with unbalanced data [Chen et al., 2004]: each tree of the forest is trained on a subset of the data randomly drawn; the probability law for drawing is adapted so that the probability of drawing non-avalanche or avalanche situations are identical. This second step acts as an oversampling of the minority class.

3.2.6 Evaluation methods

Evaluation process

We evaluated the model performance with a leave one year out approach (LOYO). The snowpack completely melts in summer, and new snowfall in autumn occurs on bare ground. Therefore, there is no memory between winter seasons and they are exchangeable. This is not the case between successive days during the winter season, with highly correlated snowpack characteristics. A simple leave one out (i.e., leave one day out) would yield better scores but would be less relevant. For each of the 58 seasons between 1960 and 2018, an evaluation set is composed of one winter season and a learning set of the remaining 57 seasons. This leads to 58 sets of trained random forests, each one being evaluated on one year. On a single winter season, there are not enough avalanche situations to be statistically relevant. Therefore, the confusion matrix of 58 evaluation years were aggregated to compute scores with all information available. This leave one year out approach is used for all evaluations presented.

We also quantified the statistical uncertainty related to the sample size. As we used 58 years of evaluation data computed separately, we were able to define an uncertainty by bootstrapping evaluation years used to compute the considered score. In practice, 1000 independent draws of 58 years (with replacement) were randomly produced and the scores were computed on each draw. The 20th and 80th percentiles were used to quantify the uncertainty of the produced scores.

Scores

The Random Forest model produces the probability of being an avalanche situation, defined as a situation with at least one avalanche event, given the snow and meteorological conditions. We selected a threshold (t) on this probability to discriminate avalanche and non-avalanche situations. It is possible to construct a confusion matrix (as presented in Table 3.2) based on this threshold. We derived three scores from the confusion matrix. The true positive rate (TPR) or recall is the ratio between correctly predicted avalanche situations divided by the number of observed avalanche situations. This score is also called probability of detection (POD). It quantifies how many avalanche situations have been correctly predicted. The false positive rate (FPR), also called false alarm ratio (FAR), is the ratio between the number of false positives (non-avalanche situations that are identified as avalanche situations) and the total number of non-avalanche situations. It corresponds to the probability that a false alarm will be raised. These two complementary indicators are interesting but do not fully characterize the performance of a binary classifier in case the two classes are unbalanced (which is the case here). We used a third score to represent how many predicted avalanche situations are really observed as such. This score is called precision and is defined as the ratio between correctly predicted avalanche situations and the number of predicted avalanche situations. The definition of these scores is summarized in Table 3.3.

Table 3.2: Confusion matrix: observed and predicted avalanche situations ("Avalanche") and non-avalanche ones ("Non-avalanche").

		Predicted	
		Avalanche	Non-avalanche
Observed	Avalanche	AA	AN
	Non-avalanche	NA	NN

Table 3.3: Scores derived from the confusion matrix.

Name	expression
True positive rate or recall	$\frac{AA}{AA+AN}$
False positive rate	$\frac{NA}{NA+NN}$
Precision	$\frac{AA}{AA+NA}$

We also mention the specificity ($1 - \text{FPR}$), to be compared with the true positive rate. Finally, we also compute the balanced precision, which is the precision we would have considering balanced positive and negative classes (avalanche and non avalanche situations), computed as $\frac{AA}{AA+NA*(AA+AN)/(NA+NN)}$.

Scores presentation

These scores can be computed for any threshold t on the avalanche situation probability. The impact of this threshold on the overall scores can be represented with two graphs: the ROC (Receiver Operating Characteristic) curve and the precision-recall graph. The ROC curve shows the true positive rate as a function of the false positive rate for all possible thresholds between 0 and 1. When the threshold is equal to 0, all situations are considered avalanche situations (true positive rate is 1, false positive rate is close to 1). When the threshold is 1, all situations are considered non-avalanche situations (true positive rate is 0 and false positive rate is close to zero). A perfect classifier would have a threshold value for which the true positive rate is 1, and the false positive rate is 0. Random classification is usually associated with the diagonal in the ROC diagram. A standard measure derived from this curve is the area between the first bisector and the ROC curve (the area under curve or AUC) [Bradley, 1997]. The AUC quantifies how good the model is compared to a random classifier. We used the AUC value to compare different classifier configurations. In addition, recall is also plotted as a function of precision to capture the model capacity to identify avalanche situations (precision) while limiting the number of false positives (recall). In this graph, the optimal point would be (1,1) i.e., a 100% precision and a 100% recall.

Importance of variables

The importance of variables was estimated through the separative power of each variable in the trees by computing the normalized mean decrease of impurity (also called Gini importance) on nodes where the given variable is used to separate the data in two groups [Breiman, 2001]. A variable importance of zero means that the variable could be removed without reducing model performance and a high value denotes a high separative capacity (between avalanche and non-avalanche situations) of the variable. If two variables contain similar information, each variable will be picked randomly in the tree construction and these variables will consequently share out the importance of the common information [Breiman, 2001]. This first approach is commonly used with random forest but only provides a first rough insight into variable importance. We thus use a more robust discrimination of the importance of variables by using different subsets of variables (see subsets in Table 3.1). The performance difference between more independent groups gives an idea of the importance of the variables present in each group.

3.3 Results

3.3.1 Overview of random forest output

The trained Random Forest model provides the probability of being an avalanche situation for each day, aspect and elevation. Figure 3.3 provides an overview of the output for a specific season (1998–1999), elevation (2400 m) and two aspects (NW and SE). We observe a high variability of the output between days. The time series differs between aspects, which gives a rough idea of the interest of the selected

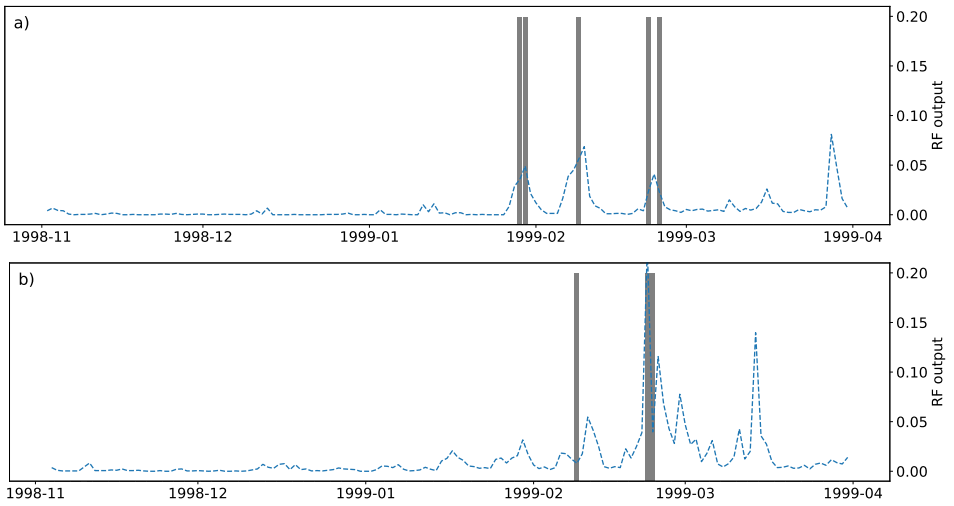


Figure 3.3: Random forest model output (trained with all variables) for winter 1998-1999 at 2400 m for aspects (a) NW and (b) SE. The grey bars represent the days for which avalanches were observed in the selected aspect and elevation range. The base rate of avalanche situations in the full dataset is 0.011. Dates represent the beginning of months.

spatial scale. When considering the observations, most peaks of the random forest output correspond to observed avalanche activity. The random forest thus provides, in this example, a relevant image of the expected avalanche activity. There are also false positives (such as late March in NW aspect) or false negatives (such as early February in SE). This first overview is insufficient for an evaluation of the performance of the model that must be conducted over longer periods, all aspects and elevations.

3.3.2 Model performance

The ROC curve of the model trained with all input variables at our spatial resolution and evaluated independently on each winter season since 1958 is shown in Figure 3.4a. Fortunately, the model is far better than a random classifier (ROC curve above the first diagonal) but it also remains far from an optimal classifier (no points close to (0, 1)). The uncertainty around the ROC curve is very low, which indicates that a sufficient amount of data is available to constrain the model and that the evaluation is not sensitive to the choice of the winter season. The optimal threshold, defined as the threshold which leads to the ROC point closest to (0, 1), is here 0.01. In other words, a situation is considered an avalanche situation when the model probability is larger than 0.01. For this threshold, we provide the corresponding confusion matrix in Table 3.4, classifying situations between observed and predicted avalanche and non-avalanche situations in all elevation and aspect bands. The corresponding scores are 75.3% for the true positive rate or recall, 23.6% for the false positive rate and 3.4% for precision. The balanced precision is 76.2%. These scores mean that more than three-quarter of the observed avalanche situations are correctly identified but avalanches were actually observed only on 3.4% of the situations when avalanches were predicted. The recall (75.3%) and sensitivity (complementary of the false positive rate, here 76.4%) are similar, indicating similar performances on observed avalanche and non-avalanche situations. An alternative point of view is to consider precision and recall rather than true and false positive rates (Figure 3.4b). The maximal precision that can be reached with our model is around 30% but with a very low value of recall (below 5%). With higher values of recall, the precision ranges between 2 and 10%.

3.3.3 Variable importance

As described in Section 3.2.6, the predictive power of the input variables can be estimated in two ways.

First, we computed the feature importance of all variables and aggregated (summed) them by groups, as defined in Table 3.1 (Figure 3.5). The most important variables are related to snow depth (Figure 3.5) and, in particular, the new snow amounts or snow depth variations. Variables related to dry snow stability appear to also be of large importance (13.6%) but with much more variables in the corresponding group: 25 dry stability indices, whereas there are only four variables in the new snow depth group. The depths of weak layers is also of importance (3.6%). Derivatives of dry snow indices decrease in importance with

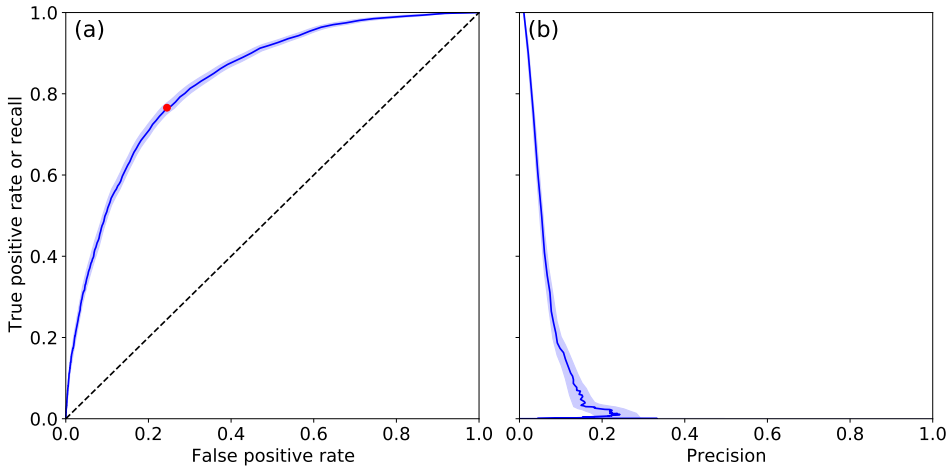


Figure 3.4: (a) ROC curve of the model trained with all input variables at our spatial resolution and evaluated independently on each winter season since 1958. The optimal point (threshold value of 0.01) is represented by a red dot. (b) Precision and recall (Table 3.3) curve. Shading represents the uncertainty based on the 20th and 80th percentile of the bootstrap on evaluation years (see methods section).

Table 3.4: Confusion matrix for the evaluation dataset: observed and predicted avalanche situations ("Avalanche") and non-avalanche ones ("Non-avalanche") summed over elevation and aspect ranges. A threshold value of 0.01 is used, i.e., predicted probabilities over 0.01 are considered to identify avalanche situations. The corresponding recall is 75.3%, the false positive rate is 23.6% and the precision is 3.4%.

		Predicted	
		Avalanche	Non-avalanche
Observed	Avalanche	1 895	623
	Non-avalanche	55 005	178 357

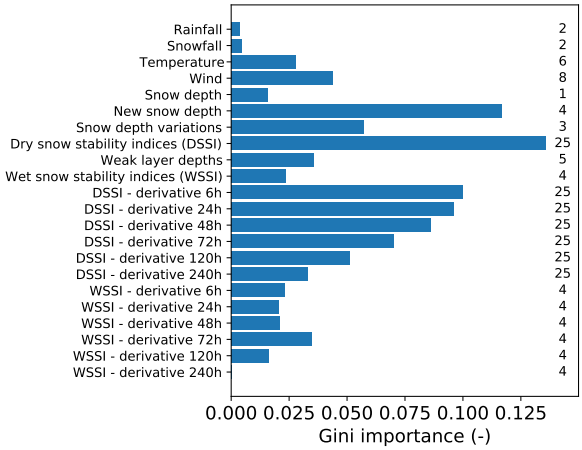


Figure 3.5: Feature importance (Gini importance) on train dataset, aggregated (summed) by groups of variables. The number of variables in each group is reported on the right.

time step, whereas for wet snow indices, the importance is more pronounced for a time step of 72 h. Temperature and wind are also important, even described with few involved variables. By contrast, snowfall and rainfall (on 24 h) are variables with low importance. The variability between years is limited (not shown), giving confidence in the robustness of these results. However, absolute values have to be taken with care as this analysis method is strictly valid only when the different variables are independent, which is far from the case we have here.

Second, we studied the importance of variable groups by removing the data related to different groups of variables before learning and observing changes in evaluation scores. Specifically, we selected six subsets of the presented variables (see Table 3.1): the meteorological variables only (Meteo), bulk variables only (Simple snow), stability variables without derivatives (Stability), stability variables and derivatives (Stability+Derivatives) and all variables (All). The ROC curves for all these subsets are presented in Figure 3.6. The associated scores for the optimal threshold are reported in Table 3.5. These thresholds are coherent with the base rate of our dataset. The ROC curve of the model trained only on meteorological variables is very close to the first bisector (Area Under the Curve AUC=0.09, Figure 3.6). In other words, this model is almost not much better than a random classifier. Using the simple snow variables (snow depth and new snow depth) allows for a first improvement in scores with an AUC of 0.19. Using the stability variables also allows for an AUC of 0.19 and combining it with the associated 174 time derivatives increases the AUC to 0.32. This result highlights the importance of time dimension in avalanche activity forecasting. The AUC of 0.32 for stability and derivatives is close to the value (AUC=0.33) obtained by using all variables. Moreover, the uncertainty linked to inter-annual variability is larger than the difference between the two latter approaches. This means that using all stability indices and their derivatives contains all relevant information available (in the context of the variables tested in this study) for discriminating avalanche and non-avalanche situations. The other scores (false positive rate FPR, recall, precision, see Table 3.3) present similar trends between groups compared to AUC. Some differences are nevertheless observed, with for instance a higher recall but a higher FPR for stability and derivatives compared to all variables, which highlights that the selection of an optimal classifier is always a question of compromise between these two scores.

3.4 Discussion

3.4.1 Machine learning for predicting avalanche activity

The model performance in the studied area decomposed into eight aspects and three elevation bands, is summarized with the confusion matrix shown in Table 3.4. Values of recall (75.3%), false positive rate (23.6%) or precision (3.4%) may seem quite low compared to current literature. Hendrikx et al. [2005] or Kronholm et al. [2006] obtained accuracy for separation around 85% with regression trees and meteorological variables or simple snow variables (snow depth or simple melting model). The accuracy of our model is 76.5% but this metric may not be the most informative when classes are highly unbalanced, as in our problem because it mainly gathers information on non-avalanche situations. Sielenou et al. [2021] reported scores above 95% for accuracy but did not exploit other metrics. Hendrikx et al. [2014]

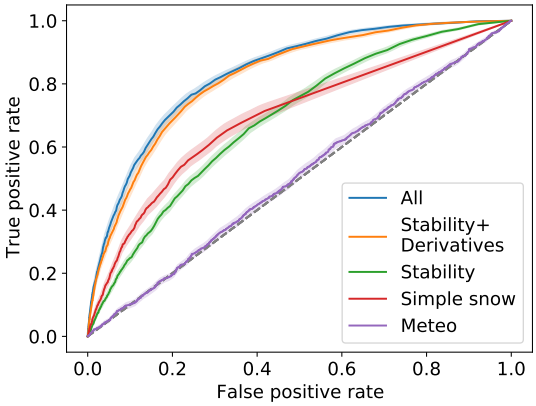


Figure 3.6: ROC curves of the model trained with different sets of variables. Shading represents the uncertainty by bootstrap on evaluation years (see methods section). Labels of subsets of variables correspond to those of Table 3.1. Scores associated to the optimal points (nearest to (0, 1)) are reported in Table 3.5

Table 3.5: Predictive performance of the model trained with different sets of variables. The scores include area under ROC curve (AUC), false positive rate (FPR), recall and precision. We also report the associated optimal threshold used to compute these scores (associated to the point of the ROC curve nearest to the optimal one). Subsets of variables correspond to those of Table 3.1

Subset	AUC	FPR (%)	recall (%)	precision (%)	threshold
Meteo	0.009	49.9	51.7	1.1	0.025
Simple snow	0.195	32.9	65.3	2.1	0.001
Stability	0.188	38.1	65.9	1.8	0.01
Stability and derivatives	0.321	26.5	76.7	3.0	0.01
All	0.334	23.6	75.3	3.4	0.01

reported a recall (focusing on observed avalanche situations) of 76 to 79%, close to our value of 75.3%. Some studies, such as Pérez-Guillén et al. [2022] or Mayer et al. [2022], did similar work using different targets (manually predicted avalanche hazard or measured stability) with accuracy also in comparable ranges (72 to 88%). Precision is highly influenced by the base rate (proportion of avalanche situations). Here, avalanche and non-avalanche situations are highly unbalanced. We nevertheless consider that the balance is representative of avalanche activity in the considered area. Moreover, low values of precision (around 3% for our model) are not uncommon for such difficult problems in related but different contexts [e.g. Rubin et al., 2012]. Eventually, to compare our results to some studies with balanced dataset, the balanced precision should be considered, which is 76.1%.

However, it remains difficult to compare scores to other studies due to differences in evaluation methods and reported scores. All studies used different methods for defining a training and an evaluation dataset. In this study, we used a robust and conservative method, consisting in isolating winter seasons for evaluation. Indeed, with the snow melting between seasons, we get rid of the snowpack memory and provide a robust separation between training and evaluation datasets, leading to trustworthy evaluation results with our method. Moreover, we discard all the situations where the snow depth is less than 10 cm in the release zone, and situations outside the winter period where avalanche release is very unlikely. Consequently, our evaluation does not include the most obvious non-avalanche situations. It is thus more strict than Sielenou et al. [2021], for instance, who used the random forest out of bag method with oversampling of the minority class. It resembles the methodology of Hendrikx et al. [2014] who selected two independent years for evaluation. Our method may be used for future benchmarks to compare competing methods on a robust and homogeneous basis. In addition, the scores reported are not homogeneous between studies either. Some of them focus on global accuracy [e.g. Kronholm et al., 2006; Pérez-Guillén et al., 2022], others on accuracy per class [e.g. Sielenou et al., 2021] and a few propose other metrics such as recall, precision or F1 score (harmonic mean of precision and recall) [e.g. Hendrikx et al., 2014]. The choice of the score depends on the goal of each study and must be adapted to it. However, limiting to a few values for summarizing the model performances limits the information available. These differences in the evaluation processes - both separation between evaluation and train sets and computed scores - limit the possibility of model comparison.

Our model predicts the probability that at least one avalanche occurs on a given day within a spatial unit corresponding to one elevation band (centred at 1800, 2400 and 3000 m) and one aspect (among 8 aspects). This spatial resolution enables to capture the spatial distribution of the expected avalanche activity in one region. This latter information is crucial to evaluate and describe the avalanche danger at regional scale [Morin et al., 2020]. This prediction goal is more demanding than a prediction at larger scales, as generally used in previous studies. Indeed, prediction at aspect-elevation resolution implies to correctly predict the avalanche activity for each aspect and elevation band and not globally at a larger scale. For instance, if one avalanche occurs one day, it implies to identify that we have one avalanche situation but also in which aspect and elevation sector to be considered a success. An avalanche predicted in an other elevation or aspect will be considered as one false negative (in the elevation-aspect it really occurred) and one false positive (in the elevation-aspect it was predicted). It inevitably leads to lower performances for similar models but provides more precise information about the spatial distribution of the avalanche hazard [Statham et al., 2018]. Indeed most studies considered avalanche activity at the scale of mountain ranges, of some thousands of km^2 [e.g. Kronholm et al., 2006; Hendrikx et al., 2014; Sielenou et al., 2021; Pérez-Guillén et al., 2022]. These approaches have the advantage of using machine learning to also aggregate information at larger scales but provide a less geographically precise indicator of avalanche activity. More local approaches have the advantage of providing a relation between snow and meteorological conditions and observed or expected avalanche activity.

3.4.2 Added value of physical modelling of snow cover, stability analysis and time-derivatives for predicting avalanche activity

We tested different input variables to train our model: meteorological variables, simple snow variables (mainly snow depth), stability indices and derivatives. We evaluated the added value of the different groups of variables with two different methods (described in Section 3.2.6). Meteorological information only was insufficient to predict avalanche activity with our method (Figure 3.6). In contrast to many other studies [e.g. Buser, 1989; Mayer et al., 2022], we did not use observed meteorological information but large-scale modelled information [Durand et al., 2009]. Thus, the meteorological information is uncertain and nearly identical for all aspects and elevations, while underlying snowpacks are generally significantly different. Therefore, we did not expect a good prediction at high spatio-temporal resolution with only meteorological information.

Most of the developed models used, at least, some basic output of snow cover models or observed

snow evolution such as snow depth [e.g. Hendrikx et al., 2014]. In our study, snow depth and new snow depth appeared as an essential variable in both methods used to estimate variable importance: its Gini importance is high (Figure 3.5) and adding it to the input variables improves a lot the model performance (Figure 3.6). This result is consistent with current literature identifying snow depth as the first statistical predictor for avalanche activity [e.g. Schweizer et al., 2003; Castebrunet et al., 2012; Sielenou et al., 2021]. Some studies used more advanced diagnostics from snow cover models [e.g. Gassner and Brabec, 2002; Pérez-Guillén et al., 2022; Mayer et al., 2022] or computed expert aggregated variables similarly to what snow cover models do from temperature and precipitations [e.g. Kronholm et al., 2006]. Snow modelling with physical models for taking into account snowpack history thus appears of high interest for automatic avalanche activity prediction.

The novelty of our model is to add a wide range of stability indices to reduce the complex information of snow cover models with the help of knowledge of physical processes and combine it with a time-dependent analysis with the use of time-derivatives of stability indices. The combination of stability indices and derivatives is crucial in our random forest model (Figure 3.5). The time dimension has been identified as a critical information. Since the first statistical forecasts, differences between time steps, for instance on temperature [e.g. Obled and Good, 1980; Navarre et al., 1987], have been used. Conway and Wilbour [1999] also have developed a stability index that explicitly uses time derivatives. We here show that the use of time derivatives, especially in a statistical system that is not able to treat simultaneously different time steps, allows for an improvement of the prediction of avalanche activity. More generally, we showed that the introduction of stability indices and time-derivatives could help identify avalanche-prone situations with machine learning models. This group of variables also gathers a great deal of information as it nearly replaces the information from other variables. Indeed, our results are quite similar when using only stability indices and their derivatives versus using all variables (Figure 3.6). This result indicates that stability indices combined with time-derivatives are a relevant way to summarize the information of meteorological and snow cover models for avalanche-prone conditions prediction, which is a new way of validating the interest of such stability indices.

Computing feature importance can drive the selection of relevant input variables but correlations between variables can affect the computed importance. Re-training the full model with a subset of input variables provides a robust estimation of their effective added value. In particular, the analysis of feature importance allows for selecting the right time steps for derivative computations in the wide range of possibilities included (last column of Table 3.1). The most important derivatives are the short-time ones (6 to 72 h) for dry snow and 72 h for wet snow (Figure 3.5). This result is consistent with the knowledge of involved processes [van Herwijnen et al., 2018], whereas it was never demonstrated so far from a statistical approach. The spontaneous release in dry snow occurs during or immediately after snowfall whereas wet snow problems are more linked to the progressive wetting of the snowpack with solar radiations (time scales of one to several days) or rain [e.g. Reuter et al., 2022b]. Variable importance allows for selecting the most relevant variables which may be kept for further work, especially on stability variables and derivatives, which our results prove to be of interest (Figure 3.6).

3.4.3 Other advantages and disadvantages of our approach

We used the EPA as the ground truth of avalanche activity. This dataset is unique in its spatial and temporal extension but mainly focuses on large avalanches often reaching valley floors. In consequence, the high-elevation avalanche activity and smaller avalanches are not reported, which leads to a limited number of avalanche situations in the dataset. Yet, for the spatio-temporal domain selected in this study (Haute-Maurienne, 1960-2018), the number of avalanche events reported in the EPA remains large enough (2518 avalanche situations). The local topography with steep slopes and the distribution of the recorded avalanche paths allow for a reasonable screenshot of the avalanche activity. However, the scarcity of reported avalanche events might become a problem in other regions as our balancing methods may become insufficient. Observation may not be possible every day (e.g., poor visibility or remote sites), and only avalanches are reported (i.e., no information on the observation that no avalanche occurred). This means that the dataset does not allow to clearly define non-avalanche situations: some situations may be identified as non-avalanche situations, while an avalanche occurred but was not reported. The data also suffers from uncertainty on the dates of avalanches. This may reduce the obtained score. Other data sources may be used to complement avalanche observation dataset, such as observations from ski resorts [e.g. Giard et al., 2018] or satellite avalanche detection [e.g. Karas et al., 2022], but no other data source has the temporal extension of EPA, except archival data that requires in-depth investigations which cannot always be undertaken [Giacona et al., 2017, 2021].

Moreover, we here trained the model with the Haute-Maurienne data. Some climatological or terrain features may lead to a predicted avalanche activity specific to the Haute-Maurienne area, especially with

a higher sensitivity of certain aspects or elevations (e.g., during easterly returns). Hence, the model may not be transferable directly to other areas without a new calibration. Finally, this study presents a binary classification as there is rarely more than one avalanche per day and spatial unit (aspect-elevation), which limits the definition of several classes of avalanche activity. In the future, such machine learning techniques may benefit from the use of other sources of data to complement EPA data and identify more avalanches, such as remote sensing [e.g. Karas et al., 2022], infrasound [Mayer et al., 2020] or seismic detection [van Herwijnen and Schweizer, 2011].

In this study, we chose to treat all avalanche types in a single learning process, including dry and wet avalanches. Some previous studies separate different avalanche activities on pre-defined time periods [e.g. Obled and Good, 1980], or by type of avalanches, restricting to dry or wet avalanches [e.g. Mayer et al., 2022; Pérez-Guillén et al., 2022]. If we assume that decision trees (or here, Random Forest) can capture dry avalanche activity on one hand and wet avalanche activity on the other hand, and if we provide information to discriminate between situations, such as liquid water content or height of wet snow in this study, then a decision tree (or an ensemble of them) will be able to be optimized on the overall avalanche activity by introducing a split in the overall tree to distinguish between dry and wet situations, if relevant. Some other studies also mix dry and wet avalanches, such as the MEPRA French operational avalanche activity indicator [Giraud et al., 2002]. Moreover, the observation dataset does not always allow to infer the processes that led to the avalanche and some situations may remain uncertain in case of a mix of dry and wet snow in the snowpack. For the forecasters, complementary information may be provided with additional tools to identify the processes or situations involved, such as Reuter et al. [2022b].

The impact of using physically-based indices of snow stability as predictors of avalanche activity instead of simpler variables was studied through a specific statistical tool, namely random forests. This method is popular due to its simple background (decision trees [Breiman et al., 1984]) which allows for in-depth analysis and interpretation to some extent and its capacity to represent non-linear phenomena [e.g. Sielenou et al., 2021; Pérez-Guillén et al., 2022; Mayer et al., 2022]. Many other statistical methods are available but random forests have been shown to be as relevant as other ones [e.g. Sielenou et al., 2021]. We introduced time-derivatives and cumulative values to represent the importance of history for snowpack-related processes. Methods in the range of recurrent neural networks are specifically designed to cope with processes having a memory of previous states [Hochreiter and Schmidhuber, 1997]. These alternative statistical methods could be further compared to our random forest approach. It may improve the prediction scores or strengthen our results on the effectiveness of combining snow physics and machine learning for predicting avalanche activity.

Our results were obtained with a reanalysis of meteorological and snow conditions, that is to say, input data that have been retroactively corrected with all available observations. This may not be completely representative of operational forecasting (prediction in the future) situation in which models are corrected by observations of the past but run unconstrained for the forecast. This transposition to the forecasting context would be the next step in terms of complexity for machine learning methods. However, the use of the reanalysis allows for a better evaluation of the capabilities of the machine learning model with fewer input errors, which was the goal of this paper.

3.5 Conclusion and outlooks

This paper combines snow cover modelling, mechanical stability indices and observational data through machine learning for avalanche activity prediction. In particular, we considered numerous stability indices and their time-derivatives. To evaluate the random forest model, we defined a robust method adapted to the specific behaviour of the snowpack (long-term memory). This evaluation was conducted on three district municipalities of the French Alps with 58 years of a comprehensive dataset of avalanche observations, with a high spatial resolution (8 aspects and 3 elevation ranges) and an extended set of variables describing both meteorological, mechanical stability variables and their time evolution.

The combination of snow physics through snowpack modelling, stability indices and their derivatives, and random forest proves to be useful for avalanche activity prediction. The snow depth and new snow depth remain the most important predictors but this study highlights the interest in using mechanical stability indices and their derivatives. This is the primary finding of our research as this had never been demonstrated with such a large variety of indices and their derivatives in previous studies [e.g. Zeidler and Jamieson, 2004; Kronholm et al., 2006; Hendrikx et al., 2014; Sielenou et al., 2021], even the rare ones using simple stability indices within machine learning models [Mayer et al., 2022]. Our results also underline the interest of physically-based snow cover models and stability indices for identifying avalanche-prone conditions.

Obtained scores of recall (75.3%), false positive rate (23.6%) and precision (3.4%) are consistent with current literature with similar goals and methods. These scores illustrate the difficulty to predict avalanche occurrence with high spatio-temporal resolution, even with the data and modelling tools currently available. Moreover, we used a rather strict evaluation method leading to lower but robust and conservative scores, which are not directly comparable to other studies [e.g. Sielenou et al., 2021]. Hence, this method may be seen as the first step for future formal comparison between approaches. More widely, with its high spatio-temporal resolution and use of physical and mechanical models, our study opens the perspective to improve modelling tools supporting operational avalanche forecasting.

We here focus on the avalanche activity reported by EPA. The method may be extended in the future to other target variables describing more precisely avalanche hazard such as release volumes or typical situations [Schweizer et al., 2003; Statham et al., 2018; Reuter et al., 2022b; Mayer et al., 2022]. Similarly, we used meteorological reanalysis for snow modelling for the quality of the data but this may not be completely representative of forecast conditions and tests have to be conducted with re-forecasts rather than reanalysis.

Chapter 4

Use of machine learning driven algorithm for operational avalanche hazard forecasting

Contents

Abstract

Avalanche forecasting is a prerequisite to warning populations and preventing dramatic accidents that avalanches can cause on infrastructures, urbanized areas or people. Countries threatened by avalanche hazard have set up operational avalanche warning services to face this issue. These services rely on different observation sources, meteorological and snow cover modeling and stability analysis from projected snowpack. The use of statistical methods for stability analysis, the so-called machine learning methods, is growing. It takes advantage of observed or modelled input data and observational datasets. Methods for evaluating such models are well identified by the machine learning community. However, the avalanche hazard has several specificities that make this application both critical and specific. Hence, special attention has to be paid to the models that are designed to be used as a decision support tool for avalanche forecasting. In this chapter, we demonstrate that the random forest model developed in Chapter 3 is able to correctly identify situations with intense avalanche activity, hence the most critical days. We show that even though not specifically trained for identifying critical days, these situations are not only correctly identified but with a probability of being an avalanche day higher compared to other days. We also evaluate more generally the output probability of the model and show that it provides reliable information on forecasted avalanche activity beyond simple binary classification. We justify the spatial scale selected for the model, that is to say working by bands of elevation and aspect, not only because of the current interest for operational forecasters but also by comparing it to larger scales. Finally, we compare the random forest model to the currently available model for French forecasters, the MEPRA model. We prove that the statistical model has better performances than the currently available model, which validates the usefulness of such a work to improve operational decision support tools for avalanche warning.

4.1 Introduction

Assessment of the snowpack stability is one of the critical element for avalanche hazard warning [Statham et al., 2018]. Field observations, and in particular stability tests, observation of instability signs such as whumpf or shooting cracks or past avalanche activity, provide useful information on the current snowpack stability at the point scale. However, they are difficult to extrapolate in time and space. In contrast, models can be applied on larger spatial domains and at a higher temporal resolution but require an additional step to convert snow profile into snowpack stability information.

Empirical expert rules have been developed to identify avalanche-prone situations from snow profiles. These rules guide the analysis of forecasters or provide a quantitative analysis of the situation in terms of stability. These rules ranges from simple thresholds on new snow amount [e.g. Schweizer et al., 2008b] to more complex sets of rules [e.g. Giraud et al., 2002]. The most common one is the so-called lemons methods [Jamieson and Schweizer, 2005], commonly used to analyze the stability of snow profiles. The model MEPRA, based on physical and expert rules, is one of the tools used by operational forecasters in France. Although these methods can be used as a red flag for forecasters in some critical situations, they also have known limits. For instance, the use of thresholds on simple variables is known to be potentially misleading in controversial situations if used without careful human control [e.g. Hastie et al., 2009].

An alternative approach that has become popular for such complex problems is using statistical or machine learning approaches to relate observed or modeled snow profiles and stability conditions. Being a method close to the work of operational forecasters, the nearest neighbor was the first method experimented in operational conditions. Obled and Good [1980] developed a nearest neighbor model to identify similar situations in the past, implemented by Navarre et al. [1987] and Buser [1983]. The observation of the current day is compared to previous observations that are associated with observed avalanche activity. This system was deployed in several ski resorts in France and was used for decades by ski patrollers. Buser [1983] developed a similar system, which was also used in ski resorts in Switzerland [Buser, 1989].

However, these statistical approaches remain not much used (see Appendix A). Only nearest neighbors have been used for a limited number of ski resorts. The reasons for this limited application of the methods may be complex and involve both scientific and sociological reasons, which study are far from the main objective of this manuscript. Several points of attention need to be addressed before proposing new systems for operational avalanche forecasting. An obvious requirement is the quality of the model results, and for instance its ability to identify critical situations. The question of the spatial scale for prediction is also crucial. The information provided is completely different whether the stability is provided at 250 m resolution or at the scale of whole mountain ranges. The information provided by the model should match the needs for forecast application and provide reliable results at this scale. Statistical models

have mainly been used for local avalanche activity assessment (no more than one ski resort) [Navarre et al., 1987; Buser, 1989]. Many attempts have been made in the literature at larger scales (massif scale, several hundreds of square kilometers) [e.g. McGregor, 1989; Hendrikx et al., 2005, 2014; Sielenou et al., 2021]. These results integrated on a large scale do not provide information on the spatial distribution of expected avalanche activity, which is one of the critical points of avalanche hazard assessment [Statham et al., 2018]. Moreover, prediction at larger scales are more difficult to interpret for the forecasters as it integrates information from potentially different areas (in terms of elevation and aspect, for instance).

In this chapter, we more deeply evaluate the results of the model presented in the previous chapter, focusing on some aspects of direct interest for operational avalanche hazard assessment. We complement the classical statistical evaluation by focusing on some questions that need to be answered before an operational use of such models. We first focus on days with high avalanche activity to define the performances of the model on critical days. In a second part, we pursue beyond the binary classification by showing the added value of the continuous value provided by the random forest classifier and showed that it can be interpreted as a probability of observing an avalanche situation. Then, the issue of the spatial scale is addressed by comparing results at two different scales. Finally, we compared the statistical results to the current model proposed to French forecasters: MEPRA [Giraud et al., 2002] (see Appendix B) to evaluate the respective performances of the models.

4.2 Material and Methods

This chapter completes the previous one with a slightly different point of view. We here briefly remind the main guidelines of the previous chapter and the changes in material and methods used. For additional details, the reader is invited to report to the corresponding section in Chapter 3,

We use the same study area as in the previous chapter, that is to say, a part of the Haute-Maurienne massif in the Northern French alps, consisting of the three district municipalities of Bessans, Bonneval-sur-Arc and Lanslevillard. The studied period is the winters between 1960 and 2018 (58 years).

4.2.1 Avalanche data and studied scales

Avalanche observations are those of *Enquête permanente sur les avalanches* (EPA) [Bourova et al., 2016]. This dataset is described in both Sections 1.2.2 and 3.2.2. The data is grouped into eight aspect sectors (from North to North-West) and three elevation bands (centered at 1800, 2400 and 3000 m). This is hereafter referred to as the AE (for eight Aspect and three Elevation bands) scales. To compare different spatial scale, we also consider the massif scale, by aggregating information over the whole massif (hereafter referred to as the massif scale).

4.2.2 Snowpack simulation and stability assessment

The SAFRAN-SURFEX/ISBA-Crocus model [Brun et al., 1989; Durand et al., 1999; Lafaysse et al., 2013; Vionnet et al., 2012] was used to simulate the snow and meteorological conditions on the Haute-Maurienne massif. Computations are done on flat terrain at the three elevations of 1800, 2400 and 3000 m and the eight aspects on slopes of 40 degrees.

From the snow profiles, the MEPRA module (see Appendix B) is used to compute a natural (R_{nat}) and an accidental (R_{acc}) hazard on the eight aspects and three elevations. The natural activity is also summarized at the scale of the whole massif by aggregating information on all aspects and elevation ranges to produce an integrated index G on an 8-level scale [Martin et al., 2001]. Aggregation is done through maximum over all aspects of the mean over altitudes of an indicator of natural hazard (see Appendix B for further details). In addition, dry and wet snow stability indices and time derivatives are computed as described in Section 3.2.4.

4.2.3 Learning procedure

Random forests are used to relate snow and meteorological conditions to avalanche activity in the two presented spatial resolutions.

Avalanche activity

Avalanche activity is based on EPA records in the selected area. Days were classified into two categories: avalanche and non-avalanche days, depending on whether one avalanche, at least, has been observed in the area (aspect sector, elevation band). In order to compare different scales, data is firstly extracted at

the AE scale (eight aspect sectors and three elevation bands). The data is then aggregated at the whole studied area scale (massif scale): a day is considered a non-avalanche day if no avalanche was observed in all aspect and elevation bands and considered an avalanche day as soon as at least one avalanche is observed for one aspect and elevation. Both numbers of avalanches and binary information (avalanche or non-avalanche day) are recorded at both scales (AE and massif).

Variables used

All computed variables used for the learning procedure are summarized in Table 3.1. These variables gather information from meteorological SAFRAN meteorological model, SURFEX-ISBA/Crocus snow model, stability indices computed from the modeled snowpack and derivatives of these variables. This set of variables is used as the input of the random forest model when studying the AE scale, associated with avalanche observation for each aspect and elevation. When dealing with the massif scale, the input consists of aggregating all these variables computed from snow simulations on flat terrain on the three elevation bands, which means three times more variables compares to the AE scale.

Machine learning algorithm

To relate observed avalanche activity and simulated snow and meteorological conditions, we used Random Forest (RF) technique [Breiman, 2001]. The two hyper-parameters are the number of trees and the tree depth. Similar to the previous work [Viallon-Galinier et al., 2022a], we let the tree fully grow and the number of trees is fixed to 3000. To limit the effect of unbalancing between the two classes (avalanche and non-avalanche day), we only considered days of the winter season and characterized by a simulated snow depth larger than 10 cm and used internal balancing of classes when data is drawn for model training [Chen et al., 2004]. The random forest algorithm provide as an output a proportion, p, of trees voting for the avalanche situation class, regularly referred to as a probability and varying from 0 to 1.

Evaluation

Presented results rely on a leave one year out approach to separate train and test sets to ensure the independence of the two datasets. To be statistically significant, the results of 58 training of random forests, evaluated on one year, are aggregated. Uncertainty is estimated by bootstrapping these 58 years: 1000 independent draws of 58 years (with replacement) were randomly produced and the scores were computed on each draw. The 20th and 80th percentiles were used to quantify the uncertainty of the produced scores.

Given the output value p of the random forest model, we consider both binary classification or direct use of the continuous output value. The output of the random forest model p could be interpreted as the probability that at least an avalanche occurs. As an alternative, it is also possible to define a threshold that discriminates between avalanche and non-avalanche days. Once a threshold is selected, a confusion matrix between our two classes can be computed and Table 3.3 scores are computed. With the different possible values for the threshold t, we compute true and false positive rate and plot ROC (Receiver Operating Characteristic) curves. We then use the area under ROC curve (AUC) as an indicator of the classifier performance.

4.2.4 Evaluations conducted

Different complementary evaluations are conducted on the trained model.

Focus on critical days

To estimate the performances on the most critical days, that is to say days which have a larger reported number of avalanches, we train the RF model with the binary information (avalanche/non-avalanche days) and then evaluate it on a subset of the test set composed of non-avalanche days and avalanche days with a minimal number of reported avalanches. Such evaluation can be conducted on the output probability of the RF model by considering output value distribution. It is also presented for a binary classification with ROC (Receiver Operating Characteristic) curves, that present the true positive rate depending on the false positive rate.

Reliability

The reliability, or calibration, is the comparison of the frequency of observation to the predicted probability frequency. We compare, for a given probability of being an avalanche day, the observed frequency of avalanches that were reported. This comparison allows for testing the interpretation of the output probability of the RF model as the probability of observing at least an avalanche in given conditions.

Spatial scale

In the previous chapter and sections, the reference spatial scale was the AE one. That is to say, meteorological, snow conditions and avalanches observations were taken in massifs by bands of aspect and elevation. However, this scale is not frequent in the literature. To compare the respective interests of both scales, we train and test the RF model on AE and massif scales. It is then possible to evaluate each model with the data at the AE scale or at the massif scale. For evaluation of the model at the massif scale while trained at the AE scale, an additional step is required. Different summary of the information were tested and we present here the simple average of values for all elevation and aspects.

Comparison to MEPRA model

We compared the results of the random forest model with the current available model providing an integrated information to the forecaster, the MEPRA model (see Appendix B). We study here the three indicators of MEPRA model, namely the natural hazard R_{nat}, the accidental hazard R_{acc} and the integrated index at massif scale G. These three indicators are computed as a post-processing of the Crocus snow and meteorological data.

4.3 Results

4.3.1 In-depth operationally-oriented evaluation

The trained random forest model provides a continuous output p between 0 and 1. p is defined as the proportion of trees of the random forest voting for the avalanche class. We here study the behavior of this output for the most critical days and check how this continuous RF output can be interpreted as a probability p.

Focus on critical days

Although the training is done with two classes, the avalanche class can cover different situations, from a single avalanche to intense avalanche activity. The EPA dataset does not report, per construction, a large number of avalanches. We hence use here only four groups: days with no-avalanches (non avalanche days), days with one, two or more than two avalanches. We compared in Figure 4.1 the output probability p for these four groups. The values of p for days with more than one avalanche have a distribution shifted right compared to days with one avalanche. In other words, the output probabilities are higher for the days with several avalanches. So the critical days are not missed by the model and identified, although the model is not specifically trained for this purpose.

With a more quantitative approach, Figure 4.2 plots the ROC curves for days with one avalanche, two avalanches, or four or more. Although the last curve may not be statistically significant due to the low number of events, it proves that the model makes fewer errors when dealing with critical days (high number of avalanches) rather than more ambiguous situations (only one avalanche). The area under the ROC curve is higher when days with higher avalanche activity are considered: 0.334 for the full dataset, 0.381 if removing days with only one avalanche from the dataset, 0.403 when also removing days with two avalanches. This reinforces the previous result, directly on the obtained scores: the more avalanches we observe, the less the model miss them. When considering days with more avalanches, the proportion of missed avalanche situations is also reduced (i.e., TPR increase) for a given probability of false alarm (FPR). Hence, the critical situations are not missed by the model.

Reliability

In the previous chapter, we do not directly consider the output of the RF model p but rather a classification in two classes depending on the value of p being below or above a given threshold. We here evaluate the information carried by the raw value, which can be interpreted as the probability of observing at least an avalanche. In Figure 4.3, we compare the probability given by the output of the model to the observed

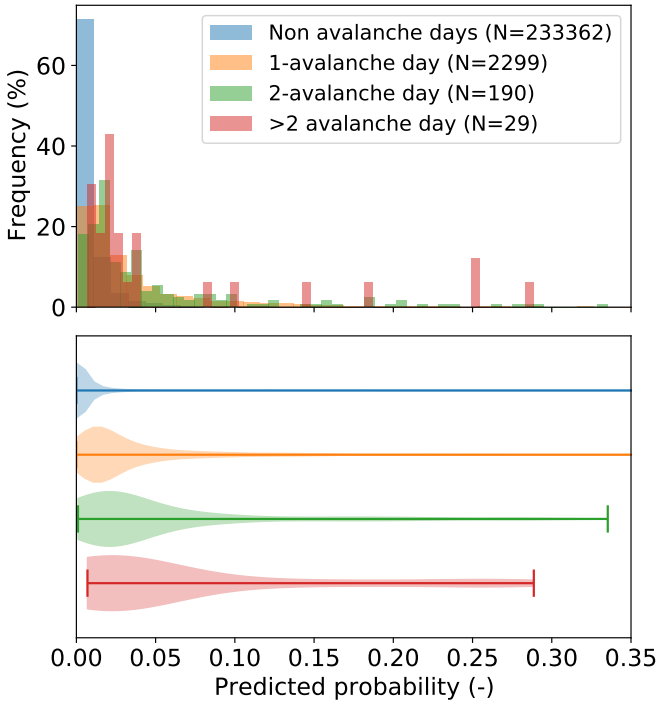


Figure 4.1: Distribution of the output values of RF model for non-avalanche days, days with one, two or more avalanche reported. Histogram (on top) and violin-plot representation (bottom) of the same data. The total number of elements in each class is reported between parenthesis in the legend.

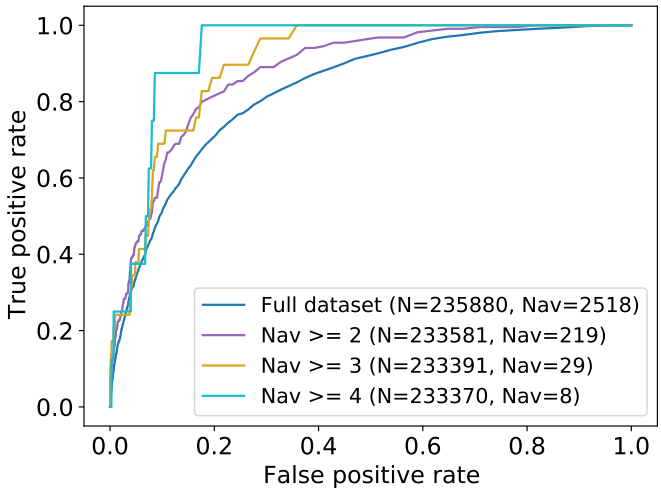


Figure 4.2: ROC curve of the same model evaluated on subsets of the evaluation dataset. Are represented evaluation for the full test dataset, for a dataset with days with only one avalanche removed and a dataset with days with one or two, or one, two or three avalanches only. For each subset, the total number of elements (N) and the number of avalanche days (Nav) are reported in the legend.

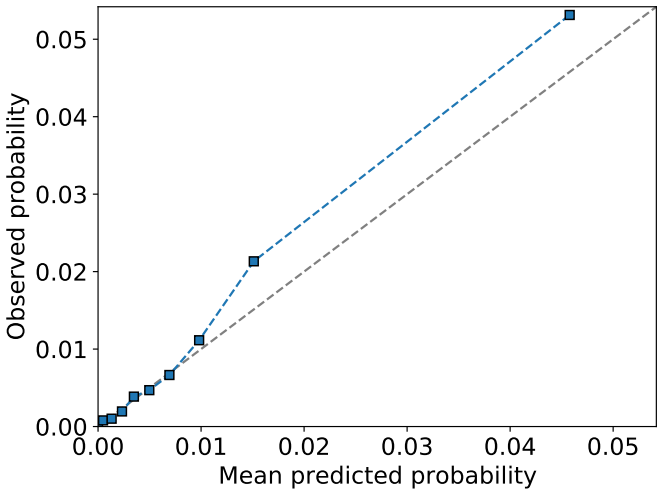


Figure 4.3: Reliability diagram of the output value of the RF model: predicted probability value compared to the avalanche observation frequency. Each point represent the same amount of data.

distribution of avalanche and non-avalanche days. Bins contain the same amount of data, and points are centered on the mean predicted and observed probability for each class. If the classifier produced a perfect probability estimation, all points would be on the diagonal. Our RF classifier shows very good performances up to the probability of 1%, with equal modelled and observed probabilities. For higher probabilities, the RF model slightly underestimates the probability value but the overall curve does not completely move apart from the 1:1 line. It means that higher modeled probabilities p are statistically associated with a higher avalanche activity, proving that the output probability value p can also be used directly as a quantitative indicator of avalanche activity.

As visible both on Figures 4.1 and 4.3, the output values of the model p does not span the full 0-1 range. Most of the values are between 0 and 0.2 (99.9% of the values between 0 and 0.21). The values are coherent with the observed frequencies of avalanches, as shown in Figure 4.3, and also consistent with the low precision observed (3.3%). Indeed, if the model predicts a high probability of avalanche (compared to its histogram) of 0.1, it means that in 90% of cases, no avalanches will be observed. The low values of RF output probability p are linked to the high unbalancing of the dataset and representative of the precision results.

4.3.2 Spatial scale

In the previous chapter and sections, we presented results at the AE scale. We here compare to the massif scale by comparing performances of the models trained and evaluated at the AE scale, trained at the AE scale and evaluated at the massif scale or, trained and evaluated at the massif scale. The results are gathered in Figure 4.4. It is first possible to compare the results of the model trained and evaluated at the AE scale with the model trained and evaluated at the massif scale. The classification has lower performance at the massif scale, with an area under curve of 0.30 compared to the model at AE scale, with an area under curve of 0.33. The slight underperformance is confirmed by the analysis of the uncertainty: both ROC curves are separated by more than the uncertainty estimated by inter-annual variability.

The previous result does not allow for discriminating between the input information provided and evaluation at different scales. In order to disentangle these two contributions, Figure 4.4 presents the ROC curve of the classifiers trained at AE scale and evaluated both at AE scale and massif scale. The area under curve is 0.31, slightly higher than for the model trained directly at massif scale but in the range of the uncertainty of the model. Other ways of converting several probabilities by elevation and aspect to a unique value are possible (in complement to the average of values used for the presented ROC curve). However, other tested aggregation functions do not give better results (e.g. maximum of values, for instance, not shown).

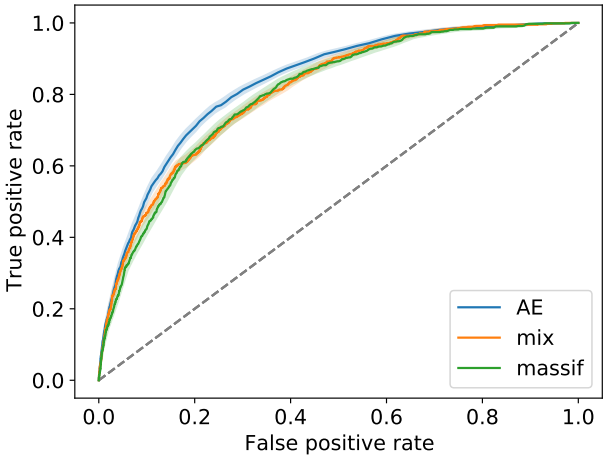


Figure 4.4: ROC curves for model at different scales. Are presented the model trained and evaluated at the AE scale, the model trained and evaluated at the massif scale and the model trained at the AE scale but evaluated at massif scale (mix).

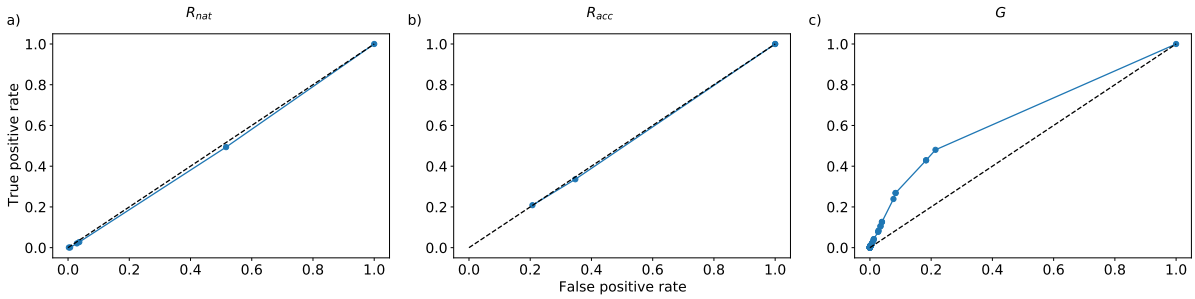


Figure 4.5: ROC curves of the three values provided by the MEPRA model: (a) natural activity at the AE scale R_{nat}, (b) triggered avalanches at the AE scale R_{acc} and (c) natural activity at the massif scale G.

4.3.3 Comparison to MEPRA

We compared the results of the three MEPRA indicators to random forest outputs. Figure 4.5 compares the ROC curve of the MEPRA model and the RF model based on EPA observations. Figures 4.5a and b show the ROC curve of the MEPRA natural (R_{nat}) and accidental (R_{acc}) models. These ROC curves are very close to the diagonal, which means that this does not bring much more information than a random classifier compared to EPA observation. The RF classifier, with its area under ROC of 0.33, is a much better classifier than MEPRA indices (area inferior to 0.02).

The results are different at the massif scale. The aggregated index G at the massif scale produced by MEPRA has more skill as the ROC curve detaches from the diagonal with an area under curve of 0.14. However, it remains far from the RF model at the equivalent scale (AUC of 0.30).

4.4 Discussion

4.4.1 Operational interest of the system

Focus on critical days

For operational purposes, both the false alarms and misses should be limited. However, some false alarms may be accepted to limit missed events, especially for rare but potentially dramatic phenomena such as avalanches. This inevitable compromise is illustrated by the selection of an optimal point in the ROC curve of the classifier. Moreover, our research relies on the EPA avalanche activity, which may not report some avalanches even in unstable conditions if avalanche activity is limited to high elevation or in a situation when small perturbations are sufficient to release the avalanche but may not release

spontaneously. This small perturbation may have occurred (in this case, it is reported as an avalanche day) or not (in this case, it may be viewed as a false alarm as no avalanches are recorded). Hence, some days may be incorrectly labelled, especially for intermediate stability. However, days with high instability are well identified with high avalanche activity. These critical days must be well identified by the model as they are the days with higher danger. Then, we evaluate the model on the most critical days in Figure 4.2 and 4.1. We show that the higher the observed avalanche activity, the more the histogram of model output values is shifted right (Figure 4.1). It means that critical days are better separated from non-avalanche days. Furthermore, the more intense is this avalanche activity (when considering at least 2 or 3 or more observed avalanches), the separation with non-avalanche days increases with the shift of the histogram. The correct identification of critical days is confirmed by the scores plotted through the ROC curves (Figure 4.2). The area under curve increases when considering only non-avalanche days and days of high activity, which indicates that these critical days are less missed than days of limited avalanche activity (one or two avalanches observed only).

Usability of the output value

Binary classification allows for easy visualization of the results through ROC curves (e.g. Figure 4.2) and contingency tables, as well as the use of simple scores (such as accuracy or class-balanced accuracy, true positive rate and true negative rate, precision, etc.), this is a strong reduction of the available information. Indeed, RF may be used to identify a class (here avalanche or non-avalanche) by majority voting but also provides the proportion of trees voting for each class which allows for a more refined analysis. In our 2-class problem, it could be interpreted as the probability of being an avalanche day. We here prove that this interpretation is reasonable by comparing it to observed distribution for ranges of probabilities (Figure 4.3). It can then be directly provided to the forecaster associated with a histogram of past observed values to precisely quantify the expected return time of the situation. It is complementary to the commonly used variables, such as thresholds on new snow [e.g. Schweizer et al., 2008b]. The main advantage is that it integrates all the processes of avalanche formation in a single indicator that may be used to compare situations easily [e.g. Dick et al., 2022] or evaluate the evolution of the avalanche hazard through time (e.g., due to climate change) [e.g. Eckert et al., 2013; Castebrunet et al., 2014; Lavigne et al., 2015].

Comparison to existing model MEPRA

MEPRA can be seen as a single decision tree, which is not statistically optimized but rather manually optimized based on expert knowledge. It is here shown to be of poor performance at the AE scale (Figure 4.5a and b), based on avalanche activity observed in EPA. Even though EPA may not be fully representative of the avalanche activity, it correctly identifies days with high natural avalanche activity, especially on the studied area (see Section 3.4.3). Such systematic and statistical evaluation of MEPRA on long periods of time is a new result as it was previously evaluated only on a limited amount of specific situations [Giraud et al., 2002].

Hence, the RF model has largely better performances. This better performance can be first explained by the statistical tools used. The RF model is statistically optimized for optimal classification, which is not the case with MEPRA. MEPRA is not statistically optimized but was produced based on expert rules selected for their relevance in combination with the Crocus model, but no systematic optimization was used, and the system was not updated while weather and snow cover evolved. Moreover, MEPRA is a single decision tree, which is known to be fairly unstable to input values due to the thresholds used. On the contrary, RF relies on hundreds of independent trees, reducing the instability of the model [Hastie et al., 2009]. Hence, MEPRA has limited skill at the AE scale compared to the RF model.

At the massif scale, MEPRA has better performances (Figure 4.5c) as it combines information on different aspects and elevations. It then does an averaging of snow and meteorological conditions around the massif. This averaging is somehow similar to the idea of the random forest that averages different decision trees. The main difference is that the variability comes from the different inputs on elevations and aspects bands rather than the decision tree itself. As these snow and meteorological conditions are different and not completely correlated, it also reduces the variance of the overall estimator G and makes it more stable than the input R_{nat}. The results are then more correlated to real avalanche activity, as shown on Figure 4.5c. However, even at the massif scale, the RF model remains better than the MEPRA model.

In conclusion, whatever the considered scale, the RF models provide more coherent results compared to the EPA dataset, which gives confidence in the usefulness of such machine learning systems to provide a useful indicator to forecasters. Moreover, re-training the system is easy so that thresholds used in each

tree can be dynamically adapted to new data (e.g., after each season, as done for instance by Navarre et al. [1987] or Buser [1989]) or to adapt to model evolutions. It is thus expected to be more reliable in time, as weather and snow cover models have yearly evolutions.

Limits of the used dataset

The operational forecasters need models that summarize the stability information on subsets of the area of interest [Statham et al., 2018]. As avalanches threaten human lives, the model should be reliable and do not miss the critical situations with a lot of avalanches reaching valleys, infrastructures and urbanized areas. We ensure here that the presented model do not miss the most critical situations. The EPA is especially well designed to identify situations for which avalanche activity is sufficiently high to come near infrastructures or houses [Bourova et al., 2016]. However, fatalities due to avalanches mainly concern mountaineers in the backcountry nowadays [EAWS, 2022]. Hence avalanche bulletins also cover the hazard in high-elevation and remote areas [Morin et al., 2020]. EPA is here considered to be the ground truth when evaluating the model but is not designed to identify high-elevation avalanche activity and small slab avalanches. The usefulness of the presented RF model is then not ensured for small high-elevation avalanche activity that concerns mainly backcountry mountaineering and complementary observations of avalanche activity would be required.

4.4.2 Spatial scale

The spatial distribution of avalanche activity is one of the key points for avalanche hazard assessment [Statham et al., 2018]. For this purpose, it is of interest to have information on expected avalanche activity at a lower resolution than the forecasting resolution (in France, below the massif resolution) to be able to evaluate whether or not all slopes are concerned by avalanche activity and, if not all the slopes are concerned, identify concerned elevations and aspects. Exploring this spatial distribution is one of the reasons why MEPRA provides, for each massif, information by bands of 300 m of elevation and for eight aspects [Giraud et al., 2002] (Figure 4.5a). In contrast, an integrated indicator is easier to visualize on larger areas (Figure 4.5b). It allows for a synthesis at the massif scale, closer to the avalanche danger [EAWS, 2022]. A unique figure is also easier to study trends over time or between areas [e.g. Roux et al., 2020; Reuter et al., 2022a] or to identify the most critical areas for a situation. These advantages of an integrated index lead to the development of a massif scale hazard level from the MEPRA system [Martin et al., 2001]. Both are used by operational forecasters in France (Figure 4.5). Both approaches (refined or AE scale, or massif scale) have their advantages depending on the goal.

From an operational point of view, the sub-massif scale has several advantages. It provides more information that can be aggregated afterwards to different areas of interest depending on meteorological situations. Hence, it is a better candidate as a decision support model for avalanche forecasting. The AE scale allows for insights into the spatial distribution of the avalanche danger, which covers two of the three key components of avalanche hazard according to Statham et al. [2018] (only avalanche size is not reported). Moreover, it is more representative of local conditions as it relates snow and meteorological conditions for a given slope and aspect to avalanche activity, whereas more global information is available in the massif approach. Finally, it allow mountaineers to reduce the risk by choosing the less exposed slopes in a geographical area.

These advantages are combined with a better prediction at this scale, as shown with Figure 4.4. This is explained by the introduction of more precise snow and meteorological conditions and also benefits from a more relevant stability analysis in its inputs as the snowpack is more representative. However, downscaling at massif scale after learning with the AE scale does not improve the results, indicating that the AE scale is more suited to avalanche study. Indeed, the stability may be very different depending on elevation and aspect, so aggregation, by itself, reduces the available information. The use of AE scale rather than massif one is also coherent with known processes of avalanches formation that include wind or solar radiations, which highly differs between aspects and elevations [Brun et al., 1989; Giraud et al., 2002].

Working at the AE scale also allows for an easier reuse of the method, as it only relates punctual information on snow and meteorological conditions to avalanche activity. Thus, it can be easily transposed to other configurations (such as spatially gridded models, similarly to weather models). However, different slopes, ground roughness or forest cover may require adaptation to local avalanche activity through re-learning or use of additional predictive variables.

All these reasons (operationally relevant scale, better scores, ability to transpose) lead us to opt for this AE scale in the rest of this thesis.

4.5 Conclusion

The Random Forest (RF) model is shown to be an interesting candidate for avalanche forecasting. The scores have already been discussed in the previous chapter. It provides high accuracy on the avalanche events despite low precision. However, this precision is also linked to the large amount of no-avalanche situations used in the evaluation dataset. There is no absolute definition of what could be a good model. In this chapter, we showed that the RF model is far more relevant than the currently available MEPRA model, which is only of interest at the massif scale and, even at this scale, remains far from RF skills (area under ROC curve of 0.14 vs. 0.30).

Beyond these large-scale evaluations, it is also interesting to check that the model does not present significant inconsistencies. We here check that the most critical situations are less missed than less active avalanche situations. We also showed that the output of the RF model could be interpreted as a probability of at least an avalanche release. The value is shown to be coherent with the observation on the EPA dataset. The correctness of its output value gives confidence in the added value of such a system as an aid to decisions for operational forecasters.

However, the results presented here are limited to the specific area of Haute-Maurienne. Before any use at a larger scale, it remains necessary to evaluate such models on other areas, with different climates and other distribution of avalanches in terms of both spatial distribution and processes involved [Reuter et al., 2022a]. EPA covers the French Alps and Pyrenees, and could be used to extend the method. However, new sources of data could also complement EPA database, especially in regions where observation network is less dense than in Haute-Maurienne. Extension to other areas and data sources would strengthen the results presented in this chapter.

Chapter 5

Generalization of the machine learning method

Contents

Abstract

The forecast of avalanche-prone conditions is critical in mountainous areas to protect human lives. To this end, forecasters rely on several tools, including snow cover simulations and statistical analyses. However, stability analyses remain mainly evaluated on restricted datasets, for specific situations (e.g. large natural avalanche cycles) and on well-defined test areas. Although this is the necessary first step in the development of new models, application to broader situations and geographical areas is needed before the operational delivery of such models. In this chapter, we use an alternative dataset of avalanche observation from the Meteo-France nivo-meteorological network to evaluate the statistical model of avalanche activity developed in Chapter 3. By using an alternative data source, we show the robustness of the chosen statistical model to the observation dataset used for training and evaluation. We then successfully extend the results to different geographical areas, giving confidence in the versatility and interest of such a model. It provides guarantees for further use in operational avalanche forecasting services where it would be required to get information on complete mountain ranges. Finally, we give preliminary results for the transposition of the method to the problem of triggered avalanches. This chapter thus clarifies perspectives of the use of statistical models for snowpack stability information derived from snow cover models.

5.1 Introduction

Avalanche warning services rely both on physical modeling of the snowpack [Morin et al., 2020] and snow or avalanche observations in the field [Schweizer and Wiesinger, 2001; Jamieson and Schweizer, 2005; Coléou and Morin, 2018]. Physical snow cover models allow for exploring a wide variety of terrains once meteorological conditions can be observed or modeled. Snow pit observations are direct snapshots of the snow conditions but are limited to a small number of safe areas. However, none of these two sources of data directly provide avalanche information. Field stability estimation can be obtained by stability tests such as compression test, extended column test [Simenhois and Birkeland, 2009] or rutschblock [Föhn, 1987a]. However, stability test results are hardly representative outside of the area they were realized [Schweizer et al., 2008a] and are limited to safe areas. Then, stability inferred from snow cover models is complementary to field observations. This stability can be summarized from snow cover model outputs with mechanically-based stability indices, sets of expert rules (see Chapter 2), or statistical models (see Chapter 3).

Avalanche releases are a good indicator of current and short-term snowpack instability. Therefore, records of avalanche releases have been organized in concerned countries. Avalanches can be observed from a distance. It is then possible to cover large areas, despite the difficulty of conducting in-place measurements. In France, the oldest network is the Enquête Permanente Avalanche (EPA) [Bourova et al., 2016] which was designed to record all avalanches exceeding defined thresholds in identified avalanche paths. The initial goal was to identify developed areas or infrastructures threatened by avalanche hazard. It mainly concerns large avalanches and is not available in real-time. To overcome these limitations, Meteo-France also performs avalanche observations in the nivo-meteorological network (NMN) [Giard et al., 2018]. These observations are reported in near real-time (up to twice a day) and cover large areas (all that is visible from the observer). Moreover, if EPA focuses mainly on natural release, the NMN network reports both natural release (NMN dataset) and artificial triggering (NMNT dataset). These two sources provide different but complementary information on past avalanche activity.

The NMN dataset was early used for statistical prediction of avalanche activity through nearest neighbor methods in ski resorts [Navarre et al., 1987]. The operational forecaster uses the observations reported by NMN observers to estimate current avalanche activity. However, the NMN dataset was no longer used for optimizing statistical models of avalanche activity after the first tries in ski resorts. The current statistical models for avalanche activity forecasting in France rely on the EPA dataset [e.g. Sielenou et al., 2021]. However, using an alternate dataset can give confidence in the robustness of the methods by ensuring the capacity of statistical models to identify correct relations between input variables and avalanche activity regardless of the precise definition of the target variable. Moreover, the use of the NMN dataset allows for insights into the information that EPA does not report. For instance, when EPA focuses on avalanches reaching infrastructures or houses, NMN can also report high-elevation and more limited avalanche activity. Therefore, NMN is well suited to test the robustness of machine learning approaches on slightly different datasets observing the same phenomenon and to extend the scope to larger areas of the French Alps and Pyrenees.

The avalanche bulletins combines in the avalanche danger both natural avalanche activity and hazard for triggered avalanches [Statham et al., 2010; EAWS, 2022]. Then, forecasters also need information

Table 5.1: Selected massifs. For each massif are reported the time period considered and the observation stations whose observations are used.

Massif	Time period	Observation stations
Haute-Maurienne	1994 - 2021	Bessans, Bonneval-sur-Arc, Val Cenis
Haute-Tarentaise	1983 - 2021	La Rosière, Les Arcs, Tignes, Val d'Isère
Vanoise	1971 - 2021	Courchevel, La Plagne, Les Menuires, Méribel, Pralognan, Valmorel
Vercors	1989 - 2021	Gresse en Vercors, Lans en Vercors, Villard de Lans
Haute-Bigorre	1983 - 2021	Barèges, Gavarnie, Luz Ardiden, Piau

on the stability of the snowpack when put under external additional loading. The NMN observers also report triggered avalanche activity, whether from preventive triggering or accidental triggering they may observe. We therefore refer to this specific dataset as the triggered avalanches reported by NMN observers as the NMNT dataset. The NMNT dataset thus provide a preliminary insight into the susceptibility of the snowpack to triggered avalanches that can be combined with machine learning in order to roughly estimate the capabilities of such model to also cover to complementary problem of triggered avalanches.

National avalanche forecast services produce bulletins for all mountainous areas of the country, which means in France, the French Alps, Pyrenees and Corsica [Morin et al., 2020]. Models for decision support should therefore provide information on all these areas. In previous chapters, we focused on a test zone in the Haute-Maurienne massif. Focus on specific regions may be useful for local avalanche forecast but do not provide information on the relevance of the method on others areas. One of the interests of snow models is to cover large areas and provide information everywhere is needed. It would be of interest to have similar properties when providing a tool to post-process snow cover model outputs to produce new indicators.

The initial machine learning methods were tested on a limited number of ski resorts [Navarre et al., 1987; Buser, 1989]. It covered both natural and triggered avalanche activity. Afterward, literature on machine learning models for avalanche activity prediction focuses on limited test areas ranging from a few to hundreds of squared kilometers [e.g. Föhn et al., 1977; Obled and Good, 1980; Dreier et al., 2016]. A few studies consider larger areas [e.g. Davis et al., 1999; Floyer and McClung, 2003; Hendrikx et al., 2014] but do not span a wide diversity of snow climates. Moreover, all these studies do not explicitly consider triggered avalanches. To our knowledge, only Sielenou et al. [2021] covered the whole French Alps and showed that the transposition of such a model is not immediate as different groups of predictors may be important depending on the climatology of the mountain ranges considered.

Moreover, avalanche forecasting services are designed to warn the authorities and the public about avalanche activity to protect urbanized areas from large natural events and mountain practitioners outside controlled areas. It means that both natural and triggered avalanche activity have to be considered by the avalanche forecaster and thus would require decision support tools.

In this chapter, we extend the method of Chapter 3 and strengthen its results by using the same statistical model on a different dataset. We first show the usefulness of the NMN dataset to replace or complement the EPA dataset when forecasting avalanche activity. We then spatially extend the results to a selection of mountain ranges across the French Alps and Pyrenees. Finally, we give preliminary results on the prediction of triggered avalanches activity, based on the NMNT dataset.

5.2 Material and methods

5.2.1 Studied area and time period

We selected five areas (so-called massifs) in French Alps and Pyrenees based on the number of available stations reporting avalanche observations and the historic period available for each station. We also selected different climates (from Alpine foothills to higher internal massifs and Pyrenees). We selected the Haute-Maurienne massif for comparison purposes with the results of Chapter 3. We added Haute-Tarentaise and Vanoise, high elevation massifs of the French Alps with numerous large ski resorts. We also added the Vercors massif to represent the Alpine foothills conditions and the Haute-Bigorre massif in the Pyrenees. Selected massifs with used observation stations are presented on the map of Figure 5.1.

The time period considered depends on the massif. We consider years for which at least three observation stations report data. The summary of the considered period for each massif is provided in Table 5.1.

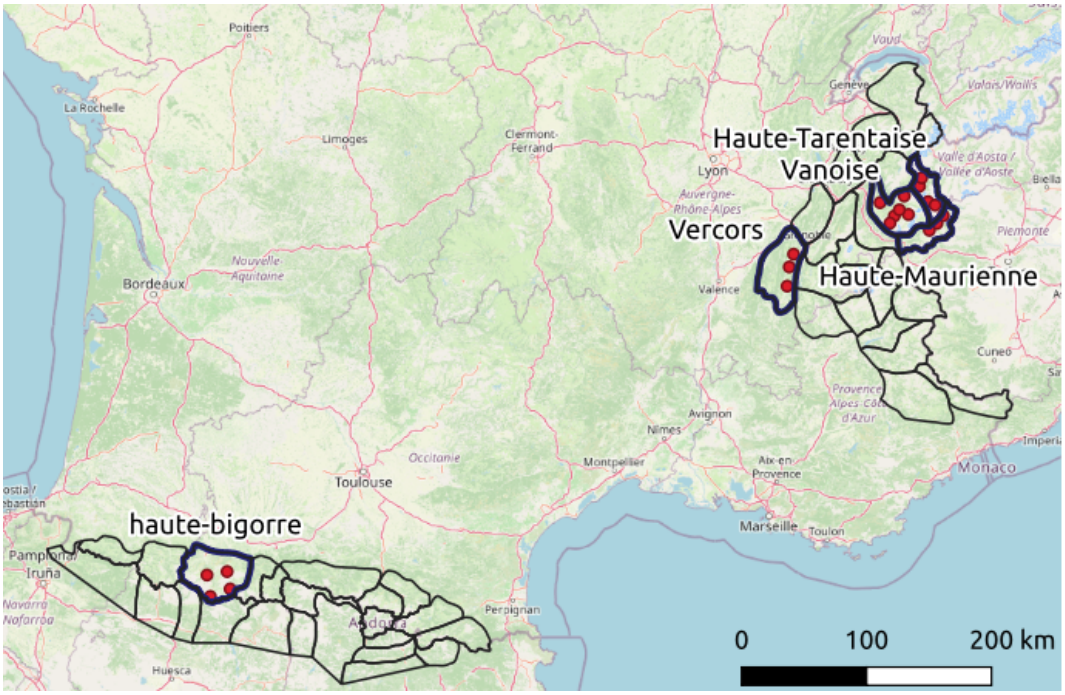


Figure 5.1: Map of the selected massifs among all French Alps and Pyrenees massifs. Red dots identifies stations reporting avalanche observations in the selected massifs.

Each massif is partitioned into classes of aspect and elevation with eight aspects (from North to North-East) and three elevation bands (centered at 1800, 2400 and 3000 meters a.s.l.). We consider the daily time step from 18 the day before to 18 p.m. We therefore use the term "situation" to refer to the conditions for a given day, in a given elevation band and aspect. In all this work, we use this reference spatial resolution, that is to say, 24 situations per day.

5.2.2 Avalanches dataset

We use the NMN dataset. The observer reports up to twice a day if avalanches were observed (whether it is on its ski resort or on opposed slopes). The number of avalanches is reported in several classes: no avalanche, one avalanche, two, three, four, five avalanches, from five to ten, ten to twenty, twenty to thirty, and more than thirty. When classes are ranges (e.g., ten to twenty avalanches), we arbitrarily choose the maximum number of possible avalanches in the class, and when it is not bounded (e.g. more than thirty), we arbitrarily choose the minimum plus 10 (here, 40). The elevation of the start zones is also reported in several bands, as well as the aspect in 8 classes from North to North-East (and all orientations). We then associate the avalanches to our elevation and aspect bands based on this information. If avalanches occur mainly in one elevation or aspect band, then the other elevation and aspects are undefined. The observer could also report a non-observation if the observation was not possible. In this case, we also consider it undefined. All the stations of the massif are aggregated to have an image of the avalanche activity by summing the number of observed of avalanches in each aspect and elevation band. If some bands remain undefined, because no observation was available this day or mainly concern other aspect or elevation ranges, then it is removed from the dataset. We then obtain a number of avalanches by classes of elevation and altitude for each day. The same process is done for the triggered avalanches reported by the same network.

The observation network consists mainly of observers from ski resorts. Hence, no observations are available before the ski resort's opening and after its closure. Therefore, we only consider days between the 20th of December and the 30th of April. We hereafter refer to this period as the winter season.

For comparison purposes, we compare the results with the dataset from Enquête Permanente Avalanche (EPA), see Chapter 3.

5.2.3 Snowpack information

For each aspect and elevation band, meteorological and snow cover information is provided by the SAFRAN-SURFEX/ISBA/Crocus model chain. SAFRAN provides meteorological information by adapt-

ing numerical weather prediction on a gridded domain to the area of interest and assimilates observed meteorological data [Durand et al., 2009]. We used the reanalysis publicly available on French Alps, Pyrenees and Corsica [Vernay et al., 2020]. The SURFEX/ISBA/Crocus model is a one-dimensional snowpack model representing snow cover evolution with a multi-layered scheme based on physical evolution laws [Brun et al., 1989; Vionnet et al., 2012]. It uses as an input the meteorological data from the SAFRAN model, and it is coupled to the soil scheme ISBA-DIF [Decharme et al., 2011] to represent energy and mass exchange at the bottom of the snowpack. Accordingly to the spatial resolution of the avalanche observations, snow conditions are computed for eight aspects and three elevation levels (1800, 2400 and 3000 m).

5.2.4 Stability analysis and machine learning methods

Stability analysis of snow conditions for each data point is done as in Section 3.2.4 with dry and wet snow stability indices and derivatives. We then use a Random Forest classifier with the same modelling approach and input variables as presented in Chapter 3.

We first compare the results of the model trained with the NMN dataset to the results of previous section, on binary classification (avalanche days and non-avalanche days). The NMN dataset nevertheless allows for more detailed insights on avalanche activity. The number of avalanches can be reported in a wide range from one to more than thirty. It makes possible to define different thresholds to consider a day as an avalanche day. In previous chapters, one avalanche was sufficient to consider an avalanche situation. We keep this definition by default but also explore the definition of an avalanche day with at least two, three or more avalanches. Note that a day still requires to have no avalanche observed to be identified as a non-avalanche day.

Finally, we use the triggered avalanches reported by the nivo-meteorological observers (NMNT dataset) rather than the natural avalanche activity, on the Haute-Maurienne massif. A day is considered an avalanche day if at least one triggered avalanche is observed.

5.2.5 Evaluation of the machine learning model

Evaluation is conducted the same way as presented in Section 3.2.6 with a leave one year out approach (LOYO).

The Random Forest model produces a continuous output value as the proportion of tree voting for an avalanche situation. It is then possible to use classical evaluation of a binary classifier by selecting a threshold (t) on this value to discriminate avalanche and non-avalanche situations. We then present the results on the form of a confusion matrix and with classical scores presented in Table 3.3. We also plot ROC (Receiving Operating Characteristic) curve as a common way to evaluate and compare binary classifiers.

Beyond the binary classification, we also consider directly the output value of the random forest. Histograms of the output value are used to evaluate advanced properties of the resulting avalanche activity model. In particular, the NMN dataset reports avalanche activity on a scale from one avalanche to more than 30. This information questions the way of identifying avalanche and non-avalanche situations. It is then possible to vary the definition of an avalanche situation considering only situations with more than n observed avalanches (by removing from the dataset situations with more than one and less than n observed avalanches) with n between 2 and 20. We also consider the histograms of output values for different groups of observed number of avalanches.

Finally, the output value of the random forest is commonly interpreted as a probability of observing a situation. We check this interpretation with the reliability diagram (or calibration curve) that compares the proportion of observed events to the output value of the random forest.

5.3 Results

5.3.1 Comparison of the datasets

The EPA and NMN datasets both observe avalanche occurrences. However, while EPA focuses on avalanches approaching valley floors or infrastructures, the NMN dataset is linked to observation from higher altitudes on larger areas. The avalanche activity reported by the two datasets are thus different. Table 5.2 compares the main characteristics of the two datasets. Being less restrictive on the avalanche activity observed, the number of avalanche days reported by the NMN dataset is nearly twice the number of avalanche days observed in the EPA dataset whereas the half period is considered (58 years for EPA

Table 5.2: Comparison of the EPA, NMN and NMNT main dataset characteristics. The periods considered (in terms of number of years and length of the winter period) are reported. The number of avalanche situations are compared to all considered situations.

Dataset	Years considered	Winter period	Avalanche situations	Total considered situations	Part of avalanche situations
EPA	1960 – 2018 (58 years)	15/10 – 15/05	2518	≈ 230 000	1.1%
NMN	1994 – 2020 (26 years)	20/12 – 30/04	4537	≈ 60 000	7.6%
NMNT	1994 – 2020 (26 years)	20/12 – 30/04	2091	≈ 60 000	3.2%

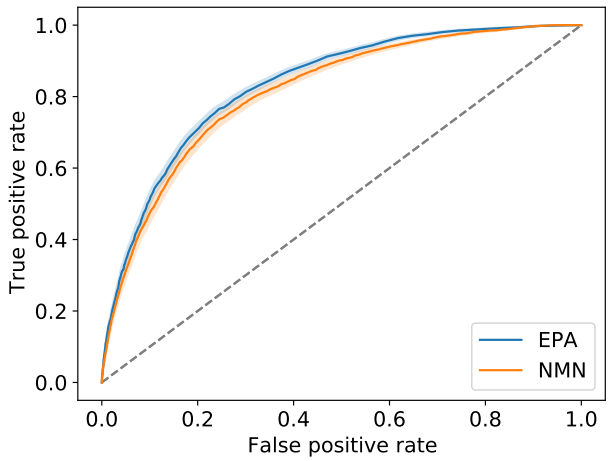


Figure 5.2: ROC curve of the model trained and evaluated with the NMN dataset (orange line), compared to the same curve for the model trained and evaluated with EPA dataset. Shading represents the uncertainty by bootstrap on test years.

dataset and 26 years for NMN dataset) and the winter season is shorter. In consequence, the NMN dataset is less unbalanced compared to the EPA one.

5.3.2 Compared performance on binary classification

The ROC curve of the model trained with NMN dataset and evaluated with a leave one year out approach on the years 1994 to 2021 is shown on Figure 5.2. It is compared on the same figure with the results of Chapter 3 with the EPA dataset. Results are close, with an area under the ROC curve of 0.315 for the NMN dataset trained and evaluated on 27 years and 0.334 for the EPA dataset on 58 years. False positive rates are very similar, with values of 23.6% and 25.5 %, respectively, for the EPA and NMN datasets. The same is observed for recall (or true positive rate), with respective values of 75.3 and 74.3%. The ROC curves are so close that the uncertainty zone around the ROC curves overlaps. The two models thus have rather similar performances.

However, the NMN data source is very different from EPA. It gathers much more avalanche days compared to the EPA dataset. In consequence, the distribution of avalanche and non-avalanche days is very different and therefore the precision indicators differ: 3.3% with the EPA dataset and 19.2% for the NMN dataset. The full confusion matrix for NMN dataset is given in Table 5.3.

5.3.3 Evaluation of the output probability

In the previous section, we consider the binary classification between avalanche and non-avalanche days. However, the random forest method provides more information as the output is a proportion of trees voting for each class. The histogram of the values of this probability is given in Figure 5.3. The histogram of the output values of the random forest is plotted for the different number of avalanches observed from

Table 5.3: Confusion matrix for the test NMN dataset: observed and predicted avalanche days ("Avalanche") and non-avalanche ones ("Non-avalanche") summed over elevation and aspect ranges. A threshold value of 0.08 is used, i.e. predicted probabilities over 0.08 are considered to identify avalanche days. The corresponding recall is 74.3%, the false positive rate is 25.5% and the precision is 19.2%.

		Predicted	
		Avalanche	Non-avalanche
Observed	Avalanche	3 371	1 166
	Non-avalanche	14 174	41 299

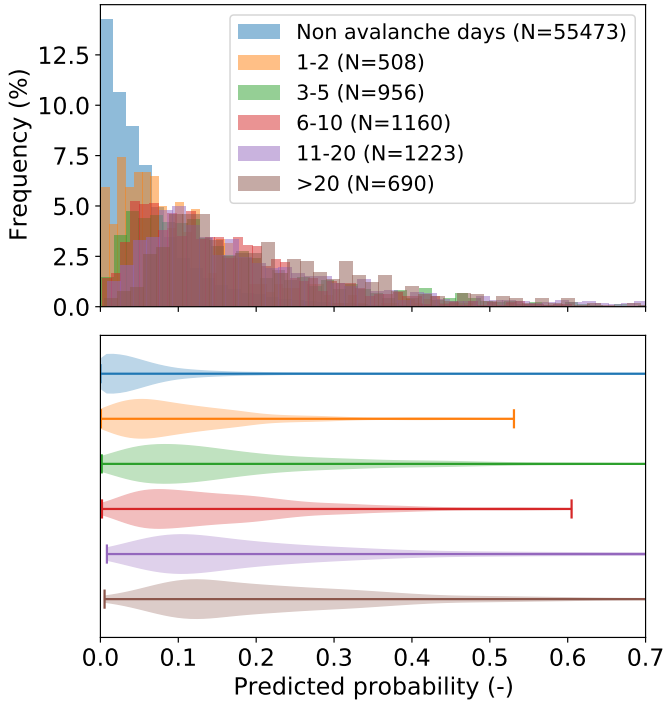


Figure 5.3: Distribution of the output values of RF model for non-avalanche days, or avalanche days categorized by the number of observed avalanches from 1 or two to more than twenty. Histogram (on top) and violin-plot representation (bottom) of the same data. The total number of elements in each class is reported between parenthesis in the legend.

zero (no-avalanche day) to more than twenty. We show that the more avalanche is observed, the more the histogram is switched right. That means that the output probability is higher for the most critical days (with a lot of observed avalanches).

The statistical quality of the output value, interpreted as a probability, is validated with the calibration curve of Figure 5.4. The output probability is compared to the observed probability of observing avalanches on the field. The probability is slightly underestimated for low values (under 0.08) and slightly overestimated for higher values (above 0.15), but there is an overall good agreement between observed and predicted avalanche activity in terms of probability.

5.3.4 Variations around the definition of avalanche day

The NMN dataset quantifies the number of observed avalanches. It allows for different partitions between non-avalanche and avalanche days. In this section, we test different definitions of the avalanche situations, by removing days with only a few avalanches observed, while the non-avalanche situation remains unchanged. Results are summarized through the ROC curves of Figure 5.5. The results for the binary classification considering days as avalanche days as soon as one avalanche is observed (NMN) are compared to the classification in which days with six avalanches or less are removed from the dataset (NMN>6). We show that the classifier has better performances when removing days with a few avalanches: the area under curve is 0.341 for NMN>6 while it is 0.315 for the full NMN dataset, which is slightly more than the reported uncertainty.

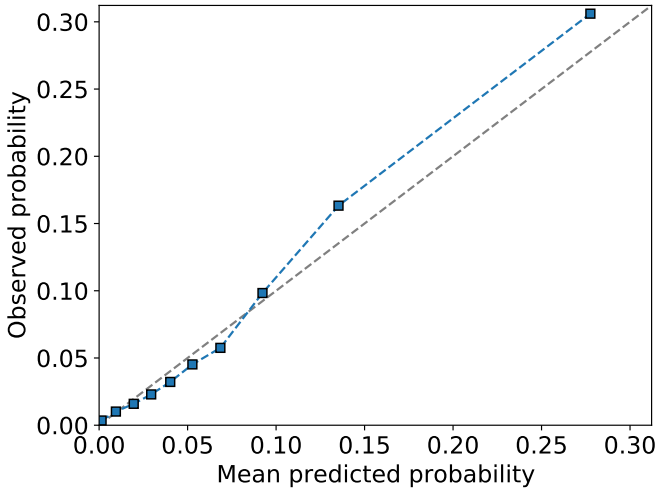


Figure 5.4: Reliability diagram (or calibration curve) of the output value of the RF model on NMN dataset: predicted probability value compared to the avalanche observation frequency. Each point represents the same amount of data.

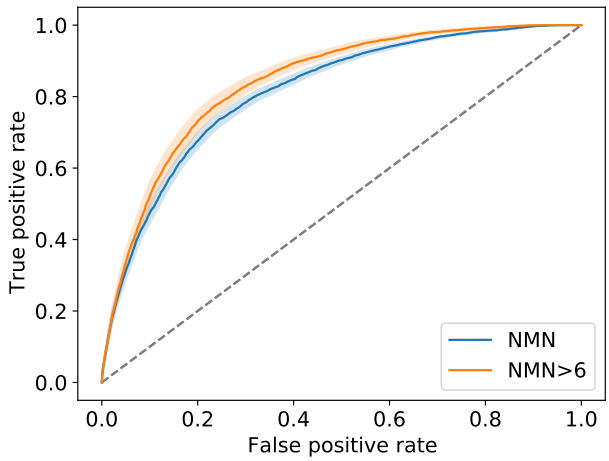


Figure 5.5: ROC curve of the mode trained and evaluated with the NMN dataset (orange line), compared to the same curve for the model trained and evaluated with the same dataset but with days with six or fewer avalanches removed (NMN>6). Shading represents the uncertainty by bootstrap on test years.

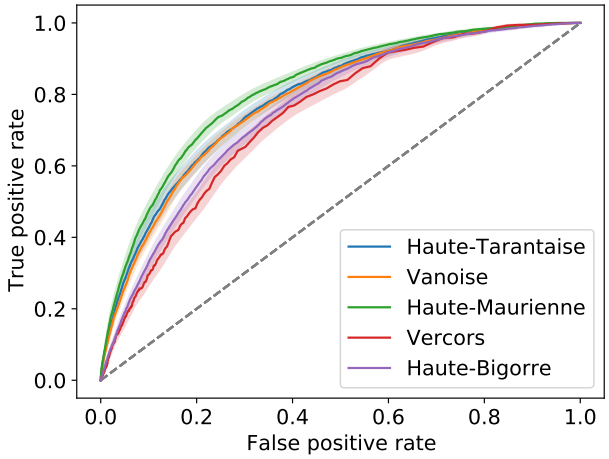


Figure 5.6: ROC curve of the mode trained and evaluated with the NMN dataset for different massifs. Shading represents the uncertainty by bootstrap on test years.

Table 5.4: Results for learning on the different massifs. The number of considered years (#Y) and number of avalanche situations (#AS) in the dataset are reported in the first columns, followed by the common scores: area under ROC curve (AUC), false positive rate (FPR), recall and precision.

Massif	#Y	#AS	AUC	FPR (%)	Recall (%)	Precision (%)
Haute-Maurienne	27	4537	0.315	24.5	73.7	19.7
Haute-Tarentaise	38	5486	0.290	30.0	73.5	18.8
Vanoise	50	6140	0.285	27.8	70.8	19.5
Vercors	31	545	0.242	33.8	70.6	2.5
Haute-Bigorre	38	6031	0.258	33.5	72.3	15.8

5.3.5 Geographical extension to different massifs

In the previous chapters and sections, we focused on a limited area of Haute-Maurienne to evaluate the statistical model. In this section, we extend the results to different massifs of the French Alps and Pyrenees (Figure 5.1). We show the results for the massifs of Haute-Bigorre, Haute-Tarentaise, Haute-Maurienne, Vanoise and Vercors on Figure 5.6. The best results are obtained on the Haute-Maurienne massif with an area under the ROC curve of 0.315. The massif of Haute-Tarentaise and Vanoise are close, with a respective area under the ROC curve of 0.290 and 0.285, and even closer if avalanche days are not defined with a minimum number of avalanches of one but rather four to six (not shown). The massifs of Haute-Bigorre follow with an area under curve of 0.258, and finally, the classifier has its lower performance on the Vercors massif (area of 0.242). The detailed results for all considered massifs, with the number of considered winters and avalanche situations are summarized in Table 5.4.

5.3.6 Extension to triggered avalanches

Finally, we extend the method to triggered avalanches by using the triggered avalanches observation of NMN dataset. The ROC curve for binary classification between avalanche days and non-avalanche days for triggered avalanches is compared to the same curve for natural avalanche activity on Figure 5.7. For the triggered avalanche activity the area under ROC curve is 0.355 compared to 0.315 for natural activity. The true positive rate or recall is 77.8% while the false positive rate is 20.3%. The precision is however lower compared to the natural avalanches, with a value around 11.5%.

5.4 Discussion

5.4.1 Interest of the NMN dataset

We show the usefulness of the NMN dataset as a qualification of the avalanche activity for a machine learning approach despite significant differences in the observed avalanche activity (Table 5.2). We show

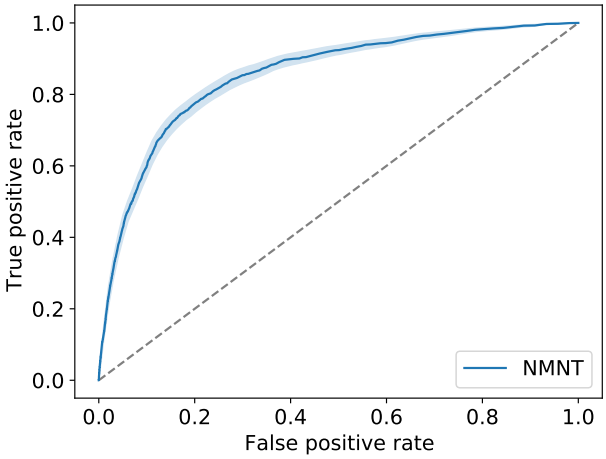


Figure 5.7: ROC curve of the mode trained and evaluated with the NMNT dataset for Haute-Maurienne massif. Shading represents the uncertainty by bootstrap on test years.

with Figure 5.2 that the results are similar to those obtained with the EPA dataset for avalanche activity prediction into binary classes. The model trained with the NMN dataset allows for a true positive rate or recall of 74.3%, a false positive rate of 25.5%, and precision of 19.2%. The in-depth evaluation also shows similar reliability on the output probability, similar to those obtained by EPA (Figure 5.4).

The NMN dataset has the advantage of clearly identifying the days without avalanches compared to the EPA dataset, which only reports days with observed avalanches and does not allow to clearly define days with no avalanches. Moreover, the NMN dataset does not have pre-defined areas for avalanche observation: the observer reports all avalanches that he can observe. Hence, it allows for more precise quantification of avalanche activity: a single avalanche observable in a high-altitude area can be reported or, on the contrary, a high number of avalanches on all aspects and elevations [Giard et al., 2018]. Figure 5.3 shows that even though the number of observed avalanches is not used for learning, the algorithm is able to produce a higher output probability for days with high avalanche activity observed rather than days with lower avalanche activity.

A first approach to use the information contained in the number of avalanches in the NMN dataset is to choose the avalanche activity we want to predict more precisely. This is done in Figure 5.5 by focusing on non-avalanche days and days with more than six observed avalanches. We showed that we obtained better results by using only days with more than six avalanches, with a recall of 77.3% and a false positive rate of 23.4%. The precision is mechanically reduced to 15.4% due to the lower amount of avalanche days. This is a first step toward better quantification of the avalanche activity. Then, further work may take advantage of the higher dynamics of the avalanche number reported by the NMN. An option is to use more than two classes: we presented here avalanche and non-avalanche days. It is, for instance, possible to cut the dataset into three classes non-avalanche, low avalanche activity and high avalanche activity, as done by Sielenou et al. [2021]. A step further would be to totally get rid of the partitioning into classes that inevitably contain uncertainty and arbitrariness. This could for instance, be done with the regression mode of Random Forest, which does not classify into classes but provide the mean value of the different leaves corresponding to the current situation of each tree and then provide a continuous value as output [Breiman, 2001; Hastie et al., 2009].

5.4.2 Robustness of the model

The use of the same random forest model with two different sources of data allows for an evaluation of the robustness of the method. We here show that the Random Forest method is comfortable with our two different sources of data: NMN and EPA, with similar skills: A recall of 75.3 and 74.3% respectively, for the EPA and the NMN dataset. Differences are in the range of the uncertainty. The choice of the dataset depends on the availability of the data and the avalanche activity that we want to predict. The EPA has the advantage of a long period of time and reports on well-identified avalanche paths [Bourova et al., 2016]. On the other side, NMN allows for recording avalanche activity on wider areas with less constraint to capture all the activity visible by the observer. The network is also planned for observing situations with no avalanches; thus, the information is different from the absence of observation that

gives better confidence on the days with no avalanches [Giard et al., 2018]. The other side of the coin is that the avalanches are less precisely located in space and avalanche activity may be confined in areas far from the observation point, for instance. The usefulness of the availability of the data on long periods of time and the homogeneity of the data in time remains to be evaluated. However, while working at the scale of mountain ranges of several hundreds of square meters, the information is still fully relevant.

5.4.3 Geographical extension to different massifs

Most statistical models for avalanche forecasting have been developed and evaluated on relatively small, delimited areas. The first models were applied to one or a few ski resorts [Navarre et al., 1987; Buser, 1989], and then mostly limited to areas of a few to hundreds of squared kilometers [e.g. Föhn et al., 1977; Obled and Good, 1980; Dreier et al., 2016]. However, from a forecasting point of view and for an operational use, such model should be applied to all the mountain ranges. For instance, in France, operational forecasting covers French Alps, Pyrenees and Corsica [Coléou and Morin, 2018]. In these operational areas, different meteorological and snow climates coexist [Sielenou et al., 2021; Reuter et al., 2022a]. Hence, if a model is relevant for a given mountain range, it does not mean it applies to another one. However, statistical models are rarely tested at larger scales.

Meteorological and snow climates varies between massifs [Sielenou et al., 2021; Reuter et al., 2022a]. Hence, if a model is relevant for a given mountain range, it does not mean it applies to another one. However, statistical models are rarely tested at larger scales. We here evaluate the model on different massifs of different climates and geographical positions (Figure 5.6). We show that we observe similar results for Haute-Maurienne, Haute-Tarentaise and Vanoise, with areas under the ROC curve between 0.285 and 0.315. The massif of Vercors has significantly lower performance with an area under the ROC curve of 0.242 (Table 5.4). The Vercors massif is a low-altitude massif with significantly lower avalanche activity compared to others considered here. Moreover, ski resorts are linearly distributed along the ridge of the eastern limit of the massif with limited visibility of the rest of the massif, which may lead to a lower quality of the avalanche activity report in the NMN dataset. The Vercors massif also has only three active observation stations, which are small stations with reduced opening periods. This is lower to the other considered massifs (e.g. four in Haute-Tarentaise, 5 in Vanoise, see Table 5.1). The lower activity combined with a lower report rate may explain the lower performances of the model on this massif. In the case where one class become very rare, other statistical models, taking into account rare events may be required for relevant predictions [e.g. Evin et al., 2021b]. The massif of Haute-Bigorre is in an intermediate position with an area under the ROC curve of 0.258. In this massif, the ski resorts have reduced opening periods compared to the high elevation Alpine ski resorts and are closed preventively in the most hazardous situations, which may also lead to a more limited avalanche information compared to the massifs of Haute-Maurienne, Haute-Tarentaise or Vanoise. Hence, the different studied massifs have similar performances with some differences for massifs with lower avalanche activity or information available.

5.4.4 Extension to triggered avalanches

The preliminary extension to triggered avalanches confirms the usefulness of both the random forest statistical model, the selection of the input variables and the NMN dataset for estimation of susceptibility of the snowpack to artificial triggering. However, NMN dataset is not fully representative of the snowpack far from ski resorts, as the triggered activity mainly includes the avalanches preventively triggered by ski resorts to protect ski tracks. As these preventive measures are taken regularly during the winter, the snowpack is, in these areas, largely modified by the previous preventive triggering. The NMN dataset can also include accidental triggering by skiers. Nevertheless, the report of accidental triggering require the observer to be aware of the event. The dataset also gather quite different triggering, whether it involves only a skier, or different explosive triggering methods. However, it provide a systematic source of information on the susceptibility of the snowpack to triggering as ski resorts will try preventive triggering as soon as they think the susceptibility of the snowpack may be sufficient. In contrast, databases linked to skier triggering are largely biased by the frequentation of the backcountry. Then, NMN provide an interesting overview of the susceptibility of the snowpack, that is largely less biased by the mountain frequentation compared to accident databases, for instance, and could be one of the information source to construct a machine learning support tool to estimate the stability of the snowpack under additional loading.

5.5 Conclusion

In this chapter, we generalize the results on the performance of the proposed random forest model for avalanche activity estimation, by taking advantage of the French nivo-meteorological network reports of avalanche activity. We highlight the capabilities of this dataset for the training of statistical models for avalanche forecasting. The performance of the trained model is shown to be similar whether it is trained with the EPA or the NMN dataset. It also shows the robustness of the chosen random forest statistical model and overall method to the observation dataset used for training and evaluation. It is also shown that even though not given to the model in the training phase, the critical days are correctly identified by the model and the probability of being an avalanche day is correctly reproduced on the test set. This allows using the developed random forest model for avalanche forecasting with slightly different data sources depending on the goal it is designed for. Here, for instance, the NMN dataset allows for a more refined analysis of the avalanche activity compared to the EPA dataset, which is focused on larger avalanches reaching valley floors. Still, the performances are similar for the two datasets.

Then, the method extensively evaluated on the Haute-Maurienne massif in this chapter and previous ones is transposed to other massifs of the French Alps and Pyrenees to cover a large variety of snow and meteorological conditions as well as observational conditions. We show that performances are similar for the massifs of Haute-Maurienne and Haute-Tarentaise, but slightly lower for the massif of Haute-Bigorre and Vercors, where avalanche report rate and avalanche activity (especially for the Vercors massif) are lower. Thereby, we strengthen the results of the previous chapters by showing that they are not limited to the Haute-Maurienne area.

We finally extend the method to triggered avalanches, also reported by the NMN observer network. The results are comparable to the results of natural avalanches, which gives encouraging perspectives for this specific but important application, as it it the main cause of the fatalities nowadays.

We here show the capabilities of the NMN dataset for binary classification of avalanche activity between non-avalanche days and avalanche days. However, the NMN dataset not only reports the observation (or not) of avalanches but also quantifies the magnitude of the avalanche cycles by counting the number of observed avalanches. It allows for going further than the binary classification between avalanche and non-avalanche days. This could include the use of more than two classes [e.g. Sielenou et al., 2021] or changing the target from one class to a raw number of expected avalanches or any other continuous quantification of avalanche activity by using the random forest for regression rather than for classification [Hastie et al., 2009]. Moreover, this chapter compares the capabilities of EPA and NMN datasets separately. However, other observations of avalanche activity may also be used in the future, such as satellite detection of avalanche deposits [e.g. Karas et al., 2022]. It is also conceivable to combine the different sources of data to define an aggregated avalanche activity.

Chapter 6

Conclusion and perspectives

Contents

6.1 Conclusion

This thesis aimed to explore one method for improving avalanche activity estimation from snow cover models by combining mechanically-based diagnosis and statistical methods. In particular, we develop and evaluate a combination of mechanically-based indices and snow variables and a random forest statistical model to predict expected avalanche activity. The long-term objective of this work is to provide a decision support system for avalanche forecasters with better performances than the current MEPRA expert model.

6.1.1 Mechanically-based stability indices

A first part of the work was dedicated to mechanically-based indicators of snowpack instability (Chapter 2). We first reviewed the existing literature on mechanically-based snowpack stability indicators. We then selected those compatible with the use in combination with snowpack modeled by detailed snow cover models in terms of vertical resolution of the processes represented and computational cost. We evaluated the usefulness of the selected snow cover models in typical situations to highlight the processes they represent and those that each model does not represent. We underlined the usefulness of combining different stability indices for an overall estimation of the stability as they complement each other representing different processes of avalanche formation.

The selected set of stability indices is designed to bring additional information for stability analysis in combination with the snow cover model Crocus and to provide a global overview of the stability from snow cover model data. The strength-stress ratios representing the initiation processes, whether it involves only the weight of snow layers (for natural activity) or an external trigger (e.g., skier, for artificial triggering), are complemented by indices related to crack propagation in the snowpack (critical crack length). The capabilities of this selection is therefore validated with the statistical approach. This first step allows for reducing the information provided by snow cover models, with the objective of retaining information specific to snowpack stability.

The stability indices require knowledge of the mechanical properties of snow layers, which snow cover models do not directly model. Hence, these mechanical properties are generally derived from other snow properties, mainly snow density, sometimes completed by grain shape. However, snow models were not designed to represent the whole processes involved in the change in mechanical behavior and density is known to be insufficient to represent the complex mechanical behavior of snow [e.g. Roch, 1966b]. This convoluted way of inferring mechanical properties inevitably introduces significant uncertainty in the estimation of the mechanical parameters and consequently in the stability estimation. Additional work would be required in snow cover models to represent the mechanical behavior of snow layers.

Some processes are not represented in the selected indicators, such as dynamic propagation of cracks into the snow [e.g. Gaume et al., 2018; Bergfeld et al., 2021b]. Even though models emerged to explain processes during dynamic crack propagation in the snowpack [e.g. Bobillier et al., 2021], these models are still very complex and not suited to be used in real-time with snow cover models. Further research is needed on these processes to determine simple indices to summarize the most important factors of instability before use in an operational way.

6.1.2 Statistical approach for avalanche activity prediction

A second part of the work uses snow and meteorological information, including the previously selected stability information, to produce an indicator of associated expected avalanche activity (Chapter 3). Among the existing machine learning techniques, we selected the Random Forest for its ability to treat non-linear problems without the requirement of a distance and because it has shown to be relevant for such problems [e.g. Sielenou et al., 2021]. The Random Forest technique is used with input data from meteorological and snow models complemented by mechanically-based stability indices. The first considered observation dataset consisted of the avalanche records of Enquête Permanente Avalanche (EPA) [Bourova et al., 2016]. The work presented here has the particularity to combine mechanically-based stability indices with other snow and meteorological data as an input, in proportions that were never presented in previous literature. Moreover, we focus on the difficult problem of identifying avalanche activity in precise areas as we work at the scale of specific elevation and aspect ranges (eight aspect ranges and three elevation bands in the area of interest).

We also developed a robust method for the evaluation of machine learning methods when related to snow and avalanches. Since snowpack accumulates information on past meteorological conditions, it is impossible to consider that two days are independent. In consequence, some results and typical techniques in machine learning science are no longer usable. For instance, it is not possible to randomly draw

observations to compose train and test sets: the dependence between observations must be considered. As we focus on seasonal snow, we can consider that snow melts in summer; therefore, days from two different winters are independent. Therefore, we separate the train and test set based on the separation between years. Systematic evaluation is then performed through a leave one year out approach. Such a method should be generalized to allow for comparison with different studies in the snow community.

In the example of the Haute-Maurienne massif, we show the usefulness of the method for classifying between avalanche and non-avalanche days based on our dataset. The model correctly identifies 75.3% of avalanche situations and 76.4% of non-avalanche situations. However, the precision remains low (around 3.3%) due to the low number of avalanche days compared to non-avalanche days in the considered dataset (Chapter 3). These scores illustrate the difficulty of predicting avalanche occurrence with a high spatio-temporal resolution, even with the current cutting-edge data and modeling tools. Yet, we showed that the developed model is already much more relevant than the currently operationally used MEPRA model that does not manage to provide information at the same scale even though it is shown to be relevant at larger scales (Chapter 4). Hence, our study opens perspectives to improve modeling tools supporting operational avalanche forecasting.

Beyond this comparison to the current operational tool, we also checked that the model appears statistically coherent (Chapter 4). We showed that the most critical days are statistically more correctly classified than days with only one avalanche, although the information is unavailable during training. Moreover, the Random Forest tool allows for producing a probability of being an avalanche day rather than simply the binary classification. We evaluated this probability and showed that it is coherent with the observations. This value is then reliable and can be used directly as a continuous indicator of expected avalanche activity. These checks prevent the rejection of the model due to failures in rare but critical situations when used in operational conditions.

The value of mechanically-based stability indices is highlighted by showing its importance in the learning process (Chapter 3). Moreover, it is shown to contain nearly all the relevant information, as it provides similar scores when used alone, compared to the complete set of input variables. These results also validate the choice of stability indices: the information contained in the reduced set of stability indices is shown to be equivalent to all tested variables in terms of capacity of prediction for observed avalanche activity. These results also and underline the usefulness of stability indices in combination with statistical methods.

6.1.3 Generalization of the statistical model

We studied extensively the usefulness the combination of physical knowledge and statistical methods on the example of the Haute-Maurienne massif in the French Alps. However, such a model aims to be used in the whole French mountains, for which avalanche bulletins are currently produced, from the Alps to Pyrenees and Corsica. The model also has been trained with a specific observation, the EPA dataset, which has its drawbacks and does not fully represent the avalanche activity. It could then introduce a bias in the results. We then examined the transferability of the modelling approach on different areas and with a different source of data (Chapter 5). We selected five different areas and used an alternative data source: the observation of avalanches by the observer network of Meteo-France (NMN dataset). We showed that the results are similar, with some discrepancies for massifs where significantly lower information is available. This validates that the previously presented results are not specific to the peculiar case mainly studied in this work but are able to transpose to different geographical areas and data sources.

6.1.4 General conclusion

By combining physical knowledge (through the mechanically-based stability indices) and statistical (so-called machine learning) methods, we developed a new tool to relate snowpack modeled by snow cover model Crocus and expected avalanche activity. This combination of machine learning techniques and physical knowledge now become common in many domains including geoscience[e.g. Karpatne et al., 2019; Kashinath et al., 2021]. We here take advantage of these tools for the avalanche research community. We highlight the usefulness of the combination of the physical and statistical approaches for avalanche activity prediction: the use of mechanically-based stability indices allows for reducing the information produced by snow cover models, focusing on snowpack stability to then enhance performances of the machine learning model with a limited number of variables.

The method presented in this work allows relating modeled snowpack stratigraphy to avalanche activity. This is useful for short-term avalanche forecasting, especially as we showed the added value compared to the currently existing tools. The method is also interesting because of its versatility. It

can be easily adapted to model changes (change of scale or physics of the different models contributing to the snowpack modeling) as long as the goal is to relate a modeled snowpack to avalanche activity. It could also be used to study the climate evolution of avalanche activity by providing a stable way to relate snowpack conditions and avalanche activity.

The developed tool is shown to be robust to the observation data and exportable to different geographical area. We also ensure the robustness of the results through a strict evaluation procedure, taking into account the specificities of snow cover. We also confirm that the model performs better for critical days with high avalanche activity. However, snow avalanches remain a complex field of research due to their scarcity and metastability in avalanche-prone conditions that may or may not result in an observable avalanche which results in a high false alarm rate when evaluated on avalanche observations.

6.2 Perspectives

6.2.1 Snow models level improvements

The avalanche activity estimation is based on snowpack stratigraphy modeled by snow cover models. These snow cover models rely on meteorological information provided by numerical weather models. Hence, the uncertainty of all the models of the model chain are combined in the final stability analysis. An alternative would be to use observed stratigraphy from snow pit [e.g. Mayer et al., 2022]. This is a complementary and necessary approach but does not allow to take advantage of the representation of the present and future snowpack at high temporal and spatial resolution allowed by snow cover models. Moreover, snow observation also contains uncertainties [e.g. Proksch et al., 2016]. Therefore, analysis of expected avalanche activity from the snow cover model is of crucial interest.

Analysis of stability based on snow cover models will be limited by the uncertainty of these models, therefore all efforts for improving snow cover model representation of the snowpack would benefit to the analysis. The use of snow observations, either measured manually in a snow pit [Viallon-Galinier et al., 2021] or measured from space by satellites [e.g. Cluzet et al., 2020], is also a way of improving snow cover model simulation by constraining the model with observations. Moreover, for stability analysis, we showed that a significant uncertainty is introduced by the lack of representation of mechanical properties of the modeled snowpack by current snow cover models (Section 2.3). Modeling of mechanical properties of snow would require a significant step in the representation of the microstructure of the snow by snow cover models [Lehning et al., 2002b] but will also significantly improve the pertinence of the mechanically-based stability models. In parallel, the improvement of the representation of physical phenomena in the snowpack would benefit to the overall representation of the stratigraphy of the snowpack and, therefore, the analysis of stability. It is especially the case for settlement (as it is crucial for the estimation of mechanical properties) or liquid water percolation (highly influencing the snow resistance).

The observation data, the statistical and the mechanical model uncertainties have been addressed in this thesis. However, it was out of the main scope of this thesis to deal with snow cover and meteorological model uncertainties. They have been considered to provide reasonable representation of the atmosphere and snow cover. However, these uncertainties are all gathered in the final product. To provide a fully usable information to the forecasters, we have to know the uncertainty on the result and consequently on the inputs of the statistical model. Lafaysse et al. [2017] provide an estimation of the uncertainty of the processes represented in snow cover model with a model ensemble. Vernay et al. [2015] uses ensemble of meteorological model outputs to drive snow cover model Crocus and represent the uncertainty from upstream meteorological models. Studies also combine these two previous ensembles to provide a better estimation of the overall uncertainty on snow cover models outputs [e.g. Cluzet et al., 2021]. The use of combination of ensemble of meteorological conditions and snow cover models would allow to estimate the uncertainty on stability estimates and would be required for the forecaster to estimate the confidence to grant to model chain implying meteorological, snow cover modelling and statistical analysis of stability.

This thesis used a reanalysis of meteorological and snow conditions. A reanalysis means that all available observations have been taken into account in the model. However, in a forecasting perspective, up to four days, no observations are available. The models outputs may be slightly different as the simulation is not constrained by any observation. An operational machine learning model should thus be evaluated and re-trained on forecast data.

6.2.2 Optimizing the use of available observational data

We used in this work the data from Enquête Permanente Avalanche (EPA) and the nivo-meteorological network of Meteo-France (NMN) to perform mainly binary classification between avalanche and non-avalanche days. However, the number of avalanches is limited in the EPA dataset as only a limited

number of pre-identified avalanche paths are observed. The NMN dataset provides more precise information with a number of avalanches that can range from zero to more than thirty. This additional information could be used to expand on binary classification between avalanche and non-avalanche days to predict a continuous avalanche activity in terms of the expected number of avalanches (eventually compared to the commonly observed avalanche activity in the area of interest). The classification problem becomes a regression one, but the Random Forest algorithm is also well suited to this problem [Breiman, 2001; Hastie et al., 2009].

We considered our two sources of data, NMN and EPA, separately. These two sources of information are complementary as they do not report the same situations. The first one is a record of all avalanches observed, mainly from ski resorts, while the second one focuses on relatively large avalanches reaching valley floors. The advantages of these two sources of information could be combined to produce an aggregated avalanche activity index used for learning of statistical methods. Moreover, we used, for each considered geographical area (each massif), only the data available in this specific massif. This limits the amount of data to treat during the learning procedure but this also limits the performance in areas where the number of avalanche observations is limited (see Section 5.3.5). It would be possible to combine information from different geographical areas in a common learning procedure to enhance the results in massifs where observations are insufficient, at the numerical cost of the extension of the model that has to treat different snow climates and avalanche situations. The combination of different avalanche information (different geographical areas or observation sources) may improve the overall quality of the resulting dataset compared to each of the individual datasets.

6.2.3 New observation sources

We used in this work the EPA and NMN datasets which are manual observations of avalanche deposits. However, new methods are currently under development for more automatic detection of avalanches which may allow to enlarge the database of observed avalanches and therefore provide complementary information on avalanche activity. Methods based on detection of the propagation of waves originating from avalanche flow were the first to be developed, whether it is seismic waves [van Herwijnen and Schweizer, 2011] or infrasound [Mayer et al., 2020]. These systems only cover limited areas and hardly detect the smallest avalanches but provide an automatic observation. More recently, methods for identifying avalanche deposits from satellite images were developed [e.g. Leinss et al., 2020; Karas et al., 2022]. Such automatic detection allows for coverage of large areas but remains under development, especially for smaller avalanches and remain limited to recent periods. All these alternative methods for recording avalanche activity could help better represent the overall avalanche activity and therefore improve avalanche activity prediction when combined with existing databases. Each dataset of avalanche observation have its advantages and drawbacks. EPA dataset is focused on avalanches reaching valley floors while NMN has no clear observation area. The combination of different sources is then required to benefit from a large dataset of avalanche observations that allow to precisely quantify the avalanche activity at all altitudes and aspects with a precise dating of the events.

The problem of triggered avalanche also require further work on the observation dataset. We proposed preliminary results for this complex problem with the triggered avalanches observed by the NMN network but it mainly covers preventive triggering by ski resorts which does not cover the wide variety of triggered avalanches and is not fully representative of the slab avalanches triggered by skiers that are the most fatal in European Alps. Avalanche observations may be insufficient for this problem as unstable snowpack may not release an avalanche if no trigger is present. Field stability tests or other measurements of stability may be required to construct a dataset suited to the problem of triggered avalanche hazard estimation.

6.2.4 Alternative statistical methods

Random forest is a generic tool for non-linear classification or regression, agnostic of the problem considered. Hence, the particularities of the specific considered problem, here the avalanche activity prediction, is not taken into account by the algorithm. From the physics of processes involved in snowpack evolution and avalanche formation, we know that the expected avalanche activity is the result of past snow and meteorological conditions. Machine learning tools exist that take into account the aggregation of information in a sort of memory [Hochreiter and Schmidhuber, 1997], such as recurrent neural network [Giles and Gori, 1998] that use both the information at a given time but also a result of the information at all past time steps. In our work, we used time-derivatives and accumulations as input variables to get this information, whereas the adaptation of the statistical model itself would allow reducing the number of input variables. Moreover, and especially for the EPA dataset, the observational data on avalanche

contains an uncertainty on the date of the event as the observer records the observation of the deposit and therefore does not necessarily know the precise date of the event. We optimized the model to give an avalanche activity on the day of the observation, whereas the avalanche may have been released the day before or maybe even earlier. It would be possible, by adapting the statistical model, to jointly optimize the prediction of avalanche and the date of the avalanche in the uncertainty range, both in the same learning procedure.

Conclusion et perspectives

Version française du Chapitre 6. French version of Chapter 6.

Conclusion

Ce travail de thèse visait à explorer de nouvelles méthodes pour améliorer l'estimation de l'activité avalancheuse à partir de la modélisation numérique du manteau neigeux en combinant des modèles de stabilité à base physique et des méthodes statistiques. L'objectif de long terme est de fournir un système d'aide à la décision aux prévisionnistes nivologues avec de meilleures performances que le système MEPRA qu'ils ont actuellement à disposition.

Indices de stabilité mécaniques

Une première partie de ce travail a été dédiée à l'étude d'indicateurs à base mécanique de la stabilité du manteau neigeux (Chapitre 2). Nous avons tout d'abord réalisé une revue de la littérature existante sur les indices de stabilité du manteau neigeux à base mécanique. Nous avons ensuite sélectionné ceux compatibles avec les sorties des modèles numériques de manteau neigeux, en terme de résolution verticale des processus représentés et de coût numérique. Nous avons enfin réalisé une première évaluation de l'intérêt de ces modèles mécaniques de stabilité sélectionnés dans des situations typiques d'avalanche pour mettre en avant les processus que chaque modèle représente ou ne représente pas. Ce travail permet de souligner l'intérêt de combiner différents indices de stabilité pour une estimation globale de la stabilité, étant donné qu'ils sont complémentaires les uns des autres, chacun représentant des processus différents de la formation des avalanches.

Les indices de stabilité ont été sélectionnés pour leur intérêt pour l'analyse de la stabilité de profils générés par la modélisation numérique du manteau neigeux par le modèle Crocus et pour donner une image globale de la stabilité du manteau neigeux. Les rapports contrainte-résistance représentent les processus d'initiation de rupture, qu'il s'agisse d'une initiation uniquement sous le poids propre des couches supérieures du manteau neigeux (activité naturelle) ou nécessite une surcharge (déclenchement provoqué, par exemple par un skieur ou un tir préventif). Ces rapports contrainte-résistance sont complétés par des indices représentatifs de la propagation des fissures dans le manteau neigeux (longueur critique de fissure). L'intérêt de cette sélection est validée a posteriori avec l'approche statistique : il est démontré que l'information contenue dans l'ensemble des indices de stabilité sélectionnés au cours du temps est équivalente à l'ensemble des autres variables utilisées en terme de capacité prédictive pour l'activité avalancheuse observée. Cette première étape permet de réduire l'information fournie par les modèles de manteau neigeux dans l'objectif d'en extraire les informations pertinentes pour l'analyse de la stabilité du manteau neigeux.

Les indices de stabilité nécessitent la connaissance de propriétés mécaniques des différentes couches du manteau neigeux, qui ne sont pas directement modélisées par les modèles numériques actuels de manteau neigeux. Ces propriétés mécaniques sont donc généralement dérivées d'autres propriétés, principalement la densité des couches de neige, parfois complétées par le type de grain. En revanche, les modèles de neige n'ont pas été conçus pour représenter l'ensemble des processus impliqués dans les variations des propriétés mécaniques et la densité est notoirement insuffisante pour représenter la complexité du comportement mécanique du matériau neige [p.ex. Roch, 1966b]. Ces déterminations indirectes des propriétés mécaniques introduisent donc inévitablement des incertitudes importantes quant à la détermination des propriétés mécaniques et en conséquence à l'estimation de la stabilité du manteau neigeux. Des travaux additionnels seront donc nécessaires pour que les modèles numériques de manteau neigeux puissent fournir des comportements mécaniques plus pertinents des différentes couches de neige qu'ils représentent.

Certains processus ne sont pas représentés par les indicateurs de stabilité sélectionnés. C'est le cas par exemple des effets dynamiques durant la propagation des fissures dans le manteau neigeux [p.ex. Gaume et al., 2018; Bergfeld et al., 2021b]. Même si des modèles commencent à émerger pour expliquer les processus à l'œuvre durant cette phase de propagation dynamique [p.ex. Bobillier et al., 2021], ces modèles restent très complexes et peu adaptés à un usage généralisé en temps réel avec des modèles numériques de manteau neigeux. Des indices simplifiés résumant les principaux facteurs d'instabilité devront être développés avant de pouvoir les inclure dans des études comme la nôtre.

Approche statistique de la prédiction de l'activité avalancheuse

Une seconde partie de ce travail utilise les informations météorologiques et nivologiques, avec les indices de stabilité précédemment sélectionnés, pour produire un indicateur relatif à l'activité avalancheuse (chapitre 3). Parmi les techniques d'apprentissage machine existantes, nous avons sélectionné la méthode des forêts aléatoires pour sa capacité à traiter des problèmes non linéaires sans nécessiter de notion de distance et parce que cette méthode a prouvé son efficacité sur des problèmes similaires [p.ex. Sielenou et al., 2021]. La technique des forêts aléatoires est utilisée avec pour entrées les données des modèles

météorologiques et nivologiques, complétées par les indices de stabilité mécanique. Le jeu de données d'observation est composé des événements rapportés par l'Enquête Permanente sur les Avalanches (EPA) [Bourova et al., 2016]. Ce travail a la particularité de combiner en entrée les indices de stabilité à base mécanique avec d'autres données météorologiques et nivologiques, dans des proportions jamais explorées dans la littérature. De plus, nous nous sommes focalisés sur le problème difficile de l'identification de l'activité avalancheuse sur des zones géographiques réduites : nous travaillons a l'échelle de bandes d'altitude et d'orientation (huit orientations et trois altitudes par massif).

Nous avons également développé une méthode d'évaluation robuste pour les méthodes d'apprentissage lorsqu'elles sont appliquées au domaine de la neige et des avalanches. Le manteau neigeux résulte de l'accumulation de neige au fil de l'hiver et accumule donc l'information des conditions météorologiques passées. Il est donc impossible de considérer que deux jours d'un même hiver sont statistiquement indépendants. En conséquence, certains résultats et techniques habituelles d'apprentissage machine ne sont plus utilisables pour de la prédiction sur la neige et les avalanches. Par exemple, il n'est pas possible de tirer au hasard dans un jeu de données complet pour séparer un jeu d'entrainement et d'évaluation : la dépendance entre les différentes conditions nivologiques et météorologiques doit être considérée. Comme nous considérons la neige saisonnière, nous pouvons considérer que le manteau neigeux fond en été et donc que deux jours d'hivers différents sont indépendants. Nous séparons donc le jeu de test et d'évaluation en séparant des années différentes. L'évaluation systématique est ensuite réalisée avec cette approche basée sur la séparation des années. Ce type de méthode pourrait être généralisé pour permettre la comparaison des différentes études utilisant de l'apprentissage machine dans le domaine de la neige.

Sur l'exemple du massif de la Haute-Maurienne, nous avons montré l'intérêt de la méthode pour classer les jours avalancheux et les jours non avalancheux sur la base du jeu de données de l'EPA. Nous avons obtenu une sensibilité de 75.3% et une sélectivité de 76.4%. La précision reste néanmoins faible (autour de 3.3%) en raison du faible nombre d'avalanches observées par rapport au nombre de situations non avalancheuses dans notre jeu de données (voir résultats complets dans le Chapitre 3). Ces scores illustrent la difficulté de prédire les occurrences d'avalanche à haute résolution spatio-temporelle, y compris avec des méthodes et données de pointe. Nous montrons néanmoins que le modèle est d'ores et déjà bien plus performant que le modèle MEPRA actuellement mis à la disposition des prévisionnistes nivologues qui ne parvient pas à apporter une information pertinente à haute résolution spatio-temporelle même s'il reste pertinent à des échelles plus larges (Chapitre 4). Cette étude ouvre ainsi des perspectives pour améliorer les outils d'aide à la décision pour la prévision opérationnelle du risque d'avalanche.

Au-delà de cette comparaison avec les outils actuels, nous avons également vérifié la cohérence statistique du modèle (Chapitre 4). Nous avons montré que les jours les plus critiques sont statistiquement mieux reconnus que les jours plus litigieux avec très peu d'avalanches observées, quand bien même cette information n'a pas été fournie lors de l'entraînement du modèle. De plus, les forêts aléatoires permettent de fournir une probabilité d'observer une activité avalancheuse, pour dépasser la classification binaire. Cette probabilité a été évaluée et nous avons montré qu'elle est cohérente avec les observations. Cette probabilité apparaît donc comme fiable et pourrait être une variable continue utilisée directement comme indicateur de l'activité avalancheuse attendue. Ces vérifications permettent d'éviter le rejet du modèle en raison d'incohérences dans des situations critiques, dans l'optique d'un usage opérationnel.

L'intérêt de l'analyse mécanique avec les indices de stabilité sélectionnés est démontré par l'importance dans le processus d'apprentissage. De plus, les indices de stabilité contiennent presque l'ensemble de l'information pertinente pour la classification, puisque l'usage de ces indices de stabilité seuls permet des scores similaires à l'usage du jeu complet de variables d'entrée. Ces résultats valident le choix des indices de stabilité retenus dans la première partie de ce travail et souligne l'intérêt de la combinaison de la physique et de l'apprentissage.

Généralisation du modèle statistique

Nous avons dans un premier temps étudié intensivement la combinaison de connaissance physique et de méthodes statistiques sur l'exemple du massif de Haute-Maurienne dans les Alpes françaises. Pourtant, un tel modèle a plutôt vocation à être utilisé sur l'ensemble des zones montagneuses. En particulier en France, il s'agit de l'utiliser sur l'ensemble des zones pour lesquelles le bulletin d'estimation des risques d'avalanche (BRA) est produit, c'est-à-dire l'ensemble des massifs alpins, pyrénéens et corses. Le modèle a également été entraîné avec des observations très spécifiques que sont le jeu de données de l'EPA, avec ses biais, et qui n'a pas été conçu pour fournir une image complète de l'activité avalancheuse. Les résultats sur ce cas d'étude peuvent donc être biaisés. Nous avons donc vérifié la pertinence du modèle développé sur différentes zones géographiques et avec plusieurs sources de données (Chapitre 5). Nous avons sélectionné cinq zones géographiques des Alpes et Pyrénées français et utilisé un jeu d'observation alternatif : les observations d'avalanche du réseau nivo-météorologique de Météo-France (jeu de données

NMN). Nous avons montré que les résultats étaient similaires, avec quelques différences pour les massifs avec une information significativement réduite. Ce travail généralise les résultats présentés précédemment, en montrant qu'ils ne sont pas spécifiques au cas d'étude sélectionné, mais sont bien transposables à différentes zones géographiques et jeux de données.

Conclusion générale

En combinant la connaissance de la physique des processus (à l'aide d'indices de stabilité basés sur la mécanique) et des outils statistiques (apprentissage machine), nous avons développé un nouvel outil pour relier un manteau neigeux modélisé par le modèle Crocus et une activité avalancheuse attendue. Cette combinaison de techniques d'apprentissage et de connaissance physique commence à se développer dans divers domaines, incluant certains domaines de géosciences [p.ex. Karpatne et al., 2019; Kashinath et al., 2021]. Par ce travail, nous apportons ces outils dans la communauté neige et avalanches. Nous soulignons l'intérêt de la combinaison des approches physiques et statistiques pour la prédiction de l'activité avalancheuse. L'utilisation d'indices de stabilité permet de réduire l'information produite par les modèles numériques de manteau neigeux, spécifiquement pour l'analyse de la stabilité du manteau neigeux. Ces indices de stabilité permettent en outre d'améliorer la performance des modèles statistiques en limitant le nombre de variables d'entrée.

La méthode présentée ici permet de relier une stratigraphie modélisée et une activité avalancheuse. Ce problème intéresse tout particulièrement la prévision à court terme de l'activité avalancheuse, d'autant plus que nous avons montré la valeur ajoutée par rapport aux outils actuellement utilisés. La méthode est également intéressante par sa capacité à être transposée à des problèmes proches ou adaptée aux évolutions des modèles météorologiques ou nivologiques dès lors qu'il s'agit de relier une stratigraphie modélisée et une activité avalancheuse. La méthode permet également de fournir un indicateur stable de l'activité avalancheuse associée à des conditions de neige et peut donc être utile pour l'étude de l'évolution climatique de l'aléa avalancheux.

L'outil que nous avons développé s'est montré robuste et transposable à d'autres zones géographiques et d'autres jeux de données d'observations avalancheuses. Nous avons également porté une attention particulière à la robustesse des résultats présentés en développant une méthode d'évaluation stricte, prenant en compte les spécificités du manteau neigeux et de son évolution. Les performances du modèle ont été confirmées pour les situations critiques, avec une forte activité avalancheuse. Néanmoins, le problème des avalanches reste un sujet de recherche complexe à la fois en raison de la rareté du phénomène et du fait qu'un manteau instable peut, ou peut ne pas, mener à une avalanche observable, ce qui donne lieu à des taux de fausses alarmes élevés lorsque ces modèles sont évalués sur des observations d'avalanche.

Perspectives

Amélioration des modèles de neige

L'estimation de l'activité avalancheuse est basée sur les stratigraphies du manteau neigeux produites par la modélisation numérique. Cette dernière s'appuie elle-même sur l'information produite par les modèles d'atmosphère. Ainsi, l'incertitude de tous les modèles de la chaîne se combine dans l'analyse finale de la stabilité. Une alternative est d'utiliser des stratigraphies observées du manteau neigeux [p.ex. Mayer et al., 2022]. C'est une approche complémentaire et nécessaire mais qui ne permet pas de bénéficier de la représentation des manteaux neigeux présents et futurs à haute résolution que permet la modélisation numérique. De plus, les observations de terrain sont elles aussi soumises à de nombreuses incertitudes [Proksch et al., 2016]. Ainsi, l'analyse de l'activité avalancheuse attendue à partir des données produites par les modèles de neige reste d'un intérêt majeur.

L'analyse de la stabilité à partir des données des modèles de neige reste limité par l'incertitude de ces modèles. Ainsi, tout effort pour améliorer la représentation du manteau neigeux permettra d'améliorer l'analyse de stabilité en aval. Nous avons montré qu'une limite importante pour l'analyse de stabilité est le manque de représentation des processus et propriétés mécaniques dans la modélisation actuelle du manteau neigeux (Section 2.3). La modélisation de ces propriétés mécaniques passe par une meilleure représentation de la microstructure de la neige [Lehning et al., 2002b] et augmentera la pertinence des indices mécaniques de stabilité. Une meilleure représentation des phénomènes physiques dans le manteau neigeux devrait également permettre de représenter la stratigraphie avec plus de précision et donc également l'analyse de stabilité. C'est en particulier le cas pour la représentation du tassement (d'autant plus que la densité est abondamment utilisée pour déterminer indirectement de nombreux paramètres mécaniques) ou la percolation de l'eau liquide (qui a une grande influence sur la résistance de la neige). L'utilisation d'observations, qu'elles proviennent de mesures de terrain [Viallon-Galinier et al.,

2021] ou de satellites [p.ex. Cluzet et al., 2020], pourra également permettre d'améliorer les résultats de la modélisation numérique.

L'incertitude sur les données d'observation, la méthode statistique et des modèles mécaniques ont été abordés dans cette thèse. Il était néanmoins au-delà du cadre de ce travail de prendre en compte l'incertitude liée aux modèles météorologiques et nivologiques. Nous avons ici considéré que ces modèles produisaient in fine une représentation raisonnable de l'atmosphère et du manteau neigeux. Leurs incertitudes sont cependant combinées dans le résultat final. Afin de donner une information pertinente aux prévisionnistes, nous devons couvrir ces incertitudes. Lafaysse et al. [2017] fournit un ensemble de modèles qui permettent d'estimer l'incertitude due à la représentations des processus dans le manteau neigeux. Vernay et al. [2015] utilise un ensemble de forçages météorologiques en entrée du modèle de neige Crocus pour représenter l'incertitude liée aux modèles atmosphériques. Des études combinent également ces deux méthodes pour donner une estimation plus globale de l'incertitude totale sur les sorties des modèles de neige [p.ex. Cluzet et al., 2021]. L'utilisation d'une combinaison d'ensembles météorologiques et de modèles de neige permettra d'estimer l'incertitude sur l'estimation de stabilité produite. Cette incertitude permettra alors au prévisionniste pour estimer la confiance à accorder à la chaîne de modèles impliquant des modèles météorologiques, nivologiques et une analyse statistique de la stabilité, pour chaque situation.

Dans cette thèse, nous nous basons sur l'usage d'une réanalyse des conditions nivo-météorologiques. Une réanalyse signifie que l'ensemble des observations disponibles ont été prises en compte pour corriger le modèle. Dans la perspective d'un usage en prévision jusqu'à quatre jours, ces observations ne seront pas disponibles. Les sorties des modèles nivo-météorologiques pourraient donc différer de ce qui est observé en réanalyse. Un modèle opérationnel d'apprentissage devra donc être ré-entraîné puis évalué sur ce type de données, en prévision.

Optimisation de l'usage des jeux de données d'observation

Nous avons ici utilisé les données de l'Enquête Permanente sur les Avalanches (EPA) et du réseau nivo-météorologique de Météo-France (NMN) afin de classifier entre les jours avalancheux et non avalancheux. Dans l'EPA, le nombre d'avalanches observées est limité, car seuls un nombre limité de couloirs sont observés. Les données du réseau nivo-météorologique en revanche apportent une information plus détaillée avec un nombre d'avalanches qui peut aller de zéro à plus de trente. Cette information additionnelle peut être utilisée pour dépasser la classification binaire entre jours avalancheux et jours non avalancheux. Il est alors possible de prédire une activité avalancheuse continue, par exemple en terme de nombre d'avalanches attendues, éventuellement ramenée à l'activité usuelle de la zone d'intérêt. Le problème de classification devient alors un problème de régression, pour lequel la méthode des forêts aléatoires peut aussi être adaptée [Breiman, 2001; Hastie et al., 2009].

Nous avons considéré séparément les deux sources de données. Ces deux sources sont complémentaires, car elles n'identifient pas les mêmes situations. Le réseau nivo-météorologique résulte d'une observation de toutes les avalanches visibles, principalement depuis des stations de ski, tandis que l'EPA se concentre sur des avalanches relativement grandes atteignant les vallées. Les avantages de ces deux sources de données pourraient être combinés pour produire un indicateur agrégé de l'activité avalancheuse, qui pourrait ensuite être utilisé pour entraîner les modèles statistiques. De plus, nous avons utilisé, pour chaque zone géographique, uniquement les données de cette zone géographique. Cette approche permet de limiter la quantité de données à traiter mais limite également la performance dans les zones où la quantité d'observations disponible est limitée. Il serait possible de combiner différentes zones dans un apprentissage unique, au prix d'une augmentation du coût numérique, le modèle devant potentiellement résumer différents climats et situations avalancheuses. La combinaison de différentes informations sur les déclenchements observés (différentes zones géographiques ou source d'observations) pourrait améliorer la qualité globale du jeu de données par rapport à chaque jeu de données individuel.

Nouvelles sources de données

Nous avons utilisé l'EPA et le réseau nivo-météorologique qui sont deux sources issues de l'observation manuelle de dépôts d'avalanche. De nouvelles méthodes sont actuellement en développement pour fournir une information complémentaire sur l'activité avalancheuse, ce qui pourrait permettre d'enrichir la base de données des observations. Certaines méthodes utilisent les ondes produites par l'écoulement d'une avalanche, que ce soient des ondes sismiques [van Herwijnen and Schweizer, 2011] ou sonores [Mayer et al., 2020]. Ces systèmes ne couvrent néanmoins que des zones limitées et peinent à identifier les petites avalanches mais fournissent des méthodes automatisées d'observation. Plus récemment, des méthodes pour identifier automatiquement des dépôts avalancheux à partir d'images satellites ont été développées

[p.ex. Leinss et al., 2020; Karas et al., 2022]. Ces méthodes de détection automatique permettent de couvrir de larges zones mais sont encore aujourd'hui en développement, en particulier pour l'identification de petites avalanches, et restent limitées à des périodes relativement récentes. L'utilisation d'archives ou de traces laissées par des avalanches passées tel que l'impact sur les forêts peut également aider à étoffer les bases de données dans le passé [p.ex. Giacona et al., 2018; Tichavský et al., 2022]. Toutes ces méthodes alternatives peuvent aider à améliorer la qualification de l'activité avalancheuse passée et donc améliorer la prédiction une fois combinées aux bases de données existantes. Chaque jeu de données possède néanmoins ses avantages et ses inconvénients. L'EPA est par exemple fait pour identifier les plus grosses avalanches atteignant des enjeux, tandis que le réseau nivo-météorologique n'a pas de zone d'observation clairement définie. Les différentes sources de données doivent donc être combinés intelligemment pour mieux quantifier l'activité avalancheuse à différentes altitudes et orientations avec une datation précise des événements.

Le problème des avalanches provoquées nécessite également un travail approfondi sur le jeu de données à utiliser. Nous avons proposé des résultats préliminaires pour ce problème complexe à l'aide de l'observation du réseau nivo-météorologique. Ces données couvrent principalement les déclenchements préventifs par les stations de sports d'hiver et à proximité et n'est donc pas représentatif de l'ensemble des déclenchements provoqués, ni des avalanches de plaque déclenchés par les skieurs qui sont les plus mortelles dans les Alpes françaises. Pour le problème des avalanches provoquées, l'observation des déclenchements pourrait être insuffisant car l'avalanche ne se déclenche pas nécessairement si l'élément perturbateur n'est pas présent. Des tests de stabilité sur le terrain ou d'autres mesures de la stabilité pourraient être nécessaires pour construire un jeu de données pertinent pour le problème de l'estimation de l'aléa lié aux avalanches provoquées.

Méthodes statistiques alternatives

La méthode des forêts aléatoires est un outil générique pour des classifications ou régressions de problèmes non linéaires, agnostique quand au problème lui-même. Les particularités liées au phénomène à prévoir, ici l'activité avalancheuse, ne sont donc pas prises en compte par l'algorithme. L'activité avalancheuse est le résultat des conditions météorologiques et nivologiques passées. Des outils d'apprentissage existent pour prendre en compte cet effet mémoire et agréger l'information temporelle [Hochreiter and Schmidhuber, 1997], comme par exemple les réseaux de neurones récurrents [Giles and Gori, 1998] qui utilisent à la fois l'information à un temps donné mais également un résumé des informations du passé. Dans notre travail, nous avons utilisé des dérivées temporelles et des sommes des conditions passées comme variables d'entrée pour prendre en compte cette spécificité du manteau neigeux mais l'adaptation des modèles statistiques eux-mêmes permettrait de réduire le nombre de variables d'entrée en prenant directement en compte cette spécificité de notre problème. De plus, et tout particulièrement pour les données de l'EPA, les observations d'avalanche contiennent une incertitude sur la date de l'avalanche. Il serait possible, en adaptant les modèles statistiques, d'optimiser conjointement la prédiction de l'activité avalancheuse et la date des avalanches, pour améliorer le modèle global.

Appendix A

Overview of machine-learning methods for short-term avalanche hazard forecasting

This appendix gives a detailed overview of the literature on statistical tools applied to forecasting avalanche hazard. Several methods have been used, with different evaluation methods, for slightly different goals, etc. All these differences makes it difficult to compare between studies. We here use a table to give an overview of previous studies, by summarizing the main characteristics of the models developed to make comparison easier.

The table summarize studies using statistical (also called machine-learning) methods to estimate the avalanche hazard for nowcasting or short-term forecasting. The precise goal is precised with methods used in the second column. The geographical scale of the study is reported in third column. The input variables of the statistical model are classified according to McClung and Schaerer [1993]: in blue the meteorological variables; in red the bulk variables that require a representation of snowpack evolution but are not specifically designed for stability analysis and the mechanically-based stability indices in green. Some indication on the performances of the model are reported in results column. However, it remains necessary to remember that the results are highly influenced by the test set and the evaluation method chosen. Finally, we report some basic indications on train and test set selection, especially the number of days considered N and the number of avalanche days N_{av} when relevant and reported. Table is ordered by statistical method used.

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Föhn et al. [1977]	Goal: Avalanche probability Method: LDA Separation of wet and dry avalanches	100km² near Davos	14 variables: • Total precipitation and snowfall on 24h • New snow water equivalent (24h) • Maximum precipitation intensity (3d) • Maximum wind speed • Incoming solar radiation • Sunshine hours • Cloudiness • Maximum, minimum air temperature • Total snow depth • Penetration depth of cone penetrometer • Temperature 10cm below snow surface • Snow-drift • Number of avalanches on previous days	Accuracy 70-80%	Train: 12 years, $N \approx 700$, $N_{av} \approx 120$ Test: $N \approx 120$, $N_{av} \approx 25$
McGregor [1989]	Goal: Avalanche probability Method: LDA	500km² in Craigieburn range (New Zealand)	12 variables: • Temperature: mean, maximum, minimum (24h) • Mean humidity (24h) • Precipitation (24h) • Maximum precipitation intensity • Total incoming radiations • New snowfall • Wind direction, speed and speed increase • Snow depth	Accuracy 75%	Train: $N \approx 150$ Test: $N = 32$

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Bois et al. [1974]	Goal: Avalanche probability Method: LDA	100km² near Davos	• Maximum difference of wind speed on several measurements • Temperature at 13:00 • Incoming solar radiations • Number of snow storms of more than 2 days since beginning of the season • Snow depth	Accuracy 74%	Train: 12 years, $N \approx 700$, $N_{av} \approx 120$ Test: $N \approx 120$, $N_{av} \approx 25$
Bois and Obled [1976]	Goal: Avalanche probability Method: LDA, QDA Separation of dry and wet avalanches. Identification of the typical weather in five groups. Separation of winter and spring.	100km² near Davos	**50 variables** derived by accumulation or differencee between days from: • New snow water equivalent • New snow depth • Snowfall intensity • Maximum wind speed • Incoming solar radiation • Temperature at 7:30 and 13:30 • Sunshine hours • Snow depth • Penetration depth of cone penetrometer • Temperature 10cm below snow surface • Number of observation of snow drift events • Past avalanche activity	Accuracy up to 80%	Train: 12 years, $N \approx 700$, $N_{av} \approx 120$ Test: 2 years, $N \approx 120$, $N_{av} \approx 25$
Bovis [1977]	Goal: Tertiary classification (dry, wet avalanches or no-avalanche day) Method: LDA	500km², Colorado, USA	16 variables: • Cumulated precipitation over 1, 2,3,5 days • Min, Mean, Max air temperature over 1, 2, 3, 5 days	Accuracy: 50-80%, with large variability	Train: 1 year Test: 1 year

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Floyer and McClung [2003]	Goal: Binary classification (dry, wet avalanches or no-avalanche day) Method: LDA	Large area with four sub areas, Bear Pass, Canada	• New snow water equivalent • New snow depth • Minimum, present and maximum temperature and difference with previous day • Wind speed • Foot penetration • Snow depth	Score 65-80%	Train: $N \approx 460$
Obled and Good [1980]	Goal: Avalanche probability Method: LDA, 40-NN	100km^2 near Davos	50 variables (meteo or bulk)	For 40-NN: Score ~ 75%, TPR 70%	2 years for test and 11 for training $N \approx 800$, $N_{av} \approx 150$
Buser [1983], Buser [1989]	Goal: Estimation of avalanche occurrence, clustering on weather type Method: LD10, 10-NN	Parsenn ski resort, switzerland	• Max precipitation intensity • Precipitation • New snow depth • Air temperature • wind speed and direction • Cloudiness • Incoming solar radiations • Sunshine hours • Snow depth • Penetration depth • Temperature 10cm below snow surface	No systematic evaluation	

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Navarre et al. [1987]	Goal: Binary classification Method: LDA, 4-NN	La Plagne ski resort, France (about 30 km²)	25 variables: • Temperature (8 and 13h the day before and 8h) • Difference in temperature since previous day • Sum of positive temperature on 3 previous days • Precipitations and cumul on 1, 2,3 days • New snow depth cumul on 1, 2, 3 days • Wind speed of present day and day before • Snowfall of 8 previous days • Forecasted temperature • Square of previously defined temeprature measurements and forecast • Temperature of snow surface • Sum of snow surface temperature • Snow depth • Snow settling since previous day • Penetration of cone penetrometer and difference with previous day	LDA: Balanced accuracy 59-78% Accuracy 59-85%	Train: $N \approx 400$ Test: $N \approx 90$
Gassner and Brabec [2002]	Goal: Binary classification Method: 10-NN	10-100km²	Variables from Obled and Good [1980] and transforms of snow temperature and snow depth near zero, snow drift index cumulated on 3 days with specific coefficients	Accuracy of 52% compared to forecast by operational forecasters on a 10 winters cross-validation (leave one year out). Depends on studied site.	

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Purves et al. [2003]	Goal: Binary classification Method: 10-NN Need of scaling and weighting of variables to reach good scores.		11 variables: • New snow depth (in 4 classes) and sum over winter • Rain at 900m (binary variable) • Air temperature • Wind speed and direction • Cloudiness • Incoming solar radiation (index) • Snow drift (binary variable) • Foot penetration • Snow temperature 10cm below surface	Accuracy 80-83% Balanced accuracy 76-81%	Test: $N \approx 1000$, $N_{av} \approx 260$
Davis et al. [1999]	Goal: Binary classification Method: Classification and regression trees Evaluation of importance of variables (first are product of snowfall or precip with wind speed).	2 areas- in USA	31 variables: • Precipitation (1d, 2d, 3d) • Average wind speed (1d, 2d, 3d) • Product of snowfall and total precipitation with wind speed (power 4) (1d, 2d, 3d) • For 8 orientations, product of snowfall ans wind speed^4 (1d) • Maximum and minimum air temperature • Degree days • Cumulative vapor pressure difference estimation • Snow depth • 1d, 2d, 3d new snow depth • Cumulative settlement • Snow depth of November 15 • Average snow depth increase since November 15	Acuracy 14-95%, depending on size of considered avalanches and studied area	

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Hendrikx et al. [2005]	Goal: Binary classification Method: Classification trees	About 200km²	48 meteorological variables: • Temperature (average, max, min) • Wind speed and direction • Precipitation (maximum and sum) • Average air pressure • Wind drift • Snow depth on period of 12, 24 and 72h.	Accuracy 78-85% Recall 86-79%	
Kronholm et al. [2006]	Goal: Binary classification Method: Classification trees	About 100km²	12 variables: • Precipitations (1, 3, 5d) • Average temperature (1d) • Degree day (5d, season) • Number of crossing of 0°C (season) • Number of days below 5°C with no precipitations (season) • Number of rain on snow events (season) • Wind speed mean and max (1, 3, 5d)	Accuracy 85%	
Hendrikx et al. [2014]	Goal: Binary classification Method: Classification trees	About 1000km²	30 variables: • Minimal and maximal temperature • Sum of new snow depth • Average and sum of preciptiation • Wind speed • Drift index (product of wind speed and precipitation) • Drift index (product of cube of wind speed and precipitation) • Average snow depth • Density all for 1, 2, 3days and: • Sum of rain on 72h • Wind direction on 24h	TPR 94% Accuracy 69% Precision 33% TPR 77% Accuracy 72% Precision 24%	$N_{av} \approx 530$ $N \approx 927$

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Schweizer et al. [2009]	Goal: Binary classification Method: Classification trees	About 1km²	• Air temperature • New snow depth (1, 2, 5, 10days) • Snow depth and difference in 3 past days	TPR 65% Precision 90%	
Mitterer and Schweizer [2013]	Goal: Binary classification Method: Classification trees, Random Forests Focus on wet-snow instability	Some squared kilometers	Several subsets of these variables: • Mean, max mean air temperature • Difference in air temperature on 3 days • Sum of positive air temperature on 3 and 5 days • Snow surface temperature (mean, max, mean) • Humidity • Min and max radiation balance • Min, max, mean energy balance • Sum of positive radiation balance on 5 days • Mean, min, max incoming, outgoing or net SW • Mean, min, max incoming, outgoing or net LW • Maximum amount of LWC within the snowpack	Classif tree: TPR 89%, TNR 80% RF: TPR 70%, FPR 34%	3 years, $N = 729$, $N_{av} = 66$
Möhle et al. [2014]	Goal: Binary classification Method: Random Forests	Some squared kilometers	14 variables: • Min, max air temperature last day and the day before • Current wind speed • Wind speed and direction on the previous day • Cloudiness • Precipitation and new snow on last 24h and day before • Snow depth	TPR 41-50% TNR 95-96% Precision 7-8%	Train: $N \approx 600$, $N_{av} \approx 40$ Test: $N_{av} \approx 10$, $N \approx 1500$

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Marienthal et al. [2015]	Goal: Binary classification Method: Classification trees, Random Forests Focus on deep slab avalanches on persistent WL. Daily and seasonal prevision.	50km^2 Bridger Bowl, USA	• Snowfall 24h • Max, min air temperature and average min and max on 2, 3, 4, 7 days • Sum of temperature above zero for max and min air temperature on 1, 2, 3, 5 days • Difference in minimum and maximum air temperature with previous day • Change in snow depth	Classif tree: FPR 50% TPR 82% Accuracy 67% RF: FPR 60% TPR 58% Precision 49% Balanced accuracy 59%	$N_{av} \approx 76$
Dreier et al. [2016]	Goal: Binary classification Method: Classification trees, Random Forests, Logistic regression Focus on glide snow avalanches. Separation between cold and warm events.	About 1km^2	• Mean, Max, Min air temperature • Difference in mean and max air temp on 24h • Difference between mean, max or min temperatures of air • Max and mean air humidity • Incoming and outgoing LW and SW radiations • Net radiations • New snow depth (1d and sum on 5d) • Snow depth and difference in last 1, 2, 3, 4days • Mean, min, max snow surface temperature • Difference between mean, max or min temperatures of snow surface	RF: AUC 0.75-0.83 TPR 65-69% TNR 72-80% Precision 51-57% Logistic regression: AUC 0.85 TPR 69-71% TNR 85-94% Precision 25-42%	10 fold cross validation $N \approx 216$ $N_{av} \approx 49$
Mayer et al. [2022]	Goal: Binary or tertiary classification Method: Random Forests Identification of the weak layer.	Local	• Snowpack output parameters of identified weak layer • Snowpack output parameters characterizing the slab • Differences in snowpack output parameters between slab and identified weak-layer • Stability indices (SK38, two versions and Sr) • Critical cut length • Stresses and strain rate	Accuracy 88-93% Precision 96% Recall 85-92%	Train: $N \approx 500$, balanced Test: $N \approx 121$ $N_{av} \approx 75$, 5-fold cross validation

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Pozdnoukhov et al. [2008]	Goal: Binary classification Method: SVM, k-NN (k=1, 10, 20, 32) Identification of 20 variables of interest out of 44. show that snowpack information is retained.	Unknow area, UK	44 variables from which a subset of 20 is identified as sufficient: • Values for the current day, previous and J-2: • New snow (on 4-scale index) and cumulative over season • Rain at 900m (binary) • Snowdrift • Cloudiness • Air temperature • Foot penetration • Snow temperature 10cm below snow surface • Wind speed and direction • South or Southeasterly wind • Change in air temperature • Avalanche activity the previous day	SVM: Accuracy 85% TPR 70% Precision 70% 20NN (better results): TPR 69% Accuracy 84% Precision 67%	Train: $N \approx 1100$, 10 years Test: $N \approx 712$ $N_{av} \approx 180$, 7 years
Singh and Ganju [2008] Binary classification	Goal: Binary classification Method: k-NN post-processed by convolutional neural networks		• Air temperature • Air temperature difference with previous day • New snow depth(24, 48, 72h) • Wind speed • Snow depth • Snow water equivalent • Snow surface temperature • RAM penetration	TPR 87-96% Precision 35-60%	3 years for learning, one for evaluation Test: $N_{av} \approx 23$ $N \approx 120$

Reference	Goal, method(s) and remarks	Scale and location	Variables	Results	Train and test sets
Dekanova et al. [2018]	Goal: Binary classification Method: Convolutional neural networks		• Maximum air temperature and temperature at 8:00 • Difference of air temperature with previous day • Wind gust speed and direction • Relative humidity • Incoming solar radiation (24h) • Precipitation (boolean) • Snow depth • New snow depth (3d) • Temperature gradient on 20^{th} first centimeters • Temperature 20cm below surface all with pre-processing for normalization	No systematic evaluation	Test on one independent year
Sielenou et al. [2021]	Goal: Tertiary classification Method: several methods Clustering of geographical areas depending of the parameters of interest for avalanche prediction highlights several climates.	Some hundreds of squared kilometers regions of the French Alps	• Minimum, maximum and mean air temperature (on one and 3 days) • Wind speed and direction • Rain and snow precipitations • Snow depth • Thickness of dry and wet snow	Accuracy 60-99%	

Appendix B

Description of the MEPRA model

Contents

This appendix aims at describing how mechanical diagnosis and stability indices are computed in Crocus snow cover model [Brun et al., 1989; Vionnet et al., 2012]. The computation of three stability indices is generally known as the MEPRA model (Modèle Expert pour la Prévision du Risque d'Avalanche).

The snow cover model Crocus computes at each time step and for each layer the following physical quantities:

- mass (kg)
- density (kg m$^-$3)
- enthalpy (from which it is possible to determine temperature and liquid water content, J)
- age of snow (days)
- complementary variables for snow microstructure that includes:
  - snow specific surface (m^2 kg$^-$1)
  - sphericity
  - dendricity
  - grain size

The MEPRA tool uses Crocus snow cover model output to compte some mechanical parameters such as penetration resistance R_p, shear strength R_c, strength-stress ratios with (S_{acc}) and without (S_{nat}) skier, and finally three hazard indicators, at the model scale for natural avalanches (R_{nat}) and triggered avalanches (R_{acc}) and a summary at the massif scale G.

B.1 Mechanical diagnostics

B.1.1 Penetration resistance

A penetration resistance is computed for each layer. It has been designed to represent the measure obtained by the rammsonde penetration resistance, commonly used in field snowpack observations [Pahaut and Giraud, 1995]. The value of penetration resistance R_p is inferred for each layer from grain characteristics (size g_s in meter, sphericity s and dendricity d), density ρ in kg m^{-3}, liquid water content LWC in kg m^{-3} and temperature T in ^{o}C. The result is expressed in kgf (1 kgf $\simeq$ 9.81 N) and given by following equation:

$$\begin{cases} \text{dendritic case: } R_p = d \cdot \max(1, 0.018\rho - 1.363) + (1-d) \cdot \max\left[2, s \cdot (0.17\rho - 31) + (1-s) \cdot (0.085\rho - 14.9)\right] \\ \text{non-dendritic case: depends on grain type} \begin{cases} \text{RG} \begin{cases} \rho < 200, \ R_p = 3 \\ \rho \geq 200, \ R_p = 0.17\rho - 31 \end{cases} \\ \text{RG+FC} \begin{cases} \rho < 200, \ R_p = 2 \\ \rho \geq 200, \ R_p = s \cdot (0.17\rho - 31) + (1-s) \cdot (0.17\rho - 31) \cdot (0.8 - g_s) + 2 \cdot g_s \end{cases} \\ \text{FC or FC+DH} \begin{cases} g_s > 0.0008, \ R_p = 2 \\ \text{else} \begin{cases} \rho < 200, \ R_p = 3 \cdot (0.8 - g_s) + 2 \cdot g_s \\ \rho \geq 200, \ R_p = 0.17\rho - 31 \end{cases} \end{cases} \\ \begin{bmatrix} \text{MF} \\ \text{MF+RG} \\ \text{MF+FC} \\ \text{MF+DH} \end{bmatrix} \begin{cases} T < -0.2 \text{ or } LWC < 5, \ R_p = \max(10, 0.103\rho - 19.666) \\ \text{else} \begin{cases} \rho < 250, \ R_p = 1 \\ 250 \leq \rho < 350, \ R_p = 2 \\ \rho \geq 350, \ R_p = 0.16\rho - 54 \end{cases} \end{cases} \\ \text{DH}, \ R_p = 2 \end{cases} \end{cases} \tag{B.1}$$

B.1.2 Shear strength

To be able to compute classical stability indexes of the snowpack, the Crocus model proposes a shear strength computed for each layer. This strength is computed from grain characteristics (g_s, s, d), liquid water content LWC in kg m^{-3} and temperature T in ^{o}C. For concision purpose, an intermediate variable is defined, which represent the ratio between liquid water content and a typical maximal value for a layer of same density $F = \frac{LWC \cdot 10^{-3}}{0.05 \cdot (1 + (LWC - \rho)/\rho_i)}$. The shear resistance R_s is expressed in kgf dm^{-2} (1 kgf dm^{-2} $\simeq$ 0.981 kPa) and given by following equation:

$$\begin{cases} \begin{bmatrix} \text{If } T \geq -0.2 \\ \text{and } LWC \geq 5 \\ \text{and grain type in } \begin{Bmatrix} \text{RG, MF, RG+FC, RG+MF,} \\ \text{DF+RG, MF+FC, MF+DH} \end{Bmatrix} \end{bmatrix} \begin{cases} \rho < 200, \ R_s = 0.1 \\ 200 \leq \rho < 320, \ R_s = 0.02\rho - 3.9 \\ \rho \geq 320, \ R_s = 0.068\rho - 18.64 \end{cases} \\ \text{else, } R_s = \max\left(0.05, C_s \cdot C_d \cdot C_{g_s} \cdot C_{LWC} \cdot C_h \cdot (\rho^2 \cdot 10^{-4} - 0.6) + 0.12\right) \end{cases} \tag{B.2}$$

where $\rho_i = 917\,\mathrm{kg\,m^{-3}}$ is the ice volumetric mass and coefficients C_s, C_d, C_{g_s}, C_{LWC} and C_h being defined as follows:

$$\begin{cases} s > 0.8 \text{ and } h \in 3,5, C_s = 1.05 \\ \text{else, } C_s = \begin{cases} 0.45 + 0.7s & s < 0.25 \\ 0.625 + 1.0 \cdot (s - 0.25) & 0.25 \leq s < 0.5 \\ 0.875 + 0.6 \cdot (s - 0.5) & 0.5 \leq s < 0.75 \\ 1.025 + 0.5 \cdot (s - 0.75) & 0.75 \leq s \end{cases} \end{cases} \tag{B.3}$$

$$C_d = \begin{cases} 1 - 0.4d & d < 0.25 \\ 0.9 - 0.4 \cdot (d - 0.25) & 0.25 \leq d < 0.5 \\ 0.8 - 0.8 \cdot (d - 0.5) & 0.5 \leq d < 0.75 \\ 0.6 - 0.6 \cdot (d - 0.75) & 0.75 \leq d \end{cases} \tag{B.4}$$

$$\begin{cases} \text{dendritic case, } C_{g_s} = 1 \\ \text{non-dendritic case} \begin{cases} g_s \leq 4 \cdot 10^{-4} - 1 \cdot 10^{-4} s, \ C_s = 1 \\ \text{else, } C_{g_s} = 1 - 530 \cdot (0.8 - 0.2s) \cdot (-4 \cdot 10^{-4} + g_s + 1 \cdot 10^{-4} s) \end{cases} \end{cases} \tag{B.5}$$

$$\begin{cases} F < 0.9, C_{LWC} = \begin{cases} 1 + F & F < 0.1 \\ 1.1 - 2.35(F - 0.1) & 0.1 \leq F < 0.3 \\ 0.63 - 0.4(F - 0.3) & 0.3 \leq F < 0.9 \end{cases} \\ \text{else, } C_{LWC} = \max\left(0.15, \min\left[0.35, (\rho - LWC) \cdot 10^{-4}\right]\right) \end{cases} \tag{B.6}$$

$$\begin{cases} h \in 0,1, \ C_h = 1 \\ F > 0.5, \ C_h = 1 \\ \text{else} \begin{cases} F = 0, \ C_h = 1.5 \cdot \left(\frac{1.15 + 0.2(1-s)}{1.15}\right) \cdot \left(1 + \frac{0.2}{C_{g_s}}\right) \\ \text{else} \begin{cases} h \in 2,3, C_h = 1 \\ \text{else, } C_h = \begin{cases} 1.5 - 2F & F < 0.1 \\ 1.3 - 0.75(F - 0.1) & 0.1 \leq F \end{cases} \end{cases} \end{cases} \end{cases} \tag{B.7}$$

B.2 MEPRA

MEPRA was firstly an optional add-on to the Crocus scheme run after the model core, designed to estimate the avalanche hazard from snowpack stratigraphy represented by Crocus, with simple mechanical diagnosis and expert rules. It has been fully implemented in the Crocus snowpack model, and its output are available with other diagnostic variables. The general idea is to use shear strength computed before, and compare to shear stress in the layer. For natural release, only weight of overlying layers are taking into account while an additional load is added to represent the accidental triggering. Expert rules are then defined to determine a hazard index from these simple mechanical stability indicators, both for natural release and accidental triggering. Expert rules were defined with the work of Giraud et al. [2002] but remained largely unpublished and evolved through versions of Crocus. Below are presented equations implemented in the current version of Crocus. Implying expert knowledge, translated into computational rules, all these equations are disputable. They are presented here only to provide an overview of Crocus management of avalanche hazard, so they are not discussed.

B.2.1 Mechanical punctual snowpack stability

At each simulation point, simple mechanical diagnosis for stability are computed by dividing the shear resistance R_s by the shear in the layer. Two values are computed, to discriminate between natural avalanche activity and human triggering. Natural avalanche activity is assumed to be due to the weight of overlying layers only while accidental triggering is due to an additional loading.

Natural release

The stability index S_{nat} for natural release is defined in each layer as follows:

$$S_{nat} = \frac{R_s}{\tau_w} \tag{B.8}$$

where τ_w is the weight of overlying layers (including considered layers), projected on the slope-parallel axis.

Accidental triggering

The stability index for accidental triggering S_{acc} is similar to equation B.8, with a supplementary term for shear stress to represent the additional load on the snowpack:

$$S_{acc} = \frac{R_s}{\tau_w + \beta\tau_a} \tag{B.9}$$

where τ_a is designed to represent the shear load induced by a skier, so it is a decreasing function of depth z, defined as a continuous piecewise linear function in following equation:

$$\tau_a(z) = \begin{cases} 4 - 15z & z < 0.1 \\ 2.5 - 10(z - 0.1) & 0.1 \leq z < 0.15 \\ 2 - 8(z - 0.15) & 0.15 \leq z < 0.2 \\ 1.6 - 4(z - 0.2) & 0.2 \leq z < 0.35 \\ 1 - 2(z - 0.35) & 0.35 \leq z < 0.5 \\ 0.7 - 1.5(z - 0.5) & 0.5 \leq z < 0.8 \\ 0 & 0.8 \leq z \end{cases} \tag{B.10}$$

and in the considered layer, β is designed to represent the bonding effect in the layers reducing the shear stress in the layers. They are defined as the mean of β_i of each overlying layer (excluding considered one, weighted by height). β_i of each layer being defined depending on grain type, temperature T and shear resistance R_s with following equation:

$$\begin{cases} \text{Grain type in } \begin{Bmatrix} \text{MF, RG+MF,} \\ \text{MF+FC, MF+DF} \end{Bmatrix} \begin{cases} T < -0.2,\ \beta_i = 0.5 \\ \text{else},\ \beta_i = 1.1 \end{cases} \\ \text{else} \begin{cases} R_s > 1.5,\ \beta_i = 1.0 \\ \text{else},\ \beta_i = 1.2 \end{cases} \end{cases} \tag{B.11}$$

B.2.2 Hazard estimation index

MEPRA analyses the previously presented mechanical indexes with other parameters (mainly grain type, temperature and liquid water content and snow heights), to determine a natural avalanche hazard R_{nat}, on a scale of 0-5 and accidental hazard R_{acc} on a scale 0-3, with a set of expert rules presented below. These hazard levels are associated with levels of instability (one of high instability, D_h, and one of moderate instability, D_m, at most). For natural release, an avalanche type (among 5) is provided when possible to identify.

Some other variables are also available, and used for analysis of the overall hazard. They are the rammsonde penetration $P_{SNOWRAM}$ at the top of the snowpack, P_{HWET} the height of wet snow and P_{HREF} the height of refrozen snow, always from the top of the snowpack. They are defined as:

$$\begin{aligned} P_{SNOWRAM} &= \text{height of continuous layers from top of snowpack satisfying } R_p \leq 2 \\ P_{HWET} &= \text{height of continuous layers from top of snowpack satisfying } LWC > 0 \\ P_{HREF} &= \text{height of continuous layers from top of snowpack satisfying } \begin{cases} h > 2 \\ \text{and } T < 0 \end{cases} \end{aligned} \tag{B.12}$$

For concision purpose, it is also defined two internal variables for each layer, accounting for total crusts thicknesses upper than considered layers: Z_c and Z_{c2}. For the first one are considered as crusts layer of MF, MF+DH, MF+FC or RG+MF which temperature is below 0.2 ^{o}C. For the second one are considered as crusts layers with same grain type but temperature below 0.2 ^{o}C or higher temperature but $LWC < 5\,\mathrm{kg\,m^{-3}}$.

Natural release

The natural release analysis relies on the analysis of the identification of a typical snowpack state at its top, called superior profile, and classified in three classes: NEW (new snow), WET (wet snowpack), FRO (refrozen snowpack) or NAN (when it could not be classified in other class). To this class is associated the height of this profile H_s (in meter, below which the snowpack is significantly different, and called inferior profile). If there is grains of type PP or DF (or mixed types with PP or DF, see Fierz et al. [2009] for grain shape descriptions) present in the snowpack, the superior profile type is always NEW. The upper profile goes down to the lower layer containing PP or DF, to which are added adjacent layers with $T \geq -0.2$ ^{o}C and $LWC \leq 5$ kg m^{-3} while $Z_{c2} < 1$ cm. If there is a layer containing grains of type MF (or mixed types with MF) in first 3 cm, the superior profile is FRO if the fist MF layer have $T < -0.2^o$C or $TEL < 5\,\mathrm{kg\,m^{-3}}$, and WET otherwise. The superior profile gather in this case all layers containing MF, mutually adjacent from the upper one. In other cases, the upper profile is identified as NAN.

The snowpack below the limit of the superior profile is called inferior profile. This inferior profile is also classified into three classes denoted SOF, HAR and NAN. If layers of the inferior profile below $0.25 \cdot H_s$ from surface have all $R_p < 8$ then it is classified as SOF, otherwise HAR, except if there is no layer under the depth threshold, so that the profile could not be determined and assigned to NAN.

With these elements it is possible to define level of instabilities (high and moderate). In this step are only considered layers of superior profile buried at least at 0.1 m. The first layer with $S_{nat} \leq 2$ is identified as high instability level and first layer (excluding the layer identified with high instability level if exists) with $S_{nat} \leq 3.05$ (in case of NEW superior profile) or $S_{nat} \leq 3.00$ (in case of WET or FRO superior profile) is identified as moderate instability level. Middle depth of these layers are stored in D_h and D_m respectively.

A first natural avalanche hazard indicator R_{nat} is defined depending on depth of instability level D_h and D_m, superior profile height H_s and superior profile type with following expert rules:

$$R_{nat_1} = \begin{cases} D_h > 0.8 \begin{cases} NEW \rightarrow 5 \\ WET \rightarrow 5 \\ FRO \rightarrow 3 \end{cases} \\ D_h > 0.4 \begin{cases} NEW \rightarrow 4 \\ WET \rightarrow 4 \\ FRO \rightarrow 3 \end{cases} \\ D_h > 0.2 \begin{cases} NEW \rightarrow 2 \\ WET \rightarrow 2 \\ FRO \rightarrow 3 \end{cases} \\ D_h > 0 \begin{cases} NEW \rightarrow 1 \\ WET \rightarrow 1 \\ FRO \rightarrow 1 \end{cases} \\ D_h \text{ not defined} \rightarrow 0 \end{cases} \quad R_{nat_2} = \begin{cases} D_m > 0.8 \begin{cases} NEW \rightarrow 3 \\ WET \rightarrow 3 \\ FRO \rightarrow 1 \end{cases} \\ D_m > 0.4 \begin{cases} NEW \rightarrow 3 \\ WET \rightarrow 3 \\ FRO \rightarrow 1 \end{cases} \\ D_m > 0.2 \begin{cases} NEW \rightarrow 3 \\ WET \rightarrow 3 \\ FRO \rightarrow 1 \end{cases} \\ D_m > 0 \begin{cases} NEW \rightarrow 1 \\ WET \rightarrow 1 \\ FRO \rightarrow 1 \end{cases} \\ D_m \text{ not defined} \rightarrow 0 \end{cases} \quad R_{nat_3} = \begin{cases} H_s > 0.8 \begin{cases} NEW \rightarrow 1 \\ WET \rightarrow 1 \\ FRO \rightarrow 0 \end{cases} \\ H_s > 0.4 \begin{cases} NEW \rightarrow 1 \\ WET \rightarrow 1 \\ FRO \rightarrow 0 \end{cases} \\ H_s > 0.2 \begin{cases} NEW \rightarrow 1 \\ WET \rightarrow 1 \\ FRO \rightarrow 0 \end{cases} \\ H_s > 0 \begin{cases} NEW \rightarrow 0 \\ WET \rightarrow 1 \\ FRO \rightarrow 0 \end{cases} \\ H_s \text{ not defined} \rightarrow 0 \end{cases}$$

$$R_{nat} = \begin{cases} \text{if } R_{nat_1} = 2 \rightarrow 2 \\ \text{else} \rightarrow \max(R_{nat_1}, R_{nat_2}, R_{nat_3}) \end{cases} \tag{B.13}$$

Note that in previous equation, the hazard associated to NAN superior profile is assimilated to NEW.

Knowing this first natural avalanche hazard, it is possible to compute the typical avalanche situation, classified in 6 classes: NEW_DRY (new snow, dry), NEW_WET (new snow, wet), NEW_MIX (new snow, mixed type), MEL_SUR (melting at surface), MEL_GRO (melting, not mainly at surface) and AVA_NAN if could not identify to an other type. The expert rules determining avalanche type from superior profile, inferior profile, temperature and liquid water content of layers of superior profile are described in following equation:

$$\begin{cases} NEW \begin{cases} \text{superior profile fully dry} \rightarrow NEW\_DRY \\ \text{superior profile fully humid} \rightarrow NEW\_WET \\ \text{else} \rightarrow NEW\_MIX \end{cases} \\ WET \begin{cases} HAR \rightarrow MEL\_SUR \\ SOF \rightarrow MEL\_GRO \\ \text{else} \begin{cases} \text{no dry layer below 10 cm in upper profile} \rightarrow MEL\_GRO \\ \text{else} \rightarrow MEL\_SUR \end{cases} \end{cases} \\ FRO \begin{cases} R_{nat} = 0 \rightarrow AVA\_NAN \\ \text{else} \begin{cases} HAR \rightarrow MEL\_SUR \\ SOF \rightarrow MEL\_GRO \\ \text{else} \begin{cases} \left[\begin{array}{l} \text{if for last layer of superior profile} \\ T \geq -0.2 \text{ and } LWC \geq 5 \\ \text{and height of layers of superior profile which top is below 0.1 m is superior to } 1/3 \cdot H_s \end{array} \right] \rightarrow MEL\_GRO \\ \text{else} \rightarrow MEL\_SUR \end{cases} \end{cases} \\ NAN \rightarrow AVA\_NAN \end{cases} \tag{B.14}$$

where dry means $T < -0.2$ or $LWC < 5$ and humid is its contrary.

Once known the avalanche type, the natural avalanche hazard is updated with following expert rules:

- If snow height is very low (under $1 \cdot 10^{-7}$m), the natural avalanche hazard is undefined.
- If superior profile is NEW and avalanche type remains unchanged from previous output time step (POTS) and total snow height is strictly inferior to POTS one and H_s is lower or equal than at POTS and avalanche type is not NEW_MIX, then avalanche hazard is updated with a function of current R_{nat} and value at POTS according to Table B.1. The same rule is applied if avalanche type is NEW_MIX but only if the height of the upper continuous ensemble of wet snow layers decreases or remains constant since POTS (wet being defined as in equation B.14).
- If superior profile is WET or FRO and superior profile at POTS was WET or FRO and inferior profile reduced at most of 5 cm from POTS:

  - If $R_{nat} = 3$ then it is reduced to 2
  - If superior profile is WET and $R_{nat} = 4$ or 5 then if previous was 3, 4 or 5, it is reduced to 3 and if previous was 1 it is reduced to 1.
- If superior profile is NAN, R_{nat} is reduced to 0.

Table B.1: Table of actualization of R_{nat} depending on currently determined R_{nat} and its value at previous output time step R_{nat_a}. - represents an undefined value.

R_{nat_a}	R_{nat}						
	0	1	2	3	4	5	-
0	0	1	2	3	4	5	-
1	0	1	1	1	4	5	-
2	0	1	1	1	4	5	-
3	0	1	1	1	3	4	-
4	0	1	2	3	3	4	-
5	0	1	2	3	3	4	-
-	0	1	2	3	4	5	-

Accidental triggering

Accidental triggering hazard is based on the identification of a slab structure in the snowpack. Then, rules are designed to identify a slab and a weak layer. The slab is identified as the first layer (from the top) verifying simultaneously the three following conditions:

- The immediately previous layer does not contain MF (pure MF or mixed type with MF)
- $R_s > 1.3$
- The grain type is DF+RG, RG, RG+FC or DF

A weak layer is the first layer below the slab which fulfill simultaneously the following conditions:

- The top of the layer is below 0.01 m (included) and above 1 m (excluded)
- Grain type is FC, DH, FC+DH, PP, PP+DF or DF.

With identified weak layers, it is possible to identify, if present, a level of high instability and a level of moderate instability. The first one is the upper weak layer with grain type FC, DH ou FC+DH and with $S_{acc} < 1.5$. The second one is the upper weak layer with grain type PP, PP+DF or DF. If there is no, it is the upper weak layer with grain type FC, DH or FC+DH and $1.5 \leq S_{acc} < 2.5$. Middle depth of these layers are recorded under variables D_h and D_m respectively.

With all these elements, the accidental hazard is defined with following expert rules:

- If there is a high instability level, with $Z_c < 1$ cm for corresponding weak layer, the accidental avalanche hazard is 3
- Else, if there is a moderate instability level, with $Z_c < 1$ cm for corresponding weak layer, the accidental avalanche hazard is 2
- Else, if there is an instability level, the accidental avalanche hazard is 1

If no slab structure is detected, the accidental avalanche hazard is 0. This could be unsatisfactory, as other mechanisms could be involved in avalanche triggering. To try to solve this problem, this hazard is compared to an equivalent of natural hazard (cast in the same range) $R_{nat_{eq}}$ defined as follows:

$$\begin{cases} \text{if total snow height} \geq 0.2\text{m} \begin{cases} S_{nat} \geq 4,\ R_{nat_{eq}} = 3 \\ 2 \leq S_{nat} < 4,\ R_{nat_{eq}} = 2 \\ \text{else},\ R_{nat_{eq}} = 1 \end{cases} \\ \text{else}, R_{nat_{eq}} = 1 \end{cases} \tag{B.15}$$

If equivalent natural hazard is superior to accidental one or if the slope is insufficient (cosine above 0.8, that is to say slope below approximately 37 degrees), the equivalent natural hazard is retained as R_{acc}, and instability level are updated with natural ones. Otherwise, accidental instability level are the only provided in the Crocus output.

A last actualization is done depending on slope, with following rules:

- If slope is sufficiently low (cosine above 0.8, that is to say slope below approximately 37 degrees), and $R_{acc} = 1$, R_{acc} is reduced to 0
- If slope is nearly flat (cosine above 0.99, that is to say slope below approximately 8 degrees), R_{acc} is undefined.

B.2.3 Aggregation at massif scale

MEPRA was initially developed as a tool for avalanche forecasters. As they have to produce bulletins at massif scale, a synthetic index was designed to summarize avalanche hazard at this scale. The chosen index is the maximum over all aspects of the mean of R_e at altitudes 1500, 1800, 2100, 2400, 2700 and 3000 m [Martin et al., 2001]. R_e is linked to the natural avalanche hazard at slope of 40 degrees, by Table B.2. This index ranges from 0 (very low hazard) to 8 (very high).

Table B.2: Table of R_e used in aggregated avalanche hazard at massif scale from R_{nat}

R_{nat}	0	1	2	3	4	5	-
R_e	0	0	1	2	4	8	0

Appendix C

Humid snow indices

Contents

C.1 Introduction

The processes in dry snow are much more studied in the literature than for wet snow, and especially when it comes to stability and avalanche release processes. The instability in wet snow mainly comes from the loss of cohesion between grains when interstitial liquid water becomes sufficient to switch from pendular to funicular regime [Denoth, 1980]. There is currently, to our knowledge, no mechanically-based stability indicator designed specifically for wet snow. However, there is indirect indicators of stability that are based on liquid water content (LWC). In this appendix, we give an overview of existing indicators of wet snow instabilities and how to use in combination with the snow cover model Crocus.

C.2 State of the art

Mitterer et al. [2013] first define an index for regional wet snow avalanche activity based on LWC, called the LWC index ($\mathrm{LWC}_{\mathrm{index}}$) and defined as the ratio between volumetric LWC (θ_v) and a threshold of 3%: $\mathrm{LWC}_{\mathrm{index}} = \frac{\overline{\theta_v}}{0.03}$. An index above one is the indication of generalized unstable conditions. The threshold of 3% is shown to be coherent for the transition between pendular and funicular regime, however the model is validated on the Snowpack model [Bartelt and Lehning, 2002] and avalanche observations rather than on the processes involved. The index can therefore be described as an optimized threshold on Snowpack LWC representation. The same index is used by Bellaire et al. [2017] to identify the days with wet snow avalanches. This choice is justified as it represents approximately the residual amount of liquid water before draining in Snowpack model.

The first index for regional wet snow avalanche activity identification was then refined for more local avalanche activity estimation. A threshold of 0.33 (to compare to the threshold of 1 for the regional avalanche activity) is identified on LWC index as a relevant threshold for identifying wet-snow avalanche activity in a single avalanche path, that is to say a LWC over about 1% in volume [Mitterer et al., 2016]. This threshold has been shown to be relevant for one instrumented avalanche path near Davos, in Switzerland, with LWC simulated by Snowpack model.

The choice of LWC as variable to identify wet snow avalanche activity is justified by Wever et al. [2016a] that identifies that LWC is more relevant than air temperature or energy balances as it combines the whole processes explaining liquid water apparition or refreezing in the snowpack. With the representation of water retention above crusts in the Snowpack model, a threshold about 5 to 6% is shown to be relevant for identifying wet avalanche days.

Previous indices were developed to identify avalanche situations. An alternative goal is to identify for a given situation whether the snowpack is rather wet or dry to identify the mechanism potentially involved in a avalanche release. This issue was addressed by Naaim et al. [2016] and in unpublished work of Mohamed Naaim. The characterization is also based on LWC and defined as the part of the snowpack that is humid and denoted I_h. A layer is considered humid if the liquid water content of the selected layer exceed a threshold. The index is used to classify between dry and wet avalanches in the recorded French avalanche activity.

In conclusion, there is two main goals which lead to significantly different indicators. The first one is to select days with high avalanche activity due to wetting of snow. The second one is to select days for which the snowpack is considered wet from days the snowpack is dry. For the first goal, thresholds will generally be higher than the second as days with high wet avalanche activity may be when the snowpack is wet but a wet snowpack is stable most of the days (because the amount of water is not sufficient for instance). However, depending on the spatial scale considered and size of avalanches considered, threshold may vary in a ratio of 1 to 6. The second goal also suffer from an existential question from its definition: how do we want to classify snowpack that are half-wet. Especially when both the wet and dry parts could be prone to avalanche independently ? The typical example is a snowfall in spring: the snowpack is fully wet before snowfall and the amount of new snow can be the trigger for gliding (glide avalanche or wet snow avalanche) of snow while the amount of new snow could be sufficient for triggering an avalanche in the new layer (new snow problem). Here we have to choose if we want to be safe for days with wet snow or if we want to eliminate all avalanche problems that are not related to wet snow.

C.3 Adaptation to Crocus snow cover model

Except the work of Naaim et al. [2016], all the other studies were designed to be used with LWC computed by the Snowpack snow cover model. Even though initial ideas were similar, the current representation of the liquid water percolation are quite different between Snowpack and Crocus. The indices calibrated for Snowpack then have to be adapted to be used with Crocus model. Both Crocus and Snowpack model

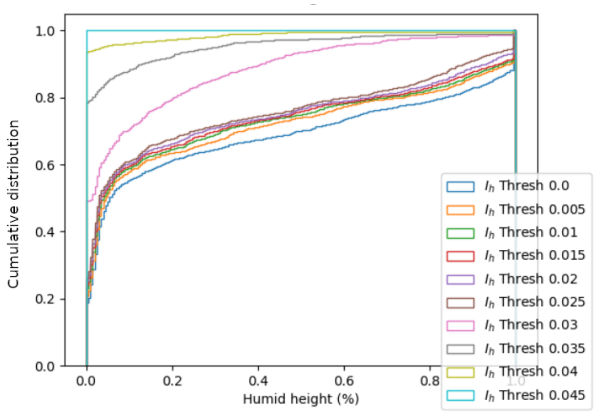


Figure C.1: Histogram of I_h values depending on the threshold of LWC chosen to determine what is a humid layer. LWC data from Crocus reanalysis on Chartreuse, Haute-Tarentaise and Thabor massifs at 1500, 2100 and 2700 m.

represent liquid water percolation based (by default, other implementations are available [Lafaysse et al., 2017; Wever et al., 2015, 2016b]) on a simple bucket scheme. For Crocus, the maximal retention is of 5% of pore volume, once reached this threshold the water flow to the layer below. For snowpack, the maximal value was originally 8% in volume [Bartelt and Lehning, 2002] (which crudely correspond to 4% of pore volume) and have been adapted to layer properties nowadays. Once reached, additional water is transferred to lower layers with a flow controlled by Darcy equation [Bartelt and Lehning, 2002]. Consequently, a threshold used on Snowpack simulation for LWC could not be immediately transferred to Crocus simulations.

For the $\mathrm{LWC}_{\mathrm{index}}$ index, the threshold or normalization on liquid water content have to be selected below 5% to be used with Crocus snow cover model. However, if the index is not used with a threshold to separate two classes, and used directly as a continuous value, the denominator is only a normalization factor and does not matter when the threshold is therefore optimized as in the Random Forest method used in this work. On the contrary, I_h rely on a threshold to identify dry and humid layers. On Figure C.1, we showed the histogram of I_h values depending on the threshold of LWC chosen to determine what is a humid layer. We showed that a threshold of 0 lead to a different behavior, as well as any threshold above 3%. However, intermediate values does not significantly change the histogram. We thus selected 0, 1 and 3% in our studies.

C.4 Conclusion

Literature for identification of wet snow avalanche situations is far less developed compared to dry snow situations. However, depending whether the goal is to identify avalanche-prone situations at different scale or classify days between dry and wet snowpack days, two indices emerged. The $\mathrm{LWC}_{\mathrm{index}}$ which is the mean LWC on the whole snowpack, with normalization, have shown to be relevant for identifying wet snow avalanche situations. The I_h index is more designed to classify between dry and wet situations. Both are interesting for short-term avalanche forecasting. The second one allows to identify if the situation may involve or not wet snow problems while the $\mathrm{LWC}_{\mathrm{index}}$ give information on the criticality of the situation.

Appendix D

List of publications resulting from this thesis

D.1 Peer-reviewed publications

- **Viallon-Galinier L.**, Hagenmuller P. and Lafaysse M. (2020). Forcing and evaluating detailed snow cover models with stratigraphy observations. *Cold Regions Science and Technology*, 180, 103163.
- Sielenou P. D., **Viallon-Galinier L.**, Hagenmuller P., Naveau P., Morin S., Dumont M., Verfaillie D. and Eckert N. (2021). Combining random forests and class-balancing to discriminate between three classes of avalanche activity in the French Alps. *Cold Regions Science and Technology*, 187, 103276.
- **Viallon-Galinier L.**, Hagenmuller P., Reuter B. and Eckert N. (2022). Modelling snowpack stability from simulated snow stratigraphy: Summary and implementation examples. *Cold Regions Science and Technology*, 103596.
- **Viallon-Galinier L.**, Hagenmuller P. and Eckert N. (2022). Combining snow physics and machine learning to predict avalanche activity: does it help? *The Cryosphere Discussions*, 1-23.
- Reuter B., **Viallon-Galinier L.**, Horton S., van Herwijnen A., Mayer S., Hagenmuller P. and Morin S. (2022). Characterizing snow instability with avalanche problem types derived from snow cover simulations. *Cold Regions Science and Technology*, 194, 103462.
- Dick O., **Viallon-Galinier L.**, Tuzet F., Hagenmuller P., Fructus M., Reuter B., Lafaysse M. and Dumont M. (2023). Can Saharan dust deposition impact snowpack stability in the French Alps? *The Cryosphere*.

D.2 Other publications

- Réveillet M., Tuzet F., Dumont M., Gascoin S., Arnaud L., Bonnefoy M., Carmagnola C., Deguine A., Evrard O., Flin F., Fontaine F., Gandois L., Hagenmuller P., Herbin H., Josse B., Lafaysse M., Gaël Roux L., Morin S., Nabat P., Petitprez D., Picard G., Robledano A., Schneebeli M.,0 Six D., Thibert E., Vernay M., **Viallon-Galinier L.**, Voiron C. and Voisin D. (2021). Dépôts massifs de poussières sahariennes sur le manteau neigeux dans les Alpes et les Pyrénées du 5 au 7 février 2021: Contexte, enjeux et résultats préliminaires Version du 3 mai 2021 (Doctoral dissertation, CNRM, Université de Toulouse, Météo-France, CNRS).

D.3 International conferences

- **Viallon-Galinier L.**, Hagenmuller P. and Eckert N. (2019) Combining physical modelling and machine learning to improve avalanche risk forecasting. *ALERT Geomaterials*, 44.
- **Viallon-Galinier L.**, Hagenmuller P., Lafaysse M. (2020) Forcing and evaluating detailed snow cover models with stratigraphy observations. *Interational Snow Science Workshop*.

- Lafaysse M., Dumont M., Nheili R., **Viallon-Galinier L.**, Carmagnola C., Cluzet B., Fructus M., Hagenmuller P., Morin S., Spandre P., François T. and Vionnet V. (2020). Latest scientific and technical evolutions in the Crocus snowpack model. In *EGU General Assembly Conference Abstracts* (p. 10217).
- **Viallon-Galinier L.**, Hagenmuller P., Eckert N. and Reuter B. (2021). Mechanical stability indicators derived from detailed snow cover simulations. In *EGU General Assembly Conference Abstracts* (pp. EGU21-1341).
- Reuter B., **Viallon-Galinier L.**, Mayer S., Hagenmuller P. and Morin S. (2021). A tracking algorithm to identify slab and weak layer combinations for assessing snow instability and avalanche problem type. In *EGU General Assembly Conference Abstracts* (pp. EGU21-1578).
- Dick O., **Viallon-Galinier L.**, Hagenmuller P., Fructus M., Lafaysse M. and Dumont M. (2021). Can Saharan dust deposition impact snow stability in the French Alps? In *EGU General Assembly Conference Abstracts* (pp. EGU21-6600).
- **Viallon-Galinier L.**, Hagenmuller P. and Eckert N. (2022). Combining snow physics an machine learning to predict avalanche activity. International Glaciological Society symposium on Snow.

Bibliography

Ancey, C., Gervasoni, C., & Meunier, M. (2004). Computing extreme avalanches. *Cold Regions Science and Technology*, *39*(2-3), 161–180, doi:10.1016/j.coldregions.2004.04.004.

Anderson, T. L. (2017). *Fracture mechanics: fundamentals and applications*. CRC press.

Bartelt, P. & Lehning, M. (2002). A physical snowpack model for the swiss avalanche warning: Part i: numerical model. *Cold Regions Science and Technology*, *35*(3), 123–145, doi:10.1016/S0165-232X(02)00074-5.

Bazile, E., El Haiti, M., Bogatchev, A., & Spiridonov, V. (2001). Improvement of the snow parameterization in arpege/aladin. In *Proceedings of SRNWP/HIRLAM Workshop on surface processes, turbulence and mountain effects*, volume 22 (pp.̃24).

Bellaire, S., van Herwijnen, A., Mitterer, C., & Schweizer, J. (2017). On forecasting wet-snow avalanche activity using simulated snow cover data. *Cold Regions Science and Technology*, *144*, 28–38, doi:10.1016/j.coldregions.2017.09.013. International Snow Science Workshop 2016 Breckenridge.

Bellman, R. (1957). Dynamic programming princeton university press princeton. *New Jersey Google Scholar*.

Bergfeld, B., et al. (2021a). Crack propagation speeds in weak snowpack layers. *Journal of Glaciology*, (pp. 1–14)., doi:10.1017/jog.2021.118.

Bergfeld, B., van Herwijnen, A., Reuter, B., Bobillier, G., Dual, J., & Schweizer, J. (2021b). Dynamic crack propagation in weak snowpack layers: insights from high-resolution, high-speed photography. *The Cryosphere*, *15*(7), 3539–3553, doi:10.5194/tc-15-3539-2021.

Birkeland, K. W., van Herwijnen, A., Reuter, B., & Bergfeld, B. (2019). Temporal changes in the mechanical properties of snow related to crack propagation after loading. *Cold Regions Science and Technology*, *159*, 142–152, doi:10.1016/j.coldregions.2018.11.007.

Bobillier, G., Bergfeld, B., Dual, J., Gaume, J., van Herwijnen, A., & Schweizer, J. (2021). Micro-mechanical insights into the dynamics of crack propagation in snow fracture experiments. *Scientific Reports*, *11*(1), doi:10.1038/s41598-021-90910-3.

Bois, P. & Obled, C. (1976). Prévision des avalanches par des méthodes statistiques. aspects méthodologiques et opérationnels. *La Houille Blanche*, *62*(6-7), 509–531, doi:10.1051/lhb/1976034.

Bois, P., Obled, C., & Good, W. (1974). Multivariate data analysis as a tool for day-by-day avalanche forecast. In *Snow Mechanics Symposium*.

Boone, A. & Etchevers, P. (2001). An intercomparison of three snow schemes of varying complexity coupled to the same land surface model: Local-scale evaluation at an alpine site. *Journal of Hydrometeorology*, *2*(4), 374–394, doi:10.1175/1525-7541(2001)002<0374:AIOTSS>2.0.CO;2.

Bourova, E., Maldonado, E., Leroy, J., Alouani, R., Eckert, N., Bonnefoy-Demongeot, M., & Deschatres, M. (2016). A new web-based system to improve the monitoring of snow avalanche hazard in france. *Natural Hazards and Earth System Sciences*, *16*(5), 1205–1216, doi:10.5194/nhess-16-1205-2016.

Boussinesq, J. (1885). *Application des potentiels a l'étude de l'équilibre et du mouvement des solides élastiques*. Gauthier-Villars.

Bovis, M. J. (1977). Statistical forecasting of snow avalanches, san juan mountains, southern colorado, u.s.a. *Journal of Glaciology*, *18*(78), 87–99, doi:10.3189/s0022143000021547.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159, doi:10.1016/s0031-3203(96)00142-2.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32, doi:10.1023/a:1010933404324.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Cole Advanced Books and Software.

Brown, C. & Jamieson, B. (2008). Shear strength and snowpack stability trends in non-persistant weak layers. In *Proceedings Whistler 2008 International Snow Science Workshop September 21-27, 2008* (pp. 939).

Brun, E., David, P., Sudul, M., & Brunot, G. (1992). A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting. *Journal of Glaciology*, *38*(128), 13–22, doi:10.3189/s0022143000009552.

Brun, E., Martin, E., Simon, V., Gendre, C., & Coleou, C. (1989). An energy and mass model of snow cover suitable for operational avalanche forecasting. *Journal of Glaciology*, *35*(121), 333–342, doi:10.3189/S0022143000009254.

Brründl, M. & Margreth, S. (2021). Integrative risk management: The example of snow avalanches. In *Snow and Ice-Related Hazards, Risks, and Disasters* (pp. 259–296). Elsevier.

Buser, O. (1983). Avalanche forecast with the method of nearest neighbours: An interactive approach. *Cold Regions Science and Technology*, *8*(2), 155–163, doi:10.1016/0165-232x(83)90006-x.

Buser, O. (1989). Two years experience of operational avalanche forecasting using the nearest neighbours method. *Annals of Glaciology*, *13*, 31–34, doi:10.3189/s026030550000759x.

Calmet, C. (2018). La vigilance météorologique, une innovation pour le grand public. *La Météorologie*, *100*, 56–65, doi:10.4267/2042/65143.

Camponovo, C. & Schweizer, J. (2001). Rheological measurements of the viscoelastic properties of snow. *Annals of Glaciology*, *32*, 44–50, doi:10.3189/172756401781819148.

Capelli, A., Reiweger, I., Lehmann, P., & Schweizer, J. (2018). Fiber-bundle model with time-dependent healing mechanisms to simulate progressive failure of snow. *Phys. Rev. E*, *98*, 023002, doi:10.1103/PhysRevE.98.023002.

Castebrunet, H., Eckert, N., & Giraud, G. (2012). Snow and weather climatic control on snow avalanche occurrence fluctuations over 50 yr in the french alps. *Climate of the Past*, *8*(2), 855–875, doi:10.5194/cp-8-855-2012.

Castebrunet, H., Eckert, N., Giraud, G., Durand, Y., & Morin, S. (2014). Projected changes of snow conditions and avalanche activity in a warming climate: the french alps over the 2020–2050 and 2070–2100 periods. *The Cryosphere*, *8*(5), 1673–1697, doi:10.5194/tc-8-1673-2014.

Chalmers, T. S. & Jamieson, B. (2001). Extrapolating the skier stability of buried surface hoar layers from study plot measurements. *Cold Regions Science and Technology*, *33*(2), 163–177, doi:10.1016/S0165-232X(01)00043-X. International Snow Science Workshop 2000.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., & Picard, G. (2016). On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model. *The Cryosphere*, *10*(3), 1021–1038, doi:10.5194/tc-10-1021-2016.

Chawla, M. & Singh, A. (2021). A data efficient machine learning model for autonomous operational avalanche forecasting. *Natural Hazards and Earth System Sciences Discussions*, doi:10.5194/nhess-2021-106.

Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Editorial. *ACM SIGKDD Explorations Newsletter*, *6*(1), 1–6, doi:10.1145/1007730.1007733.

Chen, C., Liaw, A., Breiman, L., et al. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, *110*(1-12), 24.

Choubin, B., Borji, M., Mosavi, A., Sajedi-Hosseini, F., Singh, V. P., & Shamshirband, S. (2019). Snow avalanche hazard prediction using machine learning methods. *Journal of Hydrology*, *577*, 123929, doi:10.1016/j.jhydrol.2019.123929.

Cluzet, B., Lafaysse, M., Cosme, E., Albergel, C., Meunier, L.-F., & Dumont, M. (2021). CrocO_v1.0: a particle filter to assimilate snowpack observations in a spatialised framework. *Geoscientific Model Development*, *14*(3), 1595–1614, doi:10.5194/gmd-14-1595-2021.

Cluzet, B., et al. (2020). Towards the assimilation of satellite reflectance into semi-distributed ensemble snowpack simulations. *Cold Regions Science and Technology*, *170*, 102918, doi:10.1016/j.coldregions.2019.102918.

Coléou, C. & Morin, S. (2018). Vingt-cinq ans de prévision du risque d'avalanche à météo-france. *La Météorologie*, *100*, 79–84, doi:10.4267/2042/65147.

Conway, H. & Wilbour, C. (1999). Evolution of snow slope stability during storms. *Cold Regions Science and Technology*, *30*(1), 67–77, doi:10.1016/S0165-232X(99)00009-9.

Davis, R. E., Elder, K., Howlett, D., & Bouzaglou, E. (1999). Relating storm and weather factors to dry slab avalanche activity at alta, utah, and mammoth mountain, california, using classification and regression trees. *Cold Regions Science and Technology*, *30*(1-3), 79–89, doi:10.1016/s0165-232x(99)00032-4.

De Crécy, L. (1965). Cadastre et statistique d'avalanches. *Revue forestière française*, (1), 24–33.

Decharme, B., Boone, A., Delire, C., & Noilhan, J. (2011). Local evaluation of the interaction between soil biosphere atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *Journal of Geophysical Research: Atmospheres*, *116*(D20), doi:10.1029/2011JD016002.

Dekanova, M., Duchon, F., Dekan, M., Kyzek, F., & Biskupic, M. (2018). Avalanche forecasting using neural network. In *2018 ELEKTRO*: IEEE.

Denoth, A. (1980). The pendular-funicular liquid transition in snow. *Journal of Glaciology*, *25*(91), 93–98.

Dick, O., et al. (2022). Can saharan dust deposition impact snowpack stability in the french alps? *The Cryosphere Discussions*, doi:10.5194/tc-2022-219.

Douville, H., Royer, J. F., & Mahfouf, J. F. (1995). A new snow parameterization for the météo-france climate model. *Climate Dynamics*, *12*(1), 21–35, doi:10.1007/BF00208760.

Dreier, L., Harvey, S., van Herwijnen, A., & Mitterer, C. (2016). Relating meteorological parameters to glide-snow avalanche activity. *Cold Regions Science and Technology*, *128*, 57–68, doi:10.1016/j.coldregions.2016.05.003.

Durand, Y., Giraud, G., Brun, E., Mérindol, L., & Martin, E. (1999). A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting. *Journal of Glaciology*, *45*(151), 469–484, doi:10.3189/S0022143000001337.

Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., & Mérindol, L. (2009). Reanalysis of 44 yr of climate in the french alps (1958–2002): Methodology, model validation, climatology, and trends for air temperature and precipitation. *Journal of Applied Meteorology and Climatology*, *48*(3), 429–449, doi:10.1175/2008JAMC1808.1.

EAWS (2022). Eauropean Avalanche Warning Services website. `www.avalanches.org`. Last accessed 2022-08.

Ebert, P. A. & Milne, P. (2022). Methodological and conceptual challenges in rare and severe event forecast verification. *Natural Hazards and Earth System Sciences*, *22*(2), 539–557, doi:10.5194/nhess-22-539-2022.

Eckert, N., Coleou, C., Castebrunet, H., Deschatres, M., Giraud, G., & Gaume, J. (2010a). Cross-comparison of meteorological and avalanche data for characterising avalanche cycles: The example of december 2008 in the eastern part of the french alps. *Cold Regions Science and Technology*, *64*(2), 119–136, doi:10.1016/j.coldregions.2010.08.009.

Eckert, N. & Giacona, F. (2022). Towards a holistic paradigm for long-term snow avalanche risk assessment and mitigation. *Ambio*, *52*(4), 711–732, doi:10.1007/s13280-022-01804-1.

Eckert, N., Keylock, C., Castebrunet, H., Lavigne, A., & Naaim, M. (2013). Temporal trends in avalanche activity in the french alps and subregions: from occurrences and runout altitudes to unsteady return periods. *Journal of Glaciology*, *59*(213), 93–114, doi:10.3189/2013JoG12J091.

Eckert, N., et al. (2018). Repenser les fondements du zonage règlementaire des risques en montagne « récurrents ». *La Houille Blanche*, *104*(2), 38–67, doi:10.1051/lhb/2018019.

Eckert, N., Naaim, M., & Parent, E. (2010b). Long-term avalanche hazard assessment with a bayesian depth-averaged propagation model. *Journal of Glaciology*, *56*(198), 563–586, doi:10.3189/002214310793146331.

Eckert, N., Parent, E., Faug, T., & Naaim, M. (2009). Bayesian optimal design of an avalanche dam using a multivariate numerical avalanche model. *Stochastic Environmental Research and Risk Assessment*, *23*(8), 1123–1141, doi:10.1007/s00477-008-0287-6.

Evin, G., Lafaysse, M., Taillardat, M., & Zamo, M. (2021a). Calibrated ensemble forecasts of the height of new snow using quantile regression forests and ensemble model output statistics. *Nonlinear Processes in Geophysics*, doi:10.5194/npg-2021-18.

Evin, G., Sielenou, P. D., Eckert, N., Naveau, P., Hagenmuller, P., & Morin, S. (2021b). Extreme avalanche cycles: Return levels and probability distributions depending on snow and meteorological conditions. *Weather and Climate Extremes*, *33*, 100344, doi:10.1016/j.wace.2021.100344.

Favier, P., Eckert, N., Bertrand, D., & Naaim, M. (2014). Sensitivity of avalanche risk to vulnerability relations. *Cold Regions Science and Technology*, *108*, 163–177, doi:10.1016/j.coldregions.2014.08.009.

Fierz, C., et al. (2009). The international classification for seasonal snow on the ground. *IHP-VII Technical Documents in Hydrology*, *83*.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188, doi:10.1111/j.1469-1809.1936.tb02137.x.

Floyer, J. A. & McClung, D. M. (2003). Numerical avalanche prediction: Bear pass, british columbia, canada. *Cold Regions Science and Technology*, *37*(3), 333–342, doi:10.1016/s0165-232x(03)00074-0.

Föhn, P. M. (1987a). The rutschblock as a practical tool for slope stability evaluation. *IAHS Publ*, *162*, 223–228.

Föhn, P. M. (1987b). The stability index and various triggering mechanisms. *IAHS publication*, *162*, 195–214.

Föhn, P., Good, W., Bois, P., & Obled, C. (1977). Evaluation and comparison of statistical and conventional methods of forecasting avalanche hazard. *Journal of Glaciology*, *19*(81), 375–387, doi:10.3189/S0022143000029403.

Gassner, M. & Brabec, B. (2002). Nearest neighbour models for local and regional avalanche forecasting. *Natural Hazards and Earth System Science*, *2*(3/4), 247–253, doi:10.5194/nhess-2-247-2002.

Gaume, J. (2012). *Prédétermination des hauteurs de départ d'avalanches. Modélisation combinée statistique-mécanique*. Theses, Université de Grenoble.

Gaume, J., Chambon, G., Eckert, N., & Naaim, M. (2013). Influence of weak-layer heterogeneity on snow slab avalanche release: application to the evaluation of avalanche release depths. *Journal of Glaciology*, *59*(215), 423–437, doi:10.3189/2013jog12j161.

Gaume, J., Gast, T., Teran, J., van Herwijnen, A., & Jiang, C. (2018). Dynamic anticrack propagation in snow. *Nature Communications*, *9*(1), 3047, doi:10.1038/s41467-018-05181-w.

Gaume, J. & Reuter, B. (2017). Assessing snow instability in skier-triggered snow slab avalanches by combining failure initiation and crack propagation. *Cold Regions Science and Technology*, *144*, 6–15, doi:10.1016/j.coldregions.2017.05.011. International Snow Science Workshop 2016, Breckenridge.

Gaume, J., Schweizer, J., Herwijnen, A., Chambon, G., Reuter, B., Eckert, N., & Naaim, M. (2014). Evaluation of slope stability with respect to snowpack spatial variability. *Journal of Geophysical Research: Earth Surface*, *119*(9), 1783–1799, doi:10.1002/2014jf003193.

Gaume, J., van Herwijnen, A., Chambon, G., Birkeland, K. W., & Schweizer, J. (2015). Modeling of crack propagation in weak snowpack layers using the discrete element method. *The Cryosphere*, *9*(5), 1915–1932, doi:10.5194/tc-9-1915-2015.

Gaume, J., Van Herwijnen, A., CHAMBON, G., Wever, N., & Schweizer, J. (2017). Snow fracture in relation to slab avalanche release: critical state for the onset of crack propagation. *The Cryosphere*, *11*(1), 217–228, doi:10.5194/tc-11-217-2017.

Gauthier, D. & Jamieson, B. (2006). Towards a field test for fracture propagation propensity in weak snowpack layers. *Journal of Glaciology*, *52*(176), 164–168, doi:10.3189/172756506781828962.

Gauthier, D. & Jamieson, B. (2008). Evaluation of a prototype field test for fracture and failure propagation propensity in weak snowpack layers. *Cold Regions Science and Technology*, *51*(2), 87–97, doi:10.1016/j.coldregions.2007.04.005. International Snow Science Workshop (ISSW) 2006.

Gerling, B., Löwe, H., & van Herwijnen, A. (2017). Measuring the elastic modulus of snow. *Geophysical Research Letters*, *44*(21), 11088–11096, doi:10.1002/2017GL075110.

Giacona, F., et al. (2021). Upslope migration of snow avalanches in a warming climate. *Proceedings of the National Academy of Sciences*, *118*(44), doi:10.1073/pnas.2107306118.

Giacona, F., et al. (2018). Avalanche activity and socio-environmental changes leave strong footprints in forested landscapes: a case study in the vosges medium-high mountain range. *Annals of Glaciology*, *59*(77), 111–133, doi:10.1017/aog.2018.26.

Giacona, F., Eckert, N., & Martin, B. (2017). A 240-year history of avalanche risk in the vosges mountains based on non-conventional (re)sources. *Natural Hazards and Earth System Sciences*, *17*(6), 887–904, doi:10.5194/nhess-17-887-2017.

Giard, D., et al. (2018). L'approche participative au service de l'observation météorologique. *La Météorologie*, (100 Spécial Anniversaire 25 ans), 105, doi:10.4267/2042/65152.

Giles, C. L. & Gori, M. (1998). *Adaptive Processing of Sequences and Data Structures*. Springer Berlin Heidelberg.

Giraud, G. & Navarre, J. (1995). Mepra et le risque de déclenchement accidentel d'avalanches. In *Les apports de la recherche scientifique à la sécurité neige glace et avalanche, Actes de Colloque, ANENA*.

Giraud, G., Navarre, J.-P., & Coléou, C. (2002). *Estimation du risque avalancheux dans le système expert MEPRA*. Technical report, CNRM - Centre national de recherches météorologiques. Unpublished work.

Glass, B., Huet, P., Rat, M., & Tordjeman, R. (2000). Retour d'expérience sur l'avalanche du 9 février 1999 à montroc, commune de chamonix. *Inspection Générale de l'Environnement. France, Rapport d'expertise*.

Griffith, A. A. (1921). VI. the phenomena of rupture and flow in solids. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *221*(582-593), 163–198, doi:10.1098/rsta.1921.0006.

Habermann, M., Schweizer, J., & Jamieson, J. B. (2008). Influence of snowpack layering on human-triggered snow slab avalanche release. *Cold Regions Science and Technology*, *54*(3), 176–182, doi:10.1016/j.coldregions.2008.05.003.

Hafner, E. D., Techel, F., Leinss, S., & Bühler, Y. (2021). Mapping avalanches with satellites – evaluation of performance and completeness. *The Cryosphere*, *15*(2), 983–1004, doi:10.5194/tc-15-983-2021.

Hagenmuller, P. (2014). *Modélisation du comportement mécanique de la neige à partir d'images microtomographiques*. PhD thesis, Sciences de la terre et de l'univers, et de l'environnement Grenoble. Thèse de doctorat dirigée par Naaim, Mohamed.

Hagenmuller, P., Chambon, G., & Naaim, M. (2015). Microstructure-based modeling of snow mechanics: a discrete element approach. *The Cryosphere*, *9*(5), 1969–1982, doi:10.5194/tc-9-1969-2015.

Hagenmuller, P., Theile, T. C., & Schneebeli, M. (2014). Numerical simulation of microstructural damage and tensile strength of snow. *Geophysical Research Letters*, *41*(1), 86–89, doi:10.1002/2013gl058078.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York.

Heierli, J., Gumbsch, P., & Zaiser, M. (2008). Anticrack nucleation as triggering mechanism for snow slab avalanches. *Science*, *321*(5886), 240–243, doi:10.1126/science.1153948.

Hendrikx, J., Murphy, M., & Onslow, T. (2014). Classification trees as a tool for operational avalanche forecasting on the seward highway, alaska. *Cold Regions Science and Technology*, *97*, 113–120, doi:10.1016/j.coldregions.2013.08.009.

Hendrikx, J., Owens, I., Carran, W., & Carran, A. (2005). Avalanche activity in an extreme maritime climate: The application of classification trees for forecasting. *Cold Regions Science and Technology*, *43*(1-2), 104–116, doi:10.1016/j.coldregions.2005.05.006.

Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780, doi:10.1162/neco.1997.9.8.1735.

International Association of Hydrological Sciences (1981). *Avalanche atlas; illustrated international avalanche classification*. UNESCO.

Jamieson, B. & Johnston, C. D. (2001). Evaluation of the shear frame test for weak snowpack layers. *Annals of Glaciology*, *32*, 59–69, doi:10.3189/172756401781819472.

Jamieson, B. & Schweizer, J. (2005). Using a checklist to assess manual snow profiles. *Avalanche News*, *72*, 57–61.

Jamieson, J. & Johnston, C. (1990). In-situ tensile tests of snow-pack layers. *Journal of Glaciology*, *36*(122), 102–106, doi:10.3189/S002214300000561X.

Jamieson, J. & Johnston, C. (1998). Refinements to the stability index for skier-triggered dry-slab avalanches. *Annals of Glaciology*, *26*, 296–302, doi:10.3189/1998AoG26-1-296-302.

Jeudy, J.-M. (2006). *Montagne maudite, montagne apprivoisée*. Fontaine de Siloé.

Jomelli, V., et al. (2007). Probabilistic analysis of recent snow avalanche activity and weather in the french alps. *Cold Regions Science and Technology*, *47*(1-2), 180–192, doi:10.1016/j.coldregions.2006.08.003.

Jones, A., Jamieson, J., & Schweizer, J. (2006). The effect of slab and bed surface stiffness on the skier-induced shear stress in weak snowpack layers. In *Proceedings ISSW* (pp. 157–164).

Jordan, R. (1991). *A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM. 89.* Technical report, Cold Regions Research and Engineering Lab Hanover NH.

Karas, A., Karbou, F., Giffard-Roisin, S., Durand, P., & Eckert, N. (2022). Automatic color detection-based method applied to sentinel-1 SAR images for snow avalanche debris monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–17, doi:10.1109/tgrs.2021.3131853.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2019). Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, *31*(8), 1544–1554, doi:10.1109/tkde.2018.2861006.

Kashinath, K., et al. (2021). Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2194), 20200093, doi:10.1098/rsta.2020.0093.

Keeler, C. & Weeks, W. (1968). Investigations into the mechanical properties of alpine snow-packs. *Journal of Glaciology*, *7*(50), 253–271, doi:10.3189/S0022143000031038.

Kern, H., Eckert, N., Jomelli, V., Grancher, D., Deschatres, M., & Arnaud-Fassetta, G. (2021). Brief communication: Weak control of snow avalanche deposit volumes by avalanche path morphology. *The Cryosphere*, *15*(10), 4845–4852, doi:10.5194/tc-15-4845-2021.

Keylock, C. J., McClung, D. M., & Magnússon, M. M. (1999). Avalanche risk mapping by simulation. *Journal of Glaciology*, *45*(150), 303–314, doi:10.3189/S0022143000001805.

Kirchner, H., Michot, G., & Schweizer, J. (2002). Fracture toughness of snow in shear and tension. *Scripta Materialia*, *46*(6), 425–429, doi:10.1016/S1359-6462(02)00007-6.

Kronholm, K., Vikhamar-Schuler, D., Jaedicke, C., Isaksen, K., Sorteberg, A., & Kristensen, K. (2006). Forecasting snow avalanche days from meteorological data using classification trees; grasdalen, western norway. In *Proceedings of the International Snow Science Workshop, Telluride, Colorado* (pp. 1–6).: Citeseer.

Köchle, B. & Schneebeli, M. (2014). Three-dimensional microstructure and numerical calculation of elastic properties of alpine snow with a focus on weak layers. *Journal of Glaciology*, *60*(222), 705–713, doi:10.3189/2014JoG13J220.

LaChapelle, E. R. (1977). Snow avalanches: A review of current research and applications. *Journal of Glaciology*, *19*(81), 313–324, doi:10.3189/s0022143000215633.

Lafaysse, M., Cluzet, B., Dumont, M., Lejeune, Y., Vionnet, V., & Morin, S. (2017). A multiphysical ensemble system of numerical snow modelling. *The Cryosphere*, *11*(3), 1173–1198, doi:10.5194/tc-11-1173-2017.

Lafaysse, M., et al. (2020). Latest scientific and technical evolutions in the crocus snowpack model. In *EGU General Assembly Conference Abstracts* (pp. 10217).

Lafaysse, M., et al. (2013). Towards a new chain of models for avalanche hazard forecasting in french mountain ranges, including low altitude mountains. In *Proceedings of International Snow Science Workshop Grenoble–Chamonix Mont-Blanc*, volume 7 (pp. 162–166).

Lavigne, A., Eckert, N., Bel, L., & Parent, E. (2015). Adding expert contributions to the spatiotemporal modelling of avalanche activity under different climatic influences. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *64*(4), 651–671, doi:10.1111/rssc.12095.

LeBaron, A. M. & Miller, D. A. (2016). Micromechanical analysis of energy release in snow fracture. In *Proceedings International Snow Science Workshop, Breckenridge, Colorado, 3* (pp. 658–663).

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324, doi:10.1109/5.726791.

Lehning, M., Bartelt, P., Brown, B., & Fierz, C. (2002a). A physical snowpack model for the swiss avalanche warning: Part iii: meteorological forcing, thin layer formation and evaluation. *Cold Regions Science and Technology*, *35*(3), 169–184, doi:10.1016/S0165-232X(02)00072-1.

Lehning, M., Bartelt, P., Brown, B., Fierz, C., & Satyawali, P. (2002b). A physical snowpack model for the swiss avalanche warning: Part ii. snow microstructure. *Cold Regions Science and Technology*, *35*(3), 147–167, doi:10.1016/S0165-232X(02)00073-3.

Lehning, M., Fierz, C., Brown, B., & Jamieson, B. (2004). Modeling snow instability with the snow-cover model snowpack. *Annals of Glaciology*, *38*, 331–338, doi:10.3189/172756404781815220.

Leinss, S., Wicki, R., Holenstein, S., Baffelli, S., & Bühler, Y. (2020). Snow avalanche detection and mapping in multitemporal and multiorbital radar images from TerraSAR-x and sentinel-1. *Natural Hazards and Earth System Sciences*, *20*(6), 1783–1803, doi:10.5194/nhess-20-1783-2020.

Loth, B. & Graf, H.-F. (1998). Modeling the snow cover in climate studies: 1. long-term integrations under different climatic conditions using a multilayered snow-cover model. *Journal of Geophysical Research: Atmospheres*, *103*(D10), 11313–11327, doi:10.1029/97JD01411.

Marienthal, A., Hendrikx, J., Birkeland, K., & Irvine, K. M. (2015). Meteorological variables to aid forecasting deep slab avalanches on persistent weak layers. *Cold Regions Science and Technology*, *120*, 227–236, doi:10.1016/j.coldregions.2015.08.007.

Martin, E., Giraud, G., Lejeune, Y., & Boudart, G. (2001). Impact of a climate change on avalanche hazard. *Annals of Glaciology*, *32*, 163–167, doi:10.3189/172756401781819292.

Mayer, S., van Herwijnen, A., & Schweizer, J. (2021). A random forest model to assess snow instability from simulated snow stratigraphy. In *EGU General Assembly Conference Abstracts* (pp. EGU21–12259).

Mayer, S., van Herwijnen, A., Techel, F., & Schweizer, J. (2022). A random forest model to assess snow instability from simulated snow stratigraphy. *The Cryosphere Discussions*, doi:10.5194/tc-2022-34.

Mayer, S., van Herwijnen, A., Ulivieri, G., & Schweizer, J. (2020). 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*, 102962, doi:10.1016/j.coldregions.2019.102962.

McClung, D. (2023). *The Avalanche Handbook*. Mountaineers Books, 4th edition edition.

McClung, D. & Schaerer, P. (1993). The avalanche handbook: The mountaineers. *Seattle, WA*.

McClung, D. & Schweizer, J. (1999). Skier triggering, snow temperatures and the stability index for dry-slab avalanche initiation. *Journal of Glaciology*, *45*(150), 190–200, doi:10.3189/S0022143000001696.

McClung, D. M. (1979). Shear fracture precipitated by strain softening as a mechanism of dry slab avalanche release. *Journal of Geophysical Research: Solid Earth*, *84*(B7), 3519–3526, doi:10.1029/JB084iB07p03519.

McGregor, G. R. (1989). Snow avalanche forecasting by discriminant function analysis. *Weather and Climate*, *9*(2), 3–14.

Mede, T., Chambon, G., Hagenmuller, P., & Nicot, F. (2018). Snow failure modes under mixed loading. *Geophysical Research Letters*, *45*(24), doi:10.1029/2018gl080637.

Mellor, M. (1974). *A review of basic snow mechanics*. US Army Cold Regions Research and Engineering Laboratory.

Mitterer, C., Heilig, A., Schmid, L., van Herwijnen, A., Eisen, O., & Schweizer, J. (2016). Comparison of measured and modelled snow cover liquid water content to improve local wet-snow avalanche prediction. In *International Snow Science Workshop Proceedings*.

Mitterer, C. & Schweizer, J. (2013). Analysis of the snow-atmosphere energy balance during wet-snow instabilities and implications for avalanche prediction. *The Cryosphere*, *7*(1), 205–216, doi:10.5194/tc-7-205-2013.

Mitterer, C., Techel, F., Fierz, C., & Schweizer, J. (2013). An operational supporting tool for assessing wet-snow avalanche danger. In *Proceedings ISSW*, volume 33.

Moigne, P. L. (2012). Surfex scientific documentation.

Monti, F., Gaume, J., van Herwijnen, A., & Schweizer, J. (2016). Snow instability evaluation: calculating the skier-induced stress in a multi-layered snowpack. *Natural Hazards and Earth System Sciences*, *3*, 4833–4869, doi:10.5194/nhessd-3-4833-2015.

Morin, S., et al. (2020). Application of physical snowpack models in support of operational avalanche hazard forecasting: A status report on current implementations and prospects for the future. *Cold Regions Science and Technology*, *170*, 102910, doi:10.1016/j.coldregions.2019.102910.

Mosavi, A., et al. (2020). Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping. *IEEE Access*, *8*, 145968–145983, doi:10.1109/access.2020.3014816.

Möhle, S., Bründl, M., & Beierle, C. (2014). Modeling a system for decision support in snow avalanche warning using balanced random forest and weighted random forest. In *Artificial Intelligence: Methodology, Systems, and Applications* (pp. 80–91). Springer International Publishing.

Naaim, M., Eckert, N., Giraud, G., Faug, T., Chambon, G., Naaim-Bouvet, F., & Richard, D. (2016). Impact du réchauffement climatique sur l'activité avalancheuse et multiplication des avalanches humides dans les alpes françaises. *La Houille Blanche*, (6), 12–20, doi:10.1051/lhb/2016055.

Nadreau, J.-P. & Michel, B. (1986). Yield and failure envelope for ice under multiaxial compressive stresses. *Cold Regions Science and Technology*, *13*(1), 75–82, doi:10.1016/0165-232X(86)90009-1.

Narita, H. (1980). Mechanical behaviour and structure of snow under uniaxial tensile stress. *Journal of Glaciology*, *26*(94), 275–282, doi:10.3189/S0022143000010819.

Navarre, J. P., Guyomarc'h, G., & Giraud, G. (1987). Un modèle statistique pour la prévision locale des avalanches. In *Proceedings of the Davos Symposium*, volume 162 (pp. 571–580).

Nishimura, K., Baba, E., Hirashima, H., & Lehning, M. (2005). Application of the snow cover model snowpack to snow avalanche warning in niseko, japan. *Cold Regions Science and Technology*, *43*(1), 62–70, doi:10.1016/j.coldregions.2005.05.007. Snow and Avalanches.

Obled, C., Bontron, G., & Garçon, R. (2002). Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach. *Atmospheric Research*, *63*(3-4), 303–324, doi:10.1016/s0169-8095(02)00038-8.

Obled, C. & Good, W. (1980). Recent developments of avalanche forecasting by discriminant analysis techniques: A methodological review and some applications to the parsenn area (davos, switzerland). *Journal of Glaciology*, *25*(92), 315–346, doi:10.3189/S0022143000010522.

Pahaut, E. & Giraud, G. (1995). La prévision du risque d'avalanche en france : bilan et perspectives. *La Météorologie*, *8*(12), 46, doi:10.4267/2042/52005.

Pérez-Guillén, C., et al. (2022). Data-driven automated predictions of the avalanche danger level for dry-snow conditions in switzerland. *Natural Hazards and Earth System Sciences*, *22*(6), 2031–2056, doi:10.5194/nhess-22-2031-2022.

Perla, R., Beck, T., & Cheng, T. (1982). The shear strength index of alpine snow. *Cold Regions Science and Technology*, *6*(1), 11–20, doi:10.1016/0165-232X(82)90040-4.

Podolskiy, E., Chambon, G., Naaim, M., & Gaume, J. (2013). A review of finite-element modelling in snow mechanics. *Journal of Glaciology*, *59*(218), 1189–1201, doi:10.3189/2013JoG13J121.

Pozdnoukhov, A., Matasci, G., Kanevski, M., & Purves, R. S. (2011). Spatio-temporal avalanche forecasting with support vector machines. *Natural Hazards and Earth System Sciences*, *11*(2), 367–382, doi:10.5194/nhess-11-367-2011.

Pozdnoukhov, A., Purves, R., & Kanevski, M. (2008). Applying machine learning methods to avalanche forecasting. *Annals of Glaciology*, *49*, 107–113, doi:10.3189/172756408787814870.

Proksch, M., Rutter, N., Fierz, C., & Schneebeli, M. (2016). Intercomparison of snow density measurements: bias, precision, and vertical resolution. *The Cryosphere*, *10*(1), 371–384, doi:10.5194/tc-10-371-2016.

Purves, R., Morrison, K., Moss, G., & Wright, D. (2003). Nearest neighbours for avalanche forecasting in scotland—development, verification and optimisation of a model. *Cold Regions Science and Technology*, *37*(3), 343–355, doi:10.1016/s0165-232x(03)00075-2.

Puzrin, A. M., Faug, T., & Einav, I. (2019). The mechanism of delayed release in earthquake-induced avalanches. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *475*(2227), 20190092, doi:10.1098/rspa.2019.0092.

Pörtner, H.-O., et al. (2019). Ipcc special report on the ocean and cryosphere in a changing climate. *Cambridge University Press*, doi:10.1017/9781009157964.001.

Reisinger, A., et al. (2020). The concept of risk in the ipcc sixth assessment report: A summary of cross-working group discussions. *Intergovernmental Panel on Climate Change*, (pp.1̃5).

Reiweger, I. & Schweizer, J. (2010). Failure of a layer of buried surface hoar. *Geophysical Research Letters*, *37*(24), doi:10.1029/2010gl045433.

Reuter, B. & Bellaire, S. (2018). On combining snow cover and snow instability modelling. In *Proceedings of the 2018 International Snow Science Workshop, Innsbruck, Austria.*

Reuter, B., Hagenmuller, P., & Eckert, N. (2022a). Simulating avalanche problem types to assess avalanche climate zones in the french alps. In *EGU General Assembly*: Copernicus GmbH.

Reuter, B., Proksch, M., Löwe, H., Van Herwijnen, A., & Schweizer, J. (2019). Comparing measurements of snow mechanical properties relevant for slab avalanche release. *Journal of Glaciology*, *65*(249), 55–67, doi:10.1017/jog.2018.93.

Reuter, B. & Schweizer, J. (2018). Describing snow instability by failure initiation, crack propagation, and slab tensile support. *Geophysical Research Letters*, *45*, 7019–7027, doi:10.1029/2018GL078069.

Reuter, B., Schweizer, J., & van Herwijnen, A. (2015). A process-based approach to estimate point snow instability. *The Cryosphere*, *9*(3), 837–847, doi:10.5194/tc-9-837-2015.

Reuter, B., Viallon-Galinier, L., Horton, S., van Herwijnen, A., Mayer, S., Hagenmuller, P., & Morin, S. (2022b). Characterizing snow instability with avalanche problem types derived from snow cover simulations. *Cold Regions Science and Technology*, *194*, 103462, doi:10.1016/j.coldregions.2021.103462.

Richter, B., Schweizer, J., Rotach, M. W., & van Herwijnen, A. (2019). Validating modeled critical crack length for crack propagation in the snow cover model snowpack. *The Cryosphere*, *13*(12), 3353–3366, doi:10.5194/tc-13-3353-2019.

Roch, A. (1966a). Les déclenchements d'avalanche. *IAHS Publ.*, *69*, 86–99.

Roch, A. (1966b). Les variations de la résistance de la neige. *IAHS Publ.*, *69*, 86–99.

Rosendahl, P. L. & Weißgraeber, P. (2019a). Modeling snow slab avalanches caused by weak layer failure – part i: Slabs on compliant and collapsible weak layers. *The Cryosphere Discussions*, *2019*, 1–28, doi:10.5194/tc-2019-86.

Rosendahl, P. L. & Weißgraeber, P. (2019b). Modeling snow slab avalanches caused by weak layer failure – part ii: Coupled mixed-mode criterion for skier-triggered anticracks. *The Cryosphere Discussions*, *2019*, 1–25, doi:10.5194/tc-2019-87.

Roux, E. L., Evin, G., Eckert, N., Blanchet, J., & Morin, S. (2020). Non-stationary extreme value analysis of ground snow loads in the french alps: a comparison with building standards. *Natural Hazards and Earth System Sciences*, *20*(11), 2961–2977, doi:10.5194/nhess-20-2961-2020.

Roux, E. L., Evin, G., Eckert, N., Blanchet, J., & Morin, S. (2021). Elevation-dependent trends in extreme snowfall in the french alps from 1959 to 2019. *The Cryosphere*, *15*(9), 4335–4356, doi:10.5194/tc-15-4335-2021.

Rubin, M. J., Camp, T., Herwijnen, A. V., & Schweizer, J. (2012). Automatically detecting avalanche events in passive seismic data. In *2012 11th International Conference on Machine Learning and Applications*: IEEE.

Saloranta, T. M. (2012). Simulating snow maps for norway: description and statistical evaluation of the seNorge snow model. *The Cryosphere*, *6*(6), 1323–1337, doi:10.5194/tc-6-1323-2012.

Scapozza, C. (2004). *Entwicklung eines dichte-und temperaturabhängigen Stoffgesetzes zur Beschreibung des visko-elastischen Verhaltens von Schnee*. PhD thesis, ETH Zurich.

Schirmer, M., Lehning, M., & Schweizer, J. (2009). Statistical forecasting of regional avalanche danger using simulated snow-cover data. *Journal of Glaciology*, *55*(193), 761–768, doi:10.3189/002214309790152429.

Schweizer, J. (1999). Review of dry snow slab avalanche release. *Cold Regions Science and Technology*, *30*(1-3), 43–57, doi:10.1016/s0165-232x(99)00025-7.

Schweizer, J. (2017). On recent advances in avalanche research. *Cold Regions Science and Technology*, *144*, 1–5, doi:10.1016/j.coldregions.2017.10.014. International Snow Science Workshop 2016 Breckenridge.

Schweizer, J., Bellaire, S., Fierz, C., Lehning, M., & Pielmeier, C. (2006). Evaluating and improving the stability predictions of the snow cover model snowpack. *Cold Regions Science and Technology*, *46*(1), 52–59, doi:10.1016/j.coldregions.2006.05.007.

Schweizer, J., Bruce Jamieson, J., & Schneebeli, M. (2003). Snow avalanche formation. *Reviews of Geophysics*, *41*(4), doi:10.1029/2002RG000123.

Schweizer, J. & Föhn, P. M. B. (1996). Avalanche forecasting — an expert system approach. *Journal of Glaciology*, *42*(141), 318–332, doi:10.3189/s0022143000004172.

Schweizer, J. & Jamieson, J. B. (2007). A threshold sum approach to stability evaluation of manual snow profiles. *Cold Regions Science and Technology*, *47*(1-2), 50–59, doi:10.1016/j.coldregions.2006.08.011.

Schweizer, J. & Jamieson, J. B. (2010). Snowpack tests for assessing snow-slope instability. *Annals of Glaciology*, *51*(54), 187–194, doi:10.3189/172756410791386652.

Schweizer, J., Kronholm, K., Jamieson, J. B., & Birkeland, K. W. (2008a). Review of spatial variability of snowpack properties and its importance for avalanche formation. *Cold Regions Science and Technology*, *51*(2-3), 253–272, doi:10.1016/j.coldregions.2007.04.009. International Snow Science Workshop (ISSW) 2006.

Schweizer, J. & Lütschg, M. (2001). Characteristics of human-triggered avalanches. *Cold Regions Science and Technology*, *33*(2-3), 147–162, doi:10.1016/s0165-232x(01)00037-4.

Schweizer, J., Michot, G., & Kirchner, H. O. (2004). On the fracture toughness of snow. *Annals of Glaciology*, *38*, 1–8, doi:10.3189/172756404781814906.

Schweizer, J., Mitterer, C., Reuter, B., & Techel, F. (2021). Avalanche danger level characteristics from field observations of snow instability. *The Cryosphere*, *15*(7), 3293–3315, doi:10.5194/tc-15-3293-2021.

Schweizer, J., Mitterer, C., & Stoffel, L. (2008b). Determining the critical new snow depth for a destructive avalanche by con-sidering the return period. In *International Snow Science Workshop*.

Schweizer, J., Mitterer, C., & Stoffel, L. (2009). On forecasting large and infrequent snow avalanches. *Cold Regions Science and Technology*, *59*(2-3), 234–241, doi:10.1016/j.coldregions.2009.01.006.

Schweizer, J., Mitterer, C., Techel, F., Stoffel, A., & Reuter, B. (2020). On the relation between avalanche occurrence and avalanche danger level. *The Cryosphere*, *14*(2), 737–750, doi:10.5194/tc-14-737-2020.

Schweizer, J. & Reuter, B. (2015). A new index combining weak layer and slab properties for snow instability prediction. *Natural Hazards and Earth System Sciences*, *15*(1), 109–118, doi:10.5194/nhess-15-109-2015.

Schweizer, J., Reuter, B., Van Herwijnen, A., & Gaume, J. (2016a). Avalanche release 101. In *Proceedings ISSW* (pp. 1–11).

Schweizer, J., Reuter, B., van Herwijnen, A., Richter, B., & Gaume, J. (2016b). Temporal evolution of crack propagation propensity in snow in relation to slab and weak layer properties. *The Cryosphere*, *10*(6), 2637–2653, doi:10.5194/tc-10-2637-2016.

Schweizer, J., van Herwijnen, A., & Reuter, B. (2011). Measurements of weak layer fracture energy. *Cold Regions Science and Technology*, *69*(2), 139–144, doi:10.1016/j.coldregions.2011.06.004. International Snow Science Workshop 2010 Lake Tahoe.

Schweizer, J. & Wiesinger, T. (2001). Snow profile interpretation for stability evaluation. *Cold Regions Science and Technology*, *33*(2), 179–188, doi:10.1016/S0165-232X(01)00036-2.

Shapiro, L., Johnson, J., Sturm, M., & Blaisdell, G. (1997). *Snow Mechanics: Review of the State of Knowledge and Applications*. Technical report, Defense technical information center.

Sielenou, P. D., et al. (2021). Combining random forests and class-balancing to discriminate between three classes of avalanche activity in the french alps. *Cold Regions Science and Technology*, (pp. 103276)., doi:10.1016/j.coldregions.2021.103276.

Sigrist, C. (2006). *Measurement of fracture mechanical properties of snow and application to dry snow slab avalanche release*. PhD thesis, ETH Zurich.

Sigrist, C. & Schweizer, J. (2007). Critical energy release rates of weak snowpack layers determined in field experiments. *Geophysical Research Letters*, *34*(3), doi:10.1029/2006GL028576.

Sigrist, C., Schweizer, J., Schindler, H.-J., & Dual, J. (2006). The energy release rate of mode ii fractures in layered snow samples. *International Journal of Fracture*, *139*(3), 461–475, doi:10.1007/s10704-006-6580-9.

Simenhois, R. & Birkeland, K. W. (2009). The extended column test: Test effectiveness, spatial variability, and comparison with the propagation saw test. *Cold Regions Science and Technology*, *59*(2), 210–216, doi:10.1016/j.coldregions.2009.04.001. International Snow Science Workshop (ISSW) 2008.

Simson, A., Löwe, H., & Kowalski, J. (2021). Elements of future snowpack modeling – part 2: A modular and extendable eulerian–lagrangian numerical scheme for coupled transport, phase changes and settling processes. *The Cryosphere Discussions*, *2021*, 1–33, doi:10.5194/tc-2021-73.

Singh, A. & Ganju, A. (2008). Artificial neural networks for snow avalanche forecasting in indian himalaya. In *Proceedings of 12th International Conference of International Association for Computer Methods and Advances in Geomechanics, IACMAG*, volume 16.

Smith, J. L. (1965). *The elastic constants, strength and density of Greenland snow as determined from measurements of sonic wave velocity*, volume 167. US Army Cold Regions Research & Engineering Laboratory.

Srivastava, P. K., Chandel, C., Mahajan, P., & Pankaj, P. (2016). Prediction of anisotropic elastic properties of snow from its microstructure. *Cold Regions Science and Technology*, *125*, 85–100, doi:10.1016/j.coldregions.2016.02.002.

Statham, G., et al. (2010). The north american public avalanche danger scale. In *2010 International Snow Science Workshop* (pp. 117–123).

Statham, G., et al. (2018). A conceptual model of avalanche hazard. *Natural hazards*, *90*(2), 663–691, doi:10.1007/s11069-017-3070-5.

Stethem, C., Jamieson, B., Schaerer, P., Liverman, D., Germain, D., & Walker, S. (2003). Snow avalanche hazard in canada – a review. *Natural Hazards*, *28*(2/3), 487–515, doi:10.1023/a:1022998512227.

Techel, F., et al. (2016). Avalanche fatalities in the european alps: long-term trends and statistics. *Geographica Helvetica*, *71*(2), 147–159, doi:10.5194/gh-71-147-2016.

Thumlert, S. & Jamieson, B. (2014). Stress measurements in the snow cover below localized dynamic loads. *Cold Regions Science and Technology*, *106-107*, 28–35, doi:10.1016/j.coldregions.2014.06.002.

Tichavský, R., Fabiánová, A., Koutroulis, A., & Spálovský, V. (2022). Occasional but severe: Past debris flows and snow avalanches in the helmos mts. (greece) reconstructed from tree-ring records. *Science of The Total Environment*, *848*, 157759, doi:10.1016/j.scitotenv.2022.157759.

Trottet, B., Simenhois, R., Bobillier, G., van Herwijnen, A., Jiang, C., & Gaume, J. (2021). From sub-rayleigh to intersonic crack propagation in snow slab avalanche release. In *EGU General Assembly Conference Abstracts* (pp. EGU21–8253).

van Herwijnen, A., Gaume, J., Bair, E. H., Reuter, B., Birkeland, K. W., & Schweizer, J. (2016). Estimating the effective elastic modulus and specific fracture energy of snowpack layers from field experiments. *Journal of Glaciology*, *62*(236), 997–1007, doi:10.1017/jog.2016.90.

van Herwijnen, A., Heck, M., Richter, B., Sovilla, B., & Techel, F. (2018). When do avalanches release: Investigating time scales in avalanche formation. In *Proceedings, International Snow Science Workshop*.

van Herwijnen, A. & Heierli, J. (2010). A field method for measuring slab stiffness and weak layer fracture energy. In *International Snow Science Workshop ISSW, Lake Tahoe CA, USA*, volume 17 (pp. e22).

van Herwijnen, A. & Jamieson, B. (2007). Fracture character in compression tests. *Cold Regions Science and Technology*, *47*(1), 60–68, doi:10.1016/j.coldregions.2006.08.016. A Selection of papers presented at the International Snow Science Workshop, Jackson Hole, Wyoming, September 19-24, 2004.

van Herwijnen, A. & Schweizer, J. (2011). Monitoring avalanche activity using a seismic sensor. *Cold Regions Science and Technology*, *69*(2-3), 165–176, doi:10.1016/j.coldregions.2011.06.008.

Vernay, M., Lafaysse, M., Hagenmuller, P., Nheili, R., Verfaillie, D., & Morin, S. (2020). The s2m meteorological and snow cover reanalysis in the french mountainous areas (1958 - present), version 2020.2.

Vernay, M., et al. (2022). The s2m meteorological and snow cover reanalysis over the french mountainous areas: description and evaluation (1958–2021). *Earth System Science Data*, *14*(4), 1707–1733, doi:10.5194/essd-14-1707-2022.

Vernay, M., Lafaysse, M., Mérindol, L., Giraud, G., & Morin, S. (2015). Ensemble forecasting of snowpack conditions and avalanche hazard. *Cold Regions Science and Technology*, *120*, 251–262, doi:10.1016/j.coldregions.2015.04.010.

Viallon-Galinier, L., Hagenmuller, P., & Eckert, N. (2022a). Combining snow physics and machine learning to predict avalanche activity: does it help? *The Cryosphere Discussions*, doi:10.5194/tc-2022-108.

Viallon-Galinier, L., Hagenmuller, P., & Lafaysse, M. (2021). Forcing and evaluating detailed snow cover models with stratigraphy observations. *Cold Regions Science and Technology*, *180*, 103163, doi:10.1016/j.coldregions.2020.103163.

Viallon-Galinier, L., Hagenmuller, P., Reuter, B., & Eckert, N. (2022b). Modelling snowpack stability from simulated snow stratigraphy: Summary and implementation examples. *Cold Regions Science and Technology*, *201*, 103596, doi:10.1016/j.coldregions.2022.103596.

Vionnet, V., et al. (2012). The detailed snowpack scheme crocus and its implementation in surfex v7.2. *Geoscientific Model Development*, *5*(3), 773–791, doi:10.5194/gmd-5-773-2012.

Walters, D. J. & Adams, E. E. (2014). Quantifying anisotropy from experimental testing of radiation recrystallized snow layers. *Cold Regions Science and Technology*, *97*, 72–80, doi:10.1016/j.coldregions.2013.09.014.

Wautier, A., Geindreau, C., & Flin, F. (2015). Linking snow microstructure to its macroscopic elastic stiffness tensor: A numerical homogenization method and its application to 3-d images from x-ray tomography. *Geophysical Research Letters*, *42*(19), 8031–8041, doi:10.1002/2015GL065227.

Wever, N., Schmid, L., Heilig, A., Eisen, O., Fierz, C., & Lehning, M. (2015). Verification of the multi-layer snowpack model with different water transport schemes. *The Cryosphere*, *9*(2), 2655–2707, doi:10.5194/tc-9-2271-2015.

Wever, N., Vera Valero, C., & Fierz, C. (2016a). Assessing wet snow avalanche activity using detailed physics based snowpack simulations. *Geophysical Research Letters*, *43*(11), 5732–5740, doi:10.1002/2016GL068428.

Wever, N., Würzer, S., Fierz, C., & Lehning, M. (2016b). Simulating ice layer formation under the presence of preferential flow in layered snowpacks. *The Cryosphere*, *10*(6), 2731–2744, doi:10.5194/tc-10-2731-2016.

Wilhelm, C., Wiesinger, T., & Ammann, W. (2000). *The avalanche winter 1999 in Switzerland-an overview*. Technical report, WSL/SLF Davos.

Zeidler, A. & Jamieson, B. (2004). A nearest-neighbour model for forecasting skier-triggered dry-slab avalanches on persistent weak layers in the columbia mountains, canada. *Annals of Glaciology*, *38*, 166–172, doi:10.3189/172756404781815194.

Zeidler, A. & Jamieson, B. (2006). Refinements of empirical models to forecast the shear strength of persistent weak snow layers part a: Layers of faceted crystals. *Cold Regions Science and Technology*, *44*(3), 194–205, doi:10.1016/j.coldregions.2005.11.005.

Zgheib, T., Giacona, F., Granet-Abisset, A.-M., Morin, S., & Eckert, N. (2020). One and a half century of avalanche risk to settlements in the upper maurienne valley inferred from land cover and socio-environmental changes. *Global Environmental Change*, *65*, 102149, doi:10.1016/j.gloenvcha.2020.102149.

Zgheib, T., Giacona, F., Granet-Abisset, A.-M., Morin, S., Lavigne, A., & Eckert, N. (2022). Spatio-temporal variability of avalanche risk in the french alps. *Regional Environmental Change*, *22*(1), doi:10.1007/s10113-021-01838-3.