

We thank the reviewer for the careful assessment of our manuscript and the constructive comments. Our point-by-point responses are provided below. For clarity, the reviewer's comments are shown in blue italics, followed by our responses in black.

General comments

This manuscript presents meteorological, snow-depth, and avalanche observations from the Galongla Valley, a region where systematic avalanche data remain relatively scarce. The distinction between dry- and wet-snow avalanches is appropriate, and the study could be useful for local avalanche risk management.

The results suggest that snow depth prior to a snowfall or snowmelt episode contains useful information. However, some aspects of the sampling design and validation currently limit the interpretation of the reported probabilities and forecasting performance.

The manuscript would also benefit from a clearer statement of its original contribution to avalanche forecasting. Moreover, the observations cover only two snow seasons. This relatively short period may limit the generalisability of the conclusions, as the results may partly reflect snowpack characteristics specific to those seasons. This issue is particularly relevant because snow depth is the only snowpack variable considered in the analysis.

Response: Thank you for your constructive assessment of our manuscript. We have clarified that the contribution of this study lies in an observation-based assessment of dry- and wet-snow avalanche occurrence in a maritime mountain region with limited in-situ observations. By comparing models with and without pre-event snow depth, we assessed how its inclusion affects the discriminatory performance of the dry- and wet-snow avalanche models. In addition, the models revealed negative conditional associations between snowfall or snowmelt duration and avalanche occurrence, providing new regional observational evidence on the relationship between process duration and dry- and wet-snow avalanche occurrence. These findings contribute to an observation-based understanding of avalanche occurrence in the study area, while the type-specific models provide a basis for further assessment of their potential use in local avalanche-risk assessment and road management.

We have also clarified the retrospective sampling procedure and the interpretation of the model outputs and validation results. LOOCV assesses discrimination within the constructed dataset. To further evaluate model performance across snow seasons, we conducted an additional exploratory validation using the 2024 season for model fitting and the independent 2025 season for testing, with classification thresholds determined solely from the training data. Both model formulations retained discriminatory ability for dry- and wet-snow avalanches in the independent season. We acknowledge that the two-season observation period limits the broader applicability of the findings and that snow depth alone cannot characterise snowpack stratigraphy or stability. We have therefore confined our conclusions to the observed regional conditions. The cross-season validation provides evidence of temporal transferability between the two observed seasons, without establishing long-term forecasting reliability.

Major comments

1. Sample construction and interpretation of probabilities

Section 2.4 is central to the study but remains difficult to reconstruct. Please clarify how avalanches were assigned to monitoring stations, how positive and negative process blocks were identified, whether several samples could belong to the same meteorological episode, and whether the same process could be selected more than once. A schematic time series showing the positive block, excluded periods, buffer, and retained negative blocks would be very helpful.

The datasets contain three selected negative samples per positive sample. Avalanche prevalence is therefore determined by the sampling design and does not represent prevalence among all snowfall or snowmelt processes.

The authors should define the intended prediction unit and explain the consequences of this sampling design. If possible, the model should be calibrated using all eligible processes or corrected for the sampling fractions. Otherwise, the outputs should be described as relative classification scores rather than operational occurrence probabilities.

Response: Thank you for this important comment. We agree that the sample construction, prediction unit, and interpretation of the model outputs should be stated explicitly. For each recorded avalanche, meteorological and snow-depth observations were obtained from the geographically closest monitoring station. Within the preceding 3-day window, qualifying snowfall or snowmelt process blocks were identified. The most recent qualifying process before the recorded avalanche time was selected as the positive block, while earlier qualifying processes were considered candidate negative blocks. Candidate negative blocks within the 12 h exclusion buffer before the start of the positive block were excluded. Up to three temporally separated negative blocks were then retained from the remaining candidates. Each segmented candidate block contributed at most one retained sample, so no individual block was selected repeatedly. However, multiple retained blocks could originate from the same broader snowfall or snowmelt episode.

The prediction unit is an identified snowfall or snowmelt process block, characterised by its intensity, duration, and pre-event snow depth and assigned an event/non-event binary label. Because three negative samples were selected per positive sample, the resulting 1:3 event ratio reflects the sampling design and does not represent the natural avalanche prevalence among all eligible snowfall or snowmelt processes. Accordingly, the model-derived probabilities are interpreted as relative classification scores for the sampled process blocks and not as calibrated absolute avalanche occurrence probabilities. The sample-selection procedure is illustrated in Figure R1 below. Based on an actual record from the modelling dataset, the figure shows the positive block, retained negative blocks, 3-day retrospective window, 12 h exclusion buffer, and recorded avalanche occurrence time on a common time axis.

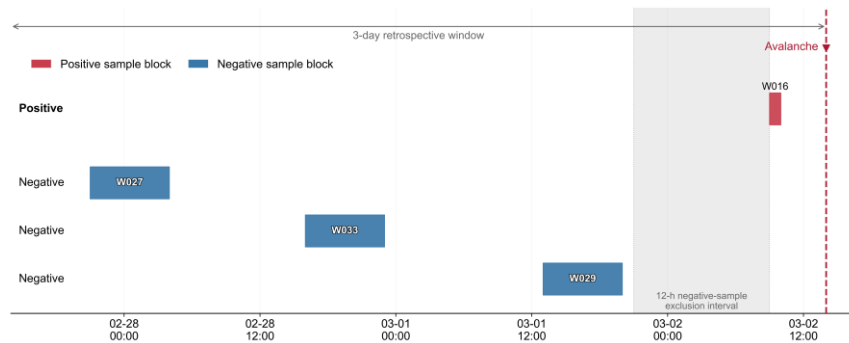


Figure R1. Illustration of positive and negative process blocks. The red block denotes the positive process block, and the blue blocks denote the retained negative process blocks. Labels within the blocks indicate sample IDs. The red dashed line marks the avalanche occurrence time. The horizontal arrow indicates the 3-day retrospective window, and the shaded interval indicates the 12 h exclusion buffer.

2. Interpretation of the forcing-duration coefficients

The negative coefficients for SFD and SMD carry considerable interpretive weight. Section 4.1 discusses them at length in terms of settlement, sintering, viscous creep, preferential flow, and restabilisation, while l. 340 presents the role of forcing duration as a relatively unexplored aspect of avalanche initiation. I have three concerns about this interpretation.

First, intensity and duration are mathematically related in the present framework: SFI and SMI are defined as cumulative amount divided by the corresponding duration (l. 117–118 and l. 134–136). Consequently, interpreting the fitted coefficients as fully separate effects of forcing rate and

duration is not straightforward. The $\ln(1 + x)$ transformation further means that there is no simple exact equivalence between the current formulation and a model expressed in terms of cumulative amount and rate. Nevertheless, the interpretation of the duration coefficient may depend on the chosen parameterisation. It would therefore be helpful if the authors could examine this relationship and compare the present model with an alternative formulation using cumulative amount and rate as predictors. This would clarify whether the negative duration effect is robust or partly dependent on how the predictors are defined.

Second, the definitions at l. 114–115 and l. 132–133 do not appear to impose a maximum process duration, whereas the non-event durations reported in the deposited dataset span a comparatively narrow range. Could the authors clarify whether a fixed window or maximum duration was applied when extracting candidate non-event blocks? If event and non-event durations were determined using different observation rules, part of the fitted duration effect could reflect the sampling procedure rather than snowpack behaviour.

Third, the statement at l. 340 that the role of forcing duration has rarely been examined would benefit from qualification. The relationship between loading rate and strength development in buried weak layers is discussed, for example, by Schweizer et al. (2003a) and within the stability-index approach of Conway and Wilbour (1999). I suggest placing the present results more explicitly within this literature and clarifying how the treatment of duration in this study differs from previous approaches.

Finally, the negative association between forcing duration and avalanche occurrence may be specific to the maritime snow climate and snowpack conditions represented in this study. It may not necessarily apply to snowpacks containing persistent weak layers, where cumulative loading over longer periods may play a different role. Marienthal et al. (2015), already cited in the manuscript, report an association between deep-slab avalanche days and precipitation accumulated over the preceding seven days. This provides a useful contrast and could help define the limits within which the present results may be interpreted. Because the response variable is binary and contains no information on avalanche size, the conclusions should remain limited to occurrence probability and should not be extended to event magnitude.

Response: Thank you for raising these closely related concerns. To examine the dependence on predictor parameterisation, we compared the current rate–duration formulation with an alternative cumulative amount–rate formulation using exactly the same samples, labels, modelling procedure, and LOOCV scheme. The two formulations showed broadly comparable discrimination (Table R1). When duration was replaced by cumulative amount, the fitted cumulative-amount coefficients were also negative; however, their bootstrap 95% confidence intervals included zero in three of the four model formulations. The consistent negative coefficient direction under the alternative formulation indicates that the observed association is not unique to the duration-based parameterisation. However, the weaker statistical support in the amount-based formulation also shows that the strength of this association depends on predictor parameterisation. We therefore interpret the negative SFD and SMD coefficients as conditional empirical associations between the temporal characteristics of the identified processes and avalanche occurrence, rather than as evidence of an independent causal effect of duration.

Regarding the comparatively narrow range of non-event durations in the dataset, this pattern partly reflects a 6 h segmentation criterion applied when constructing candidate non-event process blocks. This criterion was used to segment candidate negative blocks and did not represent a physical maximum duration of the underlying snowfall or snowmelt process. To assess its influence, we reconstructed the underlying continuous negative processes without this segmentation criterion and repeated the modelling as a diagnostic sensitivity analysis. The median durations of the reconstructed negative processes were 10 h for dry snow and 13.5 h for wet snow, compared with 6 h in the current negative samples. Because reconstruction merges segmented blocks belonging to the same underlying continuous process, the composition of the negative-

process dataset changes; we therefore do not use this analysis for a direct comparison of predictive performance. Nevertheless, the fitted duration coefficients remained negative, with bootstrap 95% confidence intervals excluding zero, in both the SMLM and MOLM for both avalanche types (Table R1). These results show that the negative coefficient direction persists under reconstruction, although changes in the duration range and sample composition prevent the isolated effect of segmentation from being determined.

We also agree that the statement that the role of forcing duration has rarely been examined is too broad. Previous studies have considered loading-rate and time-dependent snow-stability effects, including the studies identified by the reviewer. We have therefore qualified this statement and placed our results in the context of this previous work.

Finally, we agree that the negative association between forcing duration and avalanche occurrence should be interpreted within the snow-climate conditions represented by the present observations. We therefore describe the result as a negative conditional association between snowfall or snowmelt process duration and avalanche occurrence in the study area, providing new regional observational evidence on the relationship between process duration and dry- and wet-snow avalanche occurrence. This interpretation is limited to avalanche occurrence and is not extended to persistent-weak-layer conditions or avalanche magnitude.

Table R1. Sensitivity of model discrimination and fitted coefficients to predictor parameterisation and negative-process reconstruction.

Avalanche type	Model	Model discrimination (LOOCV AUC)		Fitted coefficient, β (bootstrap 95% CI)		
		Duration-based	Amount-based	Duration (Duration-based)	Amount (Amount-based)	Duration (Reconstructed negatives)
Dry	SMLM	0.83	0.85	-0.62 (-1.36, -0.02)	-0.19 (-0.77, 0.35)	-1.18 (-1.57, -0.76)
	MOLM	0.72	0.71	-0.56 (-1.36, 0.12)	-0.12 (-0.71, 0.51)	-1.22 (-1.61, -0.79)
Wet	SMLM	0.94	0.94	-1.49 (-1.97, -1.10)	-0.57 (-1.05, -0.11)	-1.17 (-1.46, -0.83)
	MOLM	0.85	0.82	-0.93 (-1.63, -0.44)	-0.01 (-0.56, 0.47)	-1.02 (-1.37, -0.69)

Note: Duration-based models include forcing rate and duration, whereas amount-based models include forcing rate and cumulative amount. SMLM additionally includes snow depth (SD), whereas MOLM does not. For dry and wet avalanches, the forcing variables refer to snowfall and snowmelt, respectively. Reconstructed negatives denote the duration-based models refitted using reconstructed continuous negative processes without the 6 h segmentation criterion.

3. Validation and uncertainty

Leave-one-observation-out cross-validation treats each row as independent. However, samples may be grouped by station, season, avalanche path, or meteorological episode. Closely related observations could therefore occur in both the training and test sets, potentially leading to optimistic performance estimates.

I recognise that the limited number of avalanches and the availability of only two snow seasons make fully independent validation difficult. This limitation should be stated clearly, and the results should be described as internal discrimination rather than generalisable forecasting performance. If feasible, an exploratory comparison in which one winter is used for training and the other for testing could provide some indication of temporal transferability, although the outcome would need to be interpreted cautiously because two seasons are insufficient for a robust validation.

The dry-snow model contains only 13 positive events and three predictors. Confidence intervals should therefore be reported for the AUC and threshold-dependent metrics, as well as for differences between the models. Confusion-matrix counts would also help readers interpret the reported rates.

Response: Thank you for raising this important point regarding validation and uncertainty. We agree that the current LOOCV evaluates discrimination within the constructed dataset and

therefore does not represent long-term generalisable forecasting performance. To further examine temporal transferability, we conducted an exploratory cross-season validation, using the 2024 snow season for model training and the 2025 snow season for validation. Classification thresholds were selected using only the 2024 training data by maximizing the Youden index ($J = \text{TPR} - \text{FPR}$) within an internal LOOCV, and the resulting thresholds were subsequently used to classify the 2025 validation predictions. For dry-snow avalanches, the MOLM and SMLM yielded AUC values of 0.74 and 0.75, respectively, with a TSS of 0.37 for both models. For wet-snow avalanches, the corresponding AUC values were 0.82 and 0.86, with TSS values of 0.56 and 0.58. The models showed retained discriminatory ability in the 2025 validation season. Given that observations are available for only two snow seasons, we regard this cross-season validation as exploratory evidence of temporal transferability rather than as a robust estimate of long-term generalisable forecasting performance.

We also quantified uncertainty in the LOOCV performance estimates using 1,000 paired bootstrap resamples of the out-of-fold prediction records. Percentile 95% confidence intervals are reported for AUC, TPR, FPR, and TSS, together with confusion-matrix counts for the threshold-dependent evaluation (Table R2). Paired bootstrap confidence intervals were additionally calculated for differences between SMLM and MOLM. For dry snow, the SMLM–MOLM differences were 0.11 for AUC (95% CI: 0.01 to 0.24) and 0.05 for TSS (95% CI: -0.08 to 0.19). For wet snow, the corresponding differences were 0.09 for AUC (95% CI: 0.02 to 0.17) and 0.21 for TSS (95% CI: 0.01 to 0.38).

Together, these analyses provide uncertainty estimates for the LOOCV performance and an exploratory cross-season validation of temporal transferability. We accordingly distinguish the internal discrimination assessed by LOOCV from the temporal transferability examined by the cross-season validation and limit the interpretation of both to the available observations.

Table R2. LOOCV performance estimates, 95% bootstrap confidence intervals, and confusion-matrix counts for the dry- and wet-snow avalanche models.

Avalanche type	Model	AUC (95% CI)	TPR (95% CI)	FPR (95% CI)	TSS (95% CI)	TN/FP/FN/TP
Dry	SMLM	0.83 (0.71, 0.94)	0.92 (0.75, 1.00)	0.36 (0.20, 0.51)	0.56 (0.33, 0.77)	25/14/1/12
	MOLM	0.72 (0.56, 0.86)	0.92 (0.75, 1.00)	0.41 (0.25, 0.57)	0.51 (0.27, 0.71)	23/16/1/12
Wet	SMLM	0.94 (0.88, 0.99)	0.92 (0.77, 1.00)	0.13 (0.06, 0.21)	0.79 (0.64, 0.91)	63/9/2/22
	MOLM	0.85 (0.76, 0.92)	0.79 (0.64, 0.95)	0.21 (0.12, 0.31)	0.58 (0.39, 0.77)	57/15/5/19

4. Interpretation and operational scope of snow depth

I support replacing “background snow depth” with “pre-event snow depth” or “prior snow depth.” Snow depth is an observable state variable, but it does not represent snowpack conditions in the broader sense because it contains no information on stratigraphy, weak layers, liquid-water content, or stability.

Its association with recorded avalanches may also reflect avalanche size, terrain roughness, the ability of an avalanche to reach the monitored corridor, season, elevation, or detection conditions. These alternative interpretations should be discussed, and the mechanical interpretation in Sect. 4.1 should be moderated.

The study is retrospective and internally validated. The manuscript states that predictors for positive samples are calculated up to the avalanche occurrence time. Please clarify how the model would be applied prospectively to an ongoing snowfall or snowmelt process, including how the observation window would be defined before the outcome is known. Until prospective validation is available, terms such as “candidate forecasting model” or “retrospective classification model” may be more appropriate than claims of demonstrated operational forecasting performance.

Response: We appreciate this important qualification. We agree that “pre-event snow depth” more accurately describes the role of this variable, and we have adopted this terminology in place of “background snow depth”. Pre-event snow depth is an observable state variable and does not

provide a complete description of snowpack structure or stability because it contains no direct information on stratigraphy, weak layers, liquid-water content, or snow stability. We also acknowledge that its association with recorded avalanche occurrence may partly reflect factors such as avalanche size and runout, terrain characteristics, elevation, season, and detection conditions. We have therefore moderated the mechanical interpretation in Sect. 4.1 and interpret the observed snow-depth effect as an empirical association within the conditions represented by the present dataset.

For prospective application, the model would be evaluated at predefined times without reference to a future avalanche occurrence time. At each evaluation time, only observations available up to that time would be used to identify the current or most recent qualifying snowfall or snowmelt process. For an ongoing process, its duration and intensity would be calculated up to the evaluation time, while pre-event snow depth would be taken from the onset of that process. These variables would then be used to obtain the model-derived avalanche-occurrence score. Thus, the prospective formulation does not require knowledge of the subsequent avalanche occurrence time. Because the present study is retrospective and has not yet undergone prospective validation, we have also moderated statements concerning operational forecasting performance and describe the SMLM as a candidate forecasting model. Prospective evaluation will be required before its operational forecasting performance can be established.

Minor comments

1. Please provide concise summary statistics and distributions of the predictors for events and non-events, together with their units and correlations.

Response: Thank you for this suggestion. We have added Table R3 to provide concise numerical summaries of the predictors for avalanche events and non-events, including sample size, median [Q1–Q3], and units. The corresponding distributions are shown in Figure R2, which presents individual observations together with boxplots for SFI, SFD, and SD in the dry-snow dataset and SMI, SMD, and SD in the wet-snow dataset. For the dry-snow dataset, Pearson’s r was -0.03 for SFI–SFD, 0.39 for SFI–SD, and 0.06 for SFD–SD. For the wet-snow dataset, the corresponding values were -0.31 for SMI–SMD, 0.48 for SMI–SD, and 0.003 for SMD–SD.

Table R3. Summary statistics of predictor variables for avalanche events and non-events.

Avalanche type	Sample	n	Rate (SFI/SMI; mm h ⁻¹)	Duration (SFD/SMD; h)	SD (m)
Dry snow	Event	13	7.94 [4.39–9.46]	4.00 [2.00–8.00]	1.04 [0.97–1.72]
	Non-event	39	3.49 [1.46–5.58]	6.00 [4.00–6.00]	0.15 [0.07–0.74]
Wet snow	Event	24	1.74 [1.00–2.84]	3.00 [1.75–7.00]	2.35 [1.03–2.69]
	Non-event	72	0.32 [0.03–1.18]	6.00 [6.00–6.00]	0.28 [0.05–0.82]

Note: Values are medians [Q1–Q3], where Q1 and Q3 denote the 25th and 75th percentiles, respectively.

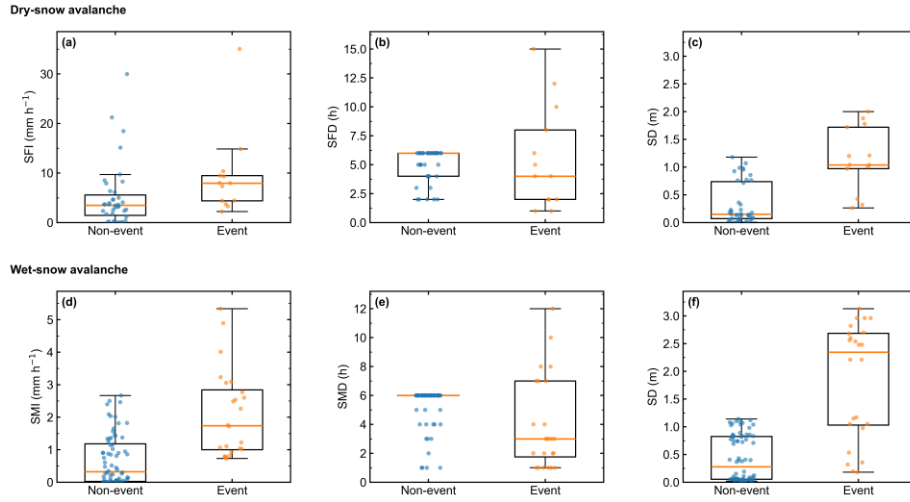


Figure R2. Distributions of predictor variables for avalanche events and non-events. Panels (a–c) show SFI, SFD, and SD for dry-snow samples, and panels (d–f) show SMI, SMD, and SD for wet-snow samples. Boxes indicate the interquartile range, horizontal lines indicate the median, whiskers extend to 1.5 times the interquartile range, and points represent individual observations.

2. Some events appear to have zero melt intensity. Please clarify how a zero melt intensity can occur in samples defined as snowmelt events.

Response: Thank you for pointing this out. We rechecked the wet-snow samples and confirmed that $SMI = 0$ occurs only in retained non-event samples; none of the wet-snow avalanche-event samples has $SMI = 0$. This occurs because non-event samples were selected from candidate snowmelt-related process blocks, whereas SMI was calculated only from the snow-depth change within the retained sample interval. Therefore, a retained non-event block could originate from a period with recent snowmelt-related conditions, while no further snow-depth decrease occurred within the retained interval itself, resulting in $SMI = 0$. We have clarified this distinction by replacing “event blocks” with “snowmelt-related process blocks” in the sample-selection description.

3. Please provide additional information in the caption of Fig. 6 and explain the meaning of the red curve.

Response: Thank you for requesting this clarification. The red curve in the original Fig. 6 is a cubic-spline interpolation of the combined dry- and wet-snow avalanche frequencies calculated for each half-month period. The bars show the observed frequencies of dry- and wet-snow avalanches separately, whereas the red curve provides a visual representation of the seasonal variation in their combined frequency. We have expanded the figure caption to clarify the calculation and meaning of the red curve.

4. Statements that snow depth has the “largest coefficient magnitude” are not directly meaningful when predictors have different units and scales. Standardised effects or odds ratios for meaningful increments would be more informative.

Response: Thank you for pointing out this important issue regarding the interpretation of regression coefficients. We agree that coefficient magnitude should be interpreted cautiously when predictors have different scales. We have therefore removed the statements referring to the SD coefficient as having the “largest magnitude” and no longer use coefficient magnitude to compare predictors. Instead, coefficients are interpreted based on their direction and uncertainty, while the contribution of SD is evaluated through the comparison between SMLM and MOLM.

5. In the abstract, “pre-event snowpack conditions provides” should read “pre-event snowpack conditions provide.” More generally, “pre-event snow depth” should replace “pre-event snowpack conditions” wherever SD is the only snow-related predictor.

Response: Thank you. We have made the corresponding revisions.

Recommendation

The study is based on a valuable observational effort in an underrepresented mountain region. I would encourage the authors to revise the manuscript, with particular attention to the construction of positive and negative samples, the interpretation of probabilities under the sampling design, the interpretation of the forcing-duration coefficients, the limitations associated with the small dataset and short observation period, and the operational interpretation of snow depth. Addressing these points would substantially strengthen the scientific interpretation and potential practical relevance of the work.

Response: Thank you for your comments and recommendations, which have helped us improve the manuscript.