Journal of Glaciology

Article

Cite this article: Hendrick M, Techel F, Volpi M, Olevski T, Pérez-Guillén C, Herwijnen A, Schweizer J (2023). Automated prediction of wet-snow avalanche activity in the Swiss Alps. *Journal of Glaciology* **69**(277), 1365–1378. https://doi.org/10.1017/jog.2023.24

Received: 4 November 2022
Revised: 17 March 2023
Accepted: 31 March 2023
First published online: 18 May 2023

Keywords:
Avalanches; snow; snow mechanics

Corresponding author:
Martin Hendrick,
Email: martin.hendrick@slf.ch

cambridge.org/jog

Automated prediction of wet-snow avalanche activity in the Swiss Alps

Martin Hendrick[1], Frank Techel[1], Michele Volpi[2], Tasko Olevski[2], Cristina Pérez-Guillén[1], Alec van Herwijnen[1] and Jürg Schweizer[1]

[1]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland and [2]Swiss Data Science Center, Zurich, Switzerland

Abstract

Wet-snow avalanches are triggered by the infiltration of liquid water which weakens the snowpack. Wet-snow avalanches are among the most destructive avalanches, yet their release mechanism is not sufficiently understood for a process-based prediction model. Therefore, we followed a data-driven approach and developed a random forest model, depending on slope aspect, to predict the local wet-snow avalanche activity at the locations of 124 automated weather stations distributed throughout the Swiss Alps. The input variables were the snow and weather data recorded by the stations over the past 20 years. The target variable was based on manual observations over the same 20-year period. To filter out erroneous reports, we defined the days with wet-snow avalanches in a stringent manner, selecting only the most extreme active or inactive days, which reduced the size of the dataset but increased the reliability of the target variable. The model was trained with weather variables and variables computed from simulated snow stratigraphy in 38° slopes facing the 4 cardinal directions. While model development and validation were done in nowcast mode, we also studied model performance in 24-hour forecast mode by using input variables computed from a numerical weather prediction (NWP) model. Overall, the performance was good in both nowcast and forecast mode (f1-score around 0.8). To assess model performance beyond the stringent definition of wet-snow avalanche days, we compared model predictions to wet-snow avalanche activity over the entire Swiss Alps, based on the raw data over 8 winters. We obtained a Spearman correlation coefficient of 0.71. Hence, our model represents a step toward the application of support tools in operational wet-snow avalanche forecasting.

1. Introduction

Liquid water from snow melt or rain modifies the properties of the snowpack. Both surficial wetting and infiltration of liquid water may reduce snowpack stability and lead to the release of wet-snow avalanches on inclined, snow-covered slopes (e.g. Kattelmann, 1985; Conway and Raymond, 1993). Despite the high damage potential of wet-snow avalanches, their formation mechanisms are less well understood compared to dry-snow avalanches (e.g. Baggi and Schweizer, 2009). This is partly due to a lack of in-situ observations such as reliable measurements of liquid water content (LWC) in the snowpack and to the largely unknown influence of water content on mechanical properties such as strength. Moreover, wet-snow instability is a highly transient and spatially-variable phenomenon related to the water transport in snow (Schneebeli, 2004). Once the snowpack becomes partly wet, the probability of avalanche release may rapidly increase (e.g. Conway and Raymond, 1993), but determining the peak and end of a period of high wet-snow avalanche activity is particularly difficult (Techel and Pielmeier, 2009).

The lack of process-based models motivated different data-based approaches based on weather and snowpack properties (Romig and others, 2004; Baggi and Schweizer, 2009; Helbig and others, 2015) and simulated LWC (Mitterer and others, 2013; Wever and others, 2016; Bellaire and others, 2017). While air temperature was clearly related to wet-snow instability (Peitzsch and others, 2021), it is a poor predictor on its own (Romig and others, 2004; Baggi and Schweizer, 2009) and causes many false alarms. As infiltrating meltwater is a prime driver of wet-snow instability, the energy balance at the snow surface needs to be considered, ideally complemented with the LWC in the snowpack. Hence, Mitterer and Schweizer (2013) considered melt water production at the snow surface using the 1-D numerical snow cover model SNOWPACK (Bartelt and Lehning, 2002). A simple proxy of water infiltration is the mean LWC of the snowpack, which motivated the introduction of the liquid water content index (LWC_{index}) (Mitterer and others, 2013; Wever and others, 2016; Bellaire and others, 2017). While indices based on LWC thresholds are generally effective in detecting the onset of wet-snow avalanche cycles, as these coincide with rapid increases in LWC, such indices are not well-suited to predict the end of avalanche periods. The transient nature of wet-snow avalanching was highlighted in several studies (e.g. Durand and others, 1999; Techel and others, 2011; Wever and others, 2016). The notion of 'first wetting' is therefore often used in the context of the onset of wet-snow avalanching and is a term best associated with the first wetting of previously dry, weak layers in the snowpack.

The use of downscaled NWP output variables for probabilistic wet-snow avalanche forecasting has been considered by, for example, Helbig and others (2015) and Bellaire and others (2017). In particular, Bellaire and others (2017) pointed out discrepancies between measured and forecast weather variables and consequently snow related properties, and developed mitigation strategies to deal with these differences. They adapted the LWC-based thresholds above which a day is considered as a wet-snow avalanche day depending on the nowcast or forecast nature of the input data.

More generally, and not restricted to the topic of wet-snow avalanches, statistical avalanche forecasting has been widely studied. A comprehensive overview was recently presented in Sielenou and others (2021). In the last few years, machine learning methods have been increasingly used in data-based approaches. In particular, various random forest (RF) models (ensemble of decision trees) have been developed to support snow avalanche forecasting. Among others, Pérez-Guillén and others (2022) successfully set up a RF model based on weather and snowpack variables to predict the avalanche danger level (EAWS, 2022a), Mayer and others (2022) implemented a RF model to assess snowpack stability, and Viallon-Galinier and others (2022) developed a RF model based on avalanche observations to predict avalanche activity. Machine learning methods have also been implemented for modeling avalanche debris susceptibility (e.g. Choubin and others, 2020).

In line with the concept of considering different avalanche problem types (Statham and others, 2018), we developed a data-driven model specifically for predicting wet-snow avalanche activity. We trained a RF model based on a dataset of more than 20 years of weather and snow measurements carried out by automated weather stations (AWS) covering the Swiss Alps. These measurements are used to simulate the snowpack for north-, east-, south- and west-facing slopes, providing crucial and temporally highly-resolved, information about changes in LWC and its impact on snowpack layering. Based on manual observations of wet-snow avalanche activity in the Swiss Alps, we defined a binary target variable, avalanche day and non-avalanche day and evaluated model performance in both nowcast and 24-hour forecast modes.

2. Data

In the following, we describe the avalanche observation data (Section 2.1), the weather and snowpack input variables (Section 2.2) and how avalanche observations (target) and input variables were linked (Section 2.3).

2.1 Avalanche data

Throughout the Swiss Alps, about 80 trained observers report avalanches in their area to the national avalanche warning service at the WSL Institute for Snow and Avalanche Research SLF, which is responsible for issuing the avalanche warnings for all regions in the Swiss Alps and Jura mountains. Observations are made on a daily basis during the forecasting season (typically from December to April). Avalanche observations include, among other parameters, an approximate location and the date of the avalanche release - and sometimes the time, the elevation and slope aspect of the release area, the LWC of the snow in the release area of the avalanche (dry or wet) (SLF, 2020), and the size of the avalanche, which is estimated according to the European avalanche size classification ranging from 1 (small) to 5 (extremely large) (Table 1; EAWS, 2022b). In most cases, the exact time of release is unknown and estimated, as the actual avalanche release was not observed. Avalanches can be reported as an individual avalanche event, or by reporting several avalanches in an avalanche summary report. If no recent avalanches were observed, this is also reported.

Table 1. Avalanche size s according to EAWS (2022b) and weight w used for the calculation of the avalanche activity index (Eqn. (2))

Size (s)	Label	Typical volume (m^3)	Weight (w)
1	small	100	0.01
2	medium	1 000	0.1
3	large	10 000	1
4	very large	100 000	10
5	extremely large	>100 000	10

The avalanche data base includes avalanche records since 2001. During this time, the reporting system changed once. Initially, avalanche records used to be linked to the place of observation rather than the actual location of the avalanche release (IFKIS platform until summer 2019; Bründl and others, 2004). Today the reporting system is map-based (SLFPro, since Oct 2019). A further important difference between previous and present recording standards concerns the way avalanche summaries are reported. Before October 2019, the summary form permitted to report both dry- and wet-snow avalanches in the same report. As a consequence, in case of a summary report, it is not possible to unambiguously extract the number of avalanches of a certain size and liquid water content, nor the aspects and elevation ranges where these avalanches released from the data base (see also Mitterer and Schweizer, 2013).

For this study, we considered avalanches that released between 1 December and 30 April.

2.2 Weather and snowpack data

We built the training and test dataset based on data from AWS used by the operational avalanche warning service in Switzerland (Section 2.2.1). We also investigated the influence of using NWP output as input data for wet-snow avalanche predictions. NWP data are briefly described in Section 2.2.2. Further details about AWS and NWP setups are provided in Pérez-Guillén and others (2022).

2.2.1 Measurement-based SNOWPACK simulations ('nowcast')

Snowpack simulations for the purpose of avalanche forecasting for the Swiss Alps use a station-based setup (Lehning and others, 1999; Morin and others, 2020). Currently, the network of automatic weather stations (AWS) of the Intercantonal Measurement and Information System (IMIS) network (Fig. 1) consists of 186 AWS, located at the elevations of potential avalanche starting zones. We only used, for our work, the 124 stations that measure snow properties in addition to weather properties. The measurements are recorded every 30 minutes (Fig. 1; see Appendix A for a list of parameters measured). These measurements are used to drive the simulation of snow stratigraphy with the 1-D snow cover model SNOWPACK (Bartelt and Lehning, 2002). We used the operational SNOWPACK setting (Snowpack DEFAULT version 20210925.e77eeeb; Wever and others, 2014; Morin and others, 2020) where liquid water percolation is modeled with a bucket-type approach (Lehning and others, 1999; Vionnet and others, 2012; Lafaysse and others, 2017). Alternatively, water transport can be modeled with the Richards' equation. However, this transport scheme is more demanding and not used in the operational setting for avalanche forecasting in Switzerland.

Stratigraphic profiles are simulated every 3 hours at the location of each station, including simulations of the snowpack for

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

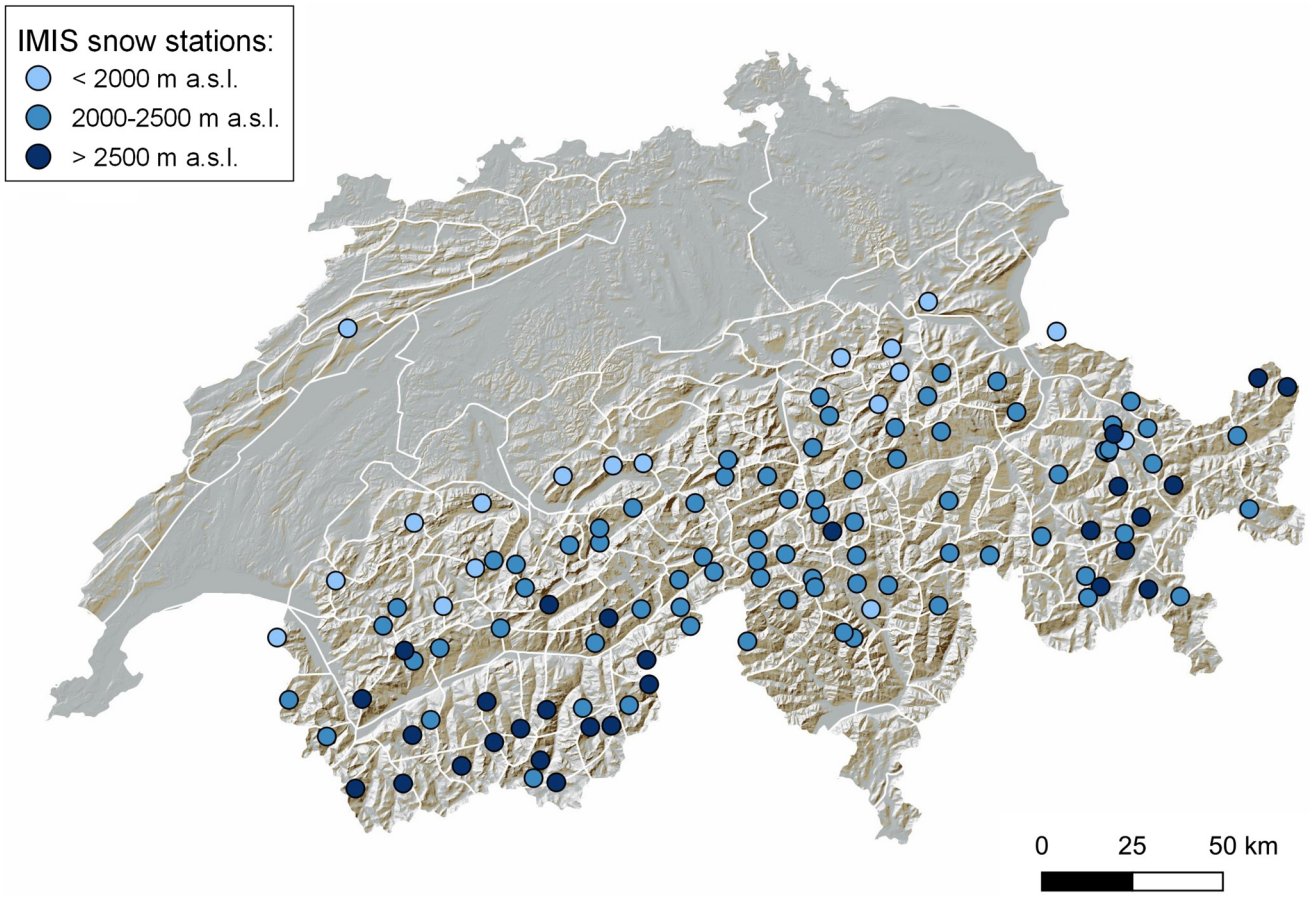

Figure 1. Location and elevation of the 124 automated weather stations (IMIS network) that measure snow properties.

four virtual slopes (aspects *asp*: N, E, S, W) with an incline of 38° (Morin and others, 2020), resulting in 8 profiles per day and aspect.

If not stated otherwise, we always refer to this nowcast setup.

2.2.2 NWP model-based SNOWPACK simulations ('24-hour forecast')

For the purpose of operational avalanche forecasting, SNOWPACK simulations are also driven with output from the NWP model COSMO (Consortium for Small-scale Modeling; Cosmo 2021) allowing the simulation of snowpack stratigraphy for the following 30 hours. To this end, the COSMO-1 model output with 1 km resolution is downscaled to the local topography, referred to as COSMO-OSHD in the following (Griessinger and others, 2019). As for the nowcast setup, the snow stratigraphy is simulated for four virtual slopes every 3 hours. This downscaling setup provides more realistic simulations of the snowpack properties than driving SNOWPACK directly with the COSMO model output of the grid point nearest to the AWS (Bellaire and others, 2017).

The NWP predictions and the simulated profiles are not stored but overwritten every 3 hours. Therefore, a historic forecast dataset is not available for model development.

2.2.3 Variable definition

From the snowpack simulations, we extracted weather parameters and properties describing the state of the snowpack. As the temporal resolution of the avalanche observations was often restricted to the release day rather than an exact time, we derived 24 h averaged weather and snowpack features (00:00–24:00 local time, LT), as well as weather and snowpack properties corresponding to the time of the wettest profile of the day (Fig. 2). We defined the wettest profile as the profile with the highest LWC index (LWC_{index}; described below); in most cases, the profile at 15:00 LT was the wettest. If the LWC_{index} was zero for all time-steps, we used the data from the simulated profile at 15:00 LT. We chose this setup to account for daily weather and snowpack properties and to have a snapshot of the situation, when the snowpack was wettest (similar to Wever and others, 2018). In addition, we computed the differences to the values describing snowpack properties 1, 2 and 3 days before. The complete list of input features used for model development is reported in Appendix A. LWC-related features used were mostly based on Wever and others (2018). We extracted in total 146 of features (Tables 5 and 6). Those included 36 features related to LWC, 9 of them were extracted from the wettest profile of the day and the remaining 27 describe changes in LWC over the last 1, 2 and 3 days. The other 110 features were meteorological and snowpack variables describing either daily averages (flagged with the subscript daily) or actual values at the time of the wettest profile (see Fig. 2).

From the simulated profiles, the LWC of each layer can be derived. It allows to compute the so-called LWC index, LWC_{index}, a feature that has been suggested to predict wet-snow avalanches (Mitterer and others, 2013). The index is the average of the LWC, in percent, of the entire snowpack weighted by layer height and divided by 3 %, *i.e.*

$$\mathrm{LWC}_{\mathrm{index}} = \frac{1}{3}\frac{\sum_{\mathrm{layer}} h(\mathrm{layer})\mathrm{LWC}(\mathrm{layer})}{\mathrm{HS}} \quad (1)$$

where the sum carried out over all snowpack layers, with $\mathrm{HS} = \sum_{\mathrm{layer}} h(\mathrm{layer})$ the snowpack height, h the layer thickness

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

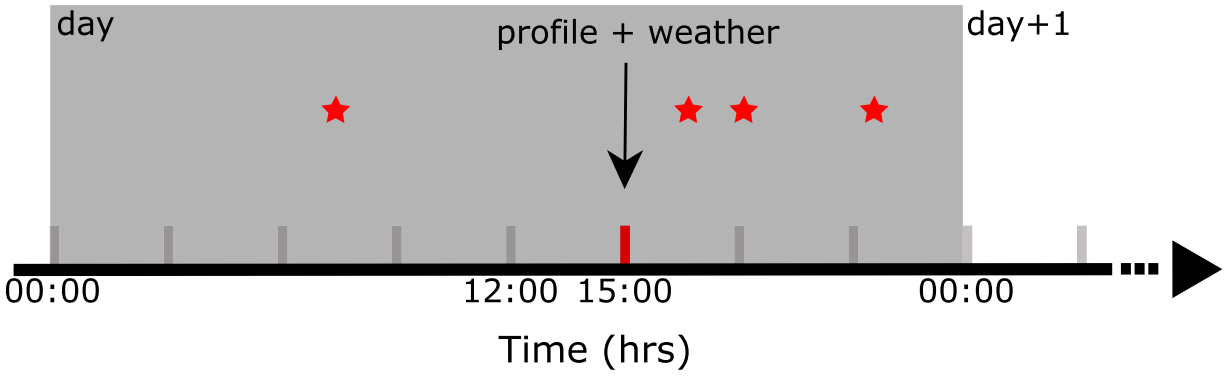

Figure 2. Temporal characteristics of the input features: 24 h averaged weather features (gray area), weather and snowpack features at the time of the wettest profile (for example at 15:00 LT). Red stars indicate observed wet-snow avalanches.

and LWC the LWC in percent volume. Above a LWC of 3 %, it is assumed that excess water drains to the layers below since it is no longer held between grains by capillary forces, i.e. the threshold of 3 % corresponds to the transition from the pendular to the funicular regime. Therefore, $LWC_{index} \geq 1$ may indicate a particular loss of strength (Mitterer and others, 2013), or a snowpack wet throughout.

2.3 Linking avalanche observations to SNOWPACK simulations

The occurrence of wet-snow avalanches depends on the snowpack conditions, which may differ strongly between slope aspect and elevation (Mitterer and Schweizer, 2013; Wever and others, 2018). Therefore, we linked an observed avalanche in the vicinity of a station *st* and SNOWPACK simulation for aspect *asp* if:

1. The avalanche was classified as a wet-snow avalanche, a size was estimated and the aspect and elevation were reported.
2. The avalanche released within an elevation band of ±250 m of the elevation of the station, H_{st}. In case of avalanche summary reports indicating an elevation range (av_{elev}.min-max), avalanches were assigned to this station relative to the overlapping proportion between this elevation band (H_{st} ± 250 m) and the elevation band indicated in the avalanche summary report. For example, if H_{st} was at 2000 m (± 250 m: 1750 to 2250 m), and av_{elev}.min-max ranged from 2000 to 2500 m, the overlapping elevation band is 2000 to 2250 m, which is 50 % of the indicated av_{elev}.min-max. In this case, each of the avalanches in the summary report was weighted with 0.5.
3. The avalanche was within a distance of at most 8.9 km, 17.8 km and 39.9 km corresponding to circle-shaped areas around the location of the station of 250 km^2, 1000 km^2 and 5000 km^2.
4. The avalanche released in the same slope aspect as the simulation was performed for. For example, a wet-snow avalanche observed on a north-facing slope was only linked to the simulated stratigraphy on the virtual slope of northern aspect. If an avalanche was observed on a northeast-facing slope, this avalanche was assigned with half its weight *w* to both northern and eastern aspect slope simulations.

Figure 3 illustrates the matching requirements 2 to 4.

The avalanche summary reports prior to October 2019 often did not allow to extract the number and size of avalanches by aspect, as dry-snow and wet-snow avalanches could be reported in the same form. Therefore, we could rarely use these data in our analysis.

To quantify avalanche activity in a region, the avalanche activity index (AAI) was used (Schweizer and others, 2003). For each station *st*, we determined the AAI for several spatial scales *a* (area) for the four slope aspects *asp* as:

$$AAI_{st}^{asp}(a) = \sum_{s=1}^{5} w(s) N_s^{asp}(a), \quad (2)$$

with avalanche size $s \in \{1, 2, 3, 4, 5\}$, weight $w(s)$ as defined in Table 1, the number of avalanches of size *s*, $N_s^{asp}(a)$ reported in the area $a \in \{250, 1000, 5000\ km^2\}$ and slope aspect $asp \in \{north, east, south, west\}$. The computation of AAI was constrained by restricting it to a specific elevation band around a station: ±250 m.

3. Random-forest model

3.1 Definition of the target variable: avalanche day and non-avalanche day

We defined a binary target variable to discriminate days with wet-snow avalanche activity (avalanche days, *AvD*) from days

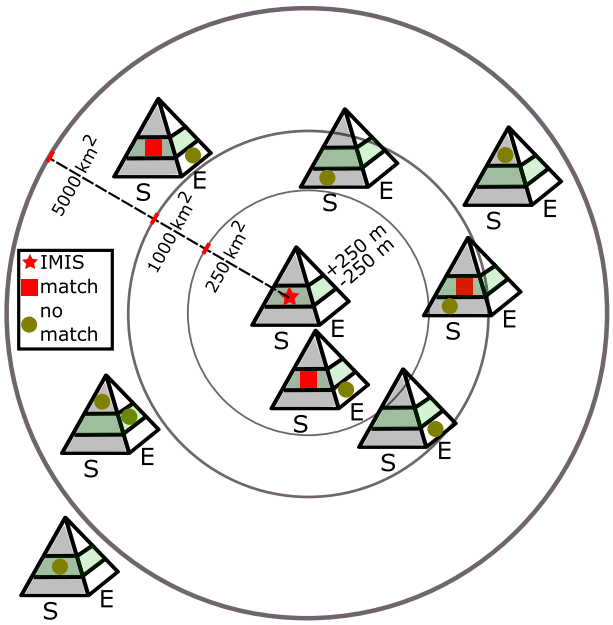

Figure 3. Schematic representation of how observed wet-snow avalanches on south facing slopes (red squares) were linked to input features derived from SNOWPACK simulations on a virtual south-facing slope at the location an AWS (station; red star). The pyramids represent different elevation bands and aspects for each observed avalanche and the virtual slopes at the station. The elevation bands show the ±250 m tolerance band. The circles represent the borders of areas surrounding the station. Green dots represent avalanches not assigned to the station because they did not satisfy the elevation and/or aspect criteria.

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

without any activity (non-avalanche days, *nAvD*). Even though avalanche observations are reported from throughout the Swiss Alps, there are regions where rarely any observations are reported. And even in the areas generally covered by observers, observations are not always possible. Therefore, the absence of an avalanche in the data base does not necessarily indicate that no avalanche occurred. Beside the challenge of detecting days when no wet-snow avalanches occurred, we are aware of two main sources of error or uncertainty relating to avalanche observations:

1. The classification of the avalanche as either a wet-snow or a dry-snow avalanche. According to the observational guidelines (SLF, 2020), the presence of liquid water should be estimated based on the conditions in the release area. However, since the avalanche release area is often not accessible, the LWC of the snow in the release area is often an estimate from the valley floor.
2. The release date. The uncertainty in the reported release date depends on the frequency of observations in an area, on weather conditions (e.g. visibility), but also on the overall quality of the observations provided by a specific observer.

To address these points, which are highly relevant for this study, we opted for an approach allowing us to check the plausibility of the release date and avalanche type classification by considering observations independently made on the same day by observers in neighboring areas. More specifically, we define an avalanche day or non-avalanche day, our target variable Y_{st}^{asp}, associated with data from a station *st* for a given slope aspect *asp* as:

$$Y_{st}^{asp} = \begin{cases} 0 & \text{if } AAI(SwissAlps) = 0 \text{ for all elevations and aspects} \\ 1 & \text{if } AAI_{st}^{asp}(250\,\text{km}^2) \geq 0.1 \;\&\; AAI_{st}^{asp}(1000\,\text{km}^2) \geq 0.4 \;\& \\ & AAI_{st}^{asp}(5000\,\text{km}^2) \geq 2 \;\&\; \text{gap check} \\ \text{NaN} & \text{otherwise,} \end{cases} \tag{3}$$

with NaN 'not a number'. The gap check requirement was $AAI_{st}^{asp}(5000\,\text{km}^2) > AAI_{st}^{asp}(1000\,\text{km}^2)$ & $AAI_{st}^{asp}(1000\,\text{km}^2) > AAI_{st}^{asp}(250\,\text{km}^2)$, ensuring that avalanche activity increased with increasing area and was not a local phenomenon.

The choice for selecting an *AAI* threshold of 0.1 was guided by the fact that an $AAI_{st}^{asp}(250\,\text{km}^2) \geq 0.1$ either means that there was at least one potentially life-threatening size 2 event (Eqn. (2); Table 1) or 10 avalanches of size 1. Moreover, an area *a* of 250 km^2 corresponds approximately to the mean size of the warning regions, the smallest spatial units used for communication of avalanche danger in the Swiss avalanche bulletin (e.g. Techel and others, 2018), and 5000 km^2 is approximately the size of a snow-climate region (e.g. Techel, 2020, p. 50). Within a warning region there is often at most one observer providing regular avalanche observations; within a snow-climatological region there are typically between 5 and 20 observers. With the definition given above, we ensure that local avalanche activity with $AAI_{st}^{asp}(250\,\text{km}^2) \geq 0.1$ is representative of avalanche activity in a larger area. We chose this approach to ensure that the target variable was not a local property, but included several independent observations, thus reducing noise due to erroneous dating or wetness classification. We considered this step as important, as AAI_{st}^{asp} only considers observations within a specific slope aspect and elevation band. We used a similarly strict criterion to define a non-avalanche day (*nAvD*: $Y_{st}^{asp} = 0$) as the uncertainty related to observations of no avalanches is much higher. Thus, a day is considered only as *nAvD* if none of the observers in the Swiss Alps reported a wet-snow avalanche. Furthermore, we considered only *nAvD* for station-aspect combinations that had at least one *AvD* in the same season.

With this definition for the target variable, we wanted to separate days with widespread avalanche activity (*AvD*) of a certain magnitude from days with absolutely no avalanches (*nAvD*), and excluded days with rather local avalanche activity (NaN values). With regard to the development of the model, our strict definition led us to train and test the classifier on rather extreme cases, which are however comparatively reliable in terms of the quality of the class label.

3.2 Labeling and splitting the dataset

Input data were labeled as *AvD* and *nAvD* according to Eqn (3). We split the data into three uncorrelated datasets according to the winter the avalanches occurred (Fig. 4): Dataset1 contained *AvD* ($Y = 1$) and *nAvD* ($Y = 0$) labeled data for the winter 2019–2020 and historical *AvD* events between 2001 and 2014. Dataset2 contained data for winter 2020–2021. Operational tests were carried out during the winter 2021–2022 (dataset3). The dataset characteristics are summarized in Table 2. The stringent manner, we defined the target variable (Eqn. (3)), considerably reduced the number of AvD events. For instance, dataset2 and dataset3 originally contained 2595 and 1696 records with AAI $\geq$ 0.1, respectively. However, only 223 (dataset2) and 97 (dataset3) were labeled as AvD according to our target variable definition (Table 2).

In the following, we name 'test set' the complementary of the training set, for example, if we used dataset1 as the training set, the test was dataset2.

3.3 Random forest model development

We trained a RF model, as it is one of the most common and robust machine learning models to classify tabular data (Breiman, 2001). We used the scikit-learn library in Python (Pedregosa and others, 2011).

The datasets are heavily imbalanced and contain only relatively few *AvD* events (Table 2). As the datasets are also rather small, we did not use a standard approach of splitting the data in a training, a validation and a test set. Instead, we only used a training and a test set.

We relied on the out-of-bag (OOB) score to assess model performance (Probst and others, 2019). Each tree of a RF model was trained separately using a training subset randomly sampled with replacement. These subsets are named bootstrap samples. The left out samples, i.e. samples not contained in the boostrap samples, are named the OOB samples and were used to compute OOB scores, such as the f1-score (see Appendix C for metrics definitions). The OOB samples size was around 1/3 of the size of the training set (Breiman, 2001).

The RF hyperparameters were tuned to maximize the out-of-bag f1-score of class *AvD* of the training set using a grid search (Table 7).

In order to reduce the complexity of the model and remove non-relevant input variables, we applied recursive feature elimination (RFE; Guyon and others, 2002). First, we fitted a model keeping all the original 146 features and computed impurity-based features importance ranking (for implementation details see Breiman, 2017). Then we discarded the less important features, re-fitted the model and repeated the process until only 4 features remained. At each step, we computed the OOB f1-score of *AvD*, which allowed us to select the best model as a function of the number of features. The best performance was obtained using 40 features (Fig. 11). We used this set of features for the remaining of the work.

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

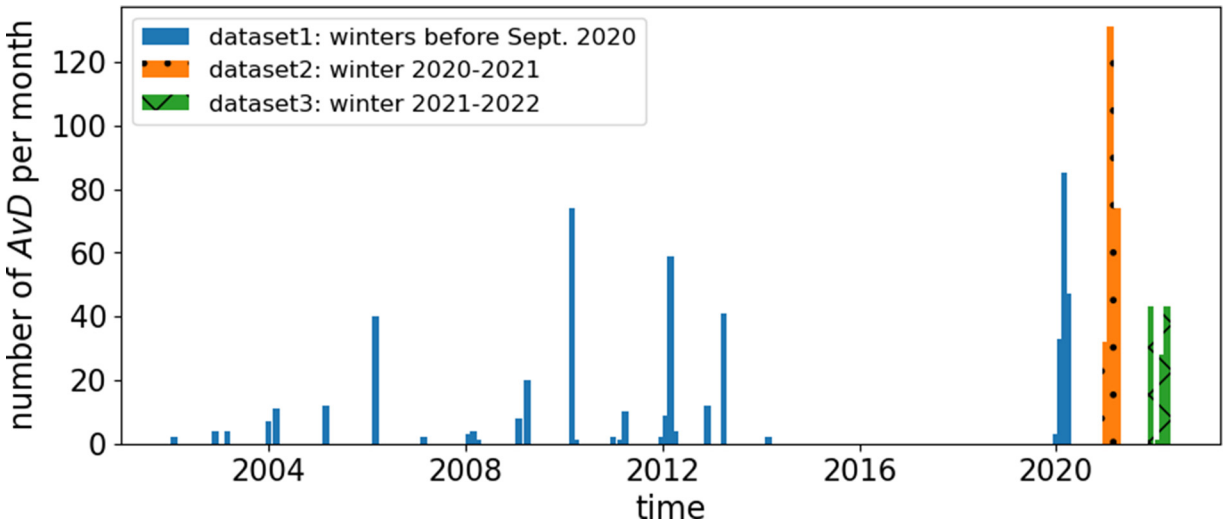

Figure 4. Number of days labeled as *AvD* for dataset1 (blue), dataset2 (orange) and the dataset3 (green).

Table 2. Number of cases in the datasets used for training and testing

	Dataset1 winters before Sept. 2020	Dataset2 winter 2020–2021	Dataset3 winter 2021–2022
nAvD	3322 (172)	5238 (42)	2151 (40)
AvD	413 (60)	223 (22)	97 (12)

In parentheses, the number of individual calendar days are given; for example, during winter 2021–2022 there were 97 *AvD* events on 12 different calendar days.

The final prediction of the RF model is obtained by taking the majority vote of the predictions made by each individual decision tree, and the probability of a particular class, here *AvD* or *nAvD*, is determined by the percentage of trees that have voted for that class. In the following, we used also the *AvD* probability to assess model performance.

4. Results

4.1 Model performance for historic data

We trained three distinct RF models: on dataset1 (winters before Sept. 2020), on dataset2 (winter 2020–2021), and on dataset12 (winters before Sept. 2021). Models are named respectively, RF1, RF2 and RF12. The OOB f1-score of RF12 was highest, while performance metrics of RF2, only trained with winter 2020–2021, were lowest (Table 3). The f1-score of the test sets for RF1 and RF2 were close to the OOB f1-scores. This finding suggests that the trained models did not over-fit. Model performance metrics (precision and recall) for days with no wet-snow avalanche activity (*nAvD*) were all very close to 1 and are not reported in Table 3.

We further evaluated the performance of the models on historical winters. Since slope aspect was not consistently reported prior to the winter 2019-2020 in the IFKIS data base, many winters were not included in the dataset. There were winters without any avalanche event satisfying the definition of the target variable (Eqn. (3)), or when the aspect information was recorded in an inconsistent format (Fig. 4). Nevertheless, for these winters, we could still determine a daily wet-snow avalanche activity for Switzerland, regardless of slope aspect, and compare it with daily average models outputs (all locations, all aspects; Table 4). For example, we showed that the daily mean *AvD* probability correlated well with avalanche activity for all aspects and all stations (Table 4 and Fig. 5a).

Table 3. Model performance for avalanche days (*AvD*) for different models and test sets

Training			Test					
Dataset	Model	OOB f1	Dataset	Mode	p_{thres}	Precision	Recall	f1-score
1	RF1	0.78	2	nowcast	0.5	0.97	0.74	0.84
2	RF2	0.72	1	nowcast	0.5	0.68	0.88	0.77
12	RF12	0.79	3	nowcast	0.5	0.75	0.91	0.82
			3	forecast	0.5	0.90	0.71	0.79
			3	forecast	**0.36**	0.87	0.87	0.87

OOB scores were computed on the training set. Tests were carried out on data not used for training, either dataset1 or dataset2 (Table 2), or the data from the dataset3 (test winter 2021–2022).

Table 4. Model performance computed for winters not included in the training set

	Correlation model/observation		
Winter	RF1	RF2	RF12
2002–2003	0.75	0.72	0.76
2006–2007	0.74	0.66	0.78
2014–2015	0.70	0.62	0.72
2015–2016	0.62	0.52	0.69
2016–2017	0.64	0.52	0.68
2017–2018	0.73	0.70	0.70
2018–2019	0.63	0.62	0.67
2020–2021	0.74	–	–
2021–2022	0.70 (0.67)	0.67 (0.65)	0.71 (0.69)

Spearman correlation coefficient (r_s) between the mean model-predicted *AvD* probability and the wet-snow avalanche activity index, *AAI*(*Swiss Alps*), regardless of elevation and aspect. In addition, correlations are shown for winter 2021–2022 when the model was run in nowcast and forecast mode. The latter is shown in brackets. All p-values are below 10^{-6}.

4.2 Detailed model performance for winter 2021–2022 (nowcast and 24-hour forecast mode)

We tested the RF12 model in a live setting during the winter 2021–2022 in nowcast and in forecast mode, as described in Section 2.2.2.

97 events were labeled as wet-snow avalanche days *AvD* and 2151 events as non-avalanche days *nAvD* (Table 2). Model performance for model RF12 in nowcast and forecast mode are summarized in Fig. 6 and Table 3. The f1-score of the nowcast model was 0.82, similar to the OOB f1-score of the model (0.79; Table 3).

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

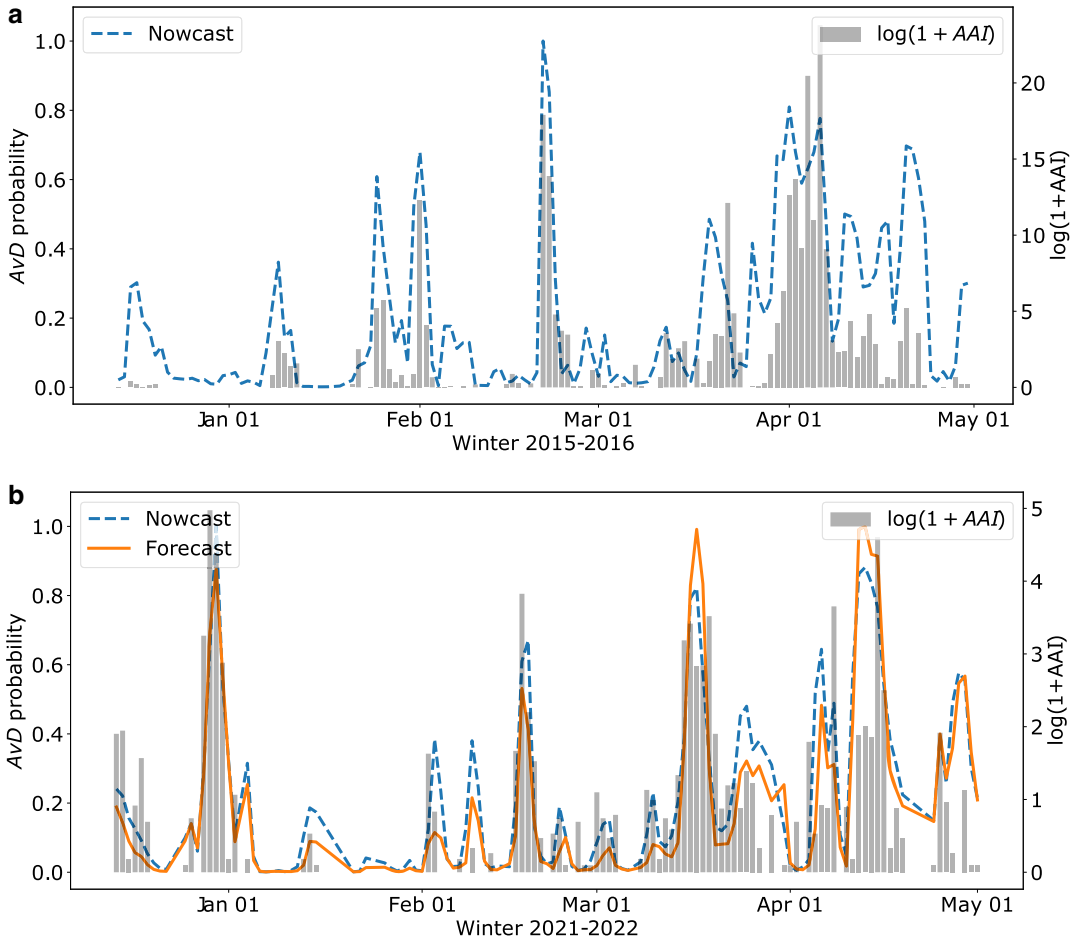

Figure 5. Daily average of (logarithm of) observed wet-snow avalanche activity for the entire Swiss Alps (*AAI*(*SwissAlps*), gray bars) compared to daily averaged probability for AvD (all stations and aspects) provided by the RF12 model for (a) winter 2015-2016 and (b) winter 2021–2022. Correlations are provided in Table 4.

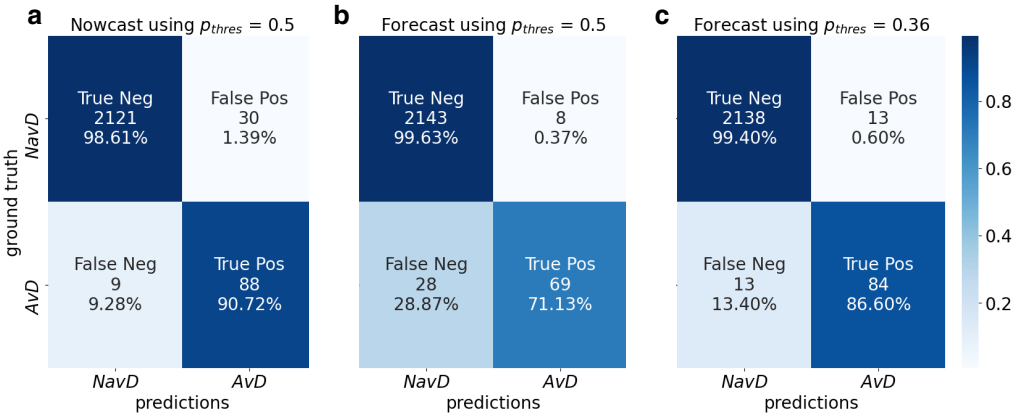

Figure 6. Performance statistics of RF12 model during live testing in winter 2021–2022 for (a) nowcast and (b) forecast mode. (c) shows the respective statistics for the model in forecast mode using an adjusted threshold $p_{\text{thres}} = 0.36$. The proportions indicated in the figure (coloring, percentage values) describe row-wise proportions.

Nowcast and forecast output probabilities were strongly correlated ($r_s = 0.96$, Fig. 5b), yet forecast f1-score was slightly lower compared to the nowcast score due to relatively low value of recall. The model using forecast inputs predicts less *AvD* (77 *AvD*) than using nowcast data (118 *AvD*; Fig. 6a and b). However, large activity was well captured by the model based on forecast input, although predicted probability for *AvD* was generally lower than the probability obtained with nowcast input (Fig. 6). This performance difference is due to discrepancies between nowcast (based on measurements) and forecast (based on NWP model) input data. This difference is clearly noticeable by studying the model behavior in function of p_{thres}, the probability threshold above which *AvD* is defined. The model was trained with nowcast input to maximize the f1-score of *AvD* (Section 3) using a default $p_{thres} = 0.5$. We observed that in nowcast mode, the model performed well with the default threshold of 0.5, *i.e.*

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

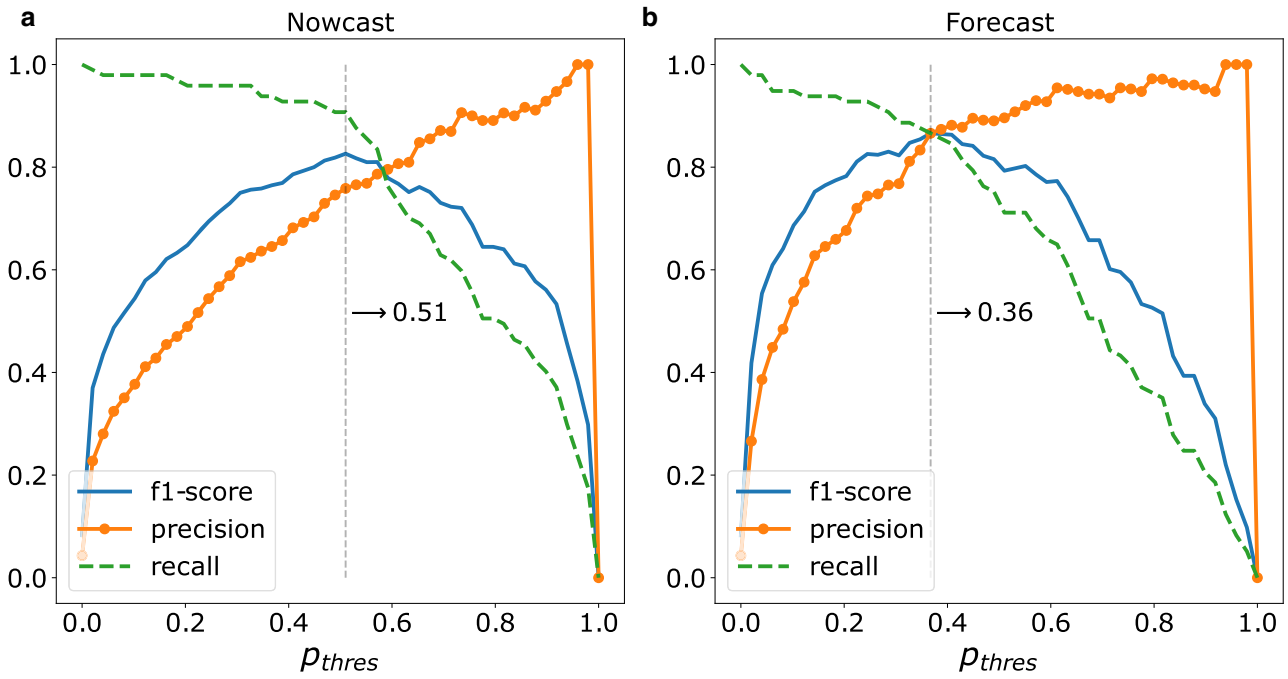

Figure 7. F1-score, precision and recall for *AvD* for different values of p_{thres} for the model using (a) nowcast and (b) forecast data as input. p_{thres} is the probability threshold above which *AvD* is defined. The best f1-score was obtained with $p_{thres} = 0.51$ in nowcast mode and $p_{thres} = 0.36$ in forecast mode.

the maximum of the f1-score of *AvD* was very close to $p_{thres} = 0.5$ (Fig. 7a). However, in forecast mode the maximum f1-score was obtained for $p_{thres} = 0.36$ (Fig. 7a, Table 3).

The target variable was designed to separate avalanche days from non-avalanche days, by reducing the influence of observation errors and, thus, providing robust training labels. However, by doing this, many wet-snow avalanches were discarded and flagged as NaN, such as, for example, an isolated large wet-snow avalanche. To assess model performance without the potential filtering bias introduced by our definition of the target variable, we also compared the aspect-specific mean daily model predictions with all avalanche observations in these slopes. In Fig. 8, the size of observed wet-snow avalanches at different elevations are plotted with the predictions of the RF12 model and the LWC_{index} for the 2021–2022 winter. Overall, the predicted probability of an *AvD* increased at the onset of wet-snow avalanche cycles and decreased afterwards. Moreover, differences between south- and north-facing slopes were generally well reproduced (compare Fig. 8a and 8b). The model performed also well with regards to the maximum elevation at which avalanche activity was observed (Fig. 8). The average absolute difference (mean absolute error) between predicted maximum elevation in both nowcast and forecast mode (using $p_{thres} = 0.36$) and observations, excluding *nAvD*, was around 300 m. Toward the end of the season, the model overestimated wet-snow activity, especially for south-facing slopes (Fig. 8e).

The LWC_{index}, also plotted in Fig. 8, is a parameter that has been suggested to predict the onset of wet-snow avalanche periods (e.g. Mitterer and others, 2013; Bellaire and others, 2017). It has also been used operationally by the Swiss avalanche warning service to estimate snowpack wetting and wet-snow avalanche activity. However, the LWC_{index} generally did not decrease when wet-snow avalanche activity decreased, as observed for south-facing slopes in the second half of April 2022 (Fig. 8). LWC_{index}-based threshold values will therefore lead to many false alarms as can be noticed in Fig. 9a (low precision means high false alarm ratio). While models including the temporal change of LWC_{index} resulted in a better performance (Fig. 9b).

4.3 Feature importance

Impurity-based feature importance (Breiman, 2017) for RF12 model suggests that snow surface conditions and the presence of liquid water in the snowpack were important to predict wet-snow avalanche activity (Fig. 10). The two most important features were related to snow surface temperature (TSS) and sum of the LWC in the top 15 cm (sum_up) of the snowpack. From the 20 most important features, 10 relate to the LWC of the snowpack, with 5 of these features describing changes in LWC over 1, 2 and 3 days (Fig. 10). The set of 20 most important features included only 11 different features, provided that the temporal characteristics of the variables (computed at the time of the wettest profile, daily averaged or temporal changes) were not considered.

5. Discussion

We implemented a RF model to predict wet-snow avalanche days. The model was trained on nowcast data from station measurements. Despite the small size and the highly imbalanced nature of the training datasets, our model reproduced the observed wet-snow avalanche activity rather well (Figs. 5, 6 and 8; Table 4). However, the operational avalanche forecasting process is based on 24-hour NWP data (in Switzerland the COSMO-OSHD data, Section 2.2.2). We therefore also investigated model performance in forecast mode, highlighting some discrepancies. Specifically, because air temperature sensors at the AWSs are not ventilated, there is a significant temperature offset compared to COSMO-OSHD forecast data, especially during sunny and calm afternoons (Bellaire and others, 2017). This impacts the downstream SNOWPACK simulations, in particular also the simulated LWC. Profiles based on NWP data therefore generally appeared dryer than nowcast profiles. Despite this difficulty, our results suggest that the model performed reasonably well in forecast mode (Figs. 5b and 6). A better approach would certainly be to train the model directly with NWP data rather than station data. However, the forecasted weather and snow

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

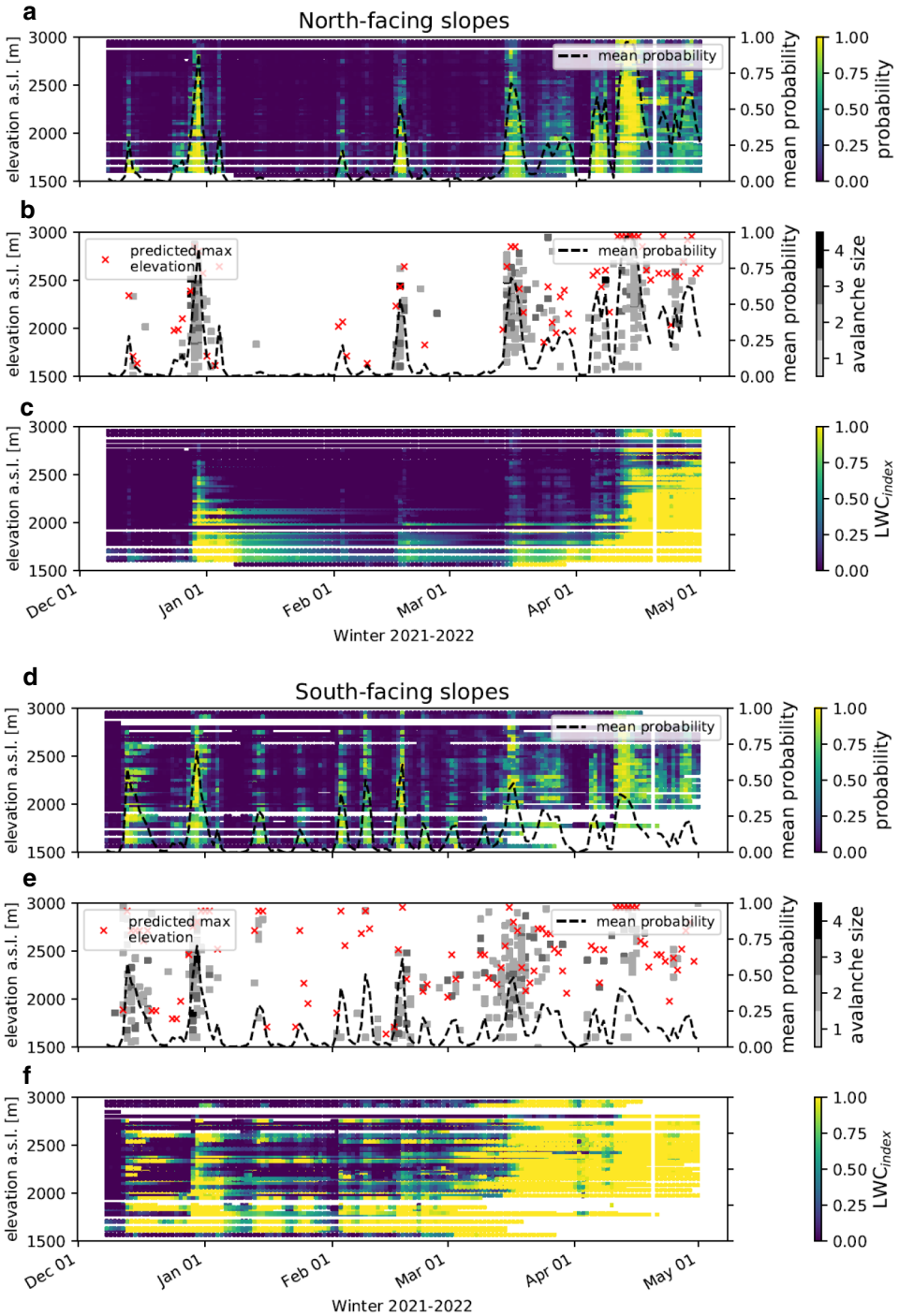

Figure 8. Model predictions, avalanche observations and LWC_{index} by elevation for winter 2021–2022 for north-facing (a-c) and for south-facing slopes (d-f). (a, d) Probability (RF12) for a wet-snow avalanche day (*AvD*) as predicted at the elevation of the stations. (b, e) Observed avalanches (by size) and highest elevation, for which an avalanche day was predicted. The dashed dark line is the daily 'mean probability' for *AvD* (all stations). (c, f) LWC_{index} as simulated at the elevation of the stations. (a, c, d, f) White stripe correspond to missing data due to lack of snow, a defective AWS or an elevation band not covered by the AWS network.

properties produced for the operational warning service in Switzerland are not stored but overwritten at each update.

A direct comparison between previous models (Mitterer and others, 2013; Bellaire and others, 2017) and our RF model is not feasible, though these models also aimed to predict days with wet-snow avalanche activity. Indeed, the definition of the target variable is different for the three approaches. Nevertheless, our RF model performed better than a simple LWC_{index} threshold approach as presented by Bellaire and others (2017) (Fig. 9). Mitterer and others (2013) mainly used weather and energy balance variables as model input. While we also used these variables, we additionally introduced variables characterizing the state and the temporal changes of snowpack conditions. Temporal changes in LWC were ranked as important variables in our model.

5.1 Feature importance

The feature importance of the model provides insight into the main drivers of wet-snow avalanche release, as it was trained without any prior expert knowledge. As expected, the snow surface temperature (TSS) and LWC at the surface (sum_up) were key predictors, as these are closely linked to snow melting, resulting in a higher LWC. Snow depth (HS) changes were also decisive

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

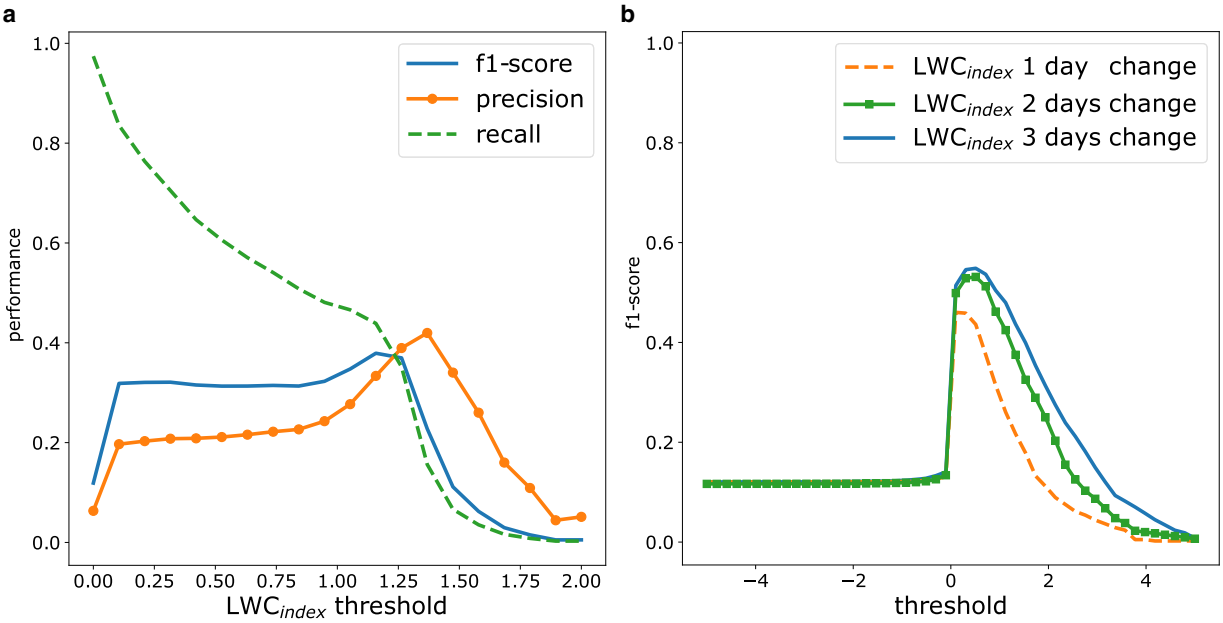

Figure 9. Performance statistics for two nowcast models, where (a) absolute values of the LWC_{index} and (b) temporal changes of LWC_{index} were used to predict avalanche days. A day is considered *AvD* if $LWC_{index} \geq$ a specific threshold (x-axis), *nAvD* otherwise. In (b), negative (resp. positive) threshold values correspond to decreasing (resp. increasing) of LWC_{index}.

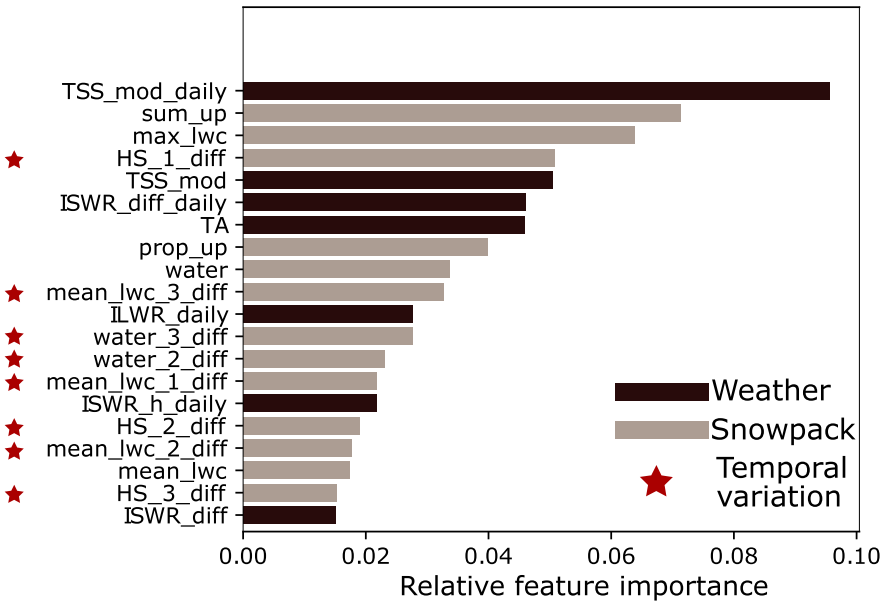

Figure 10. Relative importance of the 20 most important features for model RF12. The subscript '_daily' refers to 24-hour mean values, the remaining variables were derived from the wettest profile. The subscript '_diff' refers to temporal changes, the digit before indicates the number of days considered to calculate the change. Table 5 contains a short descriptions of each features.

for the model as already shown in previous studies (Baggi and Schweizer, 2009). As highlighted in several studies (e.g. Techel and others, 2011; Wever and others, 2016, 2018), temporal changes in the LWC parameters were also crucial. In our model, almost half of the 20 most important features describe temporal changes. This temporal sensitivity of the model allows to reproduce the onset characterized by increases in LWC, but also the end of periods with wet-snow avalanches when LWC decreased or remained constant. The capability of the model to detect the end of a cycle is an important improvement compared to previous LWC_{index}-based thresholds approaches (Mitterer and others, 2013; Bellaire and others, 2017), (Fig. 9). One of the main limitations related to the feature importance ranking scheme we used (the mean decrease impurity algorithm) is that it is biased by the variability of the different wet-snow avalanche situations contained in the training set. For example, there are only few avalanches that were caused by rain. Most avalanches released during dry weather periods when surface melting produced liquid water. Features related to rain were therefore ranked as less important. However, rain increases the LWC of the snowpack, and is therefore well represented by LWC-related input features. Alternative approaches to determine feature importance, such as Permutation Based Feature Importance (Breiman, 2001) or SHAP feature importance (Lundberg and Lee, 2017), may provide different results.

5.2 Target variable definition

One of the strengths of the developed approach is the restrictive definition for avalanche days and non-avalanche days (Eqn. (3)) allowing us to use observational data. Thus, we reduced potential errors related to the release date and avalanche type (wet- or

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

dry-snow). However, this approach came at the cost of substantially reducing the amount of data available to train the model. Consequently, the model was developed likely using a smaller variety of *AvD* and *nAvD* situations. For example, most of the wet-snow avalanche events we have used for training were associated with surface melt rather than rain. Moreover, the definition of the target variable only selected the most 'extreme' *AvD* and *nAvD* days. Being aware of this limitation of our approach, we tested whether the resulting model was oversimplified by carrying out a target-independent check. We compared model output with unfiltered, raw avalanche observations (Figs. 6 and 8; Table 4). By doing so, we returned to the original dataset keeping its intrinsic errors, but allowed for studying model behavior over entire seasons rather than for a few disparate *AvD* and *nAvD*. This 'beyond the target definition' verification was a crucial test that provided, for example, insight into model performance in an operational setting. Observed wet-snow avalanche cycles were well correlated with model output in both nowcast and forecast mode (Fig. 6).

In preliminary tests, we also explored other target variable definitions. For example, among many others, we prototyped a model based only on 'local' constraint i.e. computing *AvD* and *nAvD* from Eqn. (3) only considering the activity in the 250 km^2 AWS surrounding. However, the resulting models were characterized by a high false alarm rate. It is difficult to assess if this was due to intrinsic limitations of the RF model (i.e. difficulties to deal with 'intermediate' events) or because the quality of the input data and/or the target variable were insufficient. Therefore, it might be of interest to study the influence of the target definition (using different thresholds, radius and also exploring multiclass classification) on the performance of the model.

5.3 Spatial information

To evaluate model performance beyond the strict definition of our target variable (Section 3.1), we used avalanche observation data aggregated at the scale of the entire Swiss Alps (Figs. 6 and 8; Table 4). Alternatively, the model predictions at the individual locations of the AWS may be used to establish maps showing the spatial variations in predictions. Furthermore, this would allow for further assessing model performance beyond the binary classification and for different spatial scales.

5.4 Model limitations

Despite going back 20 years, the necessity to obtain high quality labels forced us to train the model using a comparably small dataset. We suspect that a larger dataset, covering a wider variety of avalanche conditions, may increase the model performance further. For instance, we noticed that the model over-predicted avalanche activity during the second half of April 2022 (Fig. 8). The reason for this behavior is not clear. One possible explanation is that the training data did not permit to learn the complex interaction between continued wetting of an already wet snowpack sometimes leading to avalanches, though often not. Another reason might be that the model did not include previous wet-snow avalanche activity as an input variable. After all, previous avalanche activity is highly relevant information taken into account by avalanche forecasters. Furthermore, after a period with wet-snow avalanches or extensive wetting of the snowpack down to basal layers, loading by new snow or rain may be needed to lead to another period of wet-snow avalanches. Potentially, this could be addressed by including additional input variables having longer temporal dependency (for example the sum of new snow of the last 7 days). Therefore, we primarily see two paths to further improve model performance: first, increasing the dataset available for training and testing. However, we doubt whether this should be done at the cost of a less restrictive definition of avalanche and non-avalanche days. Second, it might prove beneficial to couple model predictions with avalanche event data. This might require completely new approaches to modeling wet-snow stability.

5.5 Further perspectives

The presented wet-snow avalanche model provides an indication of the likelihood whether wet-snow avalanches will occur in a region. However, in its current setup, it does not include information on the magnitude of the event, i.e. the number and size of avalanches. Since this information is of great importance for avalanche forecasters, future work should aim to develop a model that also provides such information. In this context, the depth below the surface where liquid water is present could be used to provide a rough estimate of avalanche size, as suggested by Wever and others (2018).

To develop a model with a higher spatial resolution, one could use the gridded COSMO-OSHD data in an extended way (Section 2.2.2). Indeed, in the presented model, COSMO-OSHD data were only used to get 24-hour forecasts at the locations of the AWSs. However, by using all the 1 km grid cells that tile the Swiss Alps, would provide predictions for the entire Swiss Alps.

Finally, we believe that the model may also be suitable to study the influence of climate change on wet-snow avalanche activity, using climate scenarios such as presented in Fischer and others (2022).

6. Conclusion

We developed a RF model to predict wet-snow avalanche activity as a function of slope aspect based on a 20-year dataset of snow stratigraphy simulations for virtual 38° slopes facing north, east, south and west at the location of 124 automated weather and snow stations in the Swiss Alps. To make the best use of manual avalanche observations, we developed a rigorous approach to define avalanche days, thereby strongly reducing the potential errors in the dating and wetness classification of avalanches and, hence, increasing the quality of the labels for training and testing the model. Despite a strongly imbalanced dataset, with more than ten times more non-avalanche days compared to avalanche days, the model performance was satisfactory leading to an OOB f1-score of 0.79 (Table 3). Because of the stringent definition of our target variable, our model only learned from the most extreme active or inactive days. Therefore, we paid particular attention to study the behavior of the model for 'intermediate' days (the NaN provided by Eqn. (3)). We assessed the model performance using all avalanche observations from the Swiss Alps for eight winters not included in the training dataset, and without any of the filtering applied for training the models, yielding a correlation between model predictions and avalanche activity of around 0.71 (Spearman correlation) (Table 4).

In addition to the model development and validation using historic data, we also carried out a 'live' test during winter 2021–2022 (Fig. 8). This allowed to evaluate the model performance in both a nowcast and 24-hour forecast mode (Fig. 6). The results showed that even in forecast mode, which was not available for model training, the aspect-dependent occurrence of wet-snow avalanches, or their absence, was generally well captured (f1-score about 0.8). The promising results obtained during the semi-operational testing during winter 2021–2022 suggests that the model may be useful to support operational avalanche forecasting.

Data. The datasets used in this paper is available on https://www.envidat.ch/dataset/data_wet_aval_model and the scripts on https://gitlabext.wsl.ch/richter/rfmodel_wetsnow.

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

Acknowledgements. We thank Marc Ruesch and Mathias Bavay who provided access to avalanche forecast and weather station data and to the SNOWPACK simulations. This project was partly funded by the Swiss Data Science Center under grant C18-05 "DEAPSnow".

References

Baggi S and Schweizer J (2009) Characteristics of wet-snow avalanche activity: 20 years of observations from a high alpine valley (Dischma, Switzerland). *Natural Hazards* **50**(1), 97–108.

Bartelt P and 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.

Bellaire S, van Herwijnen A, Mitterer C and Schweizer J (2017) On forecasting wet-snow avalanche activity using simulated snow cover data. *Cold Regions Science and Technology* **144**, 28–38.

Breiman L (2001) Random forests. *Machine Learning* **45**(1), 5–32.

Breiman L (2017) *Classification and Regression Trees*. CRC Press. https://books.google.ch/books?id=CmK4tAEACAAJ.

Bründl M, and 5 others (2004) IFKIS - a basis for managing avalanche risk in settlements and on roads in Switzerland. *Natural Hazards and Earth System Sciences* **4**(2), 257–262.

Choubin B, Borji M, Hosseini FS, Mosavi A and Dineva AA (2020) Mass wasting susceptibility assessment of snow avalanches using machine learning models. *Scientific reports* **10**(1), 1–13.

Conway H and Raymond C (1993) Snow stability during rain. *Journal of Glaciology* **39**(133), 635–642.

Cosmo (2021) Consortium for small-scale modeling. https://www.cosmo-model.org/. [Online; last access: 23-October-2022].

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

EAWS (2022a) European Avalanche Danger Scale (2018/19). https://www.avalanches.org/standards/avalanche-danger-scale/, [Online; last access: 23-October-2022].

EAWS (2022b) Standards: avalanche size. https://www.avalanches.org/standards/avalanche-size/, [Online; last access: 23-October-2022].

Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognition Letters* **27**(8), 861–874.

Fischer A, and 9 others (2022) Climate scenarios for Switzerland CH2018 – approach and implications. *Climate Services* **26**, 100288.

Griessinger N, and 5 others (2019) Implications of observation-enhanced energy-balance snowmelt simulations for runoff modeling of alpine catchments. *Advances in Water Resources* **133**, 103410.

Guyon I, Weston J, Barnhill S and Vapnik V (2002) Gene selection for cancer classification using support vector machines. *Machine Learning* **46**(1), 389–422.

Helbig N, van Herwijnen A and Jonas T (2015) Forecasting wet-snow avalanche probability in mountainous terrain. *Cold Regions Science and Technology* **120**, 219–226.

Kattelmann R (1985) Wet slab instability. In *Proceedings International Snow Science Workshop, Aspen, Colorado, U.S.A., 24-27 October 1984, ISSW 1984 Workshop Committee, Aspen CO, U.S.A.*, pp. 102–108.

Lafaysse M, and 5 others (2017) A multiphysical ensemble system of numerical snow modelling. *The Cryosphere* **11**(3), 1173–1198.

Lehning M, and 5 others (1999) Snowpack model calculations for avalanche warning based upon a new network of weather and snow stations. *Cold Regions Science and Technology* **30**(1-3), 145–157.

Lundberg SM and Lee SI (2017) A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* **30**, 4–8.

Mayer S, van Herwijnen A, Techel F and Schweizer J (2022) A random forest model to assess snow instability from simulated snow stratigraphy. *The Cryosphere* **16**(11), 4593–4615. doi:10.5194/tc-16-4593-2022

Mitterer C and 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.

Mitterer C, Techel F, Fierz C and Schweizer J (2013) An operational supporting tool for assessing wet-snow avalanche danger. In Naaim-Bouvet, F., Y. Durand and R. Lambert, (eds.), *Proceedings ISSW 2013. International Snow Science Workshop, Grenoble, France, 7-11 October, 2013*, ANENA, IRSTEA, Météo–France, Grenoble, France. pp. 334–338.

Morin S, and 9 others (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.

Pedregosa F, and 15 others (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830.

Peitzsch EH, Pederson GT, Birkeland KW, Hendrikx J and Fagre DB (2021) Climate drivers of large magnitude snow avalanche years in the US northern Rocky Mountains. *Scientific Reports* **11**(1), 1–13.

Pérez-Guillén C, and 8 others (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.

Probst P, Wright MN and Boulesteix AL (2019) Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **9**(3), e1301.

Romig JM, Cluster S, Birkeland K and Locke W (2005) March wet avalanche prediction at Bridger Bowl ski area, Montana. In Elder, K. (ed.), *Proceedings ISSW 2004. International Snow Science Workshop, Jackson Hole WY, U.S.A., 19-24 September 2004*, pp. 598–607.

Schneebeli M (2004) Mechanisms in wet snow avalanche release. In *Proceedings ISSMA-2004, International Symposium on Snow Monitoring and Avalanches. Snow and Avalanche Study Establishment, India, Manali, India*, pp. 75–77.

Schweizer J, Jamieson JB and Schneebeli M (2003) Snow avalanche formation. *Reviews of Geophysics* **41**(4), 1016.

Sielenou PD, and 7 others (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.

SLF (2020) *SLF-Beobachterhandbuch (observational guidelines)*. 55 p.; [Online; last access: 23-October-2022].

SNOWPACK (2022) Snowpack repository. https://snowpack.slf.ch/, [Online; last access: 23-October-2022].

Statham G, and 9 others (2018) A conceptual model of avalanche hazard. *Natural Hazards* **90**(2), 663–691.

Techel F (2020) *On Consistency and Quality in Public Avalanche Forecasting: A Data-driven Approach to Forecast Verification and to Refining Definitions of Avalanche Danger*. Ph.D. thesis, Department of Geography, University of Zurich, Zurich Switzerland.

Techel F and Pielmeier C (2009) Wet snow diurnal evolution and stability assessment. In Schweizer, J. and A. van Herwijnen (eds.), *International Snow Science Workshop ISSW, Davos, Switzerland, 27 September - 2 October 2009*, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, 256–261.

Techel F, and 6 others (2018) Spatial consistency and bias in avalanche forecasts – a case study in the European Alps. *Natural Hazards and Earth System Science* **18**(10), 2697–2716. doi:10.5194/nhess-18-2697-2018

Techel F, Pielmeier C and Schneebeli M (2011) Microstructural resistance of snow following first wetting. *Cold Regions Science and Technology* **65**(3), 382–391.

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* **2022**, 1–23. doi:10.5194/tc-2022-108

Vionnet V, and 7 others (2012) The detailed snowpack scheme crocus and its implementation in surfex v7. 2. *Geoscientific Model Development* **5**(3), 773–791.

Wever N, Fierz C, Mitterer C, Hirashima H and Lehning M (2014) Solving Richards equation for snow improves snowpack meltwater runoff estimations in detailed multi-layer snowpack model. *The Cryosphere* **8**(1), 257–274.

Wever N, Vera Valero C and Fierz C (2016) Assessing wet snow avalanche activity using detailed physics based snowpack simulations. *Geophysical Research Letters* **43**(11), 5732–5740.

Wever N, Vera Valero C and Techel F (2018) Coupled snow cover and avalanche dynamics simulations to evaluate wet snow avalanche activity. *Journal of Geophysical Research: Earth Surface* **123**(8), 1772–1796. doi:10.1029/2017JF004515

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

Appendix A

Table 5. Variables used for training the random forest algorithm

Feature description	Unit	Feature name	Selected
Measured variables			
Reflected short wave radiation	W/m^2	OSWR	
Air temperature	°C	TA	WP
Snow surface temperature	°C	TSS	
Ground temperature	°C	TSG	
Relative humidity	–	RH	WP, Daily
Wind velocity	m/s	VW	
Wind direction	°	DW	
Snow height	cm	HS_meas	
Simulated meteorological variables			
Incoming long wave radiation	W/m^2	ILWR	Daily
Incoming short wave radiation	W/m^2	ISWR	
Incoming short wave on horizontal	W/m^2	ISWR_h	Daily
Direct incoming short wave	W/m^2	ISWR_dir	Daily
Diffuse incoming short wave	W/m^2	ISWR_diff	WP, Daily
Solid precipitation rate	kg/m^2/s	MS_Snow	
24 h height of new snow	cm	HN24	
Rain rate	kg/m^2/s	MS_Rain	WP, Daily
Rain energy	W/m^2	Qr	
Simulated snowpack variables			
Snow surface temperature	°C	TSS_mod	WP, Daily
Bottom temperature	°C	T_bottom	
Snow temperature at 0.25 m	°C	TS0	
Snow temperature at 0.5 m	°C	TS1	
Snow temperature at 1 m	°C	TS2	
Sensible heat	W/m^2	Qs	WP, Daily
Latent heat	W/m^2	Ql	WP, Daily
Ground heat at soil interface	W/m^2	Qg0	
Outgoing long wave radiation	W/m^2	OLWR	
Net long wave radiation	W/m^2	LWR_net	
Net short wave radiation	W/m^2	Qw	WP, Daily
Parametrized albedo	–	pAlbedo	WP
24 h snow drift	cm	wind_trans24	
Wind velocity drift	m/s	VW_drift	
Erosion mass loss	kg/m^2	MS_Wind	
Snow water equivalent	kg/m^2	SWE	
Total amount of liquid water	kg/m^2	MS_Water	
Virtual lysimeter	kg/s^2/h	MS_SN_Runoff	
Sublimation mass	kg/m^2	MS_Sublimation	Daily
Evaporated mass	kg/m^2	MS_Evap	
Snow height	cm	HS_mod	
Hoar size	cm	hoar_size	
Stability class	–	Sclass2	
Deformation rate stability index	–	Sd	
Depth of deformation rate stability index	cm	zSd	
Natural stability index	–	Sn	WP
Depth of natural stability index	cm	zSn	WP
Sk38 skier stability index	–	Ss	
Depth of Sk38 skier stability index	cm	zSs	
Structural stability index	–	S4	
Depth of structural stability index	cm	zS4	

The three types of features are: Measured (Meas) meteorological variable, modeled (Mod) meteorological variables by SNOWPACK. The column 'Selected' indicates the features used for the model (Section 3), WP refers to variable extracted at the time of the wettest profile, 'Daily' indicates 24-h average. For further information see Morin and others (2020) and SNOWPACK (2022).

Table 6. Variables computed from the wettest profiles of the day

Feature description	Unit	Feature name	Selected
LWC_{index} (Eqn. (1))	–	mean_lwc	WP, 1,2,3_diff
Sum of LWC over all layers	%	water	WP, 1,2,3_diff
Max of LWC among all layers	%	max_lwc	WP
Depth of the deepest wet layers	cm	lowest	2,3_diff
Proportion of the snowpack that is wet	–	prop_wet	3_diff
Proportion of the lowest 30 cm that is wet	–	prop_base	
Sum of LWC in the lowest 30 cm	–	sum_base	
Proportion of the top 15 cm that is wet	–	prop_up	WP
Sum of LWC in the top 15 cm	%	sum_up	WP
Snow height	cm	HS	WP, 1,2,3_diff

The LWC variables are based on Wever and others (2016) and Mitterer and others (2013). In addition, 1, 2 and 3 days temporal changes are computed for all the listed variables. In the text these variables are indicated by a subscript _1_diff, _2_diff or _3_diff after the feature name. The column 'Selected' indicates the features used for the model (Section 3); WP refers the value computed from the wettest profile.

Appendix B

We tuned the hyperparameters of the model using a grid search procedure. The range hyperparameters explored and the obtained results for RF12 (model trained on dataset 1 and 2) are presented in Table 7.

Table 7. Hyperparameters explored by the grid search

Hyperparameters	Values
number of trees in the forest	[100, ***1000***, 2000]
criterion	[gini, ***entropy***]
max depth of the tree	[2, 5, 10, ***15***, 20]
min number of sample required to split	[2, 4, 6, ***8***, 10, 12, 14]
min number of samples required to be a leaf	[2, ***4***, 6, 8, 10, 12, 14]
number of features for best split	[none, ***sqrt***, log2]

Bold italic characters indicate the hyperparameters of the best model trained on dataset 1 and 2.

We reduce the number of features used to train the model using recursive feature elimination (Guyon and others, 2002). The results of the procedure are shown in Fig. 11.

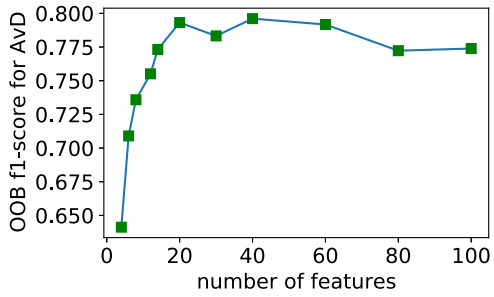

Figure 11. OOB f1-score for *AvD* obtained with recursive feature elimination for a model trained with the combined datasets 1 and 2.

Appendix C

As the target variable was binary: $Y = 0$ ($nAvD$) or $Y = 1$ (AvD), metrics used to optimize and assess model performance were based on the confusion matrix as defined in Table 8, for more details see e.g. Fawcett (2006). The f1-score f_1 is the harmonic mean between recall R and precision P, *i.e.*,

$$f_1 = 2\frac{PR}{P+R}. \quad \text{(C1)}$$

Table 8. Binary confusion matrix, the subscripts $_p$ and $_{gt}$ hold for predicted (model prediction) and ground truth, respectively (target computed from Eqn. (3))

		Predicted class	
		$Y_p = 0$	$Y_p = 1$
Ground truth	$Y_{gt} = 0$	true neg. (correct rejection)	false pos. (false alarm)
	$Y_{gt} = 1$	false neg. (miss)	true pos. (hit)

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press

The recall R (or hit rate) is:

$$R = \frac{\text{number of true positives}}{\text{number of true positives + number of false negatives}}. \quad \text{(C2)}$$

Precision (also called positive predictive value), P, is defined as:

$$P_{Y=1} = \frac{\text{number of true positives}}{\text{number of true positives + number of false positives}}, \quad \text{(C3)}$$

True/false and positive/negative are defined according to Table 8.

https://doi.org/10.1017/jog.2023.24 Published online by Cambridge University Press