EVALUATION OF THE TRANSFERABILITY TO THE PYRENEES OF MACHINE LEARNING MODELS DEVELOPED FOR AVALANCHE FORECASTING IN THE SWISS ALPS

Montse Bacardit[1,*], Cristina Pérez-Guillén[2], Jordi Gavaldà[1], Ivan Moner[1], Martin Hendrick[2], Fabiano Monti[3]

[1] *Centre Lauegi Aran - Conselh Generau Aran, Val d'Aran, Spain*
[2] *WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland*
[3] *Alpsolut S.L.R, Livigno, Italy*

ABSTRACT: Avalanche forecasting is crucial for ensuring safety and mobility in snow-covered mountainous regions worldwide. While traditional avalanche forecasts relied solely on human expertise, machine-learning models have been developed in recent years, and they are already in operation for avalanche forecasting in Switzerland. Such models can rapidly assimilate large amounts of complex datasets, providing objective and high-resolution predictions on the avalanche danger level, the snowpack instability, and wet-snow avalanche activity, assisting forecasters with a "second opinion". These models were originally trained on a historical database containing 20 years of Swiss meteorological measurements, snow cover simulations, danger level assessments and avalanche observations. The performance of the models was comparable to the experienced-based avalanche forecasts in the Swiss Alps, although some variations were observed in different climatic regions. For the future operational application of the models, there is the challenge of addressing the spatial variation of predictive performance between different snow climate regions, and hence verifying the capability to transfer such models into other mountain regions outside of Switzerland. In this study, we test for the first time the avalanche danger and wet-snow avalanche activity models using a database recorded during the last winter seasons in Val d'Aran, a small county on the Atlantic watershed of the Spanish Central Pyrenees. The snow climate has oceanic influence, with an average annual snowfall of 5-6 meters and the snowpack reaching 1-3 meters in high altitudes during mid-winter. The Aran Avalanche Center issues a daily public regional avalanche forecast based on EAWS standards. Additionally, for the last winter seasons, the SNOWPACK model has provided snow cover simulations for two automated weather stations for the flat field and for four mountain slope aspects (north, east, south and west). These data are used as input to run the danger level and avalanche activity models. Even though Val d'Aran is a relatively small mountain region, the high-quality data collected there, including daily avalanche danger level forecasts, snow-cover simulations, snowpack stability field assessments, and avalanche activity observations, makes it a perfect study site to test and validate the predictions of the Swiss machine-learning models in a region with different snow climate conditions.

Keywords: machine learning, avalanche forecasting, avalanche danger, avalanche activity, SNOWPACK model.

1. INTRODUCTION

Avalanche forecasting is crucial for ensuring safety and mobility in snow-covered mountainous regions worldwide. Operationally, many countries issue a daily avalanche bulletin during the winter season to communicate the avalanche hazard to the public and local authorities. In European countries, the key component of these bulletins is the avalanche danger level, defined according to a five-level, ordinal danger scale (EAWS, 2022a). Each danger level is determined by a combination of snowpack stability, frequency distribution, and expected avalanche size (Techel et al., 2022). Avalanche forecasters analyze large data sources such as weather forecasts, meteorological observations, snowpack model outputs, and field observations of snow cover stability and avalanche activity, which are usually limited since natural avalanches often release under poor visibility conditions (Schweizer et al., 2020). While traditional avalanche forecasts relied solely on human expertise, data-driven models have been recently increasingly applied to assist avalanche forecasting (Horton et al., 2020; Viallon-Galinier et al.; Techel et al., 2022; Herla et al., 2023), e.g.

In Switzerland, machine learning models have been developed to predict the avalanche danger levels for dry-snow conditions (Pérez-Guillén et al., 2022), snowpack instability (Mayer et al., 2022), and avalanche activity (Hendrick et al., 2023; Mayer et al., 2022). These models perform similarly to the Swiss avalanche human forecasts, providing objective predictions with higher spatiotemporal resolution (Maissen et al., 2024; Pérez-Guillén et al., 2024). In this study, we evaluate for the first time the transferability to the Pyreenes of the avalanche

*Centre Lauegi Aran - Conselh Generau Aran, 25538 Casau, Val d'Aran, Spain; tel: +34 973641801
email: m.bacardit@aran.org

danger (Pérez-Guillén et al., 2022) and wet-snow avalanche activity (Hendrick et al., 2023) models used operationally by the Swiss warning service since 2021-2022 winter season. The models are validated using a database of meteorological data, snowpack simulations, avalanche danger assessments and observations collected during the last three winter seasons in Val d'Aran.

2. TEST SITE AND DATA

2.1 Avalanche forecast in Val d'Aran

Val d'Aran is a relatively small county of 625 km^2 located in the northwestern part of the autonomous community of Catalonia, Spain. Since it is located in the Atlantic watershed of the Spanish Central Pyrenees (Figure1-inserted map), the snow climate is strongly influenced by the ocean, resulting in an average annual snowfall of 5-6 meters and snowpack heights reaching 1-3 meters above the treeline during mid-winter. For historical reasons explained by this geographic setting, Val d'Aran has a local government responsible for various mandates, including civil protection and emergencies, as well as the avalanche warning service in the territory. A severe avalanche crisis in 2003 triggered the official constitution of the Val d'Aran Avalanche Centre. Then, a local avalanche danger forecast was issued for two key roads during winter, followed by a daily public regional avalanche danger forecast since 2013 (Figure1). The avalanche forecast is issued at 17:00 LT (local time), valid from 00:00 LT to 00:00 LT the following day, with an update at 08:00 LT when necessary. The bulletin is published using the multilingual ALBINA framework (Lanzanasto et al., 2018), which follows best practices and standards of the European Avalanche Warning Services.
The Aran forecast domain is split into three warning regions (Figure 1): North and Center, Southern Limit, and Molières and Conangles. These regions are then clustered into danger regions based on the expected avalanche conditions. Each elevation and aspect of the region is assigned a danger level based on the European Avalanche Danger Scale (EAWS, 2022a, levels: 1–Low, 2–Moderate, 3–Considerable, 4–High, and 5–Very High). The bulletin also includes detailed information on the identified avalanche problems, descriptions of the danger, snowpack stability, weather conditions, and trends. The avalanche problems are communicated in the avalanche bulletin using the five main Typical Avalanche Problems (EAWS, 2022b):

- The *New snow* problem (NS) indicates that the most recent snowfall can be released.
- The *Wind drift* problem (WD) is given when fresh or recently wind-drifted snow is of concern.
- The *Persistent weak layer(s)* problem (PWL) highlights that persistent weak layer(s) within the snowpack are potentially prone to triggering.
- The *Wet snow* problem (WS) indicates the weakening of the snowpack due to the presence of liquid water (melt or rain).
- The *Glide snow* problem (GS) indicates the entire snowpack is prone to trigger on the ground.

The NS, WD and PWL problems apply for dry-snow conditions, whereas WS and GS apply for wet-snow conditions. Typically, one or more avalanche problems are identified in the avalanche bulletin for each danger region (Figure1). The aspects and elevation limits where each avalanche problem is prevalent are indicated. If no avalanche problem is dominant, this information may be omitted, and the avalanche situation is reported as "no distinct avalanche problem".

2.2 SNOWPACK and avalanche activity data

For operational avalanche forecasting, a network of automatic weather stations (AWS) provides hourly weather measurements, while snowpack and avalanche observations are collected by webcams and field observers (including avalanche forecasters, ski patrols, professional guides, and advanced backcountry recreationists). Additionally, since the three last winter seasons, based on the measurements provided by the AWS of Sasseuva (SASS) and Comalada (CMLD) (Figure1), snow cover simulations with the 1D physically-based, multi-layer model SNOWPACK (Lehning et al., 1999, 2002) are performed automatically throughout the winter, providing *nowcast* simulations for the flat field (Figure2a) and four virtual mountain slope aspects (north, east, south and west) at the station's location. In *forecast* mode, snow cover simulations are driven with input data from the numerical weather prediction models Meteoblue and WFR.
This study used the *nowcast* snowpack simulations from CMLD and SASS stations during the last three winter seasons to extract the input features to run the danger level model for dry-snow conditions (Section 3.1) and the wet-avalanche activity model (Section 3.2). The daily danger level assessments available for 436 days and the avalanche observations collected in the Val d'Aran region during this period were used to validate the predictions of the machine learning models developed in Switzerland. The distribution of forecast danger levels for dry-snow conditions in the public bulletin was: 47.71 % for Level 1-Low, 36.01 % for Level 2-Moderate, 14.45 % for Level 3-Considerable and 1.83 % for Level 4-High; whereas for wet-snow conditions was: 67.23 % for Level 1-Low, 26.46 % for

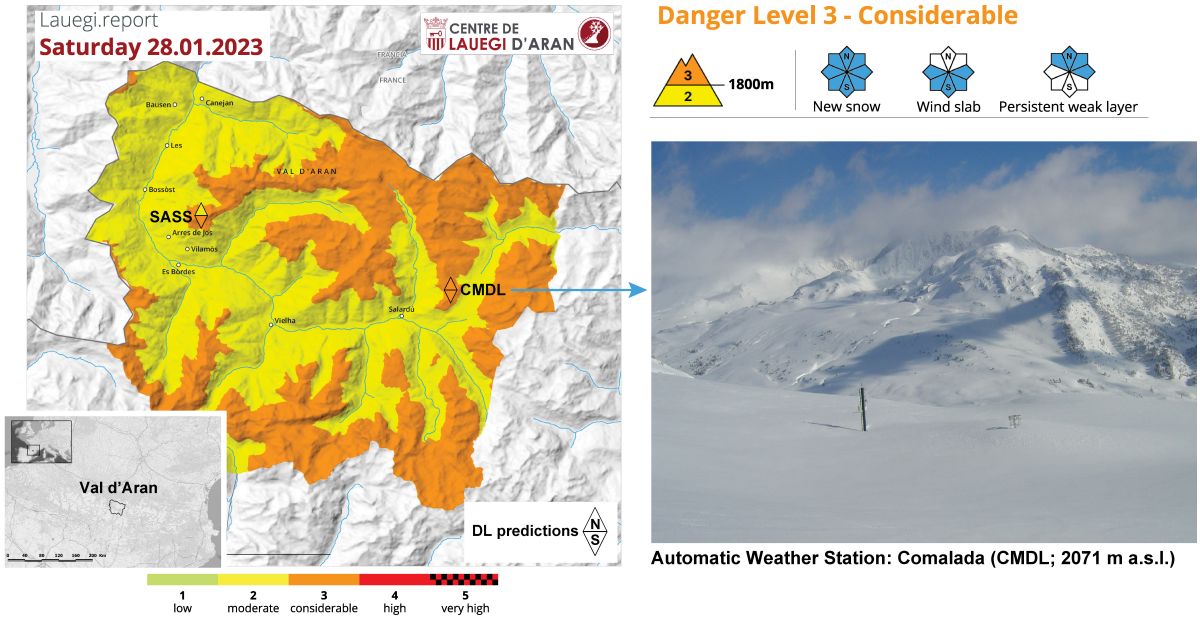

Figure 1: Left panel: Avalanche danger map published in the bulletin issued on 28 January 2023 by the Val d'Aran Avalanche Centre. The locations of the two automatic weather stations, Sasseuva (SASS) and Comalada (CMLD), are shown together with the danger level model predictions at these locations. Predictions are represented by upward triangles for the North (N) aspect and downward triangles for the South (S) aspect, with the colour representing the danger level. Left-bottom inset map: Val d'Aran location in the Pyrenees mountains and Europe. Right panel: The main danger level issued, with the critical elevations and slope aspects, and the main avalanche problems specified. Picture of the AWS in Comalada (2071 m a.s.l.).

Level 2-Moderate, 5.51 % for Level 3-Considerable and 0.81 % for Level 4-High. Avalanche observations in Val d'Aran fulfill at least one of the following criteria: human-triggered events, avalanches reaching transportation routes or affecting slopes above them, and visually observed avalanches of size 3 or larger. All these events are mapped and included in the avalanche database of Catalonia (BDAC).

3. MODELS

3.1 Dry-avalanche danger level model

Dry-snow avalanches are primarily driven by rapid snow accumulation due to snowfall or wind, combined with an unstable snowpack (Schweizer et al., 2003). To automatically predict danger levels for dry-snow conditions in Switzerland, Pérez-Guillén et al. (2022) developed two Random Forest classifiers trained with a 20-year historical data set of measured meteorological data and the output of snow cover simulations driven with these input weather data, and danger ratings from Swiss forecasts. The first model was trained using as a target variable the forecast danger levels published in the avalanche bulletin (RF1 in Pérez-Guillén et al. (2022)), while the second model used a quality-controlled subset of danger level labels (RF2 in Pérez-Guillén et al. (2022)). The accuracy of the models was approximately 70 % (Pérez-Guillén et al., 2022; Pérez-Guillén et al., 2024), reaching a performance close to the Swiss avalanche forecasts (Techel and Schweizer, 2017).

The input features for the models comprise a set of meteorological variables, averaged over a 24-hour time window similar to the avalanche forecast, and features extracted from the simulated snow cover profiles. Here, we extracted the input features from the snow-cover simulations of the two weather stations in Val d'Aran every hour, providing *nowcast* danger level predictions with this time resolution. The output of the models included a danger level from 1 to 4 as well as an expected danger value computed with the probabilities (Maissen et al., 2024; Pérez-Guillén et al., 2024) for each AWS and slope aspect. To evaluate the performance of the models, we considered the station elevation and virtual slope aspects of the input data compared with the critical elevations and aspects specified in the bulletin (Figure 1).

3.2 Wet avalanche activity model

Wet-snow avalanches are released by liquid water infiltration into the snowpack (Schweizer et al., 2003). To predict the local wet-snow avalanche activity at the locations of automated weather stations, Hendrick et al. (2023) developed a Random Forest classifier. Model input consisted of measured meteorological data and SNOWPACK variables computed for 38-degree slopes facing the 4 mountain aspects. The model was trained using wet avalanche observations in Switzerland from 2001 to 2021 to discriminate days with wet-snow avalanche activity from days without activity. Overall, the model

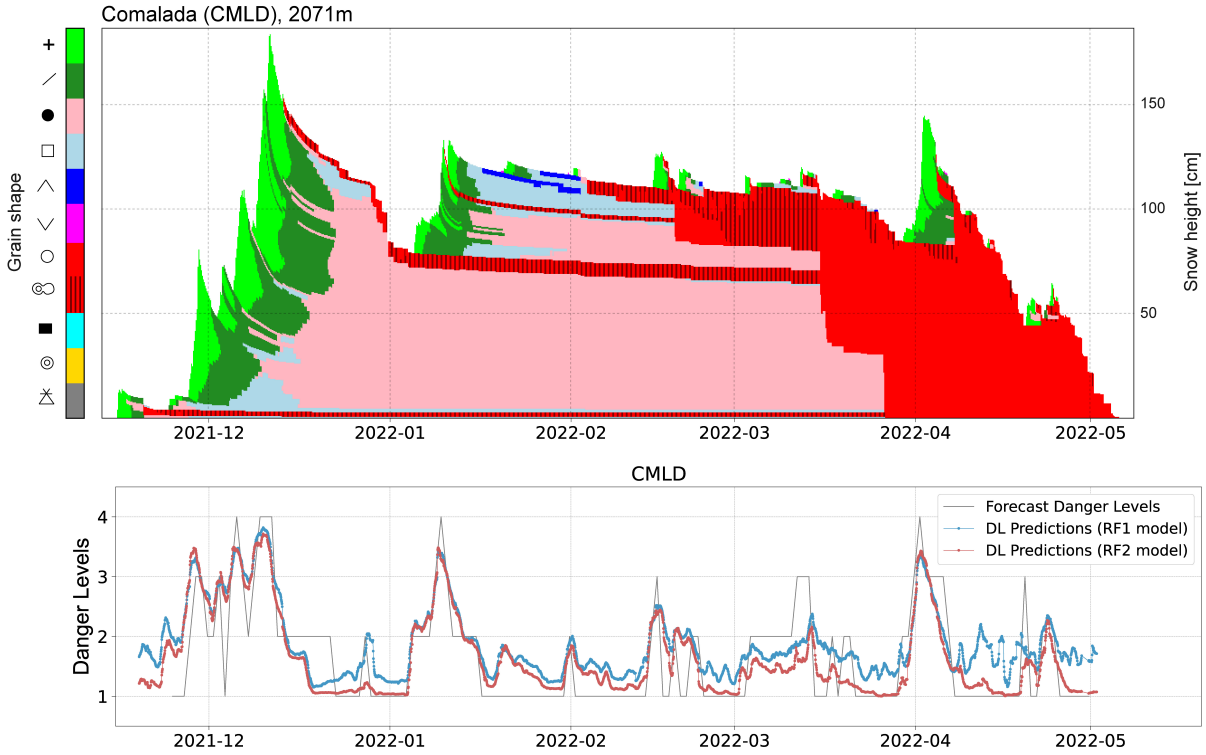

Figure 2: (up) SNOWPACK simulation for the flat field of the location of the automatic weather station of Comalada (CMLD, Figure 1) for winter 2021–2022, showing the evolution of snow grain shape. (down) Time series of the danger level forecast in the bulletin and the expected danger levels, computed with the flat predictions from the danger level model at this location. The red line shows the prediction of the RF1 model, and the blue line of the RF2 model (Pérez-Guillén et al., 2022).

achieved a f1-score of about 0.8 and model predictions were highly correlated (Spearman correlation coefficient of 0.71 $p < 0.01$) with wet-snow avalanche activity over the entire Swiss Alps. For this study, we extracted the model's input features and provided daily avalanche activity predictions for the four aspects at the two stations in Val d'Aran. Model output is a numerical index, from 0 to 1, considering that 0.5 is the critical value above that wet avalanche events are likely to occur.

4. RESULTS

4.1 Performance of the danger level model

The time series of the flat field snow cover simulations in Comalada during the winter season 2021-2022 and the expected danger value predictions of the two danger level models, compared with the regional danger level forecast in the public bulletin, are shown in Figure 2. In general, both models reproduce the dynamics of the forecast danger levels well, showing increases in danger level predictions during the most significant snowfall episodes and decreases during periods without snowfall. Comparing both models, RF2 often predicted a slightly higher danger level when the forecast level was 4 and correlated better when the forecast was at danger level 1, whereas RF1 predicted danger levels 2 and 3 more accurately. This is in line with the evaluation using the entire dataset (Table 1), where RF1 showed better performance in the intermediate danger levels, whereas RF2 performed better at the extreme levels of the scale. When evaluating the overall performance, both model predictions achieved a very similar agreement rate with the bulletin forecast, ranging from 59% (North aspect) to 71% (South aspect) as shown in Table 1). Comparing the agreement rates per aspect and danger level, North aspect predictions achieved the highest agreement rates for danger levels 2 to 4, while they were lowest for danger level 1, which caused the overall performance decrease. Conversely, the flat and other aspect predictions showed the highest agreement rates for danger levels 1 and 3.

Additionally, the confusion matrices of the RF1 predictions for the flat field are shown in Figure 3, first displaying the most critical North aspect and the opposite South aspect. The diagonal of each matrix represents the percentage of correct predictions for each danger level, indicating the agreement rate with the bulletin forecast, and shows the highest agreements for danger levels below 4. Misclassifications are within one danger level on the scale, showing that both models tended to predict one danger level lower on the scale for danger levels 3 and 4, and one danger level higher for danger levels 1 and 2. Misclassifications were two danger levels on the scale in a few cases for danger level 1 (< 10%),

Aspect	Support					DL model	Agreement rate [%]				
	Total	N1	N2	N3	N4		Overall	DL1	DL2	DL3	DL4
F	667	324	213	114	16	RF1	68	79	53	67	48
						RF2	69	92	34	75	25
N	565	220	223	108	14	RF1	59	41	65	85	36
						RF2	60	60	51	79	71
E	511	195	191	109	16	RF1	65	75	49	77	31
						RF2	64	89	32	77	31
S	274	102	78	78	16	RF1	71	78	62	79	25
						RF2	62	86	27	74	25
W	474	222	148	88	16	RF1	66	72	51	80	31
						RF2	64	86	30	75	25

Table 1: Summary of the performance of the two danger level models, RF1 and RF2, for each aspect (F: flat; N: north; E: east; S: south; W: west), using snowpack data from three winter seasons and two stations. The Support column shows the total number of predictions and the number of predictions per danger level. The agreement rate is estimated both overall and per danger level.

and none for the other danger levels.

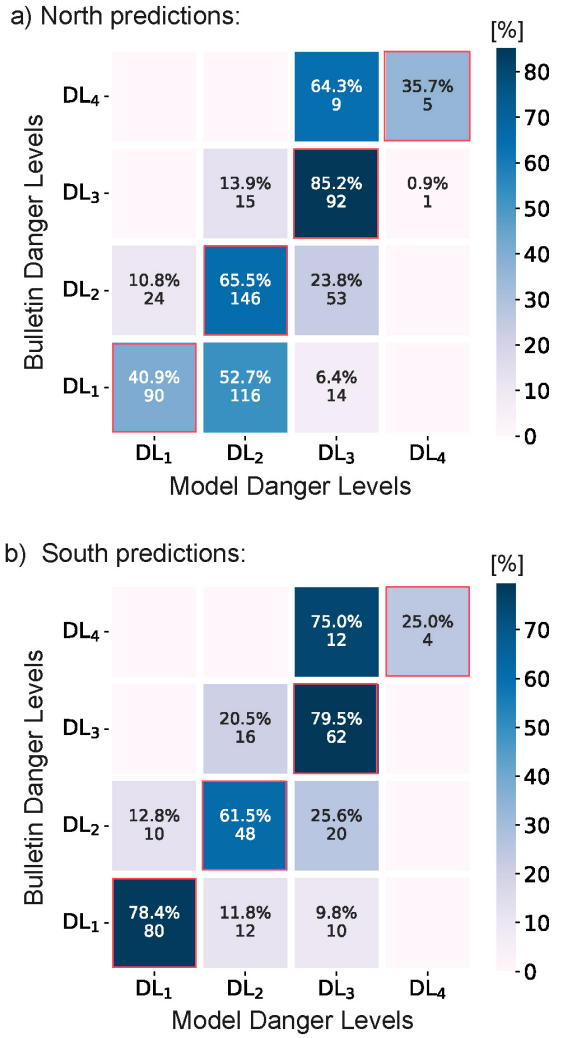

Figure 3: Confusion matrices of the comparison of the danger levels forecast in the public bulletin and the danger level predictions of the RF1 model for two stations and the aspects: a) North and b) South. The color map shows the percentage of samples per danger level, calculated row-wise. Number of samples per case is indicated below percentages.

4.2 Performance of the wet avalanche activity model

Given that avalanche events are human-based observations, periods without observed avalanches do not necessarily indicate the absence of avalanches in the Val d'Aran region. Therefore, to validate the performance of the wet avalanche activity model, we correlated the model's output probabilities with the danger level forecasts for wet-snow conditions provided in the bulletin and compared them with wet-snow avalanche observations, as shown in Figure 4 at the Comalada station during the 2021-2022 winter season. The comparison shows that peaks in the modelled wet avalanche activity frequently coincide with, or often have a time delay from, the observed wet avalanche releases. These time lags are possible since the exact timing of the avalanche release is usually inaccurate. Despite differences in scale between the model and danger level, at the Comalada station, the correlation between the wet avalanche activity probability and the danger level forecast during this winter season is moderate for all the aspects (Figure 4). Overall Spearman correlation coefficients (r_s in Table 2; $p < 0.01$) range from 0.28 (season 2022-2023) to 0.45 (season 2021-2022). Unexpectedly, higher correlations were observed for the North aspect instead of the South, where the wet-snow problem is usually more prevalent. However, the number of predictions for this aspect is much lower due to the relatively low elevation of the AWS stations and higher radiation at the south slope aspects, which favours faster snowmelt.

5. DISCUSSION AND CONCLUSIONS

5.1 Transferability of the models

This study shows that to transfer machine learning models to another region, it is necessary first to have the required input variables for running the predictions, as well as a high-quality dataset to comprehensively validate the model's output. Our evaluation demonstrates that the danger level models (RF1

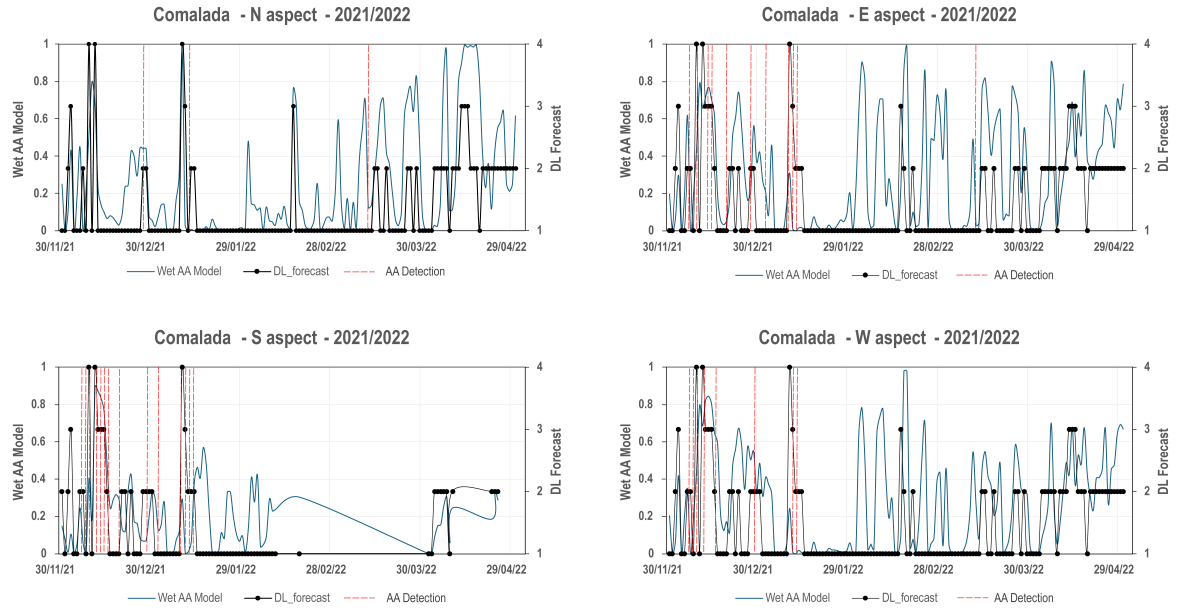

Figure 4: Time series of the wet avalanche activity predictions per aspect (blue line) compared with the danger level forecast for wet-snow conditions (black line) using data from the AWS in Comalada during the winter season 2021-2022. The dashed red lines indicate the dates of recorded wet-snow avalanche releases in particular aspect in the area of the AWS.

Aspect	Winter season	r_s	N
All	2021-2022	0.38	1096
	2022-2023	0.28	927
	2023-2024	0.45	955
N	2021-2022	0.43	304
	2022-2023	0.33	262
	2023-2024	0.56	272
E	2021-2022	0.38	304
	2022-2023	0.26	261
	2023-2024	0.45	267
S	2021-2022	0.18	184
	2022-2023	0.29	143
	2023-2024	0.28	147
W	2021-2022	0.37	304
	2022-2023	0.26	261
	2023-2024	0.42	269

Table 2: Summary of the correlations of the wet avalanche activity model with the danger level forecast for wet problems, aspect (All: average of four aspects; N: north; E: east; S: south; W: west) and winter season. The Spearman correlation coefficient: r_s is presented for all predictions with a p-value of $p < 0.01$. The Support column shows the total number of predictions.

and RF2) predicted avalanche danger levels in the Val d'Aran region with overall agreement rates with the public forecasts (70%; Table 1), similar to those observed in Switzerland (Pérez-Guillén et al., 2022; Pérez-Guillén et al., 2024). Compared to Switzerland, the agreement rates for danger levels 2 and 4 decreased (more than 10%), especially for danger level 4. However, the support of predictions for this high level in this test set is very low. Here, the aspect predictions of RF2 achieved better performance at the extreme levels of the scale, while RF1 performed better for the intermediate danger levels, in line with (Pérez-Guillén et al., 2024). Despite the danger level models are tested against human predictions made by different forecasters, which introduces bias, yet the performance is satisfactory. Therefore, both danger level models could potentially be used to make predictions with input data from Val d'Aran without re-training the models with data from the site.

Validating the performance of the wet avalanche model is more challenging since the frequency and exact timing of avalanche observations are usually inaccurate. Hence, we explored for the first time the relationship between the danger level ratings for wet-snow conditions in the public bulletin and model predictions, showing correlations ranging from low to moderate (Table 2). The transferability of this model is less straightforward due to the difficulty in validating its output. Nevertheless, the output of this model can be employed operationally as a new index assisting in the forecast of wet snow problems.

5.2 Future Use

This study demonstrates the feasibility of transferring machine learning models developed for avalanche forecasting in the Swiss Alps to the Pyrenees, broadening their application across different snow climates. The models showed a predictive performance comparable to those in Switzerland. Both models have open-source code (Pérez-Guillén et al., 2024; Hendrick et al., 2023), making them easily implementable in other countries that use

SNOWPACK operationally for avalanche forecasting. However, comprehensive validation of their performance is recommended before practical application. Optimizing the probability thresholds using a test set from the new location could be a solution to improve the models' performance in new regions.

ACKNOWLEDGEMENT

We thank our institutions, Conselh Generau d'Aran and WSL Institute for Snow and Avalanche Research SLF, for supporting this study. We also recognise Alpsolut's contribution beyond the Snowpack model operational relationship.

References

EAWS: European Avalanche Danger Scale, https://www.avalanches.org/wp-content/uploads/2022/09/European_Avalanche_Danger_Scale-EAWS.pdf, [Online; last access 2-June-2022], 2022a.

EAWS: Avalanche Problems, Edited, EAWS - European Avalanche Warning Services, https://www.avalanches.org/wp-content/uploads/2022/09/EN_EAWS_avalanche_problems.pdf, [Online; last access 2-June-2024], 2022b.

Hendrick, M., Techel, F., Volpi, M., Olevski, T., Pérez-Guillén, C., van Herwijnen, A., and Schweizer, J.: Automated prediction of wet-snow avalanche activity in the Swiss Alps, Journal of Glaciology, pp. 1–14, doi:10.1017/jog.2023.24, 2023.

Herla, F., Haegeli, P., Horton, S., and Mair, P.: A Large-scale Validation of Snowpack Simulations in Support of Avalanche Forecasting Focusing on Critical Layers, EGUsphere, 2023, 1–38, 2023.

Horton, S., Towell, M., and Haegeli, P.: Examining the operational use of avalanche problems with decision trees and model-generated weather and snowpack variables, Natural Hazards and Earth System Sciences, 20, 3551–3576, doi:10.5194/nhess-20-3551-2020, 2020.

Lanzanasto, N., Boninsegna, A., Cestari, P., Kriz, K., Nell, D., Pucher, A., and Mitterer, C.: Project ALBINA: The technical framework for a consistent, cross-border and multilingual regional avalanche forecasting system, in: ISSW-International Snow Science Workshop, Innsbruck, pp. this issue, 2018.

Lehning, M., Bartelt, P., Brown, B., Russi, T., Stöckli, U., and Zimmerli, M.: SNOWPACK model calculations for avalanche warning based upon a new network of weather and snow stations, Cold Reg. Sci. Technol., 30, 145–157, doi:10.1016/S0165-232X(99)00022-1, 1999.

Lehning, M., Bartelt, P., Brown, B., Fierz, C., and Satyawali, P.: A physical SNOWPACK model for the Swiss avalanche warning: Part II. Snow microstructure, Cold Reg. Sci. Technol., 35, 147–167, doi:10.1016/S0165-232X(02)00073-3, 2002.

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

Mayer, S., van Herwijnen, A., Techel, F., and Schweizer, J.: A random forest model to assess snow instability from simulated snow stratigraphy, The Cryosphere, 16, 4593–4615, doi:10.5194/tc-2022-34, 2022.

Pérez-Guillén, C., Techel, F., Volpi, M., and van Herwijnen, A.: Assessing the performance and explainability of an avalanche danger forecast model, EGUsphere, 2024, 1–29, doi:10.5194/egusphere-2024-2374, 2024.

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

Schweizer, J., Kronholm, K., and Wiesinger, T.: Verification of regional snowpack stability and avalanche danger, Cold Reg. Sci. Technol., 37, 277–288, doi:10.1016/S0165-232X(03)00070-3, 2003.

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

Techel, F. and Schweizer, J.: On using local avalanche danger level estimates for regional forecast verification, Cold Reg. Sci. Technol., 144, 52–62, doi:10.1016/j.coldregions.2017.07.012, 2017.

Techel, F., Mayer, S., Pérez-Guillén, C., Schmudlach, G., and Winkler, K.: On the correlation between a sub-level qualifier refining the danger level with observations and models relating to the contributing factors of avalanche danger, Nat. Hazards Earth Syst. Sci. Discussions, pp. 1–29, doi:10.5194/nhess-2022-54, 2022.

Viallon-Galinier, L., Hagenmuller, P., and Eckert, N.: Combining snow physics and machine learning to predict avalanche activity: does it help?, doi:10.5194/tc-2022-108.