ATM

An open-data avalanche terrain model

Methods, design assumptions and preliminary evaluation

developed with generative AI

Timo Braun

topo^alpin

Document version 1.0.1

2 November 2026

Contents

1 Status of this version 4
2 AI-assisted development 5
3 Introduction 6
4 Related approaches 6
5 Design scenario and rules 7
5.1 What CAT shows 7
5.2 Design scenario 7
5.3 Origin of values 8
5.4 How variants are chosen 8
6 Data and regions 9
7 Terrain model for the Alpine map 10
8 Method 12
8.1 Release potential 13
8.2 Release areas and single avalanches 15
8.3 Flow 17
8.4 Outline and map 19
9 Parameters and their origin 20
10 Evaluation design 21
11 Reproducibility and provenance 24
12 Development results of ATM 1.0 25
12.1 Technical verification 25
12.2 Release and runout 27
12.3 Where the runout differs from CAT 27
12.4 Forest mask and sign of the gap term 29
12.5 Flow ablation and the choice of cohesion and deposition 32
12.6 Release geometry 35
12.7 Size diagnostics 35
12.8 Coverage of mapped avalanches up to size 3 during two extreme periods 36
12.9 Mapped release areas (Davos) 37
12.10 Accidents 37
12.11 Grid-phase envelope 38
12.12 Numerical robustness 38
13 Held-back test in Glarus 39
14 Correction: ATM 1.0.1 41
15 National computation and accidents 44
15.1 Computation 44
15.2 Accidents in Switzerland and Austria 45
15.3 Comparison with CAT in Switzerland 46
15.4 Potential at accident points 46
15.5 Potential at mapped release areas 47

15.6 Presentation of the map 48
16 Evaluation without the rock masks 50
17 What was tried and not adopted 52
18 Completed and planned work 53
19 Open questions 54
20 Limitations 55
Bibliography 56

1 Status of this version

ATM 1.0.1 corrects a release-fragment assignment defect. On eight previously evaluated regions, the correction changed map extent only slightly and did not change the recorded observation hits. In a newly evaluated region near Disentis, the prespecified non-inferiority criterion was not met conclusively: the paired difference from CAT was −1.3 percentage points, with a nominal 95% interval of −5.0 to +2.3 percentage points. This neither shows non-inferiority nor shows a disadvantage beyond the margin.

This is the methods paper of ATM. The method is experimental. It describes how ATM is built, which choices are taken from the literature and which are our own, what has been checked and what has not. It provides a preliminary, scenario-specific evaluation of ATM. The report includes the original held-back evaluation of ATM 1.0 in Glarus and the subsequent evaluation of ATM 1.0.1 near Disentis, whose reference-exposure limitation is documented. These evaluations do not establish general reliability across Switzerland or the Alps. The solver agrees closely with the MoT-Voellmy reference in two benchmark cases and meets every idealised check fixed in advance, and the frozen chain of ATM 1.0 ran without adjustment in a region not used before and met the predeclared primary criterion there. ATM 1.0.1 corrects a defect found afterwards; in a further new region its primary result is unclear (Section 14). The Alpine computation is complete and descriptive results for accidents in Switzerland and Austria are reported (Section 15); they do not change the unclear result. Eight statements frame the version:

1. ATM uses one fixed main configuration. Variants that only move the placement of the 10-m delimitation grid are evaluated separately as a sensitivity envelope.
2. Published parameters, adapted methods and our own choices are marked separately (Section 5.3).
3. Size and runout-length measures diagnose the design scenario; they do not by themselves accept or reject a physical extension.
4. The comparison with SPOT6 outlines tests coverage of avalanches up to size 3 during two extreme avalanche periods; it does not establish a return period.
5. Non-inferiority to CAT is judged only against a tolerance fixed in advance and with its uncertainty interval; unclear results stay unclear.
6. A Swiss region not used before tests the spatial transfer of the frozen chain. In Glarus, the pooled primary endpoint over the mapped avalanches of both periods met the predeclared nominal non-inferiority criterion relative to CAT, with a provisional margin of 2 percentage points (Section 13). Transfer outside Switzerland is not covered.
7. At accident points in Switzerland, ATM and CAT hit about equally often at similar area; no tolerance was set, and the reading of CAT rock changes the sign at the cell of the point (Section 15). Austria has accident points only, and the other Alpine countries are computed but not evaluated.
8. ATM 1.0.1 corrects the assignment of small release fragments to avalanche objects. The intended modelling rule and parameter values are unchanged. Besides changing object grouping, this restores simulation of some fragments that were previously omitted when their remainder object fell below the minimum-area threshold. Object-level and map-level results have been re-evaluated because the correction may affect release geometry, volume classification and derived travel-length statistics.

“Main configuration” denotes the configuration frozen as **ATM 1.0** before the held-back-region test: deposition threshold 2 kPa, cohesion 50 Pa, no entrainment, forest mask, corrected solver, 5-

m computational grid, elevation bands, all terrain unchannelled. Its values form one record with a fixed checksum; the manifest of the frozen version lists it with the code version and all inputs. The term is used for ATM 1.0 only. Earlier configurations appear as development references or sensitivities, and each isolated experiment names the configuration it used. ATM 1.0.1 has the same values; Section 14 gives what its correction changes.

Use of the map. ATM depicts modelled potential avalanche terrain for a prescribed design scenario. It does not predict current avalanche conditions or the probability of release. Uncoloured terrain must not be interpreted as avalanche-safe. The scenario covers the dense part of dry avalanches, not all avalanche types. ATM has no layer for remote triggering, and a 5-m grid does not mean a positional accuracy of 5 m. Near the edge of a computed region the map can be incomplete: release areas outside the region are not simulated, and runout leaving it is cut. The held-back test concerns one defined end point on a selected set of events, not the reliability of each single delimitation. The four grid placements show one segmentation sensitivity, not a full uncertainty envelope of the map.

2 AI-assisted development

Generative AI tools were used extensively throughout the development of ATM: to discuss model formulations and sources; develop the implementation and its tests; design and execute experiments and apply decision rules; analyse results and debug the implementation; and prepare this report and the development record. Two tools were used with distinct roles (Table 1): one primarily for development, the other for critical review of successive drafts of the methods paper, including the methods and results reported in them. The review tool did not examine code or raw data.

The decision rules were proposed by the development tool and applied by it under a general mandate from the author. Most model choices therefore followed from applying these rules without case-by-case approval by the author; applying a rule sometimes required judgement, for example which variant is closer to a source procedure. The author decided selected questions of principle, such as the forest data source, and authorised the freeze and the held-back test.

Source checking, including locating and checking cited parameter values in the publications, was also performed using AI tools. These source checks were not independently repeated by the author.

Tool	**Period**	**Role**
Claude Code (Anthropic), model Claude Opus 5.5	throughout development until 2 November 2026	discussion of formulations and sources, implementation, tests, experiments, analysis, drafts of this report and of the development record
Perplexity	throughout development until 2 November 2026	critical review of the methods paper in successive rounds, with follow-up questions
Perplexity	1 November 2026	review of version 1.0 in a new session, with the paper and a neutral review brief supplied as the review materials, without an account of earlier development discussions

Table 1: AI tools and their roles.

AI-generated suggestions and review comments were treated as development inputs, not as independent validation or authoritative sources. The methodological reviews referred to in this report were AI-assisted reviews; they do not replace review by avalanche experts, which has not yet taken place. Responsibility for the adopted model formulations, the implementation, the methodological

decisions and the reported results rests with the author. Evidence concerning model performance comes from the technical verification and the empirical evaluations described in this report, not from the judgements of the AI tools.

3 Introduction

The Classified Avalanche Terrain map (CAT) of the SLF separates potential release areas from the runout of avalanches up to size 3 on a 5-m grid (Harvey et al.[1,2]). Each potential release area is simulated with RAMMS::EXTENDED. Comparable avalanche-terrain products differ in geographic coverage, model structure, availability and reproducibility (Section 4). For CAT, the published information does not provide all data and implementation details required for an independent reproduction; the tools of the workflow are not openly accessible (Spannring[3]), and the friction values of its simulations are locally adapted (Spannring[3]).

ATM (Avalanche Terrain Model) addresses the same two questions — where can an avalanche of the design scenario release, and how far can it run — with published methods and openly available data. Numerical parameter values are taken from published sources or are explicitly declared modelling assumptions. Reference maps and observations are used for model development and for choosing between variants, but not for direct numerical fitting of parameters. Combining published components does not by itself make a validated method; how far this combination holds has to be shown separately, and this version shows only part of it.

The report describes the chain (Section 8), marks the origin of every value (Section 9), separates technical verification, comparison with CAT and comparison with observations (Section 10), and reports development results of the main configuration (Section 12), together with what was tried and rejected, the planned work and the open questions.

4 Related approaches

Automated avalanche terrain maps are built by several institutions. They share a structure — release areas from terrain, runout from a simulation, a map from both — and differ in how each step is parameterised (Sykes et al.[4]).

CAT and ATH (SLF, Switzerland). Potential release areas from terrain and forest, each simulated with RAMMS::EXTENDED for avalanches up to size 3; the map separates release, runout and remote triggering (Harvey et al.[1,2]). ATH adds consequences as a continuous hazard surface; a third product recasts the results onto the ATES classes (Sykes et al.[4]). CAT and ATH are public in Switzerland, France and Austria (Sykes et al.[4]).

Hazard indication mapping (SLF, canton of Grisons). Automated release areas and RAMMS simulations for the whole canton, in eight scenarios from frequent to very rare, with and without forest (Bühler et al.[5]). Several values of ATM come from this work.

AutoATES (Norway, Canada). Release areas with the fuzzy method of Veitinger et al.[6], runout with the data-based routing model Flow-Py (D'Amboise et al.[7]), and a rule-based classification into the classes of the Avalanche Terrain Exposure Scale (Statham & Campbell[8], Toft et al.[9]). It was validated and adjusted in western Canada (Sykes et al.[10,11]).

Adaptations of AutoATES. For the Austrian Alps (Huber et al.[12]); according to Sykes et al.[4], this adaptation calibrated sub-models against reference data, tuned forest friction and adjusted avalanche-size parameterisations (Hesselbach et al.[13]). Forest along the track was studied with

com4FlowPy (Huber et al.[14]). In Bulgaria the classifier was replaced by a model learned from expert maps (Markov et al.[15]).

NAKSIN (Norway). Release areas from terrain, climate and forest, release probabilities, and runout with the depth-averaged model MoT-Voellmy, for regional hazard indication maps (Issler et al.[16], Issler[17]). ATM takes only selected parts of this system.

Comparisons. AutoATES and CAT were compared near Davos and in the Sellrain valley; release areas agreed closely, and Flow-Py ran longer than RAMMS::EXTENDED (Spannring [3]). Five products were compared with an expert map at Rogers Pass in Canada; all agreed with it about as well as the experts agreed with each other, and the comparison does not rank them (Sykes et al.[4]).

Where ATM stands. ATM combines a fuzzy release model in the manner of AutoATES with release filters of the SLF, a depth-averaged flow model of the MoT-Voellmy type with the friction table of the SLF, and an output in the two layers release and runout. This combination is a hybrid of its own; it is neither the release model of Veitinger, nor AutoATES, nor the SLF workflow. ATM produces no ATES classes.

5 Design scenario and rules

5.1 What CAT shows

CAT distinguishes classes of potential release area, classes of runout and classes of steep rock (Harvey et al.[1]). Runout classes are coloured only inside the simulated runout. Since 2024, the simulations use an updated RAMMS::EXTENDED that "includes enhancements to entrainment processes", a 5-m terrain model, release polygons from an object-based approach and a fracture depth of 0.6 m; slopes smaller than about 200 m² are treated with a simple slope-gradient approach instead of a simulation (Harvey et al. [2, sec. 2.3]). The kernel for remote triggering rests on 75 remotely human-triggered avalanches (Harvey et al.[2, sec. 2.2]).

For the comparison at Rogers Pass, the SLF supplied CAT and ATHc products for design scenarios of 2, 10 and 30 years (Sykes et al.[4]). The relation between these products and the public Swiss CAT layer has not been established here.

ATM computes the same two quantities as two of these layers and is compared with them: its release layer with the CAT release classes, its runout with the runout classes. It has no counterpart to remote triggering.

5.2 Design scenario

CAT focuses on avalanches up to size 3 (Harvey et al.[1]). EAWS describes size 3 with a "typical" length of "several 100 m" and a volume of 10,000 m³, and names size 2 as "the typical skier-triggered avalanche" (EAWS[18]). ATM follows the target of CAT and represents a prescribed small-to-medium design avalanche intended to approximate terrain affected by avalanches up to size 3. It is a frequent-design parameterisation, not a scenario of a stated return period. Its defining assumptions are:

- a fracture depth of 0.6 m, measured normal to the slope, for all release areas (Harvey et al.[2]);
- the friction values of the 10-year column of the SLF table (SLF[19]);
- a largest volume per avalanche of 31,623 m³. This is our own choice: the geometric mean of the typical volumes of sizes 3 and 4 given by EAWS[18]. It is not an official boundary between the sizes.

These are modelling assumptions, not a deterministic implementation of the EAWS size classification.

Size diagnostics. For simulated avalanches we report the projected travel length and the affected area against the ranges for size 3 used by Huber et al.[12, Tab. 2]: 200 to 1,000 m travel length ("based on EAWS") and 10,000 to 80,000 m² affected area (after Bühler et al.[20]). Huber et al. also use 50 to 250 kPa peak pressure; a pressure diagnostic is planned but not reported in this version, since peak speeds have not converged at 5 m (Section 12). The area classes of Bühler et al. are a mapping convention that relies "on area only" (Bühler et al.[20]). These measures diagnose the design scenario; they are not an acceptance criterion on their own.

5.3 Origin of values

Every value of the chain carries one of four labels:

Taken over	same definition, unit and comparable context as in the source
Adapted	a published method with a changed scale, input or application
Own choice	a decision that does not follow directly from a source
Experimental	not decided for the main configuration

The label describes how a value relates to its source, not its validation for ATM. Origin and selection are separate: a published value may be taken over unchanged and still have been chosen among several source values on development data, as the deposition threshold (Section 12.5).

5.4 How variants are chosen

Variants are proposed from sources and judged by rules written down before the computation. Changes were judged so far in four decision regions, against CAT, against the size diagnostics and against observations. This is data-informed model selection; the regions and observations used for it are development data and not independent tests. Three further regions were added later and have also been looked at. One Swiss region not used before is held back (Section 10).

Until this version, a change was rejected if in any decision region the skill score of the whole map fell by 1 percentage point or the share of travel lengths within 200 to 1,000 m fell by 2 percentage points. The second criterion turns the size diagnostic into a rejection rule; it is replaced by the four levels of Section 10.

Six rules or measures were changed or added after results had been seen; one further rule was corrected before any computation on data, when a unit test showed that the narrowest gullies would have been classified as open (search limit for the contour angle, Section 17). Each is reported in its original and its changed form:

- The limit on the size diagnostic was changed from a fixed share of 85 % to a drop of more than 2 percentage points. Under the new limit, the lower deposition threshold failed by about 0.1 point in three of four regions. Under the original limit it could not decide, because two of the regions were below 85 % without any change.
- The share of runs above 1,000 m was replaced by their count and by the number of release objects with such a run, and re-evaluated for all affected variants.
- In the width-scaling sensitivity, the minimum observed width for relative errors was not fixed in advance; the more cautious of the two readings is reported.
- A release slope from 30° was withdrawn after measurement, under the limit valid at the time.

- Settings taken from the authors' script for the release potential were withdrawn after visual inspection of the maps.
- The rule for an open Pareto front (take the candidate closest to the sources) was added after the flow candidates had been compared and the front had stayed open; it was fixed before the grid tests and the final choice.

The count of runs longer than 1,000 m entered two later decisions (no deposition, compact release pieces) as one measure among several. It was read as a shift away from the design scenario of avalanches up to size 3, not as evidence that such runs are physically wrong; the range of Huber et al. is diagnostic, not a hard limit of size 3. Table 23 lists, for every variant rejected after a measurement, the rule applied, the decisive difference and its kind.

Terms. Four kinds of class appear and are kept apart:

Class	Bounds	Used for
Volume class	tiny < 5,000, small < 25,000, medium < 60,000 m³	friction values of the SLF table (Section 8.3)
Release-area class	below 1,000, 1,000 to 10,000, above 10,000 m²	grouping in the grid tests (Section 12.12)
Release-area range of Huber et al.	10,000 to 100,000 m²	selection for the size diagnostics (Section 5)
Observed size	size 3 at most, by the area classes of the dataset	selection of SPOT6 outlines (Section 6)

Pooled percentages are taken over all single items (simulations, outlines, points or cross-sections) unless stated; regions are then weighted by their number of items.

6 Data and regions

Terrain. swissALTI3D, the national lidar terrain model of Switzerland, delivered at 0.5 m and resampled bilinearly to the 5-m grid; not the national 5-m product. Which terrain model the Alpine map uses, and why, is set out in Section 7.

Forest. Copernicus High Resolution Layer Tree Cover Density 2024 at 10 m, resampled bilinearly to the 5-m grid. Resampling adapts the layer to the grid; it adds no forest structure at 5 m. Tree cover density is defined as the "vertical projection of tree crowns to a horizontal earth's surface", per pixel (Copernicus Land Monitoring Service[21, sec. 7.1], version 2.6, which covers the 2024 results). The layer has no forest definition or minimum mapping unit applied; those belong to the separate Forest Type layer (FAO definition, at least 10 % cover and 0.5 ha). It excludes, among others, "Dwarf pine / green alder in alpine areas", shrubland and dwarf-shrub areas (Copernicus Land Monitoring Service [21, Tab. 7-1]). The manual states no minimum tree height. The product provides canopy-cover estimates, not the vegetation-height and stand-structure information required by the original protective-forest workflow, and its values depend on phenology and on the availability of satellite data (same section). This publication has been prepared using European Union's Copernicus Land Monitoring Service information.

Regions. Seven Swiss regions of about 400 km² each (Table 2). Four were used for decisions from the start, three were added later.

Reference and masks. CAT and ATH were obtained as map tiles; in all seven regions every tile is present, so the reference has no missing data inside the regions. A cell of a valid tile that CAT does not colour is valid background, a negative for every CAT class. Two kinds of cell are kept out of the

evaluation as **rock**: cells in the CAT classes for steep rock, and cells that ATH shades grey as rock. The **compared area** of Table 2 consists of the cells with terrain data, outside both rock masks. It is the evaluation area for the map comparison (Table 6), the SPOT6 outlines and the accidents. The map comparison is repeated on all cells with terrain data, with CAT rock as negative, to show the effect of the rock masks. Cells without terrain data have no ATM result and are never counted. The rock masks are those of the reference maps and exist only in Switzerland; the model itself has no rock class. What the evaluation gives without them is reported in Section 16.

Region	**Terrain**	**Compared**	**ATH rock**	**CAT rock**	**Excluded**	**Elevation (m) 1 / 50 / 99 %**	**Cover ≥ 50 %**
Davos	421.6	412.3	7.2	8.2	9.3	1,162 / 2,070 / 2,773	22.5 %
Leventina	415.1	364.1	47.8	36.0	51.0	289 / 1,492 / 2,794	53.4 %
Scuol	414.0	386.9	23.9	25.3	27.1	1,147 / 2,205 / 2,996	24.9 %
Les Diablerets	403.9	385.2	17.0	15.7	18.7	472 / 1,552 / 2,798	37.2 %
Val de Bagnes	404.4	375.5	24.3	26.3	28.9	867 / 2,402 / 3,518	16.9 %
Engelberg	406.8	370.4	33.5	28.2	36.4	465 / 1,513 / 2,790	28.3 %
Alpstein	412.9	397.8	13.7	12.7	15.1	628 / 1,125 / 2,104	33.7 %
Glarus (held back)	412.1	371.3	37.9	31.8	40.8	447 / 1,512 / 2,541	30.8 %

Table 2: Regions; areas in km². The first four were decision regions from the start. Rock: grey rock of ATH, steep-rock classes of CAT, and their union within the terrain data, which is excluded. "Compared" is the terrain data minus excluded rock; elevation and tree cover refer to it. Version 0.4 listed only the CAT rock outside ATH rock (1.4 to 4.6 km²) as excluded rock. Tree cover from the cover layer used by ATM 1.0; the value for Davos was corrected from 18.6 %, which came from an earlier cover layer. The row for Glarus was computed after its evaluation.

Observations, used for development evaluation and variant selection in the seven regions; the observations of the held-back region were reserved for its single frozen evaluation:

- **SPOT6 avalanche outlines** of 24 January 2018 (Hafner & Bühler[22], described by Bühler et al. [20]) and of 16 January 2019 (Hafner & Bühler[23]). Both document periods of very high avalanche danger (level 5). Used: outlines that are fully or mostly visible, size 3 at most by the area classes of the dataset.
- **Mapped release areas** of three ski areas around Davos (Bühler et al.[24]).
- **Avalanche accidents** of the SLF since 1970/71 (SLF[25], version 2.0, file of 15 December 2025): ski tours and off-piste, at least one person caught, accuracy of the start-zone point at most 100 m. The coordinate is the "start zone point"; its accuracy is given in metres.

7 Terrain model for the Alpine map

ATM 1.0 was developed and tested on the Swiss national terrain model, resampled once from 0.5 m to 5 m (Section 6). A map of the Alps cannot use it beyond the Swiss border. It uses one terrain model for the whole Alps instead, assembled from the national and regional lidar models of the Alpine countries: each source is resampled bilinearly, once, to a common 5-m grid in an equal-area projection, and neighbouring sources are blended over a narrow band along the borders. Within Switzerland this Alpine model holds the same 5-m terrain, resampled a second time into the common grid. This chapter explains why the Alpine map uses the Alpine model and not the finer Swiss model, and what the second resampling changes.

Why one Alpine model. The reasons are operational, not claims about accuracy:

1. **One grid without seams.** An avalanche path does not end at a border. Computing each country in its own national grid would give as many grids as countries and a seam at every border, where release areas and runout of neighbouring computations would have to be joined.
2. **No finer source everywhere.** For parts of the Italian Alps the openly available terrain model has a cell size of only 5 m. A common 5-m grid standardises the computational geometry, but not the effective resolution or quality of the source terrain models.
3. **The chain is computed at 5 m.** The predeclared criterion for a finer computational grid was not met (Section 12.12), so the computation stays at 5 m. This decides the grid, not how often and in which way the terrain is resampled to reach it; that is what the measurement below examines.

The Alpine model already exists, assembled from the 5-m terrain models of the countries, and serves other map layers as well. Rebuilding its Swiss part directly from the 0.5-m source would remove the second resampling; it was not done, and the measurement below asks whether the second resampling matters for ATM.

What the second resampling changes. The heights hardly change: the Swiss part of the Alpine model differs from the Swiss 5-m terrain by 5 mm in the median and by less than 4 cm for 99 % of the cells. Its slopes are slightly smoother. In a 10 × 10 km section of steep terrain near Davos, the ruggedness and curvature filters that delimit release areas (Section 8.2) remove 11.5 % of the steep cells on the Alpine model against 14.1 % on the Swiss terrain (ruggedness 7.2 against 8.2 %, curvature 6.4 against 8.7 %). Smoother terrain therefore keeps more release area. Whether this matters was measured before the Alpine map was computed.

Measurement. The Davos development region was computed with ATM 1.0, unchanged, on the Alpine model in its own grid: the same geographic rectangle, the terrain cut from the Alpine model without further resampling, the cover layer resampled to the common grid as for the Swiss runs. The resulting map was transferred to the Swiss grid by nearest neighbour and evaluated there as the main run (Section 10). Two effects besides the terrain enter such a comparison: the transfer between grids, and the edges of the two computational domains, which do not coincide. Carrying the frozen Davos map into the common grid and back, without any computation, lowers the share of SPOT6 lowest points in the map for Davos, both periods, from 91.6 to 88.6 % and the skill score of the whole map from 79.5 to 75.1 points, mainly along a strip at the region edge that the rectangle in the common grid does not cover. Missing release areas near an edge can also act differently in a new computation than in a map that is only cut.

Two evaluations were therefore made, each with its rule fixed before it was computed. The first compares, over the whole region, the computation on the Alpine model with the transferred frozen map as control. The second, added after review, uses only the interior: cells at least 3.5 km from both computational boundaries. This exceeds the largest simulated travel length in either computation (3.4 km) and is intended to reduce boundary effects. It does not establish a strict upper bound on the reach of avalanches from unmodelled release areas outside the domains. On this interior the computation on the Alpine model is compared directly with the frozen Davos map; the transferred control shows whether the transfer still matters there. In both, a difference counts as relevant if the skill score of the whole map against CAT or the share of SPOT6 lowest points in the map (both periods) changes by at least 1 percentage point. This threshold is half the provisional tolerance of Section 10; it is a development threshold, not a demonstrated bound of practical irrelevance.

Measure (Davos)	**Frozen map**	**Transferred control**	**Alpine model**	**Difference**
Whole region, compared area (412 km², 729 outlines); difference to the control				
Skill score, whole map (pp)	79.5	75.1	75.1	**0.0**
SPOT6 lowest points, both periods (%)	91.6	88.6	89.4	**+0.8**
SPOT6 lowest points, 2018 / 2019 (%)	93.2 / 86.7	90.3 / 83.4	90.7 / 85.6	+0.4 / +2.2
Skill score, release / runout (pp)	87.2 / 71.2	82.1 / 67.1	81.8 / 67.7	−0.3 / +0.6
Accident points, release layer, main radius (%)	84.7	81.5	80.4	−1.1
Interior, 3.5 km from both edges (166 km², 340 outlines); difference to the frozen map				
Skill score, whole map (pp)	79.0	78.9	78.7	**−0.3**
SPOT6 lowest points, both periods (%)	93.5	93.5	93.5	**0.0** (−2.0, 2.0)
Skill score, release / runout (pp)	87.0 / 71.8	87.0 / 71.7	86.5 / 71.9	−0.5 / +0.1
Release / runout area (km²)	43.9 / 31.8	43.9 / 31.8	43.6 / 32.2	−0.3 / +0.4

Table 3: ATM 1.0 in Davos on the Swiss terrain (frozen map) and on the Alpine model, evaluated in the Swiss grid after a nearest-neighbour transfer. The control is the frozen map carried into the common grid and back. Bold: the two measures of the rules fixed in advance; the other rows are descriptive. Interval: paired, as in Table 12; on the interior, 4 outlines are reached only on the Alpine model and 4 only on the frozen map, and the control differs from the frozen map in none. Values differ from Table 6 and Table 12 because of the evaluation area and the transfer.

Result. Over the whole region, neither of the two predeclared measures differs by 1 percentage point or more from the transferred control; this comparison includes transfer and edge effects. On the selected interior, the round-trip transfer changed no lowest-point hit and changed the whole-map skill score by 0.1 points; the larger differences over the full region are therefore concentrated outside this interior, and the respective contributions of domain clipping and raster transfer were not fully separated. On the interior, the computation on the Alpine model differs from the frozen map by −0.3 points in the skill score of the whole map and 0.0 points in the lowest points. Its release layer is slightly smaller there (−0.3 km²), so the runout rows of Table 3 are scored on slightly different cells (Section 10). The predeclared threshold was not crossed by either measure in either evaluation. This is a development decision for this comparison, not a demonstration of equivalence or of negligible effects across endpoints and regions: the interval of the lowest points on the interior, ±2 points with 340 outlines, does not exclude differences of that size, and over the whole region the 2019 outlines and the accident points differ by +2.2 and −1.1 points. The Alpine map uses the Alpine model.

Limits of the measurement. One region, north of the main ridge, and one terrain source. Terrain and forest are not separated: the cover layer is also resampled into the common grid. Where the source itself is coarse, as in parts of Italy, the question is a different one, the detail missing from the source, which the measurement does not address. Swiss results of ATM 1.0 remain those of the Swiss terrain.

8 Method

The chain has five steps: release potential, release areas and single avalanches, flow, outline, map. The values are listed with their origin in Section 9.

Figure 1 shows the steps on a section of one development region; the figures in this chapter illustrate single steps on idealised terrain. All are computed with the ATM 1.0 implementation and values; the evidence remains the measurements reported in Section 12.

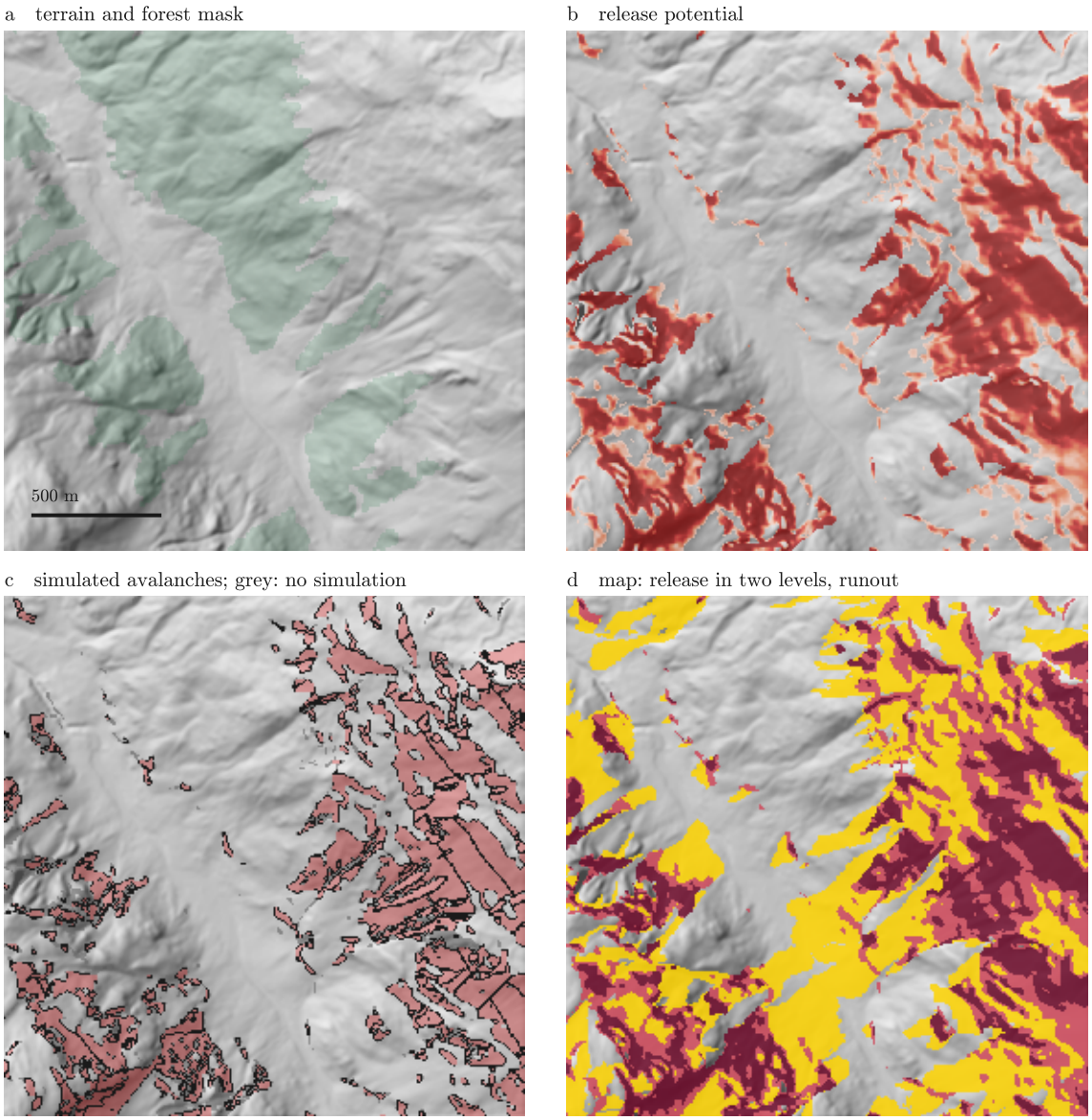


Figure 1: *Illustration, not evaluation.* A 2 × 2 km section of the main ATM 1.0 run of the Davos development region. (a) Shaded terrain and forest mask (green). (b) Release potential, from white (0) to dark red (1), shown where above zero. (c) Simulated release objects with their boundaries; grey: cells of the mapped release layer that start no simulation. (d) The map: release in two levels split at a potential of 0.77, runout in yellow (see the presentation of the map below). The levels are for display only; the computation has no release classes, and the colours do not establish equivalence with CAT classes.

8.1 Release potential

Slope and wind shelter are computed on a 10-m grid obtained by averaging the 5-m terrain model and are refined back to 5 m. Each is turned into a membership between 0 and 1 by a generalised bell function (Veitinger et al.[6, sec. 2.2]): slope with $a = 8$, $b = 3$, $c = 40°$ and zero outside 28° to 60°; wind shelter with $a = 2$, $b = 3$, $c = 2$. Wind shelter follows Plattner et al. as used by Veitinger et al., with the third quartile instead of the maximum, a radius of 60 m (Toft et al.[9]) and no preferred wind direction. The two memberships are combined with the fuzzy AND of Veitinger et al.[6, eq. 18–19]:

$$\mu = \gamma \min(\mu_s, \mu_w) + (1 - \gamma)\frac{\mu_s + \mu_w}{2}, \quad \gamma = 1 - \min(\mu_s, \mu_w).$$

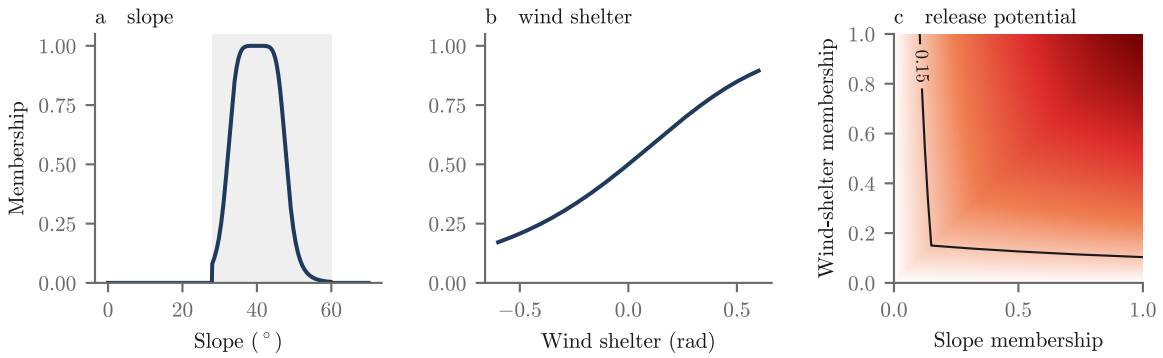


Figure 2: *Illustration, not evaluation.* Memberships of the release potential. (a) Slope; the grey band is the range 28° to 60° outside which the membership is zero. (b) Wind shelter; the index in radians enters the bell function directly, without rescaling. (c) The fuzzy AND of the two memberships (not of slope and shelter themselves); the line marks the threshold 0.15. Forest suppression follows after this step.

The snow-depth-dependent roughness membership of Veitinger et al. is omitted: it needs a 2-m terrain model and a snow depth for which we have no source for the design scenario. A separate, static terrain filter (Section 8.2) is retained.

Forest at release. Release potential is set to zero where a canopy-cover-based approximation of the logistic model of Bebi et al.[26] indicates protection for the frequent scenario:

$$\text{logit } P = -6.17 + 0.18s - 0.03c + 0.05w, \quad \text{protected if } 100P < 65,$$

with slope s in degrees from the terrain model smoothed with a Gaussian kernel of 10 m standard deviation on the 5-m grid, an adaptation of the 10-m terrain model of Bebi et al., and held within 28° to 48°, crown cover c in percent as the mean of the cell and of its 25-m (5 × 5 cells) window, and gap width w in metres. A gap is a connected area of at least 500 m² of cells with less than 50 % tree cover; its width is the length along the contour through the cell, searched up to 300 m. Bebi et al. average over "a 10 m environment" without stating its shape; ATM averages the widths over a window of 3 × 3 cells (15 × 15 m) in the plane, not along the contour; cells outside a gap enter this mean with width 0, so the averaged width is smaller at the margin of a gap and extends one cell into the adjacent forest. Intercept, slope and cover coefficients, the slope range and the threshold of 65 for the frequent scenario are those of Bebi et al. Bebi et al. derive cover and gaps from a vegetation height model and correct the index for surface roughness and shrubs; ATM has neither, so the threshold of 65 is applied to a reduced index. The gap definition by cover, the 50 % and the 500 m² are our adaptation.

Where the forest model applies. Since version 0.9, the equation is evaluated only inside a forest mask, as in the original workflow (Bebi et al.[26]; Bühler et al.[5, sec. 3.2]): the Copernicus forest-type layer (FAO definition, 10 m, nearest neighbour to 5 m), with enclosed holes below 0.5 ha counted as forest. Outside the mask there is neither release suppression nor forest friction; the choice and the earlier state without mask are reported in Section 12.4. A cell without tree-cover data counts as 0 % cover. The gap search stops at the search limit, at the end of the gap and at the region boundary; the width is the count of gap cells found on both sides, so a boundary only shortens one side. Three limiting cases are checked by tests: a treeless slope of 20°, 30° or 45°, also without cover data and up to the boundary, has no protected cell; a closed stand of 100 % cover on 35° is protected throughout; a 100-m-wide gap is unprotected throughout, in the middle of a forest and along the region boundary. On a 30° slope in closed forest, gaps narrower than about 50 m stay partly protected, at 40° none from 20 m.

ATM uses a modified gap-width term with a positive coefficient. This departs from the printed equation of Bebi et al.[26], which has a minus sign, and reflects our process-based interpretation. The direction rests on earlier work from the same Swiss forest-avalanche project. The logistic model of Bebi[27, Tab. 19], fitted to 112 releases and 111 control plots in coniferous forest, has a positive gap-width coefficient (+0.105, $p = 0.001$; "the larger the gap, the higher the probability of a forest avalanche release", our translation). The field data give critical gap widths as lower bounds (Schneebeli & Meyer-Grass[28]). The critical widths of Schneebeli & Bebi[29], as reproduced by Krumm & Bebi[30], grow with crown cover and shrink with slope, which a negative coefficient cannot produce. We could not inspect the 2001 paper from which Bebi et al.[26] adapt their model (Bebi et al.[31]), and the 2021 coefficients differ from the 1999 model, so these sources support the direction of the effect but do not show that the printed sign is an error. Keeping the printed magnitude of 0.05 is our assumption. It does not reproduce the widths of these sources: at the protection threshold and 50 % cover, a gap counts as unprotected from 58 m at 30°, against about 28 m in the reproduced figure (Table 7). The effect of the sign and of the alternative coefficients on the map is reported in Section 12.4.

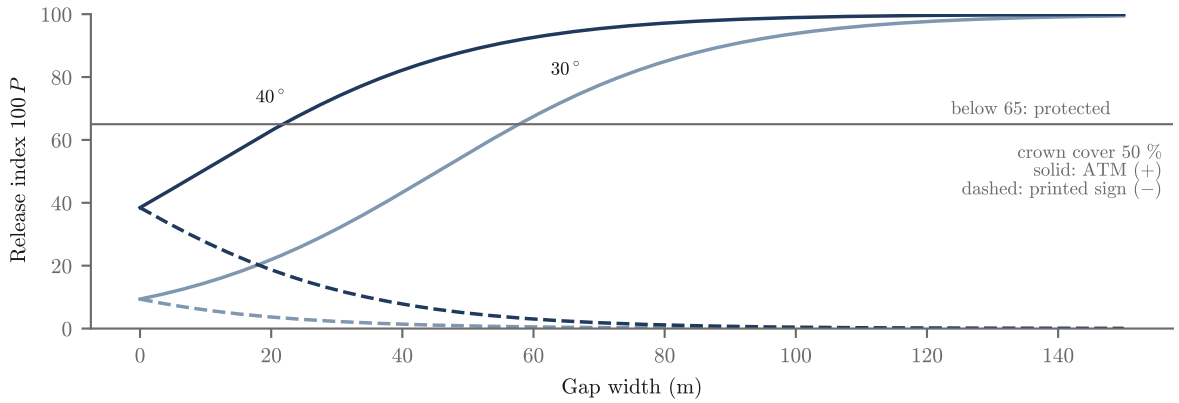


Figure 3: *Illustration, not evaluation.* The forest equation for a crown cover of 50 % at slopes of 30° and 40°, as a function of gap width. Solid: ATM, with a positive gap-width term; dashed: the printed negative sign. Cells with an index below 65 count as protected for the frequent scenario. With the printed negative coefficient, the computed index decreases as gap width increases, enlarging the range classified as protected. The curves isolate the equation's response at fixed crown cover; they do not represent a spatially explicit forest gap, in which gap cells have less than 50 % cover by definition.

8.2 Release areas and single avalanches

Three quantities are kept apart. The **release potential** is the continuous value of Section 8.1, after forest suppression. The **mapped release layer** comprises all cells where the potential reaches 0.15; it is the release layer of the map and the one compared with CAT and with accident points. The **simulated release objects** are the single avalanches for which a flow simulation is started; steps 2 to 7 below lead only to them and do not change the mapped release layer.

1. **Threshold.** A 5-m cell is release where the potential reaches 0.15. The value was set by Toft et al.[9] for their Canadian study areas, who note that it "could be adjusted depending on whether frequent or more extreme avalanche scenarios are of interest" and warn against transfer without site-specific calibration. It is a starting value, not an Alpine parameter.
2. **Static terrain filter** (Bühler et al.[24, sec. 3.2]), on the 5-m grid: cells with a vector ruggedness above 0.06 in a window of 9 cells (45 m) are removed, and cells whose plan curvature exceeds 6 in absolute value. Plan curvature follows Zevenbergen and Thorne in units of 1/100 m with the sign convention of ArcGIS reversed; removing both convex and concave cells is our reading of "highly convex or concave".

3. **Delimitation** on a 10-m grid after NAKSIN (Issler et al.[16, app. A]): a 10-m cell is release where at least half of its 5-m cells are. Cells with fewer than 2.5 neighbours (edge neighbours 1, corner neighbours 0.5) or without a neighbour on both sides of an axis are removed repeatedly. Areas larger than 100,000 m² are divided at the boundaries of drainage basins. The cut 200 to 500 m below the top of each area that NAKSIN applies is left out (own choice).
4. **Splitting by aspect.** Eight aspect sectors as in Bühler et al.[24, sec. 3.1], from the terrain smoothed over five cells. Within each area, connected cells of one sector form an avalanche; segmenting by connected sectors alone is our adaptation. Pieces smaller than 200 m² join the nearest larger avalanche of the same area (own rule). Fixed sector boundaries can separate cells of almost equal aspect and keep cells of more different aspect together; like the grid placement, this is a segmentation sensitivity.
5. **Exclusions.** Avalanches smaller than 200 m² (Harvey et al.[2]), flatter than 28° on average, or lower than 600 m on average (Bühler et al.[5, sec. 3.3]) are not simulated. Bühler et al. use 1,000 m² and set these values with cantonal experts for Grisons; the combination is an adaptation. The 600-m limit is a regional assumption.
6. **Volume.** $V = \sum_i d_n A_i / \cos\theta_i$ with the slope-normal fracture depth $d_n = 0.6$ m, the horizontal cell area A_i and the cell slope θ_i.
7. **Elevation bands.** An avalanche whose volume exceeds 31,623 m³ is cut from the top into as many bands of equal volume as needed; a band that falls apart gives one avalanche per piece. This is our own choice. It keeps the volume limit but does not necessarily give realistic single slab avalanches; a wide, flat band has the same volume as a compact release yet starts a different flow.

Cells removed by the terrain filter, the trimming or the exclusions remain in the mapped release layer but start no simulation; not simulated is not the same as not relevant. Across the full terrain-data domain of the seven regions, including areas excluded from the CAT comparison, 27.6 % of mapped release terrain in ATM 1.0 does not initiate a simulation (207 of 747 km²; 22.7 to 34.2 % per region); within the compared area of Table 2 it is 25.8 % (184 of 714 km²; 21.7 to 31.4 %): 77 % by the terrain filter, 22 % by the delimitation on the 10-m grid, under 1 % by the exclusions and the volume limit. Runout from these patches is therefore not explicitly modelled, although it may overlap runout from other simulated release objects. Simulated outlines, release included, cover 39.0 to 45.6 % of these patches per region. Coverage of these patches by other simulated outlines does not establish coverage of the runout that could originate from the patches themselves.

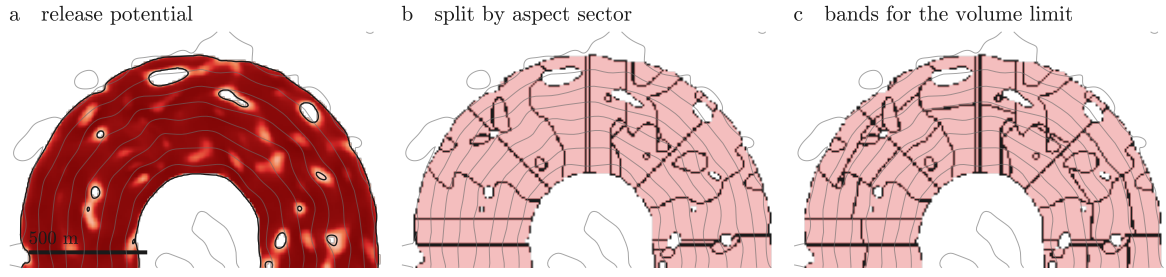


Figure 4: *Illustration, not evaluation.* Release areas and single avalanches on an idealised cirque with walls of 38° and slight random relief; thin lines are 50-m contours. (a) Release potential; the black line marks 0.15. (b) Connected cells of one aspect sector form an avalanche. (c) Avalanches above the volume limit are cut from the top into bands of equal volume; each band is simulated as a separate avalanche. One colour for all objects; lines mark their boundaries. The relief comes from a fixed random seed. On this smooth terrain almost every release cell starts a simulation; on real terrain about a quarter does not (Figure 1).

8.3 Flow

Each avalanche is simulated alone with a depth-averaged model on terrain-following coordinates after MoT-Voellmy, solved with the first-order Method of Transport (Issler[17]): mass and momentum balance, hydrostatic pressure, gravity split normal and parallel to the slope, curvature in the flow direction added to the normal force, time step by a Courant number of 0.8. There is no active or passive earth pressure; Issler[17] expects "more lateral spreading than those models". The basal shear stress is

$$S = \mu N + \rho g u^2/\xi + (1-\mu)N_0\big(1 - e^{-N/N_0}\big),$$

Voellmy friction with cohesion as in the RAMMS manual (SLF[32]), $N_0 = 50$ Pa, $\rho = 300$ kg/m³.

Friction values. μ and ξ depend on the volume class of the avalanche (tiny $<$ 5,000 m³, small $<$ 25,000 m³, medium $<$ 60,000 m³) and on the elevation band of each cell (above 1,500 m, 1,000 to 1,500 m, below 1,000 m), from the 10-year, unchannelled rows of the SLF table[19]. These values are a published baseline. Issler et al.[16, sec. 2] state that "MoT-Voellmy and RAMMS produce very similar simulation results with identical initial data and friction parameter fields" and adopt the table; their transfer to the present implementation has not been validated independently for the full design scenario. All terrain is treated as unchannelled.

Forest in the flow. In the cells classified as protected at release, and only there, tiny and small avalanches get $\mu + 0.02$ and $\xi = 400$ m/s² (SLF[19], Bühler et al.[5, sec. 3.5]); medium avalanches are assumed to destroy the forest. Forest that the equation does not classify as protected, for example a stand of moderate cover on a slope above about 40°, therefore adds no friction.

Deposition after Vicari & Issler[33, sec. 2.2.4]: where the strength of deposited snow, $\tau_d e^{-\gamma_d u}$, exceeds the basal shear stress and the flow slows down, the cell deposits at the rate $(\tau_d e^{-\gamma_d u} - \tau)/(\rho_d u)$, and volume and momentum leave the flow in the same proportion. The rate uses the speed after the friction step; below 10^{-9} m/s the whole cell deposits, and a cell never deposits more than it holds. The density of deposited snow ρ_d is set to the flow density, 300 kg/m³. ATM uses $\tau_d = 2$ kPa, the value of the back-calculated dry avalanche in the same paper (Vicari & Issler[33, Tab. 1]), and $\gamma_d = 1$ s/m from the mass-block example of that section. The scenario simulates only the dense part of dry avalanches (Bühler et al.[5]); the 5 kPa of the paper belong to the mass-block example and to a hypothetical wet avalanche. Up to version 0.7 ATM used 5 kPa; the change follows the comparison in Section 12.5, and 5 kPa is shown as a sensitivity.

No entrainment of snow.

Stop. A run ends when its total momentum falls below 5 % of its maximum (Bühler et al.[5, sec. 3.5]), or after 600 s. The 600 s are our choice; the RAMMS manual gives 300 s as default (SLF[32, sec. 4.2.6]).

Figure 5 follows a single release block on an idealised slope through the flow and its outline; Figure 6 shows the forest friction for two sizes.

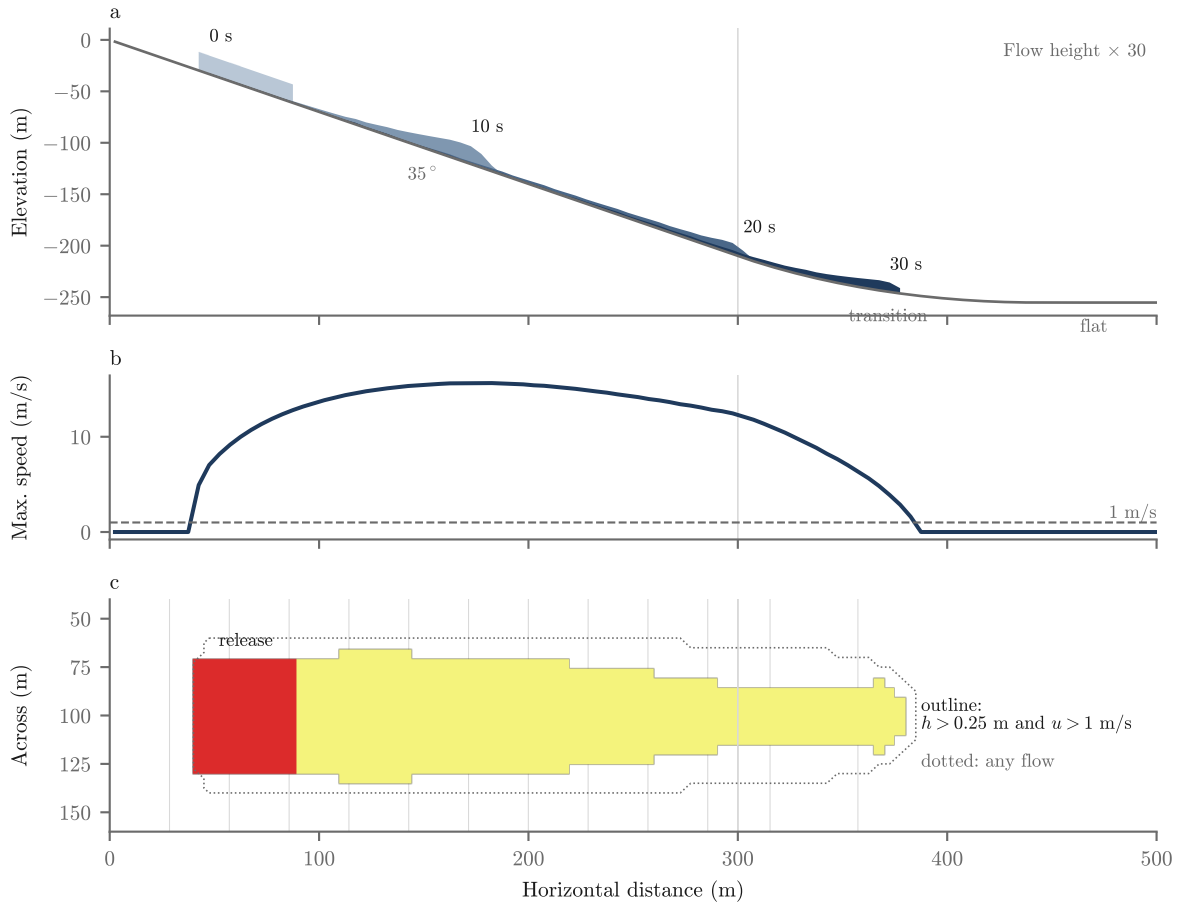


Figure 5: *Illustration, not evaluation.* A block of 60 × 50 m and 0.6 m fracture depth (about 2,200 m³, tiny class) on an idealised slope of 35° with a circular transition to flat ground, computed with the friction values for tiny avalanches above 1,500 m, cohesion and deposition of ATM 1.0. (a) Flow height along the centre line at four times, exaggerated 30 times. (b) Maximum speed along the centre line; dashed: speed threshold of the outline. (c) Plan view; red and yellow here and in Figures 6, 8 and 10 are schematic, not the display of the map: release (red) and outline (yellow); dotted: all cells the flow touched. The furthest cell meeting the outline thresholds lies 378 m (horizontal) from the top of the slope; the simulation ends after 44 s by the momentum rule; about four fifths of the volume is deposited on the way.

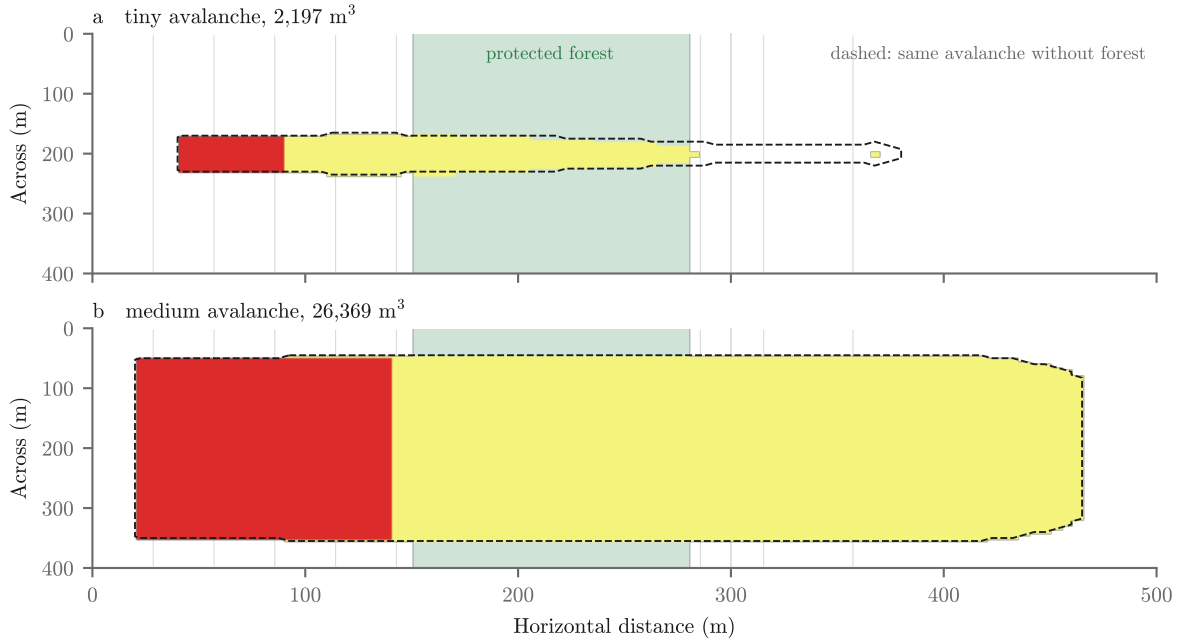


Figure 6: *Illustration, not evaluation.* The same slope with a band of protected forest (green). (a) A tiny avalanche of 2,197 m³ gets the forest friction; its furthest outline cell lies 10 m short of the run without forest (dashed). This is a property of this set-up, not a general measure of the forest effect. (b) A medium avalanche of 26,369 m³ runs the same with and without forest: this shows the implemented rule of no forest friction for medium avalanches, not a simulated destruction of the forest. Plan views are not to scale across the slope.

8.4 Outline and map

A cell is reached where the maximum flow height over time exceeds 0.25 m and, separately, the maximum speed exceeds 1 m/s (Bühler et al.[5, sec. 3.5]); the two maxima need not occur at the same time. Bühler et al. use this outline on a 10-m grid, ATM on 5 m.

The map has two layers: release, where the potential reaches 0.15 (including areas not simulated), and runout, the union of the outlines of all simulated avalanches outside the mapped release layer.

Grid-phase envelope. The 10-m delimitation grid has an arbitrary origin. The main map uses one fixed placement. Four placements shifted by one 5-m cell give a sensitivity envelope: for each cell the share C of placements that reach it. $C = 1$ is reached in all, $0 < C < 1$ depends on the placement. C is an agreement measure, not a probability, and the envelope is not one physical event.

Figure 7 needs the four placements and is therefore recomputed, unlike Figure 1, on the same 2 × 2 km section with a 500-m margin. Its release layer equals that of the main run in every cell. Its runout for the main placement differs from the main run in 482 cells (cells in only one of the two runout masks), against 47,982 runout cells of the main run in the section (1.0 %); 96 % of them lie within 100 m of the edge of the section, none farther than 140 m.

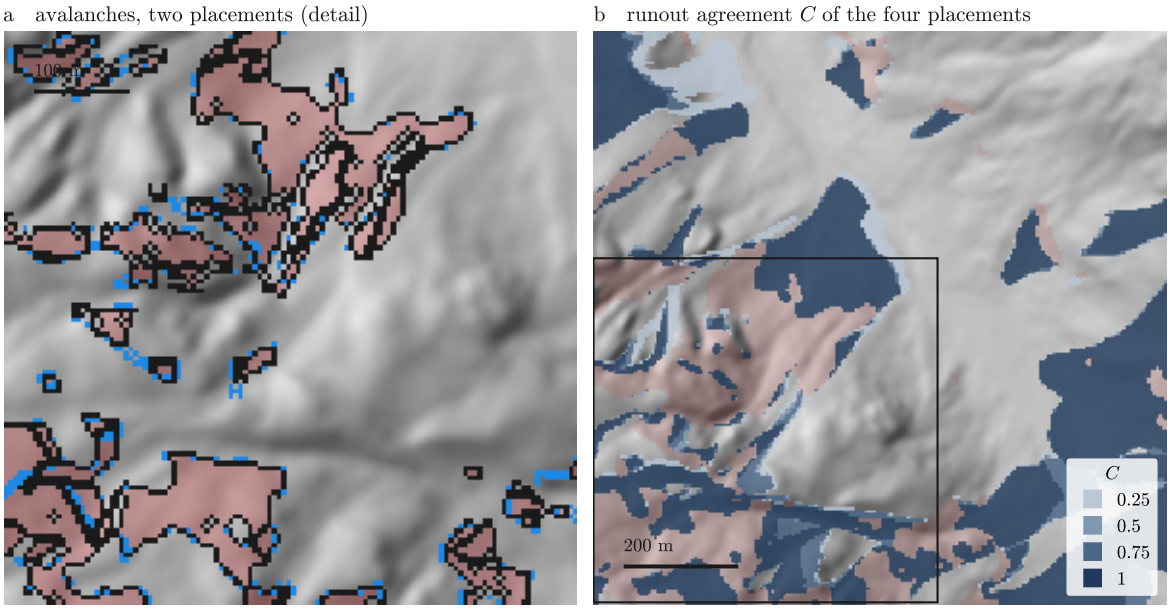


Figure 7: *Illustration, not evaluation.* Grid-phase envelope on part of the section of Figure 1. (a) Detail of the box in (b): boundaries of the simulated avalanches in the main placement (black) and in the placement shifted by one cell in both directions (blue, where it differs). (b) The mapped release layer is shown in light red and is identical in all four placements. Outside this layer, colours show the agreement C of the four runout maps. Among cells reached in at least one placement, C takes the values 0.25, 0.5, 0.75 and 1; cells reached in no placement have $C = 0$ and are not coloured. Light colours depend on the placement.

9 Parameters and their origin

Value	Setting	Label	Source
Slope bell	8, 3, 40°; 28–60°	taken over	Veitinger et al.[6, sec. 2.2.1]
Wind-shelter bell	2, 3, 2	taken over	Veitinger et al.[6, sec. 2.2.2]
Wind shelter	third quartile, 60 m, 10 m	adapted	Veitinger et al.[6]; radius: Toft et al.[9]
Fuzzy AND	eq. 18–19	taken over	Veitinger et al.[6]
Release threshold	0.15	adapted	Toft et al.[9]; Canadian value
Forest model	−6.17, 0.18, −0.03; slope range 28–48°	taken over	Bebi et al.[26]
Slope for the forest model	Gaussian smoothing, σ = 10 m, on the 5-m grid	adapted	Bebi et al.[26, Tab. 1]: slope “based on a 10-m DTM”
Mean gap width	3 × 3 cells (15 × 15 m) in the plane; width 0 outside gaps	adapted	Bebi et al.[26, Tab. 1]: “the mean over a 10 m environment”; window shape and treatment of cells outside gaps not stated there
Gap coefficient	+0.05	own choice	Bebi et al.[26] print −0.05
Protection threshold	$100P < 65$	adapted	Bebi et al.[26]; reduced index
Crown cover, gaps	cell and 25-m window; < 50 %, ≥ 500 m²	adapted	Bebi et al.[26]; from cover, not height
Missing cover data	0 %	own choice	assumption, not observed
Ruggedness filter	> 0.06, 9 cells	taken over	Bühler et al.[24, sec. 3.2]

Curvature filter	abs. plan curvature > 6	adapted	Bühler et al.[24, sec. 3.2]; absolute value our reading
Delimitation, trimming	10 m, 2.5 neighbours	adapted	Issler et al.[16, app. A]
Aspect sectors	eight	taken over	Bühler et al.[24, sec. 3.1]
Segmentation	connected cells of a sector	adapted	–
Small pieces	join nearest avalanche	own choice	–
Minimum area	200 m²	adapted	Harvey et al.[2]; Bühler et al. use 1,000 m²
Mean slope, elevation	28°, 600 m	adapted	Bühler et al.[5, sec. 3.3]; regional
Fracture depth	0.6 m slope-normal	taken over	Harvey et al.[2, sec. 2.3]
Largest volume	31,623 m³	own choice	geometric mean of typical volumes, EAWS[18]
Elevation bands	equal volume, from the top	own choice	–
μ, ξ	10-year, unchannelled	taken over	SLF[19]
Forest friction	$\mu + 0.02$, $\xi = 400$	taken over	SLF[19]; Bühler et al.[5, sec. 3.5]
Cohesion N_0	50 Pa	taken over	SLF[32]; Bühler et al.[5, sec. 3.5]
Density	300 kg/m³	taken over	SLF[32]
Deposition	2 kPa, 1 s/m	taken over	Vicari & Issler[33]: 2 kPa for the back-calculated dry avalanche (Tab. 1); 5 kPa only for the wet one and in a mass-block example (sec. 2.2.4); chosen among the source values in Section 12.5
Stop	5 % of peak momentum	taken over	Bühler et al.[5, sec. 3.5]
End time	600 s	own choice	manual default 300 s
Outline	$h_{\max} > 0.25$ m, $u_{\max} > 1$ m/s	adapted	Bühler et al.[5, sec. 3.5]; 10 m there
Channel classes	all unchannelled	own choice	two transfers of published criteria tested and rejected by visual check (Section 17)
Forest mask	Copernicus forest type 2024, holes < 0.5 ha	adapted	Bebi et al.[26]; Bühler et al.[5, sec. 3.2]: model applied within a forest mask; mask product our choice
Computational grid	5 m, the terrain resolution	own choice	tested at 2.5 and 1.25 m; the criterion fixed in advance for replacing 5 m was not met, so 5 m is retained provisionally, without a claim of grid independence (Section 12.12)

10 Evaluation design

Three kinds of evidence are kept apart.

Block	Question	Results
Technical verification	Does the implementation compute what it should?	benchmark comparison, mass balance, reference cases, regression, convergence

Map comparison	Where and why does ATM differ from CAT?	overlap, area differences, spatial pattern of differences, per region
Empirical evaluation	How well does ATM cover observed events?	release and outline coverage, lowest points, accidents; independent expert review planned

A good comparison with CAT does not replace comparison with observations; a worse one does not by itself show a physical error. Coloured but unobserved terrain is called additionally modelled, not a false alarm, because event inventories are incomplete (Bühler et al.[20]).

Map comparison. Probability of detection, of false detection and their difference, the skill score, for release, runout and the whole map, per region, on the compared area. Release and the whole map are scored on all cells of the compared area. The runout is scored only on cells of the compared area that lie in neither release layer, of ATM or of CAT; there, ATM runout is compared with CAT runout. The runout area therefore depends on both maps and is not the same for different ATM variants; release cells are neither positives nor negatives of the runout score. Changes in the runout score between variants with different release layers may reflect both changes in modelled runout and changes in the evaluation domain. Such changes are not interpreted as isolated improvements or deteriorations of the flow model. Where variants share the release layer, as for the grid placements, the additional start places and the compact release pieces, the runout score is computed on the same cells. The comparison is repeated on all cells with terrain data, with CAT rock as negative (Section 6). Version 0.4 reported the comparison over all cells of a region, including cells without terrain data; that version is kept only as a development figure in the record.

Size diagnostics as in Section 5, for the avalanches whose release area lies in the range of Huber et al.[12] (10,000 to 100,000 m²).

Observed avalanches during two extreme periods. For each SPOT6 outline: the share of its area outside the map (as in Harvey et al.[1]), and whether its lowest point lies in the map. CAT validates with the distance along the flow line (Harvey et al.[2, sec. 3.2]); ATM measures the straight distance, a simplification. Reported per region and period, per period, pooled, and leaving out one region at a time. An outline is used if at least 95 % of its cells lie in the compared area, and only those cells enter. Pooled shares outside the map are the sum of outside cells over the sum of outline cells, so larger avalanches weigh more. This is a coverage test; it does not prove a return period and gives no negative class. Because the compared area is defined with the rock masks of CAT and ATH, the reference products co-determine which events enter; the results refer to this masked set of events, not to all avalanches of a region (Table 4). Whether the excluded events would be harder or easier for either map is not known.

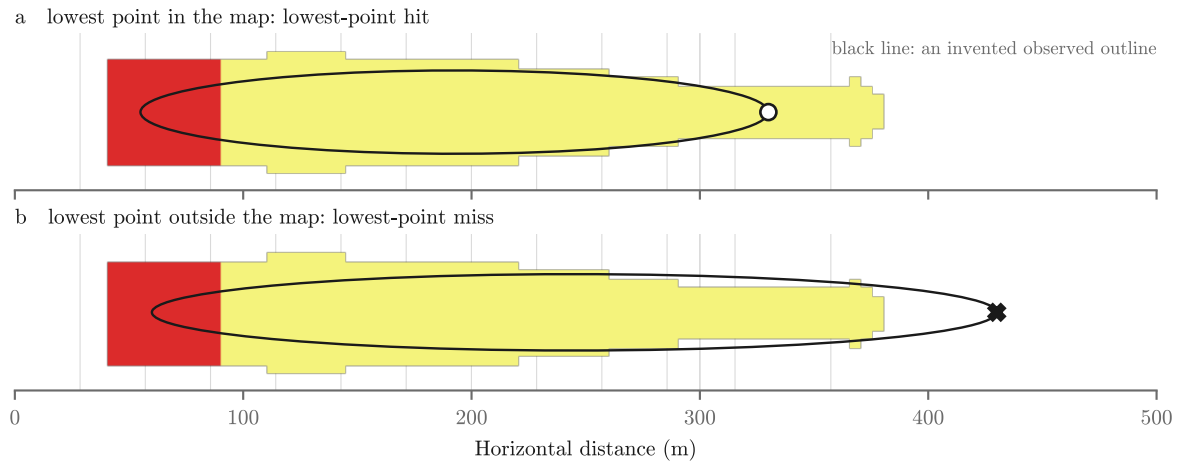


Figure 8: *Illustration, not evaluation.* The lowest-point endpoint with two invented observed outlines (black) on the flow of Figure 5. (a) The lowest point (circle) lies in the map: a lowest-point hit, although part of the outline lies outside the coloured area. (b) The lowest point (cross) lies beyond the map: a lowest-point miss. The share of outline area outside the map is reported separately.

Region	**Visible, size ≤ 3**	**Region edge or no terrain data**	**Rock masks**	**Used**	**Used without rock masks**
Davos	923	20	174	729	–
Scuol	337	4	125	208	–
Les Diablerets	161	5	69	87	–
Val de Bagnes	329	15	98	216	–
Engelberg	842	40	331	471	–
Alpstein	97	1	36	60	–
Development regions	2,689	85	833	1,771	2,604
Glarus	1,174	28	612	534	1,146

Table 4: Selection of SPOT6 outlines, both periods. "Visible, size ≤ 3": outline quality exact or estimated, size 3 at most, at least one cell in the region. The next two columns give the reason an outline fails the 95-% rule, checked first against the region with terrain data, then against the rock masks. Last column: outlines passing the rule on all cells with terrain data, including rock (Section 16). Counted after the evaluations; descriptive. Version 1.0 of this paper split the losses wrongly, assigning most rock-mask losses to the region edge (751 and 167, Glarus 591 and 44) and gave 1,169 visible outlines for Glarus from a different count; the rock masks, not the region edge, remove most outlines.

Accidents. Whether the release layer contains the start-zone point; the main radius is the documented accuracy of each point, other radii are sensitivity. Reported with a uniform-area baseline, since a map that colours everything hits everything: the share of the compared area whose cell or circle of the same radius contains the layer. This is a geometric reference based on uniformly placed points, not a model of recreationist exposure or accident occurrence. The pooled baseline is the mean over the accidents of the baseline of their region and radius, so regions are weighted by their number of accidents. The test concerns the mapped release layer, not the simulated release objects (Section 8.2).

Non-inferiority to CAT. For a paired end point, $\Delta = p_{\mathrm{ATM}} - p_{\mathrm{CAT}}$ with a two-sided 95-% score interval for paired proportions (Newcombe [34], method 10); no interval is given below ten events. Version 0.3 used a Wald interval; the change follows the AI review (Section 2) and does not alter any conclusion. With a tolerance δ fixed in advance, non-inferiority is shown only if the lower bound lies above $-\delta$; if the interval contains $-\delta$ it is not shown; if the whole interval lies below $-\delta$ there is a disadvantage beyond the tolerance. For the lowest points of SPOT6 outlines, $\delta = 0.02$ is our provisional development tolerance, not a validated safety margin; it was set after a difference of about -0.019 was known, so these data cannot confirm it. A tolerance for accident points needs

its own reasoning and is open. Region × period is a reporting unit, not an independent block; a region-clustered interval may be reported as spatial uncertainty conditional on the two periods. All intervals treat events as independent apart from the pairing of the two maps; spatial dependence and repeated avalanche paths are not modelled, so a criterion met by such an interval is met nominally, for the evaluated set of events.

Physical extensions are judged on four levels: (A) technical — mass balance, stability, grid and time-step dependence, reference cases, bounded snow supply; (B) budget — initial plus entrained mass equals moving, deposited and outflowing mass; (C) empirical benefit on unchanged release areas; (D) suitability for the design scenario. The size diagnostics inform (C) and (D).

Held-back region. One Swiss region that has never been computed is held back. Before its single held-out evaluation, code, configuration, main and sensitivity evaluations, masks, hit definitions and the treatment of an unclear or poor result are frozen. It tests spatial transfer with comparable national data; it does not test other terrain models, other forest data, other snow climates or event frequency, and if its observations come from the same two SPOT6 periods it is no new temporal test. Recomputing after a technical failure does not spoil the test; changing a method decision after its results have been seen does.

The held-back region is **Glarus** (north-alpine, about 400 km², at least 5 km from all regions used so far). It was chosen, before any computation there, as the only one of three candidate regions with SPOT6 outlines of both periods and with more than a handful of accidents; the two others lie outside the imagery. It does not cover the southern slope or low forested valleys.

11 Reproducibility and provenance

The workflow records input provenance, parameter settings and code versions. Independent reproduction additionally requires the implementation and its configuration, which are not published with this report. We distinguish internal reproducibility (shown below), a description that can be followed from outside (this report), and independent re-implementation (not possible from this report alone). This report claims the first two only; it describes the method and its evaluation, not a complete implementation specification.

The order of freezing and held-back evaluation (1 November 2026) is recorded internally by version control and by the checksums of the manifest. There is no external timestamp or public archive of the frozen state; readers have to rely on this record.

Cryptographic checksums (SHA-256) secure four points:

- **Inputs.** Each input raster of a region is recorded in a manifest with its checksum, its grid and the checksums of the sources it was derived from; a source with its own manifest must match it exactly. Before every run, each input is hashed again and compared; if one byte differs, the run does not start. Accident records and mapped release areas are recorded with their checksum when read. The SPOT6 outlines, the forest-type layer of each region and all evaluation files are recorded with their checksums in the manifest of the frozen configuration.
- **Values.** All values that determine the map form one record; its checksum is the fingerprint of a run. A test fails when a value is added to the chain without entering the fingerprint.
- **Outputs.** Every raster carries the input checksums and the fingerprint as metadata, and is written in full before it replaces a previous file.

- **Record.** Every run appends one line with time, values, fingerprint, input checksums, code version, whether the code differed from the version under version control, and all measures. A result counts only if the code was unchanged.

A small excerpt of real terrain near Davos serves as a regression test: release areas and single avalanches must be identical to the bit, the runout may differ in at most 0.5 % of cells because the last digit of floating-point arithmetic can differ between machines. A regression test shows that the implementation has not changed unintentionally; it does not show that the equations are solved correctly.

12 Development results of ATM 1.0

The tables of this section, unless a subsection names another configuration, come from ATM 1.0: one set of values, the forest mask, the corrected solver, computed from code under version control without local changes, in the seven regions of Table 2. Code version, value checksum, input and observation checksums and the evaluation files are recorded in the manifest of the frozen version. They are development results: the same regions and observations informed the choices. Experiments that isolate a single question name the configuration they used; earlier configurations appear only as development references or sensitivities. All results of this section were computed with the superseded implementation containing the fragment-assignment defect; Section 14 gives their values under ATM 1.0.1 where they change.

12.1 Technical verification

Benchmark. ISeeSnow is a model intercomparison with three test cases (Wirbel et al. [35]). We ran the two selected Voellmy cases with the inputs prescribed there (release fields, 1.5 m, fixed μ and ξ, no cohesion, deposition or forest) and compared the cells with a peak flow height above a threshold (Table 5).

Case	Reference	Area ATM / ref. (km²)	IoU	Only ATM / only ref.	Peak speed ATM / ref. (m/s)
Real topography	MoT-Voellmy	0.531 / 0.531	95 %	3 / 3 %	50.8 / 44.4
Real topography	RAMMS	0.531 / 0.499	84 %	12 / 6 %	50.8 / 45.9
Idealised	MoT-Voellmy	0.184 / 0.183	99 %	1 / 0 %	40.8 / 41.6
Idealised	RAMMS	0.184 / 0.184	89 %	6 / 6 %	40.8 / 37.9

Table 5: ISeeSnow cases at a peak flow height above 0.5 m. IoU = intersection over union; “only” areas in percent of the reference area. RAMMS = the published RAMMS::Avalanche results with Voellmy friction.

These comparisons show close agreement with the MoT-Voellmy reference outputs for the two selected Voellmy cases. Above 0.01 m flow height, ATM covers 0.80 km² on real topography against 0.68 km² for RAMMS, a larger affected area. In both cases all models reach the same lowest point of the domain, so runout length does not discriminate here. The peak speed on real topography is approximately 14 % higher than the MoT-Voellmy reference and 11 % higher than the RAMMS reference, while the median speeds of the jointly covered cells agree (26.6 against 26.4 m/s with MoT-Voellmy). The comparisons do not replace analytical verification and do not establish correctness of the additional ATM process formulations (cohesion, deposition, forest), which the benchmark switches off.

Idealised cases. Each check has a criterion fixed before the computation (in brackets); all were met:

- **rest:** a uniform layer on flat ground does not move (peak speed 0); a pile spreads and stops before the time limit, without reaching the boundary (relative volume change below 10^{-9});
- **inclined plane:** the Coulomb acceleration $g(\sin\theta - \mu\cos\theta)$ is reproduced to rounding error at 30°, 35° and 40° (criterion 1 %). The Voellmy velocity was compared with the analytical solution at the final time of 20 s; the largest relative deviation was 0.024 %, at 30° (criterion 3 %). Figure 9 shows the 35° case, with a deviation of 0.012 % at the final time; in the first seconds the deviation is larger, up to 2.8 %;
- **mass balance:** with cohesion, curvature, deposition and two friction zones in a chute, and separately with outflow across the boundary, the sum of moving, deposited and outflowing volume equals the initial volume to a relative deviation of about 10^{-14} (criterion 10^{-9}, reference: the initial volume); without entrainment the initial volume is the whole budget;
- **symmetry, rotation, shift:** mirror-symmetric terrain gives a symmetric result, rotations by 90° and 180° give the same result (criterion 10^{-6} m flow height), integer shifts of the grid an identical one;
- **refinement** from 10 m to 5 m to 2.5 m on a slope-to-flat transition: the travel of the centre of mass changes by 0.25 % (Coulomb) and 0.39 % (Voellmy) between 5 and 2.5 m (criterion: smaller change on the finer step, at most 5 %); there is no criterion for the peak speed;
- **rotation by 45°** (criterion 10 % for affected area and centre-of-mass travel): see below;
- **concave transition** without curvature: the travel angle of the centre of mass, $\tan\alpha = 0.405$, matches the friction coefficient 0.4 within 1.4 % (criterion 5 %); curvature raises it to 0.470, as the added normal force should;
- **convex transition:** curvature slightly increases the peak speed, flow height stays non-negative and the volume is kept (criterion 10^{-9}).

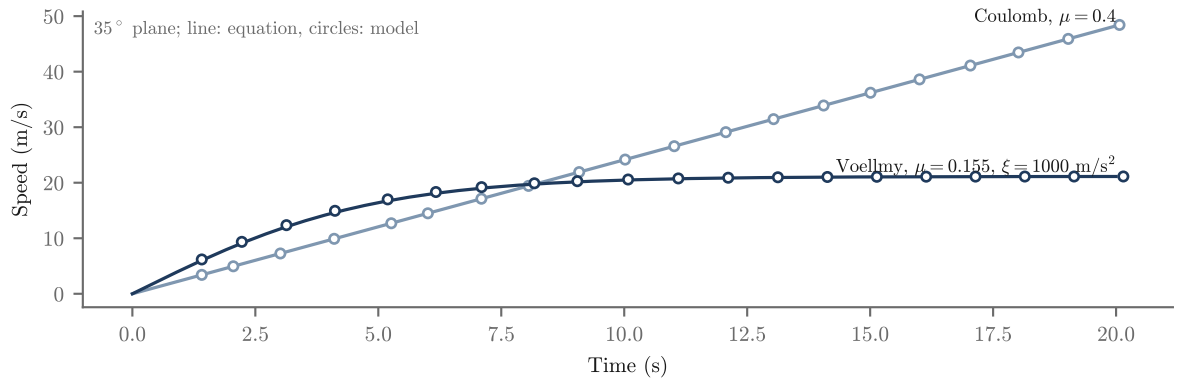


Figure 9: *Technical verification on idealised terrain, not empirical validation.* The inclined-plane case: a uniform layer of 1 m on a 35° plane, 5-m grid, no cohesion. Each circle is a separate run stopped at a time limit of 1 to 20 s; it shows the maximum speed in a cell row in the middle of the plane at the time the run actually reached (the last time step may pass the limit). On the uniform plane the maximum of this row equals the maximum over the whole field. Line: the equation. Coulomb friction ($\mu = 0.4$): deviation below 0.01 %. Voellmy friction ($\mu = 0.155$, $\xi = 1000$ m/s²): 2.8 % at the first point (1.4 s), falling to 0.01 % at 20 s; the criterion of 3 % was checked at the end time.

Two results limit the precision at 5 m. A flow running diagonally to the grid covers 9.7 % more area and its centre of mass travels 5.5 % less than the same flow along the grid, so the grid is not direction-neutral; the 45° case meets its criterion of 10 % with little margin. The peak speed decreases with refinement (48.0, 45.4 and 43.7 m/s for Coulomb friction) and has not converged at 5 m. This suggests resolution sensitivity of peak velocity, but does not by itself explain the differences from the benchmark references. The four grid placements of the delimitation (Section 8.4) do not capture these effects: shifting the segmentation and the orientation dependence of the flow solver

are different sensitivities. Cohesion and deposition are checked only for mass balance, not against an independent solution.

12.2 Release and runout

Region	Avalanches	Release km² ATM / CAT	Runout km² ATM / CAT	Skill release	Skill runout	Skill map	With rock: release / map
Davos	26,271	117.9 / 113.9	75.0 / 82.0	87.2	71.2	79.5	86.2 / 78.1
Leventina	18,671	78.6 / 98.8	35.1 / 34.0	67.6	68.6	70.4	66.1 / 67.8
Scuol	27,753	135.0 / 135.8	66.2 / 67.6	84.2	71.0	78.6	82.4 / 75.8
Les Diablerets	17,404	79.2 / 81.8	49.7 / 48.5	82.2	71.2	79.1	81.2 / 77.4
Val de Bagnes	31,992	143.8 / 141.3	76.0 / 83.4	83.7	68.7	77.1	81.6 / 72.8
Engelberg	25,486	111.2 / 109.6	61.3 / 56.8	84.4	71.7	78.5	83.0 / 75.9
Alpstein	12,651	48.4 / 48.5	30.2 / 27.2	82.8	68.9	79.5	82.1 / 78.2

Table 6: ATM 1.0 against CAT on the compared area of Table 2 (terrain data, without rock). Skill score = probability of detection minus probability of false detection, in percentage points. Last column: the same on all cells with terrain data, CAT rock counted as negative of CAT. Version 1.0 of this paper counted this column for Scuol also on cells without terrain data (81.2 / 74.0).

Simulated avalanches by volume class (tiny / small / medium) are, for example, 22,728 / 3,182 / 361 in Davos; no avalanche reaches the large class because of the volume limit. Including rock lowers the skill score by 0.7 to 2.1 pp for release and 1.3 to 4.3 pp for the whole map; the difference comes from false detection on rock, where ATM colours 2.0 to 6.6 km² of release per region. Release areas agree better with CAT than runout. Where the runout differs, and why, is examined in Section 12.3.

Against the development reference of version 0.7 (5 kPa, no forest mask, before the solver corrections), ATM 1.0 colours 0.5 to 1.0 km² more release and 1.5 to 2.3 km² more runout per region; the skill scores change by +0.1 to +0.6 pp for release, +0.4 to +1.6 pp for runout and −0.3 to +0.5 pp for the whole map. Because the release layers differ, the runout scores are computed on slightly different cells and their change is not an isolated effect of the flow (Section 10). In every region, at most one avalanche reaches the time limit of 600 s and the largest relative mass error is below $3 \cdot 10^{-14}$. Between 0.9 and 2.4 % of the avalanches of a region lose mass across the boundary of the computational domain. The domain is the region itself, without a buffer: release areas outside it are not simulated, and runout leaving it is cut. Both can reduce the modelled extent near the edge and limit comparability with CAT, which is computed nationwide. This was measured after the held-back test in two regions, Davos and Val de Bagnes, the latter with the longest runs under ATM 1.0 (that longest run was an artefact of the defect, Section 14). Each was recomputed with ATM 1.0 unchanged on a domain enlarged by 5 km on every side, and evaluated on its original cells. A control run used the same enlarged inputs but no terrain beyond the original edge, so that small differences from resampling the cover layer onto a larger grid are not counted as an edge effect. Against this control, the margin raises the skill score of the whole map by 0.2 and 0.1 points; no SPOT6 lowest point and no accident point changes. The maps differ in 0.17 and 0.19 % of the compared cells, 98 and 99 % of them within 500 m of the original edge and none farther than 2 km. The rule was fixed before the computation (relevant from 1 point); by it the missing margin is not relevant in these two regions. The other regions and Glarus were not recomputed.

12.3 Where the runout differs from CAT

The analyses of this subsection were made during development with the configuration of version 0.7 (deposition threshold 5 kPa, no forest mask) and explain where and why the maps differ; they

are not results of ATM 1.0 unless stated. How far ATM 1.0 differs from that configuration is given below Table 6.

Decomposition. Of the CAT runout on the compared area (399.6 km²), ATM misses 132.6 km² (33.2 %). Each missed cell is assigned to one category: 37.2 % lie in the ATM release layer (a change of class, not uncoloured terrain), 53.2 % are uncoloured but within 100 m of the ATM map and not markedly lower (beside a path; median distance 7 to 15 m, one to three cells), 7.8 % are within 100 m and markedly lower (below the end of a path), and 1.7 % lie farther than 100 m from any ATM cell (a missing path). The thresholds of 100 m and a drop of 0.3 times the distance are our own and only describe; nothing was tuned to them. The difference is thus almost entirely at the edges of paths, not in their length.

Additional start places. Parts of the release layer start no simulation of their own because they are lost in the 10-m delimitation, or are rough, convex or concave. Simulating these parts as additional objects, with unchanged physics and without altering the existing objects, adds up to 178 km² of simulated release area (+25 %) but recovers only 6.5 km² of the missed CAT runout (4.9 %). The lowest points of SPOT6 outlines reached rise by 0.5 pp (95-% interval 0.1 to 0.8). The additional avalanches are small (median travel length 76 m) and mostly run into terrain that is already coloured. Lost delimitation parts alone raise the runout skill score in all seven regions by 0.1 to 1.7 pp; the skill score of the whole map falls in four regions by up to 0.4 pp. At this stage of development no variant was adopted; the final decision is reported in Section 12.5 and Section 17.

Change of class. The 49.4 km² of CAT runout inside the ATM release layer form a fringe of one to three cells around release areas that CAT also maps as release: 89 % lie at most 15 m inside the release layer, 97 % at most 25 m from CAT release, on all sides and slightly more often below. They lie mostly just above the release threshold (75 % at a potential of 0.15 to 0.3, against 6 % of the release cells both maps share) and on moderate release slopes (80 % at 28 to 35°, against 35 %). Only 0.1 % of these cells lie more than 100 m from CAT release; this shows that the fringe is not release terrain far from any CAT release, not that ATM has no additional release objects closer than that. A fringe of one or two 5-m cells is of the order of the generalisation of class boundaries; this pattern is consistent with differences in class-boundary placement and generalisation, but does not by itself identify which boundary is more accurate. Whether the release threshold or the lower flank of the slope membership should change is a separate experiment with its own rule; choosing a threshold from this comparison would tune the release layer to CAT.

Cross-sections and back-calculation. For 527 well-mapped slab avalanches in six regions (SPOT6 outline quality “exact” or “estimated”, size up to 3, natural release, no contact with another outline), a centreline was derived from the outline and the terrain only, independent of any model variant, and cross-sections were placed every 25 m (6,437 usable sections). Each avalanche was recomputed from its own observed release cells. The width, end-point and width-scaling analyses below used the configuration of version 0.7 (5 kPa, no forest mask) and are development results. With ATM 1.0 all 527 avalanches were recomputed again (Section 12.4): 72.7 % reach the observed lowest point and 81.2 % of the cross-sections in track and deposit are reached; runs pass in the median 18.8 m below the observed lowest point and extend 6.3 m laterally beyond the outline. Against the outlines as mapped, the recomputed avalanche is in the median 7.5 m wider than the observed one (observed median width 52.5 m; quartiles of the difference 0 and 22.5 m), with a median lateral offset of 5 m; this holds for both outline qualities and in every region.

The outlines themselves are uncertain. Against georeferenced helicopter imagery, the satellite mapping of 2018 classified 73 % of the avalanche area correctly, missed 16 % and added 11 %, and missed 25 % in shaded terrain (Bühler et al.[20, Tab. 4]). Experts mapping the same avalanches differ more (Hafner et al.[36]). The outline edges of this sample were not delineated independently a second time,

and the following is a sensitivity analysis in the absence of a second mapping, not a correction for mapping uncertainty. We widened or narrowed every observed outline symmetrically by hypothetical width factors derived from these shares under an all-lateral-error assumption: 0.87 to 1.19, and 1.34 for shaded terrain, with 1.23 and 1.56 from the expert study as a check. The published disagreement may also come from displacement, missing parts, other end points or undetected avalanches, not only from uniform under-width.

- Under the assumed net-area width-scaling scenario (factor 1.06), the recomputation stays wider than observed (median per avalanche +5 m, +10 %).
- If all missed area lay at the edges (1.19), it is 2 to 3 % (1 to 2 m) narrower; this cannot be excluded.

The apparent lateral excess is sensitive to plausible but unverified assumptions about outline error. The analysis does not establish a systematic width bias in either direction, and it does not support increasing lateral spreading solely to match CAT. Both maps cover the observed avalanches laterally almost completely at every factor up to 1.56 (upper quartile of the under-coverage 0 m on both sides); the CAT map extends on each side about 20 m farther beyond the outline than the ATM map. The lateral difference of the decomposition therefore cannot be read directly as insufficient spreading of individual simulated avalanches.

End points. The back-calculation misses 17 to 21 % of the cross-sections in the track and the deposit. 83 % of these misses are runs that stop too early, not runs on another path, and they concentrate on a minority: 21 % of the avalanches end above the observed lowest point and carry 94 % of the missed sections, most of them in the smallest volume class. The majority runs farther than observed (median 17 m lower; about 100 m for the largest class in the sample) and places 38 % of its area outside the outline. Small avalanches thus tend to stop short and large ones to run too far. The observed volume is unknown and the release of the back-calculation is derived from the outline with a fixed depth, so both deviations may also stem from the released mass.

12.4 Forest mask and sign of the gap term

Before the mask (up to version 0.8; development state). The equation was evaluated in every cell with terrain data. Treeless terrain was not protected because it forms a gap whose width along the contour reaches the search limit, which keeps P far above the threshold; protection on cells with little or no cover occurred only in narrow gaps and in gaps below 500 m² (width 0). In the seven regions, 0.2 to 0.7 km² per region were protected although the cell itself had 0 % cover, and 0.15 to 0.41 km² with less than 10 % mean cover of cell and window, out of 66 to 221 km² of protected terrain. These figures describe the state without mask, not ATM 1.0.

With the mask. With the Copernicus forest-type layer (FAO definition, 10 m) as forest mask, small enclosed holes below 0.5 ha counted as forest, protection outside the mask is no longer applied. With the modified positive gap coefficient, the mask changes little: the release skill score against CAT is equal or slightly higher in all seven regions (+0.1 to +0.6 pp), and no treeless cell outside the mask is protected any more (before 1.5 km²). With the printed negative sign inside the mask, more terrain in gaps is protected, and the skill score falls in all seven regions (−0.5 to −2.4 pp); inside the mask almost the whole release layer is lost. Accidents in the release layer: +0.0 pp (interval −0.4 to 0.4) and −0.5 pp (−1.1 to 0.1). The printed sign finds no support here.

The forest mask also removes forest friction outside the mask. With the full chain and the reference flow settings, it was compared against no mask on the same six measures as the flow candidates. On none of the six measures does the paired 95-% interval lie entirely above or below zero (lowest point reached +0.2 percentage points, −0.4 to 0.8; runs above 1,000 m +24 of 2,364, −111 to 159, with 1 % more avalanches; both median over-extensions identical); by the selection rule of Section 12.5

neither is dropped, pooled and in every region. By the rule for an open front the forest mask is adopted, since the source applies the forest model only within a mask; its adoption rests on this structural approximation to the original procedure, not on a gain in performance. Its visible effect is on the release layer; in the evaluated flow measures no clearly detectable difference was found, which is not the same as no effect on the flow. The rule for an open front thus decided twice: for the deposition threshold it was written after the first comparison was known, for the forest mask before the comparison. Both remain development decisions on regions already used.

Sensitivity to the gap term. After a review of the sources, the whole chain was run again in the seven regions with three other sets of coefficients; threshold, mask and everything else as in ATM 1.0. All sets keep the logistic form of the forest equation and are continuous in slope, cover and gap width; the map uses only the boundary at the index of 65. The rule was written before the runs: an alternative is relevant if the whole-map skill score or the low points, pooled, change by at least 1 percentage point in absolute value.

The printed set differs only in the sign. *Bebi 1999, adapted* takes the coefficients of Bebi [27, Tab. 19] but omits its convex-slope term, which ATM does not have, and applies the threshold of 2021 to a model that did not use it. The last set is an approximate reconstruction from the figure attributed to Schneebeli & Bebi [29], reproduced in Krumm & Bebi [30]; we have not inspected the original. The figure gives critical gap width against crown cover for slopes of 30, 35, 40 and 45°, as straight lines over about 20 to 90 % cover. We read 28, 20, 12 and 5 m at 50 % cover, about 0.5 m per percent cover and about 1.6 m per degree, and so used $w = 28 + 0.5(c - 50) - 1.6(s - 30)$ with a reading accuracy of about 1 to 2 m, estimated from the grid of the figure. Two choices are ours: the figure gives no probability, so we placed this line at the index of 65, and the gap coefficient of 0.105 is taken from Bebi 1999, which sets the other coefficients but not the boundary. Outside the range of the figure the line is extrapolated, with slope held within 28 to 48° as in ATM 1.0. The line gives 4 m at 45° against the 5 m read.

Table 7 gives the definitions and results. The rule did not say how the skill score is aggregated across regions. We therefore report both the score computed from the contingency counts summed across all regions and the unweighted mean of the seven regional scores; the low points are pooled over all events. The limit is not reached either way (largest change −0.3 and −0.4 pp, both for the printed sign).

No alternative crossed the threshold for the predefined aggregate measures; individual regional changes exceeded 1 percentage point (printed sign: −1.7 pp in Leventina, −1.6 pp in Alpstein). ATM 1.0 stays as frozen. However, the alternatives substantially changed mapped release terrain in forest (0.8 to 28.9 km²) and some forest-specific coverage measures (69.1 to 80.0 % of forest cells in outlines), and with the printed sign four of 19 accidents in forest leave the release layer, with a wide interval (−21.1 pp, −41.2 to 0.5). These results do not identify the correct coefficient set and do not validate the retained coefficient of +0.05. Aggregate measures respond weakly here, because they are diluted by the large area and the many events outside forest.

	ATM 1.0	Printed sign	Bebi 1999, adapted	Reconstruction from figure, 2004
Intercept	−6.17	−6.17	−5.784	−4.736
Coefficient of slope (per °)	0.18	0.18	0.153	0.168
Coefficient of cover (per %)	−0.03	−0.03	−0.051	−0.0525
Coefficient of gap width (per m)	+0.05	−0.05	+0.105	+0.105
Unprotected from, 30° (m)	58	none	42	28
35°	40	none	34	20
40°	22	none	27	12
45°	4	none	20	4
Release in forest (km²; CAT 88.5)	22.9	0.8	22.0	28.9
Runout in forest (km²; CAT 19.1)	63.5	56.1	63.1	65.2
Whole-map skill, pooled cells	78.9	78.6	78.9	78.9
mean of regions	77.5	77.1	77.5	77.5
change per region (pp)		−1.7 to +0.7	−0.2 to +0.1	−0.3 to +0.4
SPOT6 low points (n = 1,771; pp)		−0.1 (−0.4; 0.1)	+0.1 (−0.2; 0.3)	+0.1 (−0.1; 0.4)
low point in forest (n = 65; hits)	62	62	62	62
Forest cells in outlines covered (%)	77.8	69.1	77.1	80.0
Accidents in release (n = 885; pp)		−0.5 (−1.1; 0.1)	0.0 (−0.4; 0.4)	0.0 (−0.4; 0.4)
point in forest (n = 19; lost)		4	0	0

Table 7: Sensitivity to the gap term in the seven development regions, on the compared area. Paired differences against ATM 1.0 with 95-% intervals (Newcombe). “Unprotected from” is the gap width at which the index reaches 65 at 50 % cover; the Bebi 1999 model omits its convex-slope term and uses a threshold from 2021. All values are development results.

The tested gap-term alternatives do not account for most of the discrepancy in mapped forest release area between ATM and CAT, where CAT has about four times the release (88.5 km²) and a third of the runout (19.1 km²) of ATM. This does not mean that the gap term is unimportant in forest. A missing SPOT6 outline under the canopy proves nothing, CAT is a reference map and not an observation of events, and the alternative sets come from the same forest avalanches, so they are no independent confirmation. Whether additional release or runout in forest is too much has not been measured.

Forest against CAT. After this result we tabulated where the two maps differ inside the forest mask; “forest” in this paragraph means that mask, not closed canopy. This is descriptive and was not governed by a rule set beforehand; it decides nothing and changes nothing in ATM 1.0.

Crown cover	Forest (km²)	Release, CAT (km²)	Release, ATM (km²)
Below 30 %	14.8	4.4	6.0
30 to 50 %	30.3	9.2	10.0
50 to 70 %	87.5	22.2	6.6
70 % and above	724.9	52.7	0.3
All forest	857.4	88.5	22.9

Table 8: Release inside the forest mask by Copernicus crown cover, seven development regions. Descriptive.

More than the whole difference in release lies at 50 % cover and above (Table 8); below that, ATM has slightly more release than CAT. At 70 % cover and above lies 60 % of the release of CAT, while ATM has almost none, as the forest equation implies. Harvey et al.[1] do not consider dense forest as release terrain, and Harvey et al.[2] still describe forest as either forested or non-forested. If that still holds, these cells are not dense forest in the forest layer of CAT, although the cover layer used here says they are. We do not have that layer, so this remains an inference. The table is an overlay with our cover layer; it does not show that CAT uses the same cover or forest state and rates it differently, nor which layer is closer to the forest on the ground. The difference does not prove that ATM models too little release in forest.

Runout differs in the opposite direction. Of the 13,427 release patches of CAT of at least 200 m² lying at least 95 % in forest, 53 % have runout within 100 m below their lowest point, against 82 % of patches in open terrain; for ATM the figures are 75 % and 85 %. We do not know why. In the SPOT6 outlines, ATM reaches 62 and CAT 42 of the 65 low points located in forest, and ATM covers 77.8 % of the forest cells inside outlines, CAT 67.7 %, with less coloured forest in total (86.3 against 107.6 km²). ATM thus covers a larger share of the mapped forest portions and more lowest points located in forest in this optically observable event sample than CAT. This does not establish performance beneath closed canopy or independently validate the forest-flow formulation: visibility limits the sample, the map combines release and runout, the result depends on release geometry and simulations from above, and more coverage does not show that all additional cells are right.

The same tabulation was read for Glarus after its evaluation and with the same pattern (66 % of the release of CAT in forest at 70 % cover and above; forest patches with runout below, CAT 54 %, ATM 79 %; low points in forest reached, ATM 23, CAT 11 of 29). The original held-back primary analysis remains unchanged (Section 13). This forest-stratified analysis is post hoc and does not contribute to its verdict. Glarus is no longer unexamined data for future model development.

12.5 Flow ablation and the choice of cohesion and deposition

This experiment uses the development reference: cohesion 50 Pa, deposition threshold 5 kPa, no forest mask, corrected solver, on frozen release objects (158,953 in seven regions; ATM 1.0 has 160,228, because its forest mask enlarges the release layer). No run reaches the time limit, and the largest relative mass error is below 4e−14. Paired per avalanche against this reference, the median ratios of affected area and travel length are 1.057 and 1.054 without cohesion, 1.030 and 1.022 with a deposition threshold of 2 instead of 5 kPa, and 1.172 and 1.181 without deposition. Without forest in the flow, only avalanches passing forest change (travel length 1.237); the forest stays a release barrier. Lowest points of SPOT6 outlines in the map rise by 1.8, 0.6, 2.0 and 0.5 pp (all intervals above zero); accidents do not change measurably.

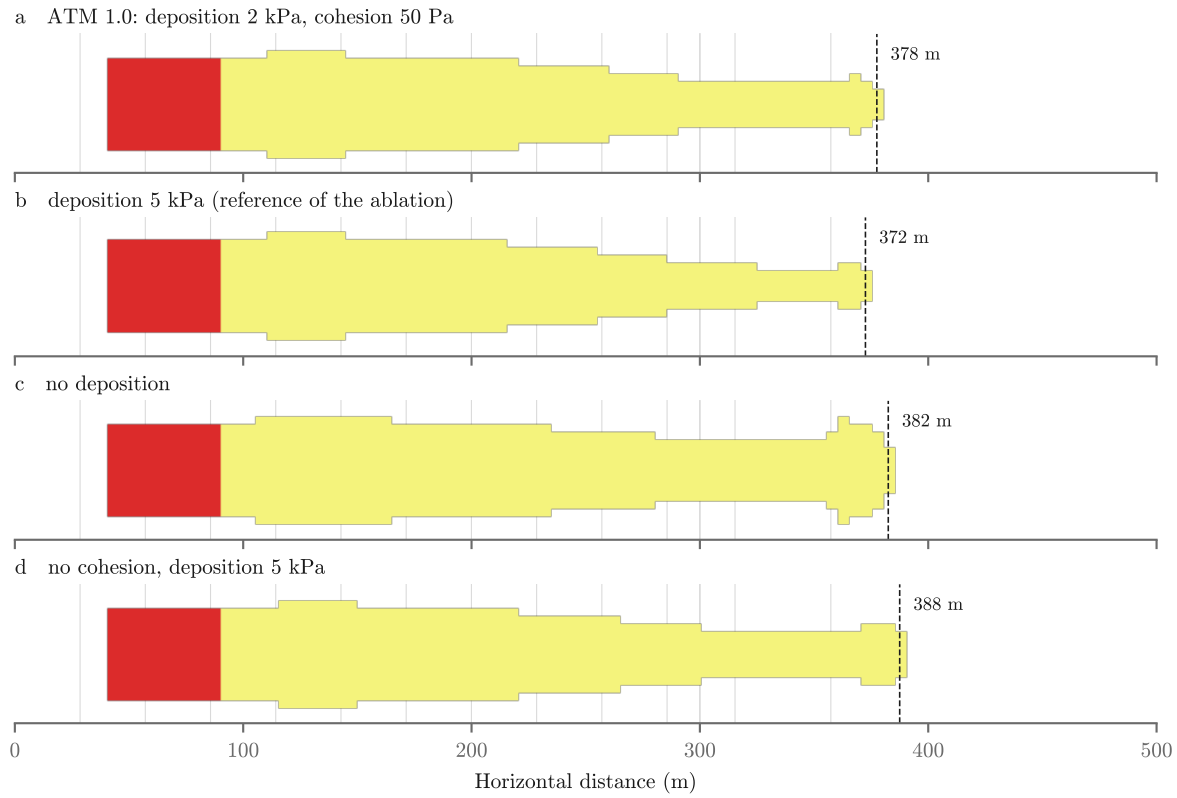


Figure 10: *Illustration, not evaluation.* The block of Figure 5 with the reference and the flow candidates of the ablation: (b) the development reference, cohesion 50 Pa, deposition 5 kPa; (a) deposition 2 kPa, which is ATM 1.0; (c) no deposition; (d) no cohesion. Each candidate changes one setting of the reference. Dashed: furthest outline cell. The differences in this idealised example are not representative of real terrain; the development comparisons are those reported in the text above.

The choice between cohesion and deposition values used six measures fixed in advance: on the back-calculated sample of Section 12.3, the share of observed lowest points reached and of cross-sections in track and deposit reached (benefit), the median height by which a run passes below the observed lowest point and the median lateral extension beyond the outline (over-extension); on the map, the share of SPOT6 lowest points inside (benefit) and the share of travel lengths above 1,000 m (design size); their computation is given in Table 9. A candidate is dropped if, against another candidate, the paired 95-% interval of at least one measure lies entirely on the favourable side of zero and none lies entirely on the unfavourable side. This is a pragmatic development selection rule, not a test of Pareto dominance: an interval that contains zero counts as neither better nor worse, which does not show that the candidates are equal on that measure. The measures are coupled — lower end points, longer runs and more cross-sections reached go together — and are not six independent confirmations. No correction for multiple comparisons was applied; the rule makes no inferential claim beyond the development choice.

Measure	Unit	Statistic	Paired 95-% interval
Lowest point reached	back-calculated avalanche	share	Newcombe, method 10
Sections reached	cross-section in track and deposit	share	Newcombe, method 10; sections of one avalanche treated as independent
Passes below	back-calculated avalanche	median height below the observed lowest point, 0 if above	bootstrap over avalanches, 2,000 draws
Lateral beyond	cross-section	median of the mean extension on both sides	bootstrap over avalanches with all their sections, 2,000 draws
SPOT6 lowest point in map	outline	share	Newcombe, method 10
Travel above 1,000 m	simulated avalanche	share (count in later re-evaluations)	bootstrap over simulated avalanches, 2,000 draws

Table 9: Measures of the flow selection. All are pooled over the regions without weighting, so regions count by their number of items. Decision boundary: zero difference; there is no tolerance.

Candidate	Lowest point reached (%)	Sections reached (%)	Passes below (m)	Lateral beyond (m)	SPOT6 lowest point in map (%)	Travel > 1,000 m (%)
Cohesion 50 Pa, 5 kPa	70.0	80.5	17.0	5.0	91.0	1.31
No cohesion	74.8	82.9	22.5	6.3	92.8	1.75
Deposition 2 kPa	72.5	81.2	18.8	6.3	91.6	1.49
No deposition	75.9	83.3	23.9	7.5	93.0	4.14

Table 10: Pooled measures of the flow candidates on frozen release objects without forest mask. First row: development reference; third row: the selected flow settings in this experiment. The full ATM 1.0 configuration additionally includes the forest mask, with which 72.7 % of lowest points are reached (Section 12.3).

No deposition was not pursued under the selection rule: its additional benefits over no cohesion were not clearly supported on the other measures, whereas the number of long runs was more than twice as high. This is a development selection decision; it does not establish statistical dominance or equivalence on the remaining measures. The former configuration, no cohesion and the 2-kPa threshold remain: each step towards higher mobility reaches more observed end points and passes farther beyond them, and the rule drops none of them. By the rule, **the choice stays open**. For this case a further rule was fixed before the grid tests of Section 12.12, but after these results were known: among the remaining candidates, the one closest to the sources for the computed scenario becomes the reference. Among the retained non-dominated candidates, 2 kPa was selected provisionally because its source application, a back-calculated dry avalanche, is closer to the intended scenario; cohesion of 50 Pa is kept as in the scenario. This does not establish overall superiority, and an open front is not a set of candidates shown to be equivalent. No cohesion and the former 5-kPa configuration are shown as sensitivities. An earlier incomplete flow ablation, in five of the seven regions and before the solver corrections, was superseded by this seven-region analysis; it is retained in the development record and is not used as evidence here. In every region the former configuration remains on the front. Most of the CAT runout additionally reached by the mobile candidates lies beside the paths, not below them. Because small avalanches tend to stop short and large ones to

run too far, a single change of mobility for all sizes moves both ends the same way; a size-dependent formulation would need its own rule and a justification from the literature.

12.6 Release geometry

This experiment uses the development reference (deposition threshold 5 kPa, no forest mask). Release objects above the volume limit are divided either into elevation bands of equal volume across the whole object (as in ATM 1.0) or into the same number of compact pieces of equal volume, by repeated division at the volume median perpendicular to the longer principal axis. Each variant was run with and without the additional start places lost in the delimitation, all four from one code state. Compact pieces do not flow wider: the median ratio of mean width (area over travel length) per divided object is 0.98 to 0.99 in every region. They reduce the CAT runout missed beside the paths by 2.6 km² (3.7 %) and raise the runout skill score by 0.2 to 1.0 pp in all seven regions. They also raise the number of runs above 1,000 m by 11.5 % (a relative increase of the count; the share rises by 0.17 percentage points): 80 release objects gain such a run and one loses it, and objects that already had one get more long pieces. Observed lowest points do not change measurably (+0.2 pp, interval −0.1 to 0.5). About a third of the effect on the lateral margin comes from pieces that change volume class and hence friction; a diagnostic run in Davos with friction classes held fixed keeps the other two thirds. The effects of geometry and start places add almost exactly. By the rule fixed in advance, compact pieces are not adopted. The share of runs above 1,000 m proved unsuitable for comparing variants with different numbers of avalanches. The additional start places add many short avalanches and lower the share without changing the count. Re-evaluated with the count and with the number of release objects that have a long run:

- the start places lost in the delimitation add no long run at all;
- all excluded terrain together adds 331, all from new release objects.

The decisions do not change. ATM 1.0 keeps elevation bands.

12.7 Size diagnostics

The range of Huber et al. selects release areas of 10,000 to 100,000 m². Measured on the simulated release objects, that is after the division into elevation bands, this is a small part of all simulations (Table 11): most objects are much smaller and, as expected, run less than 200 m.

Release area (m²)	Simulations	Share	Travel length 200–1,000 m
200–500	48,019	30.0 %	0 %
500–1,000	34,575	21.6 %	4 %
1,000–10,000	63,321	39.5 %	51 %
10,000–100,000	14,313	8.9 %	84 %

Table 11: Simulated release objects of all seven regions (160,228), by release area. Median area 925 m².

In the selection of Huber et al. (Davos 2,485, Leventina 1,436, Scuol 2,605, Les Diablerets 1,710, Val de Bagnes 2,932, Engelberg 2,240, Alpstein 905 objects), 79 to 95 % have a travel length of 200 to 1,000 m (Davos 88.4, Leventina 78.8, Scuol 79.2, Les Diablerets 86.6, Val de Bagnes 81.9, Engelberg 84.3, Alpstein 94.5 %); almost all of the rest are longer. Their affected area lies within 10,000 to 80,000 m² for 62 to 74 %. Over all regions, 2,388 runs exceed 1,000 m. Huber et al. report 57 % within the travel-length range for Flow-Py in the Austrian Alps (Huber et al. [12, Tab. 4]). This cross-study comparison is contextual only: differences in terrain, release delineation, design assumptions and measurement definitions prevent reading these percentages as a ranking of the runout models.

12.8 Coverage of mapped avalanches up to size 3 during two extreme periods

Region	Period	n	Both	ATM only	CAT only	None	In map (%) ATM / CAT	Δ pp	95 % interval	Outside (%) ATM / CAT
Davos	2018	548	489	22	26	11	93.2 / 94.0	−0.7	−3.3, 1.8	2.8 / 2.1
Davos	2019	181	153	4	12	12	86.7 / 91.2	−4.4	−9.2, 0.1	8.6 / 7.0
Scuol	2018	141	116	5	13	7	85.8 / 91.5	−5.7	−12.1, 0.4	5.1 / 4.0
Scuol	2019	67	54	6	5	2	89.6 / 88.1	1.5	−9.2, 12.2	4.0 / 4.7
Les Diablerets	2018	87	84	2	0	1	98.9 / 96.6	2.3	−2.7, 8.3	1.5 / 1.3
Val de Bagnes	2018	216	181	2	18	15	84.7 / 92.1	−7.4	−11.9, −3.4	6.0 / 4.0
Engelberg	2018	186	168	9	3	6	95.2 / 91.9	3.2	−0.7, 7.6	5.3 / 6.3
Engelberg	2019	285	261	10	5	9	95.1 / 93.3	1.8	−1.1, 4.8	6.1 / 5.5
Alpstein	2018	4	3	0	1	0	75.0 / 100.0	–	–	3.0 / 0.8
Alpstein	2019	56	52	1	1	2	94.6 / 94.6	0.0	−8.2, 8.2	4.1 / 2.8
All	2018	1,182	1,041	40	61	40	91.5 / 93.2	−1.8	−3.5, −0.1	4.1 / 3.3
All	2019	589	520	21	23	25	91.9 / 92.2	−0.3	−2.6, 1.9	6.4 / 5.7
All	both	1,771	1,561	61	84	65	91.6 / 92.9	−1.3	−2.7, 0.0	4.7 / 4.0

Table 12: Lowest point of each outline inside the map: counts for both maps, only ATM, only CAT and none; shares; area of the outline outside each map; paired difference with a score interval for paired proportions (Newcombe[34], method 10), none below ten outlines. The intervals account for the pairing of ATM and CAT, not for spatial dependence or repeated avalanche paths; they are descriptive and conditional on the evaluated event sample. Leventina lies outside the imagery; Les Diablerets and Val de Bagnes have no outlines of 2019.

Pooled over both periods, the paired difference is −1.3 pp; the lower bound of the interval, −2.7 pp, lies below the provisional tolerance of −2 pp: **non-inferiority is not shown**. The development reference of version 0.7 gave −1.9 pp (−3.3 to −0.5); that value is superseded and is not a result of ATM 1.0. Leaving out one region at a time gives differences between −0.5 (without Val de Bagnes) and −2.6 pp (without Engelberg). The deficit is concentrated in Val de Bagnes, Scuol 2018 and Davos 2019; Engelberg reaches more lowest points than CAT. Where a lowest point lies outside ATM, its median straight distance to the map is 21 m (CAT: 20 m). These distances matter for interpretation but show neither harmlessness nor correctness. Figure 11 shows the differences together with those of the held-back test.

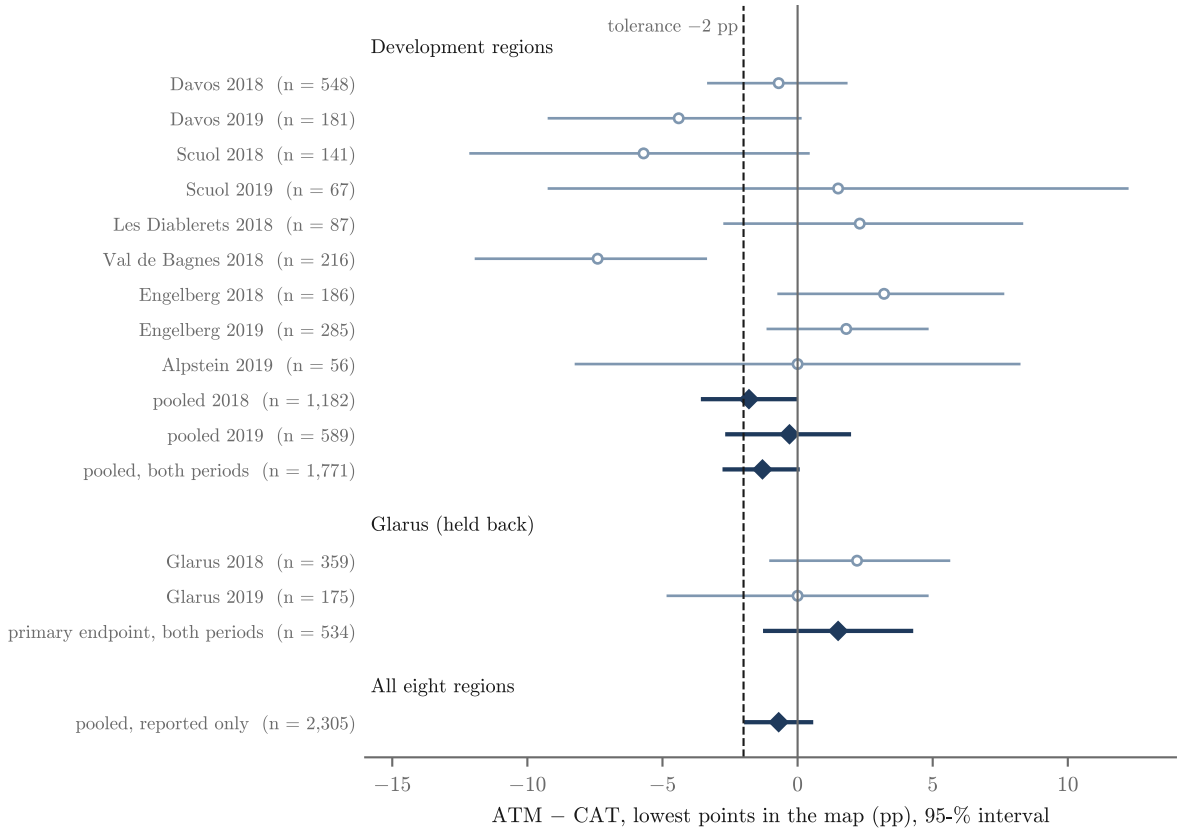


Figure 11: Lowest point of SPOT6 outlines in the map, ATM minus CAT, with paired 95-% intervals (Newcombe, method 10), from the frozen evaluation results. Development regions (Table 12; Alpstein 2018 with four outlines has no interval) and Glarus (Section 13) are kept apart; the pooled value over all eight regions is reported only. Dashed: the provisional tolerance of −2 pp. The intervals do not account for spatial dependence.

12.9 Mapped release areas (Davos)

Against the release areas mapped in three ski areas (Bühler et al.[24]): ATM release detection 98.1 %, false detection 24.3 %, skill score 73.8; CAT release classes 95.6 %, 20.8 %, 74.9.

12.10 Accidents

885 accidents in seven regions. **Fixed radii (sensitivity).** The release layer contains the start-zone point for 83.3 % (CAT 82.6 %), within 50 m for 99.0 % (CAT 98.8 %). ATM colours 26.5 % of the compared area of Table 2 as release, CAT 27.1 %; the release areas of Table 6 refer to the same area.

Main radius. For each point, the radius is its documented accuracy: 0 m (the cell of the point) for 699 accidents with a recorded accuracy of 0 m, 100 m for 186; a recorded accuracy of 0 m does not mean that the true release position is known without error. Each hit rate is set against the uniform-area baseline of Section 10 (Table 13).

Selection	n	Hit (%) ATM / CAT	Both / ATM / CAT / none	Δ, pp (95 % interval)	Baseline (%) ATM / CAT
Davos	438	84.7 / 84.2	361 / 10 / 8 / 59	0.5 (−1.6, 2.5)	35.8 / 35.1
Leventina	9	100 / 100	9 / 0 / 0 / 0	–	42.1 / 50.3
Scuol	34	88.2 / 88.2	29 / 1 / 1 / 3	0.0 (−12.2, 12.2)	46.2 / 46.5
Les Diablerets	68	97.1 / 89.7	61 / 5 / 0 / 2	7.4 (0.2, 16.3)	36.6 / 38.8
Val de Bagnes	191	88.5 / 88.0	164 / 5 / 4 / 18	0.5 (−3.0, 4.1)	50.6 / 50.2
Engelberg	86	90.7 / 90.7	77 / 1 / 1 / 7	0.0 (−5.2, 5.2)	42.6 / 42.0
Alpstein	59	81.4 / 83.1	46 / 2 / 3 / 8	−1.7 (−10.6, 7.0)	23.3 / 23.4
All	885	87.1 / 86.3	747 / 24 / 17 / 97	0.8 (−0.7, 2.3)	39.4 / 39.1
Recorded accuracy 0 m	699	83.8 / 83.0	563 / 23 / 17 / 96	0.9 (−1.0, 2.7)	29.3 / 28.7
Recorded accuracy 100 m	186	99.5 / 98.9	184 / 1 / 0 / 1	0.5 (−1.8, 3.2)	77.1 / 78.1

Table 13: Start-zone points of accidents in the release layer at the main radius. Interval as in Table 12. Baseline = uniform-area baseline: expected hit rate of points placed uniformly in the compared area with the same radii; a geometric reference, not a model of exposure.

Both maps hit the start-zone points about twice as often as the uniform-area baseline. The estimated paired difference is +0.8 pp; the nominal 95-% interval, −0.7 to 2.3 pp, includes zero. Leaving out one region at a time gives +0.2 to +1.1 pp. No equivalence or non-inferiority conclusion is drawn. The test concerns the mapped release layer, of which 25.8 % within the compared area starts no simulation (Section 8.2); it does not test the simulated release objects. The 186 points with a recorded accuracy of 100 m lie in both maps almost always, at a baseline above 76 %, and carry little information. For release or runout terrain, ATM reaches 90.1 % and CAT 91.8 % (−1.7 pp, −3.2 to −0.3; baseline 54 and 55 %). No tolerance for accident points has been set; these results are descriptive.

12.11 Grid-phase envelope

For ATM 1.0, the union over the four placements reduces the CAT runout missed by ATM by about 10 to 13 % per region (9.6 to 12.6 %). The skill score of the runout layer rises by +1.2 to +2.6 pp, that of the whole map changes by −0.2 to +0.7 pp. The share of travel lengths of 200 to 1,000 m in the selection after Huber et al.[12, Tab. 2] changes by at most −0.7 pp in any single placement. These numbers describe the sensitivity to the grid origin; the envelope is not the main map.

12.12 Numerical robustness

On a subset fixed before the computation (100 avalanches on open slopes and 100 in channels in Davos, spread evenly over the volume classes present), each avalanche was recomputed on its window at 5 m, at 2.5 m (terrain refined bilinearly, release cells split, same depth) and with the grid rotated by 30° and 45°, for the configuration of the time (5 kPa, no forest mask) and the two open flow candidates. Outline area, travel length and mean width are defined invariantly under a pure geometric rotation and are therefore compared directly; the computed values are not invariant, because the solver depends on the grid direction. Rotation changes nothing in the median (ratios 1.000 to 1.003); single avalanches scatter by about ±10 % in area and ±15 % in travel length (10th to 90th percentile). Refinement to 2.5 m lengthens the runs systematically, by 6 % in the median (open slopes 7 %, channels 5.5 %) and by almost half for one avalanche in ten, and narrows them slightly (−2.4 %). Travel lengths at 5 m are therefore not independent of the grid; this is a change between two grids, not a measured error against the true runout.

The effects of the candidates are robust by the rule fixed in advance: without cohesion, area and travel length grow by 5.7 to 5.9 % and 5.4 to 6.0 % in all four settings, with the lower deposition

threshold of 2 instead of 5 kPa by 1.8 to 1.9 % and 1.3 to 1.7 %. Their effect on width is below 1 % already at 5 m and changes sign; by the rule it is numerically weak. Both candidates make avalanches longer, not wider. The resolution effect on travel length is as large as the effect of removing cohesion, so part of the runs that stop short of observed lowest points may be due to resolution.

A second test, with its rule fixed in advance, compared three grids (5, 2.5 and 1.25 m) on 200 avalanches drawn as before, with the configuration of the time (5 kPa, no forest mask), grouped by release area (48 below 1,000 m², 42 up to 10,000 m², 110 above). A release-area class counts as converging if the change from 2.5 to 1.25 m is at most half the change from 5 to 2.5 m or at most 1 %; 2.5 m would have become the computational grid if the travel length converged in the two classes above 1,000 m². In the median, halving the cell size lengthens the runs by 6.3 % and halving it again by 2.6 %; for release areas below 1,000 m² by 30 and 14 %, from 1,000 to 10,000 m² by 10 and 4 %, above 10,000 m² by 3.0 and 1.8 %. These are changes between grids, not measured errors against the true runout; the 1.25-m solution is not the truth either. The class above 10,000 m² misses the criterion narrowly. The predeclared criterion for replacing the 5-m production grid was not met. **The 5-m grid is retained provisionally; this does not establish superior accuracy or grid independence at 5 m.** By the same criterion, width changes little in all classes and outline area keeps changing in all. The criterion compares class medians and is sensitive when the first step is already small; a later version should judge convergence behaviour, practical size in metres and the spread over single avalanches separately. In this computation the initial volume differed between grids by up to 0.47 %, because the inclined cell areas of the release change with refinement. A control fixed in advance repeated it with the release depth scaled so that the initial volume on every grid equals the 5-m volume (relative difference below 10^{-13}), for 5 kPa and for the 2-kPa threshold of ATM 1.0. The verdict is the same for both, and the class medians change by less than 0.4 pp; the numbers above are those of the control. The control was still running when ATM 1.0 was frozen. Its 5-kPa part was first run with an unintended 2-kPa setting and was corrected after the freeze and after the Glarus evaluation. The correction changed neither ATM 1.0, nor the held-back evaluation, nor the grid decision, which rested on the three-grid test above; it is a subsequent check of the numerical analysis.

A third test asked whether the candidates only compensate for the coarse grid. The 527 observed avalanches were back-calculated in their own windows at 5 and 2.5 m with each candidate. The refinement alone reaches 5.9 percentage points more observed lowest points and passes 5.7 m farther below them; its effect is about as large as removing cohesion. The gains of the candidates shrink on the finer grid (no cohesion from 4.7 to 3.2 points, 2-kPa threshold from 2.5 to 1.5) but remain supported; by the rule fixed in advance their effect carries over to the finer grid. No setting dominates another on either grid.

13 Held-back test in Glarus

ATM 1.0 was computed once in Glarus, from the frozen code and configuration, after the manifest of the frozen version had been written and with the protocol of Section 10 unchanged. No technical failure occurred; nothing was recomputed. The region is described in Table 2. 24,492 avalanches were simulated; the largest relative mass error is below $3 \cdot 10^{-14}$, and 1.4 % of the avalanches lose mass across the edge of the computational domain (development regions 0.9 to 2.4 %).

Primary endpoint. Of 534 outlines (359 of 2018, 175 of 2019), the lowest point lies inside the ATM map for 90.8 % and inside the CAT map for 89.3 % (both 456, only ATM 29, only CAT 21, neither 28). The paired difference is +1.5 pp with a 95 % interval of −1.2 to +4.2 pp. The lower bound lies above the provisional tolerance of −2 pp: by the rule fixed in advance, **ATM meets the**

predeclared nominal non-inferiority criterion relative to CAT for the pooled events of both periods in Glarus, and no more. The criterion was met for the pooled primary analysis, not separately for both periods: 2018 +2.2 pp (−1.0 to +5.6), 2019 0.0 pp (−4.8 to +4.8) (Figure 11). The interval accounts for the pairing of the two maps, but not for spatial dependence or repeated avalanche paths, and one region allows no estimate of the spread between new regions. For these 534 events the difference of +1.5 pp is a descriptive property of the evaluated set; any wider statement depends on assumptions about which events are independent. It is no evidence that ATM is better than CAT. Pooled with the seven development regions (2,305 outlines), the difference is −0.7 pp (−1.9 to +0.5); this pooled value is reported as agreed but does not carry the verdict. The tolerance itself is not validated, and Glarus does not test the southern slope, low forested valleys, other snow climates or a new time period.

The evaluated events are those lying almost fully inside the region with terrain data and outside the rock masks (Table 4). In Glarus the rock masks remove more than half of the visible outlines of size 3 at most (612 of 1,174, against 833 of 2,689 in the development regions), the region edge 28. Avalanches touching rock are therefore under-represented; without the rock masks the criterion is still met (Section 16).

Measure	**Glarus**	**Development regions**
Skill score release / runout / map (pp)	83.0 / 70.1 / 77.6	67.6–87.2 / 68.6–71.7 / 70.4–79.5
Same with CAT rock as negative, release / map	81.7 / 75.2	66.1–86.2 / 67.8–78.2
Release / runout area ATM, CAT (km^2)	100.9, 101.0 / 56.3, 55.5	–
Outline area outside the map, ATM / CAT (%)	4.0 / 6.1	1.5–8.6 / 0.8–7.0
Selection after Huber: travel 200–1,000 m (%)	90.0	78.8–94.5
Selection after Huber: area in size-3 range (%)	71.3	62.3–74.3
Runs longer than 1,000 m	191	–
Avalanches lost across the edge (%)	1.4	0.9–2.4
Avalanches stopped by the time limit	4	0–1
Accident points, main radius: ATM / CAT (%)	82.8 / 82.8	–
Accident points: Δ (pp), 95 % interval	0.0 (−13.4, 13.4)	−1.7 to +7.4
Grid placements: reduction of missed CAT runout (%)	11.8	9.6–12.6
Grid placements: change of map skill score (pp)	+0.2	−0.2 to +0.7

Table 14: Secondary measures in Glarus beside the range of the seven development regions from ATM 1.0; descriptive, without tolerance or verdict. Accident points: 29 points (23 with a recorded accuracy of 0 m, 6 of 100 m), baseline 37 %; with fixed radii of 0 and 50 m ATM reaches 82.8 and 100 %, CAT 79.3 and 100 %.

The secondary measures lie inside the ranges of the development regions, with one exception: four simulations reached the time limit of 600 s, against at most one per development region. Such a run ends at the limit, and the peak values reached until then enter its outline; whether further runout would have followed is unknown. Their influence on mapped extent and end-point coverage was not quantified in Glarus. They remain included in the frozen evaluation; no model change or rerun was

made. Section 14 reports a later examination of such runs elsewhere. With 29 points, the accident comparison allows no conclusion; its interval spans ±13 pp.

14 Correction: ATM 1.0.1

The defect. Small pieces of a release area are meant to join the nearest large avalanche of the same area. In ATM 1.0 a piece whose nearest large cell lay in a *different* area went to a remainder object instead, even when its own area had large avalanches. The defect entered three days before ATM 1.0 was frozen and so affects all its runs. It was found during the national computation, when one object spanned several kilometres. ATM 1.0.1 corrects the assignment. The intended rule and all values are unchanged. Results of ATM 1.0 stay in this paper as a record of development, marked as computed with the superseded implementation; an earlier recommendation to publish ATM 1.0 no longer stands unqualified.

Rule and tests. A piece joins only an object of its own area; no cell is lost or duplicated, each has one target, ties resolve deterministically, area and volume are conserved, and the behaviour without a large object is explicit. A regression case has a foreign large avalanche nearer than the area's own. Pieces of an area with no large avalanche still form one remainder object. This is a remaining model assumption, not an established property: it can join pieces far apart. Splitting such objects by connectivity would be a method decision of its own, and is not part of this correction.

Re-evaluation on seen regions. All eight regions were recomputed. Glarus was previously evaluated with ATM 1.0. The ATM 1.0.1 computation is a post-correction comparison on a seen region, not a new held-back evaluation. One percentage point was a threshold for closer inspection, not a rule for accepting the correction. The quantities compared at the map and observation level were fixed before these runs. The span and one-piece measures of objects were defined afterwards and are post-hoc diagnostics of the defect, as is the count of omitted pieces.

Region	Map cells differing	Runout cells gained / lost	Objects	Objects changed	Longest run (m)	Widest object in pieces (m)
Davos	899	649 / 250	26,271 → 26,248	610	2,156 → 2,156	1,770 → 745
Val de Bagnes	848	806 / 42	31,992 → 31,929	1,057	3,445 → 2,599	3,855 → 1,245
Scuol	385	366 / 19	27,753 → 27,738	445	2,511 → 2,511	1,510 → 1,100
Leventina	172	158 / 14	18,671 → 18,662	266	2,613 → 2,613	750 → 710
Engelberg	269	256 / 13	25,486 → 25,469	391	1,994 → 1,994	865 → 750
Les Diablerets	289	245 / 44	17,404 → 17,396	330	2,084 → 2,084	1,325 → 785
Alpstein	181	171 / 10	12,651 → 12,639	244	1,571 → 1,571	750 → 750
Glarus	878	362 / 516	24,492 → 24,475	454	2,250 → 2,250	1,220 → 875

Table 15: ATM 1.0 → ATM 1.0.1. Map cells differ in at most 0.006 % of the cells with terrain.

In all eight regions, the release potential and release layer are identical cell by cell. Observation hit classifications are identical for the evaluated SPOT6 lowest points and for the accident points at all radii; the share of an outline outside the map changes for 14 of 2,305 outlines. Map-comparison rates are unchanged at the reported rounding precision, except the runout in Scuol (hit rate 79.2 → 79.3 %, skill score 71.0 → 71.1); their underlying counts differ, since cells do (Table 15). The number of runs at the time limit is unchanged.

The correction is more than a reassignment: where the remainder object fell below the minimum area it was removed, so those pieces were not simulated at all. In Davos that is 965 cells in 528 pieces of at most 175 m². They start now, the area of the objects rises by 0.02 to 0.04 %, and runout mostly grows slightly. Glarus alone loses more runout than it gains; this was not examined further.

Aggregate object measures change little: the volume classes (at most 64 objects of the smallest class in a region, at most one of the largest); the selection of Huber et al. (at most 0.2 points); runs over 1,000 m in the seven development regions, 2,388 → 2,381. Single extreme objects change a lot. The longest run of Val de Bagnes under ATM 1.0 was an artefact of the defect (3,445 → 2,599 m). In the most affected computed tile, in the upper Valais, the longest run falls from 5,481 to 3,085 m and the objects whose frame exceeds 2 km from 9 to none. Travel length is measured from the highest release cell and is no flow path for an object in pieces, so objects in one piece are now reported beside it. After the correction three objects in pieces span at least 1 km in the eight regions (widest 1,245 m).

Analysis	**Uses the automatic division?**	**Status after the correction**
Back-calculation from mapped outlines: cross-sections, end points, width scaling, forest mask	No	Not affected by the defect
Main maps, CAT comparison, SPOT6, accidents, sizes	Yes	Recomputed for ATM 1.0.1 (Table 15)
Grid-phase envelope, evaluation without rock masks, Alpine terrain, margin test, gap-term variants	Yes, through the maps	ATM 1.0 only; the Davos comparison on Alpine terrain repeated, the rest not rechecked
Decomposition of the runout difference, additional start places, class changes	Yes	Not rechecked; historical
Flow ablation and Pareto choice of cohesion and deposition	Yes, frozen release objects	Not rechecked; the choice of values is not reconfirmed
Release geometry	Yes	Not rechecked; historical
Numerical robustness on rotated and three grids	Yes, except the back-calculation	Not rechecked; the back-calculation part is not affected

Table 16: Which analyses depend on the corrected step. “Historical” means a development result of ATM 1.0 not reconfirmed for ATM 1.0.1. The defect touches 1.4 to 3.3 % of the objects, an indication, not evidence for each experiment.

These historical experiments document how the parameterization was selected. Unless explicitly re-evaluated, they are not additional validation evidence for the corrected implementation.

Runs at the time limit. In 109 of the 111 examined time-limit cases, the transport diagnostics identified persistent local circulation between a few cells in grid-scale notches (cycles of two, four or six cells where opposite flanks exchange volume at constant speed). The footprint defined by

the stated height and speed thresholds did not expand between 600 and 6,000 s, though 11 runs ended before 6,000 s and peak values changed by up to 0.49 m and 2.4 m/s. This supports footprint stability over the tested time interval, but not physical validity of the local flow heights, velocities or stopping behaviour. Two cases remain unexplained. Median speed is 5.7 m/s and median flow height 7.1 m, up to 58 m, so this is not rounding noise. The stop criterion sums the magnitude of the momentum and stays above its threshold (median 8.6 % of the peak) without net outward propagation. These runs end by the time limit, not by demonstrated physical rest, and dynamic quantities in these cells are not reliable. The finding covers these 111 cases (computed before the correction with the same solver), not automatically every tile computed later; all later cases are in Section 15. Neither threshold nor limit was changed.

Technical re-test. Repeated with ATM 1.0.1 by the production route: a tile computed twice with different thread counts is identical cell by cell, with all per-avalanche figures; two overlaps of 3.96 million cells each and a contiguous comparison of 24 million cells show no difference in potential, runout or object edges; in Davos, the tiles on the Alpine terrain give a skill score of 79.1 against 79.1 on the Swiss terrain and a SPOT6 difference of −0.3 points (−1.7 to 1.1); the value that fills missing terrain changes no cell. The accelerated code matches a computation with the same correction without the acceleration on two tiles, one of them the most affected one. This is limited evidence for one machine and one software environment, not a full equivalence test of all optimisations and not a claim across platforms. Tiles from the technical trial stay in the national computation; they come from a slightly earlier code version, whose differences to the later one do not touch the computation (one added region entry, freeing memory before writing). Provenance is kept per tile.

New test region. The region around Disentis (46.68 to 46.86° N, 8.80 to 9.06° E) was selected before ATM 1.0.1 results or CAT comparisons were inspected in this area, using observation availability and exclusion of previously evaluated ATM regions. The area includes locations previously considered in a separate ATH-related project. A first pass that also excluded those locations left only areas without SPOT6 outlines, so that exclusion was relaxed; general lessons from that project may have shaped ATM, though no site-specific use is known. The selection knew the distribution of SPOT6 outlines. The protocol had been written before the reference statistics were viewed, but was committed afterwards, because of an error in a command. One reference-class summary was seen; no ATM output or model–reference comparison had been inspected. The primary endpoint, eligibility filters and decision rule, those of Glarus, were unchanged. The test is therefore prospective with this limitation and not fully blinded to the reference.

Result: **unclear; non-inferiority not shown.** Lowest point of SPOT6 outlines, both periods, 298 outlines: ATM 85.2 %, CAT 86.6 %, difference −1.3 points (95 %: −5.0 to 2.3); the interval contains the provisional margin of −2 points. 2018: −0.5 (−4.2 to 3.2, 214 outlines). 2019: −3.6 (−13.0 to 5.8, 84 outlines). Without outlines touching the ATH-related locations: −0.4 (−4.7 to 3.8, 224). Skill scores for release, runout and map 82.6, 69.2 and 76.4; 23 accident points, 82.6 % each, ±20 points. Two runs reached the time limit; the longest run is 2,538 m; no object in pieces spans 1 km. No value was changed after this result. If ATM is improved from here, this region becomes a development region and a new version needs a new one.

Deviation in the sequence. The national computation was started before completion of the planned methodological review. The operator interpreted the user instruction “start everything” as authorisation to proceed after a successful technical trial. This differed from the documented sequence. The run was stopped after a release-fragment assignment defect was detected. Outputs from this computation are quarantined and are not used as ATM 1.0.1 results. It ran on 1 November 2026 in two starts and reached 39 of 71 Swiss tiles. After this correction, the owner of the project instructed that all remaining steps run without a further separate release. The national computation

of ATM 1.0.1 started on that instruction, after the unclear result above, and is descriptive. Continuing it is an internal operating decision to produce a research product, not a pass: that only a demonstrated disadvantage stops the computation is no rule for safety or release, and an unclear interval cannot exclude a relevant disadvantage. Publication was a separate decision. More regions, if wanted for precision, need a fixed selection, a stopping rule and a joint evaluation that includes Disentis in advance; another good region would not undo this result.

15 National computation and accidents

This chapter is descriptive. No tolerance was set, so nothing here is a pass or a fail, and no value of ATM 1.0.1 was changed after these results. The rules for the accident evaluation were written down before the computation, but the evaluation ran after a partial comparison with CAT had been seen; that partial comparison does not count here. The potential bands were chosen after the question arose, and the colours of the map are a design decision without a prior rule.

15.1 Computation

The frozen chain ran over the Alpine terrain model in overlapping tiles with a 30-km core and a 5-km margin: 392 tiles and 11.66 million avalanche entries, of which 71 tiles and 3.52 million entries for Switzerland and 128 tiles and 4.10 million for Austria. Entries in overlaps are counted once per tile, so these are not counts of distinct avalanches. The tiles come from three closely spaced code versions whose differences are outside the computation; provenance is kept per tile.

Checks of the assembled result:

- All 1,426 overlaps between neighbouring tiles, away from both outer margins: no differing cell in potential or runout and no differing object edge.
- Every tile read completely and held against its record: consistent. This shows that files are whole, not that their values are right.
- Five tiles computed again from a later code version: identical cell by cell, including the per-avalanche figures. This is a regression test on five tiles, not proof for the others.
- A process killed during computation and during writing leaves no tile counted as finished. A loss of power was not tested.
- No run reaches the 5-km check. The longest run from one connected piece is 3,520 m; it is compact (0.035 km²), follows a gully and has a travel angle of 29.4°. Its position was inspected on the terrain, not checked against an observation.
- Objects in several pieces remain: 82 entries span at least 1 km, the widest 2,090 m. That one consists of two fragments with 1,225 m² in total on either side of a ridge. It is one object under the rule, but not one avalanche in a physical sense, and its travel length is not a flow path.

Runs at the time limit. The 562 entries at the time limit, about 351 distinct avalanches, were recomputed to 600 and 6,000 s on sections with a 3-km margin. Three questions are kept apart. *Stability in time:* within each section no cell entered or left a footprint between 600 and 6,000 s at the stated thresholds; 52 ended before 6,000 s. This does not show that the footprint was free of error before 600 s, or that it would be equally stable under other thresholds. *Agreement of section and tile:* on 559 sections the area equals that of the tile, on three it differs by 50 to 175 m². One of the three reaches beyond the outer edge of its tile: the footprint inside the tile is identical, and the section adds seven cells outside it. The other two show identical figures from two neighbouring tiles for an object touching the outer edge of both; the footprint differs from the section by eight cells. No differing cell lies in a core used by the map. *Dynamic quantities:* peak values changed by

up to 35 m and 4.7 m/s, far more than in the 111 cases of Section 14. Heights, speeds and pressures derived from them are not meaningful in these cells. Whether all cases show the circulation pattern was not examined.

15.2 Accidents in Switzerland and Austria

The evaluation uses accident coordinates from the Swiss record and, for Austria, a public record whose coordinate meaning and accuracy are not documented; there, the points are a spatial assignment only. Evaluated accidents exclude the development regions. Rates are compared with a uniform-area baseline, the expected rate of points placed uniformly in the land area with the same radius; it is a geometric reference and no model of exposure (Table 17). A hit means that a point lies in a layer: in Austria this is point inclusion, not detection of a release area.

Of the 2,846 Swiss accidents with terrain data, 563 lie in the development regions, which leaves the 2,283 evaluated here. The bands and the comparison with CAT also leave out the 22 points in Disentis, which leaves 2,261; two of these have no CAT in reach, which leaves 2,259 for the comparison.

Layer	**Area (km²)**	**Baseline 0 m (%)**	**Hit 0 m (%) (95 %)**	**25 m**	**50 m**	**100 m**
Switzerland, 2,283 accidents						
ATM release	6,685	16.2	75.9 (74.2, 77.7)	93.7	96.8	98.6
ATM release or runout	10,466	25.3	84.1 (82.6, 85.6)	96.4	98.1	98.9
Slope from 28°	13,347	32.3	83.7	95.5	98.0	99.1
Austria, 2,474 accidents						
ATM release	8,080	9.6	66.0 (64.2, 67.9)	85.0	92.1	98.2
ATM release or runout	12,625	15.0	75.8 (74.0, 77.4)	88.8	93.3	98.6
Slope from 28°	21,184	25.2	76.0	91.5	97.9	99.6

Table 17: Accident points in the layers of the map, national evaluation. Intervals are nominal (Wilson) and account for neither spatial dependence nor repeated avalanche paths. The two countries are not comparable, because the point records differ.

Both layers lie far above the baseline. Hits at 100 m carry little information: nearly every point is then in reach of any of these layers. The release layer takes 50 % of the area of the slope layer in Switzerland and 38 % in Austria, and misses points that the slope layer contains: −7.8 points (−8.9 to −6.6) in Switzerland and −9.9 points (−11.2 to −8.7) in Austria, at the cell of the point. Release or runout reaches about the hit rate of the slope layer with 22 % less area in Switzerland and 40 % less in Austria. This holds at accident points only and says nothing about terrain where no accident is recorded.

Strata at the cell of the point, for the release layer. Swiss points with a recorded accuracy of 0 m reach 81.8 % (n = 1,403), 100 m 77.4 % (455), 250 m 67.8 % (171), 500 m 53.8 % (93), without an entry 42.1 % (159); two points carry 1,000 and 2,000 m. Rates rise with time: 71.0 % in the 1970s and 87.6 % in the 2020s in Switzerland, 50.5 % in the 2000s and 73.9 % in the 2020s in Austria; this may reflect better coordinates and has not been separated from other causes. In Austria, records with an activity of ski touring reach 75.6 %, those without a known activity 59.8 %. Excluding the edge zone changes the rate by at most 0.1 point.

Forest. Accident coordinates inside the forest mask are rare, and most lie outside the map. The release layer holds 11.8 % (95 %: 5.5, 23.4; n = 51) in Switzerland and 3.4 % (1.6, 7.3; n = 174) in Austria, release or runout 31.4 % (20.3, 45.0) and 11.5 % (7.6, 17.1). Within the same forest area, release covers only 1.8 % of 12,544 km² and 1.1 % of 37,209 km², release or runout 6.5 and 4.7 %. Against this area the points are enriched by 6.6 (3.1, 13.2) and 3.2 (1.5, 6.7) for release and 4.8 (3.1,

6.9) and 2.4 (1.6, 3.6) for release or runout, with nominal intervals. So the layers concentrate points in forest but miss most of them. Whether these are coordinate errors, release in forest or runout from above cannot be told from the points. The canopy-based approximation is not validated for identifying release in forest (Section 20).

15.3 Comparison with CAT in Switzerland

Against the public CAT layer over the whole country. Pairs: 2,259 accidents outside the development regions and outside Disentis; 2 points without CAT in reach are not compared.

Layer, reading of CAT rock	**Hit 0 m (%), ATM / CAT**	**Both / ATM / CAT / none**	**Δ 0 m, pp (95 %)**	**Δ 25 m**	**Δ 100 m**
Release	76.0 / 74.4	1,597 / 120 / 84 / 458	1.6 (0.4, 2.8)	0.6 (0.0, 1.2)	−0.1 (−0.4, 0.2)
Release or runout, rock uncoloured	84.2 / 83.5	1,800 / 103 / 87 / 269	0.7 (−0.5, 1.9)	0.7 (0.1, 1.2)	−0.2 (−0.5, 0.1)
Release or runout, rock coloured	84.2 / 89.5	1,866 / 37 / 155 / 201	−5.2 (−6.4, −4.0)	0.5 (0.0, 1.1)	−0.2 (−0.5, 0.1)

Table 18: ATM and CAT at accident points in Switzerland, paired per accident. Interval nominal, without spatial dependence. Area in the order of the rows: ATM 6,685, 10,466 and 10,466 km², CAT 6,459, 9,878 and 11,309 km². Baseline at 0 m: ATM 16.2 and 25.3 %, CAT 15.6, 23.9 and 27.4 %.

For release, the difference of +1.6 points goes with 3.5 % more area; hits per share of area are 4.7 for ATM and 4.8 for CAT. For recorded accuracies up to 100 m (n = 1,840) the difference is +2.2 (0.8 to 3.6), in the seen regions (n = 585) +2.4 (0.2 to 4.6). The two rows for release or runout concern different targets: modelled release and runout terrain, and all CAT terrain indications including steep rock. They are not interchangeable estimates of avalanche-footprint accuracy (Section 16). From 25 m the difference is gone, which is consistent with nearby edges or coordinate uncertainty and does not show that the points lie along rock edges. The share of hits per share of area is an enrichment against uniformly placed points, neither a probability of accidents nor a proof of equivalence. This is not a non-inferiority result: no margin applies to accident points, the points reflect where people go, and misses of the map where no one was recorded cannot be seen. It does not change the unclear result of Disentis.

15.4 Potential at accident points

The computation produces a continuous release-potential index and no release classes. Whether the value of potential orders accidents was measured afterwards, with bands chosen after the question arose, against the land area (Table 19).

In Switzerland the ratio rises with potential up to 0.8 and then falls slightly; the CAT classes rise throughout. In Austria the lowest band has the highest ratio, and no order appears. The Austrian pattern is not explained; that it comes from the coordinates of that record, whose meaning is not documented, is only a hypothesis. Points with a recorded accuracy of 0 m in Switzerland show the same shape as all Swiss points. These results are exposure-weighted and post hoc. They do not establish release classes or a general ranking of danger within the layer, and they do not show that potential is without use.

Band	Share of release area (%)	Switzerland, ratio (n = 2,261)	Austria, ratio (n = 2,474)
Potential 0.15–0.30	14	2.4 (2.0, 2.9)	9.0 (8.1, 10.1)
0.30–0.50	12	3.3 (2.8, 3.8)	5.6 (4.8, 6.6)
0.50–0.70	12	4.5 (3.9, 5.1)	5.3 (4.5, 6.2)
0.70–0.80	10	6.2 (5.5, 7.0)	6.6 (5.7, 7.7)
0.80–0.90	26	5.6 (5.2, 6.1)	6.7 (6.1, 7.3)
0.90–1	26	5.1 (4.7, 5.5)	7.3 (6.8, 8.0)
CAT class 1	18	3.7 (3.2, 4.1)	–
CAT class 2	35	3.8 (3.5, 4.1)	–
CAT class 3	30	5.4 (5.0, 5.8)	–
CAT class 4	16	6.9 (6.3, 7.6)	–

Table 19: Share of accidents in a band divided by its share of the land area, at the cell of the point. Nominal intervals. CAT shares of release area are its own. A ratio is no probability of release or of an accident.

15.5 Potential at mapped release areas

Accident points depend strongly on where people go. The mapped release areas of Bühler et al.[24], from the event record of 1970 to 2016 at Parsenn, Jakobshorn and Rinerhorn, are not accident points, and they are the same kind of data from which CAT derives its release density. The evaluation rule was fixed and recorded before the computation, with the bands above: potential orders these areas if the mapped share rises from band to band at all three sites, a decrease below 2 points counting as equal. The evaluation domains are convex hulls enclosing the mapped release areas, as in our earlier use of these data; the original release polygons define the mapped class.

Band	Mapped share (%), Parsenn / Jakobshorn / Rinerhorn	Mapped share (%), together	Share of mapped area (%)
Below 0.15	0.7 / 1.4 / 0.5	0.7	1.9
0.15–0.30	13.6 / 26.7 / 11.8	15.3	3.5
0.30–0.50	36.7 / 57.6 / 34.8	39.8	9.0
0.50–0.70	41.8 / 73.1 / 39.8	46.5	12.1
0.70–0.80	46.0 / 82.5 / 46.8	52.8	11.1
0.80–0.90	63.4 / 91.8 / 55.4	67.4	34.2
0.90–1	73.4 / 98.1 / 54.1	72.1	28.3
CAT class 1	18.8 / 50.2 / 15.6	21.4	4.9
CAT class 2	45.0 / 73.6 / 38.0	47.9	24.8
CAT class 3	63.7 / 88.4 / 55.9	66.7	36.7
CAT class 4	76.7 / 93.7 / 66.1	77.4	29.1

Table 20: Share of a band's area inside mapped release areas, and share of the mapped area in the band. No intervals, because cells are spatially dependent; the sites are shown side by side.

The rule is met (Table 20). The only decrease is 1.3 points at Rinerhorn between the two highest bands. Two thirds of the mapped area lie above a potential of 0.77, 0.77 and 0.76 at the three sites, on 54 % of the ATM release area in them. CAT classes 3 and 4 hold 65.8 % of the mapped area on 55 % of the CAT release area, in line with the two thirds stated for CAT.

As an internal check, fixed before computing, the value holding two thirds of the mapped area was taken from two sites and applied unchanged to the third. It stays at 0.767, 0.770 and 0.772; above it lie 67.4, 66.6 and 65.6 % of the mapped area at the third site, on 52.8, 56.7 and 53.3 % of the release area there. The mapped share of release above and below it is 66.3 and 33.3 % at Parsenn, 93.3 and

58.0 % at Jakobshorn, 54.6 and 31.6 % at Rinerhorn. This shows stability inside this region, not transfer to other terrain or snow climates.

This is a limited, exploratory result. The three sites lie in one development region and in the area from which CAT derives its density, so neither map is tested independently here. Earlier work on these data had shown that release share falls with slope above 35°, so the data were not unseen. Release areas that were never mapped count as none. The mapped share is neither a frequency of release in time nor a calibrated probability. The local ordering against mapped release areas is not reproduced consistently by the accident-point analysis. Because the two analyses have different targets and sampling mechanisms, this does not by itself invalidate the potential index. It prevents interpreting display categories as a demonstrated general ordering of accident risk or current avalanche danger.

15.6 Presentation of the map

The map display uses three colours: a darker red for a release-potential index of at least 0.77, a lighter red for release terrain from 0.15 to below 0.77, and yellow for modelled runout under the design scenario. For display only, the index is divided into these two categories; the display threshold does not affect release selection or flow simulation. The value 0.77 was chosen after the ordering check above, as a developed display rule. The two categories and the boundary were not fixed in advance and are not independently confirmed. It is no physical threshold, and resembling the anchor stated for the two upper CAT classes (Harvey et al.[1, sec. 3.1]) does not establish equivalence with their calibration.

In the three Davos evaluation sites, the higher-potential category contains a larger proportion of mapped release terrain per unit area. The legend reads higher and lower release potential, stated as relative model values, not a danger level. That the ordering was examined only near Davos is recorded in this paper. An uncoloured cell is no all-clear. It looks the same as a cell without terrain data in the tiles. The published tiles therefore include a coverage outline derived at 200-m resolution, in which a 200-m cell counts as covered if any of its 5-m cells has terrain data. It indicates the approximate computed extent, not cell-level data availability: it leans towards too much area, and gaps that fill no whole 200-m cell do not appear. Near its boundary, an application must treat coverage as uncertain unless a more precise validity mask is available. Yellow for runout is also used by CAT; this implies no equivalence of method, coverage or safety. This describes display version 2 of ATM 1.0.1, which the current published tiles state in their record. Version 1 showed a single release colour and is superseded.

The layer lies over other maps and has to stay readable when partly transparent, including over glaciers and lakes. Figure 12 shows it at three opacities on the national map and on our base map, for the Blüemlisalp massif with the Oeschinensee and for the glaciers at Konkordiaplatz.

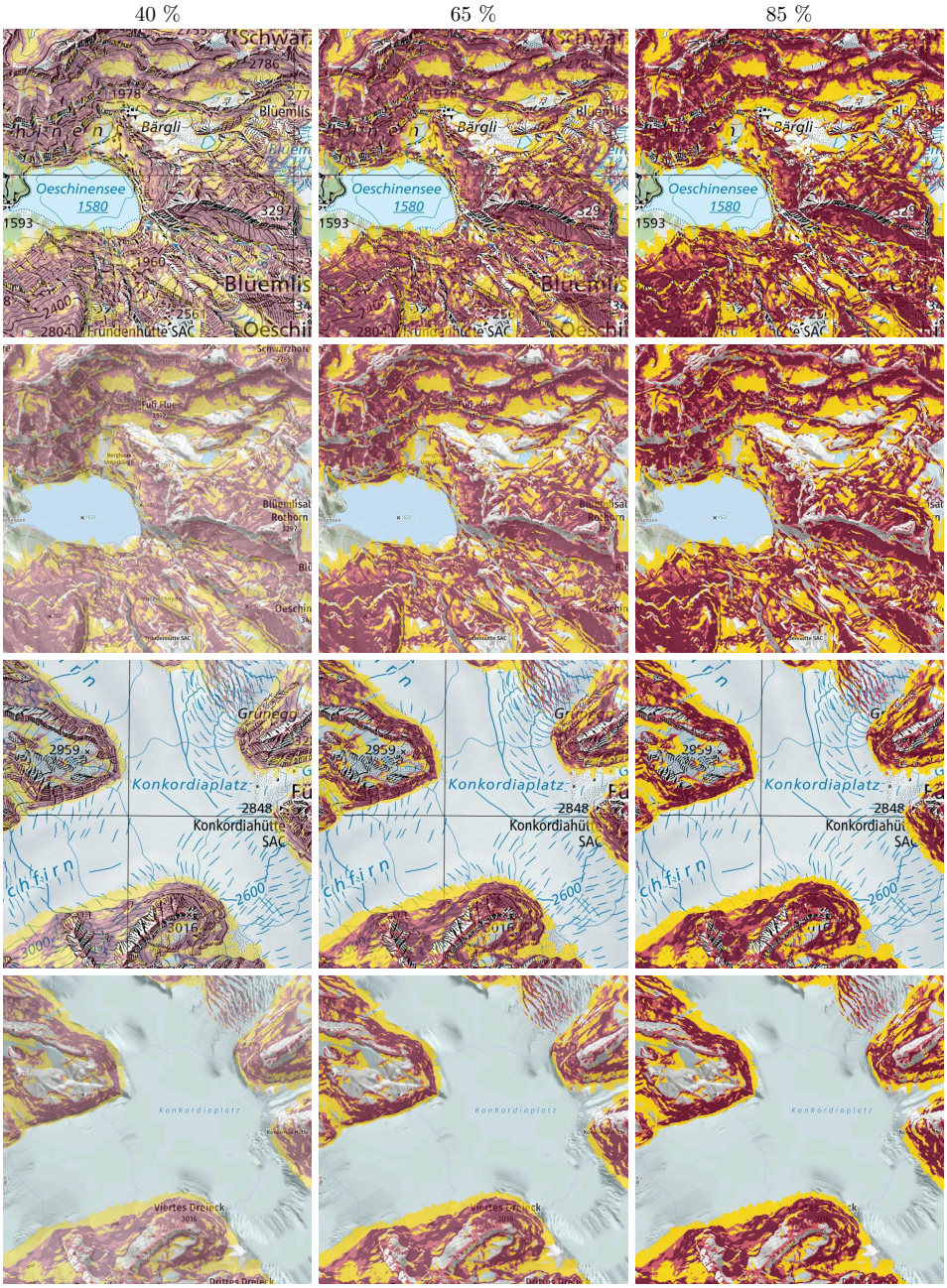


Figure 12: *Illustration, not evaluation.* The map layer at three opacities, sections of about 4.2 × 4.2 km. From top: Blüemlisalp and Oeschinensee on the national map and on the base map; Konkordiaplatz on the national map and on the base map. National map © swisstopo. Base map with data © OpenStreetMap contributors (ODbL 1.0) and swisstopo (OGD).

The colour difference CIE76 in Lab space was computed for blended colours over the actual map pixels in these sections, for normal vision and for simulated red-green deficiency (deuteranopia and protanopia, severity 1, Machado et al.[37]) (Table 21).

Opacity, vision	Runout vs map	Lower red vs map	Upper red vs map	Runout vs lower red	Lower vs upper red
40 %, normal	45–47	26–28	30–31	51	13
40 %, deuteranopia	44–46	14–17	23–25	40–41	12–13
40 %, protanopia	46–48	16–18	26–28	48–49	11
65 %, normal	69–73	43–46	49–51	77	21
65 %, deuteranopia	67–71	23–28	37–40	59–60	21
65 %, protanopia	74–77	25–28	42–45	76–77	19
85 %, normal	83–87	57–61	64–67	90–91	28
85 %, deuteranopia	80–85	31–38	48–52	67–68	28
85 %, protanopia	85–89	32–37	55–59	87–88	25

Table 21: Median colour difference (CIE76) per section and map, range over the two sections and two maps.

The tested sections show substantial median differences under the selected blending and simulations. The smallest values are those between the two reds and of the lower red against the map, at 40 % opacity. This is a technical check: median differences in two sections are no proof that every boundary can be seen on every background, CIE76 is not a contrast ratio, and the layer has not been tested with people, with or without a colour vision deficiency, or on devices in outdoor light. It is no user-based readability or accessibility validation; a legend and labelled layers keep meaning from resting on colour alone.

16 Evaluation without the rock masks

The primary regional evaluations are reported on the compared area, which excludes the rock of the two reference maps (Section 6). These masks cover only Switzerland, and for Austria, Germany, France and most of Italy no official rock mask was found. No suitable, harmonised rock mask was available for this implementation across the Alpine domain; the present Alpine map is therefore produced without rock exclusion. This chapter asks what the results in the Swiss test regions are when rock is not excluded. It does not evaluate the Alpine map itself.

What the masks did. The model has no rock class; rock enters only through the terrain. Rock with a release potential of at least 0.15 is release terrain in ATM 1.0 and starts avalanches, in Switzerland as elsewhere. The masks acted only in the evaluation. Leaving them out changes no model value and no computation: the frozen maps are evaluated again, nothing is recomputed.

Evaluation area and readings. The area is now every cell with terrain data and reference tiles, including the rock of both reference maps. CAT colours steep rock as a class of its own, which is neither release nor runout. Two readings are reported, and neither is chosen:

- **Reading A:** rock of CAT counts as uncoloured, as on the compared area. A lowest point or accident on CAT rock is a miss for CAT.
- **Reading B:** rock of CAT counts as coloured. A lowest point on CAT rock is a hit for CAT. Reading B evaluates inclusion in the release, runout or steep-rock classes; it does not treat the steep-rock class as a simulated avalanche footprint.

Reading B can only raise the hit rate of CAT against reading A or leave it unchanged; which reading is right does not follow from this. The map comparison uses reading A; the observations are reported under both. The rule was written down before the evaluation. Glarus is reported as a sensitivity only, and its verdict stays with the frozen protocol (Section 13).

Map comparison. With rock, the skill score falls in every region (Table 6, last column; Glarus in Table 14): by 0.7 to 2.1 pp for release, 0.2 to 1.2 pp for runout and 1.3 to 4.3 pp for the whole map in the development regions, and by 1.3, 0.9 and 2.4 pp in Glarus. For the runout the evaluation area itself grows with the rock (Section 10). The loss is false detection on rock: ATM colours 2.0 to 6.6 km² of release on CAT rock per region, 4.7 km² in Glarus. The largest losses of the whole map lie in Val de Bagnes (4.3 pp) and Scuol (2.8 pp).

Observed avalanches. Without the masks, 833 more outlines pass the 95-% rule in the development regions and 612 in Glarus (Table 4); the added outlines lie in the median 9 to 16 % on CAT rock. The share of lowest points in the ATM map hardly changes; the difference from CAT depends almost entirely on the reading (Table 22).

Regions	**Area**	**n**	**In map (%) ATM / CAT**	**Δ, pp (95 % interval)**
Development	compared area	1,771	91.6 / 92.9	−1.3 (−2.7, 0.0)
	with rock, reading A	2,604	91.6 / 91.0	+0.7 (−0.5, 1.9)
	with rock, reading B	2,604	91.6 / 92.9	−1.3 (−2.4, −0.2)
Glarus	compared area	534	90.8 / 89.3	+1.5 (−1.2, 4.2)
	with rock, reading A	1,146	88.5 / 82.6	+5.8 (3.7, 8.0)
	with rock, reading B	1,146	88.5 / 87.2	+1.3 (−0.8, 3.5)

Table 22: Lowest points of SPOT6 outlines in the map, both periods pooled, on the compared area (Table 12, Section 13) and on all cells with terrain data. Interval as in Table 12. Reading A: CAT rock uncoloured; reading B: CAT rock coloured.

In the development regions, leaving out one region at a time gives −1.6 to +1.5 pp under reading A and −2.9 to −0.4 pp under reading B. Measured against the provisional tolerance of −2 pp, non-inferiority is not shown on the compared area, is shown under reading A and is not shown under reading B, whose lower bound is −2.4 pp. In Glarus the lower bound lies above −2 pp in all three evaluations. The share of outline area outside the map is 7.1 % for ATM, 9.5 % for CAT under reading A and 3.5 % under reading B.

Accidents. With rock, 941 instead of 885 start-zone points enter in the development regions. In the release layer, ATM reaches 85.4 % and CAT 83.5 % (+1.9 pp, 0.3 to 3.5; baseline 39.9 and 38.9 %). For release or runout terrain, the difference is 0.0 pp (−1.7 to 1.7) under reading A and −3.6 pp (−5.3 to −2.0) under reading B. With 31 points, Glarus allows no conclusion. No tolerance for accident points has been set; these results are descriptive.

What this means for the numbers. Adding the previously excluded outlines did not markedly reduce the aggregate lowest-point hit rate of ATM in the development regions (91.6 % in both); in Glarus it decreased by 2.3 percentage points, from 90.8 to 88.5 %. This does not assess local errors, differences between kinds of event, or the correctness of the additionally coloured terrain. What the masks decide is the comparison with CAT, and there the answer depends on how CAT rock is read, which the reference map does not settle. The skill scores against CAT are lower with rock in every region, by up to 4.3 pp for the whole map, because ATM colours rock as release. The verdict of the held-back test holds under both readings. Without the masks, the development regions show non-inferiority under one reading and miss it narrowly under the other; with them, it is not shown.

Consequence for the map. The Alpine map is produced without a rock mask, since no suitable, harmonised mask was available; the evaluation without rock masks above concerns Swiss test regions, not a validation of the Alpine map. Rock in the ATM map is coloured as release where its potential reaches the threshold; the map does not distinguish rock faces from snow-holding slopes. Results on the compared area remain those of the frozen protocol; results with rock are reported beside them,

under both readings, and none of the readings is preferred. A rock class of its own would be a model change and needs its own test (Section 18).

17 What was tried and not adopted

Variant	Rule	Decisive difference	Kind
No deposition	selection rule (Section 12.5)	runs above 1,000 m 4.14 against 1.75 % for no cohesion; further benefits not supported by the intervals	scenario shift
No cohesion; 5 kPa	rule for an open front	not dropped by the selection rule; 2 kPa closer to the source application; both kept as sensitivities	source preference
Compact release pieces	geometry rule fixed in advance	lowest points not distinguishable (+0.2 pp, −0.1 to 0.5); runs above 1,000 m +11.5 %	scenario shift
Additional start places	same rule	small gain in lowest points; counted as a number, long runs unchanged; no demonstrated benefit	benefit not shown
Computation at 2.5 m	grid criterion (Section 12.12)	class above 10,000 m² not converging (1.8 against at most 1.5 %)	technical criterion
Terrain classes	visual check, 15 of 20	8 and 13 of 20 patches recognisable	plausibility
Printed gap sign	both signs compared	release skill score −0.5 to −2.4 pp in all regions	empirical, against CAT
Entrainment	size limit valid at the time	travel lengths out of the size-3 range by 10 to 42 pp	scenario shift
Release slope from 30°	limit valid at the time (Section 5)	whole-map skill score −1.1 and −1.6 pp in two regions, limit 1 pp	empirical, against CAT
Wind-shelter settings of the authors' script	visual inspection	more release area and false detection, no additional hits	empirical, against CAT

Table 23: Variants rejected or not selected after a measurement. "Scenario shift": a move away from the design scenario of avalanches up to size 3, not a demonstration that the variant is physically wrong.

Change	Observation and status
Entrainment of snow (two formulations)	Wider runout; travel lengths moved out of the size-3 range by 10 to 42 pp. Not adopted for the design scenario. A size shift alone does not show whether entrainment was implemented, parameterised or chosen wrongly; a renewed test needs a snow budget and a mass balance (Section 10, levels A and B).
No cohesion; lower or no deposition	Higher mobility; reaches more observed end points and passes farther beyond them (Section 12.5). No deposition dropped; no cohesion, 2 and 5 kPa remain on the Pareto front. 2 kPa taken as reference by the rule for an open front; no cohesion is shown as a sensitivity.
Remote triggering from terrain	Not built: the data, coefficients and rules of the CAT kernel cannot be reconstructed from the publications.
Wind-shelter settings of the authors' script	More release area and more false detection, no additional hits; not adopted.
Release slope from 30°	Measured in two regions; not adopted.

Union over grid placements as main map	Withdrawn as main map; kept as envelope.
Additional start places	Small gain in observed lowest points, little recovered CAT runout (Section 12.3); not adopted after the release-geometry test.
Computation at 2.5 m	Lengthens runs, reaches more observed lowest points and passes farther beyond them; the criterion fixed in advance for replacing the 5-m grid was not met (Section 12.12). Not adopted; this does not show that 5 m is more accurate.
Terrain classes for friction	Two transfers of published criteria to the plane, each fixed before computation: the channelisation categories of the SLF one-dimensional model (contour angle below 90° or 60° measured over half the flow width, width-to-height ratio 3:1, only over more than 80 m), and geomorphons (valley and hollow as channelled). A visual check fixed in advance required at least 15 of 20 randomly drawn channelled patches to lie in a recognisable channel, hollow or gully. The first reached about 8: measured over the width of the union of all simulated avalanches, it marks broad bowls rather than channels. The second reached 13; several patches lie across the slope on benches or slope feet. Both rejected before any performance measure; terrain stays unchannelled. The check was made by one person; the threshold of 15 of 20 is an internal screening criterion, not a demonstrated accuracy. The patches were redrawn with scale, north arrow, contour interval and fall line, and a judgement per patch is kept in the development record; for the first transfer it was written down only after the decision (6 matching, 6 unclear, 8 not matching), which does not change it.

18 Completed and planned work

Each step had its rule written down before the computation; where a rule was added or changed afterwards, Section 5 lists it.

Completed during development. The idealised cases and the benchmark of the technical verification; the decomposition of the runout difference from CAT; additional start places; cross-sections and back-calculation of observed avalanches with the end-point analysis; the width-scaling sensitivity; the forest mask, the sign of the gap term, the sensitivity to its coefficients and the descriptive comparison with CAT in forest; the flow ablation on frozen release objects with the Pareto choice of cohesion and deposition (Section 12.5); the release-geometry experiment (compact pieces not adopted); numerical robustness on rotated grids and on three grids, and the back-calculation on two grids (Section 12.12); the forest mask through the full chain (Section 12.4); two transfers of terrain classes for friction, both rejected (Section 17).

Completed before freezing. The main runs of all seven regions from the frozen code and configuration; the results of Section 12 from these runs; a manifest with code version, configuration fingerprint, input checksums including the forest-type layer, observations and evaluation files; the protocol of the held-back test (Section 10).

Completed after freezing. The single held-back test in Glarus (Section 13). The equal-volume control of the grid test, begun before freezing, was completed afterwards, with its 5-kPa part corrected (Section 12.12); it is a subsequent check and changed no decision. The decision basis for the grid was the three-grid test, available before freezing.

After the held-back test, without model change. Two development regions recomputed with a 5-km margin against a control, to measure the effect of the missing buffer (Section 12). The evaluation without the rock masks, under two readings of CAT rock (Section 16); one development region on the common Alpine terrain model, against a control for the transfer between grids and, after review, on the interior away from both computational edges (Section 7). Each rule was written down before its evaluation; none changed a value of ATM 1.0.

After ATM 1.0.1. The correction, the re-evaluation on eight seen regions, the time-limit examination, the technical re-test and the single evaluation of a new test region (Section 14).

National computation. All 392 tiles of the Alpine computation, their checks, all runs at the time limit, and the descriptive accident evaluations for Switzerland and Austria with the comparison with CAT for Switzerland (Section 15).

After the held-back test, each as a new model version that needs a new held-back region for an independent confirmation:

- terrain classes for friction with the width of single avalanches, as the one-dimensional model intends, with a new random sample for the visual check and the questions of feedback between friction, width and classification answered in advance;
- a grid-refinement criterion that separates convergence behaviour, practical size of the change in metres and its distribution over single avalanches, with a tolerance derived from the use of the map;
- the release threshold and the lower flank of the slope membership, with separate measures;
- friction depending on random kinetic energy (Bartelt et al.[38]), as a model family of its own, with velocity, flow height, deposition, width and mass balance, not only agreement with CAT;
- test regions outside Switzerland;
- a rock class of its own, from terrain or from openly available land-cover data, tested against both readings of CAT rock.

19 Open questions

1. **Width of the runout.** The CAT map is wider than ATM, but back-calculated ATM avalanches show no established width bias against observed ones (Section 12.3). What makes the CAT map wider is open: the friction law of RAMMS::EXTENDED (Christen et al.[39], Bartelt et al.[38]), the overlap of many simulated avalanches, or the generalisation of the classes. CATv2 also changed entrainment, release geometry and resolution (Harvey et al.[2]). The published values of Bartelt et al. ($\alpha = 0.1$, $\beta = 0.65$ s^{-1}, $R_0 = 5$ kJ/m^3, $\mu_0 = 0.62$, $\xi_0 = 650$ m/s^2) come from avalanches of 83,000 to 119,000 m^3; are there values for smaller avalanches?
2. **Scenario of the public CAT layer.** Which design scenario, or combination, does the publicly distributed Swiss CATv2 layer represent, and are the 2-, 10- and 30-year products of the Rogers Pass comparison comparable to it?
3. **Outline.** Which flow-height and speed thresholds are suitable for small skier-relevant avalanches and for comparison with optically mapped outlines? To be tested with several thresholds fixed in advance.

4. **Release threshold.** No directly applicable Alpine value was identified in the sources reviewed here; a sensitivity analysis around 0.15 with separate measures for observed release area, additional release, geometry and runout is planned.
5. **Forest.** Release suppression, resistance along the track, detrainment and forest destruction are different processes; how should they be combined without braking twice?
6. **Accident tolerance.** What tolerance is meaningful for the start-zone points of accidents?
7. **Terrain shape.** No directly applicable thresholds for open slopes, gullies and channels were identified in the sources reviewed here, and two transfers of published criteria failed a visual check (Section 17). Would the contour angle measured over the width of single avalanches, as the one-dimensional model intends, be an acceptable correction, or a further degree of freedom?
8. **Outside Switzerland.** At least one test region in Austria and one in Italy or France, with review by local experts, before claims beyond Switzerland.
9. **Remote triggering.** A separately named indication of connected steeper terrain, without a probability, would be a product feature of its own and needs its own test.

20 Limitations

- ATM 1.0.1 has not shown non-inferiority to CAT in its only new test region (Section 14).
- The accident evaluations are descriptive and depend on exposure and on the coordinates of the records. Elsewhere in the Alps the map is computed but not evaluated. In Austria there are accident points only, no reference map and no observed outlines.
- Inside the forest mask, most accident points lie outside the map, in the release layer 88 and 97 % (Section 15). The effect of forest is strongly simplified and not validated for release in forest; uncoloured forest is no all-clear.
- Some runs end by the time limit, not by physical rest: local numerical circulation in grid-scale notches can persist, and dynamic quantities there are not reliable. The binary footprint was stable in all 562 entries at the limit; two earlier cases are unexplained, and later cases were not examined for the pattern (Section 14, Section 15).
- Remainder objects can join pieces of a release area with no large avalanche, as a model assumption.
- The runout map is narrower than CAT's (Section 12). Whether back-calculated avalanches are too wide or too narrow depends on unverified assumptions about the error of the observed outlines, which were delineated once, not independently twice; no width bias in either direction is established. Small ones tend to stop short and large ones to run too far (Section 12.3).
- One design scenario only, without a stated return period.
- No remote triggering and no consequence layer.
- Areas below 200 m² are not simulated, and the mapped release layer is not filtered by ruggedness or curvature: 27.6 % of it (25.8 % within the compared area) starts no simulation of its own.
- Grid orientation scatters single avalanches by about ±10 to 15 % without a shift in the median. Neither peak velocities nor travel lengths are independent of the grid at 5 m. For release areas below 1,000 m², refining from 5 to 2.5 m lengthens the simulated travel lengths in the median by 30 %, a further refinement by 14 %; over all releases by 6 and 3 %. These are changes between grids, not measured errors, but they point the same way as the short runs of small avalanches against observations.

- Forest from crown cover only; no tree height, shrubs or gap height. Cells without cover data are assumed to be treeless; this affects release suppression and forest friction. The cover product excludes dwarf pine and green alder, so their protective effect is not represented.
- The forest equation keeps the printed magnitude of the gap term with a changed sign. Its limit widths at gentle slopes are about twice those of the sources (Table 7); the effect on the map is small, the printed sign remains unresolved. Whether forest release is right has not been tested against observations. CAT has about four times the release in forest, almost all of the difference at 50 % cover and above; the cause is not known. The evidence for forest runout comes only from avalanches visible from above. Testing release under closed canopy would need georeferenced release observations in forest with contemporaneous cover, tree height and gaps; ATM makes no claim of reliable release modelling there.
- All terrain is treated as unchannelled. For channels and gullies the SLF table gives higher friction (small avalanches between 1,000 and 1,500 m: μ 0.325 instead of 0.275, ξ 1,350 instead of 1,750 m/s^2). The current parameterisation may underestimate resistance relative to the channelled parameter sets; the resulting runout bias has not been established independently. Two transfers of published criteria were rejected by a visual check (Section 17).
- The 600-m limit and the minimum area combine values from different workflows.
- All regional evaluations so far are in Switzerland, partly on data from two extreme avalanche periods. The primary regional evaluations use the reference-based rock masks; additional evaluations without these masks are reported separately in Section 16. The observations evaluated are those lying almost fully inside a region and, in the primary evaluations, outside the rock masks. The original Glarus evaluation of ATM 1.0 met its nominal criterion. The subsequent Disentis evaluation of ATM 1.0.1 was inconclusive. Both concern the same two avalanche periods; the intervals do not account for spatial dependence, and the non-inferiority margin is provisional.
- No suitable, harmonised rock mask was available across the Alps; the Alpine map colours rock as release where its potential reaches the threshold. Without the rock masks, the comparison with CAT depends on how CAT rock is read (Section 16).
- The Alpine map uses a common 5-m terrain model, in Switzerland resampled once more than the terrain of the development runs. This was measured in one region only (Section 7); where a source is itself coarse, the missing detail is not tested.
- The original regional evaluation runs were computed without an external buffer: release areas outside a region are not simulated and runout leaving it is cut, so those maps can be incomplete near their edge. Separate margin tests and the production tiling procedure are described in their respective sections. In Davos and Val de Bagnes a 5-km margin changed the map only within 2 km of the edge and the reported numbers by at most 0.2 points (Section 12); the other regions were not recomputed.
- The method is described, not specified to the level of an independent re-implementation; the implementation and its configuration are not published.

Bibliography

[1] S. Harvey, G. Schmudlach, Y. Bühler, L. Dürr, A. Stoffel, and M. Christen, 'Avalanche terrain maps for backcountry skiing in Switzerland', in *Proceedings, International Snow Science Workshop, Innsbruck, Austria*, 2018, pp. 1625–1631.

[2] S. Harvey, M. Christen, Y. Bühler, C. Hänni, N. Boos, and B. Bernegger, 'Refined Swiss avalanche terrain mapping CATv2 / ATHv2', in *Proceedings, International Snow Science*

Workshop, Tromsø, Norway, 2024. [Online]. Available: https://arc.lib.montana.edu/snow-science/item/3363

[3] P. Spannring, 'Comparison of two avalanche terrain classification approaches: Avalanche Terrain Exposure Scale – Classified Avalanche Terrain', 2024.

[4] J. Sykes *et al.*, 'Characterizing differences among automated avalanche terrain mapping approaches: a multi-institution comparison at Rogers Pass, Canada', in *Proceedings, International Snow Science Workshop, Whistler, Canada*, 2026. [Online]. Available: https://arc.lib.montana.edu/snow-science/objects/ISSW2026_FS2.7.pdf

[5] Y. Bühler *et al.*, 'Automated avalanche hazard indication mapping on a statewide scale', *Natural Hazards and Earth System Sciences*, vol. 22, pp. 1825–1843, 2022, doi: 10.5194/nhess-22-1825-2022.

[6] J. Veitinger, R. S. Purves, and B. Sovilla, 'Potential slab avalanche release area identification from estimated winter terrain: a multi-scale, fuzzy logic approach', *Natural Hazards and Earth System Sciences*, vol. 16, pp. 2211–2225, 2016, doi: 10.5194/nhess-16-2211-2016.

[7] C. J. L. D'Amboise *et al.*, 'Flow-Py v1.0: a customizable, open-source simulation tool to estimate runout and intensity of gravitational mass flows', *Geoscientific Model Development*, vol. 15, pp. 2423–2439, 2022, doi: 10.5194/gmd-15-2423-2022.

[8] G. Statham and C. Campbell, 'The Avalanche Terrain Exposure Scale (ATES) v.2', *Natural Hazards and Earth System Sciences*, vol. 25, pp. 1113–1137, 2025, doi: 10.5194/nhess-25-1113-2025.

[9] H. B. Toft, J. Sykes, A. Schauer, J. Hendrikx, and A. Hetland, 'AutoATES v2.0: Automated Avalanche Terrain Exposure Scale mapping', *Natural Hazards and Earth System Sciences*, vol. 24, pp. 1779–1793, 2024, doi: 10.5194/nhess-24-1779-2024.

[10] J. Sykes, P. Haegeli, and Y. Bühler, 'Automated snow avalanche release area delineation in data-sparse, remote, and forested regions', *Natural Hazards and Earth System Sciences*, vol. 22, pp. 3247–3270, 2022, doi: 10.5194/nhess-22-3247-2022.

[11] J. Sykes, H. Toft, P. Haegeli, and G. Statham, 'Automated Avalanche Terrain Exposure Scale (ATES) mapping – local validation and optimization in western Canada', *Natural Hazards and Earth System Sciences*, vol. 24, no. 3, pp. 947–971, 2024, doi: 10.5194/nhess-24-947-2024.

[12] A. Huber *et al.*, 'AutoATES Austria — testing and application of an automated model-chain for avalanche terrain classification in the Austrian Alps', in *Proceedings, International Snow Science Workshop, Bend, Oregon*, 2023, pp. 929–936. [Online]. Available: https://arc.lib.montana.edu/snow-science/item/2989

[13] C. Hesselbach *et al.*, 'Avalanche scenario maps: from terrain identification to decision support tool', in *Proceedings, International Snow Science Workshop, Whistler, Canada*, 2026. doi: 10.15788/1790098413.

[14] A. Huber *et al.*, 'Regional-scale avalanche modeling with com4FlowPy — potential and limitations for considering avalanche-forest interaction along the avalanche track', in *Proceedings, International Snow Science Workshop, Tromsø, Norway*, 2024. [Online]. Available: https://arc.lib.montana.edu/snow-science/objects/ISSW2024_P4.1.pdf

[15] K. Markov *et al.*, 'Machine learning for automated avalanche terrain exposure scale (ATES) classification', *Natural Hazards and Earth System Sciences*, vol. 25, pp. 4375–4403, 2025, doi: 10.5194/nhess-25-4375-2025.

[16] D. Issler, K. Gisnås, and U. Domaas, 'Approaches to including climate and forest effects in avalanche hazard indication maps in Norway', 2020.

[17] D. Issler, 'Basic equations and numerical methods in MoT-Voellmy', 2025.

[18] European Avalanche Warning Services, 'Avalanche size'. [Online]. Available: https://www.avalanches.org/standards/avalanche-size/

[19] WSL Institute for Snow and Avalanche Research SLF, 'RAMMS::Avalanche 1.4: friction parameters', Davos, 2011.

[20] Y. Bühler, E. D. Hafner, B. Zweifel, M. Zesiger, and H. Heisig, 'Where are the avalanches? Rapid SPOT6 satellite data acquisition to map an extreme avalanche period over the Swiss Alps', *The Cryosphere*, vol. 13, pp. 3225–3238, 2019, doi: 10.5194/tc-13-3225-2019.

[21] Copernicus Land Monitoring Service, 'HRL Tree Cover & Forests — Product User Manual', Oct. 2026.

[22] E. Hafner and Y. Bühler, 'SPOT6 avalanche outlines 24 January 2018'. 2019. doi: 10.16904/envidat.77.

[23] E. Hafner and Y. Bühler, 'SPOT6 avalanche outlines 16 January 2019'. 2021. doi: 10.16904/envidat.235.

[24] Y. Bühler, D. von Rickenbach, A. Stoffel, S. Margreth, L. Stoffel, and M. Christen, 'Automated snow avalanche release area delineation – validation of existing algorithms and proposition of a new object-based approach for large-scale hazard indication mapping', *Natural Hazards and Earth System Sciences*, vol. 18, no. 12, pp. 3235–3251, 2018, doi: 10.5194/nhess-18-3235-2018.

[25] WSL Institute for Snow and Avalanche Research SLF, 'Avalanche accidents in Switzerland since 1970/71'. 2023. doi: 10.16904/envidat.411.

[26] P. Bebi, A. Bast, K. Helzel, G. Schmucki, N. Brozova, and Y. Bühler, 'Avalanche protection forest: from process knowledge to interactive maps', *Protective Forests as Ecosystem-based Solution for Disaster Risk Reduction (Eco-DRR)*. IntechOpen, London, 2021. doi: 10.5772/intechopen.99514.

[27] P. Bebi, 'Erfassung von Strukturen im Gebirgswald als Beurteilungsgrundlage ausgewählter Waldwirkungen', 1999. doi: 10.3929/ethz-a-002094956.

[28] M. Schneebeli and M. Meyer-Grass, 'Avalanche starting zones below the timberline – structure of forest', in *Proceedings of the International Snow Science Workshop*, Breckenridge, Colorado, 1993, pp. 176–181.

[29] M. Schneebeli and P. Bebi, 'Hydrology: Snow and Avalanche Control', *Encyclopedia of Forest Sciences*. Elsevier, 2004. doi: 10.1016/B0-12-145160-7/00271-4.

[30] F. Krumm and P. Bebi, 'Avalanches: forest interactions and risk management'. [Online]. Available: https://netriskwork.ctfc.cat/wp-content/uploads/2017/10/KRUMM-Frank_Forest-Avalanches.pdf

[31] P. Bebi, F. Kienast, and W. Schönenberger, 'Assessing structures in mountain forests as a basis for investigating the forests' dynamics and protective function', *Forest Ecology and Management*, vol. 145, no. 1–2, pp. 3–14, 2001, doi: 10.1016/S0378-1127(00)00570-3.

[32] WSL Institute for Snow and Avalanche Research SLF, 'RAMMS::AVALANCHE User Manual, version 1.8'. Davos, 2022.

[33] H. Vicari and D. Issler, 'MoT-PSA: a two-layer depth-averaged model for simulation of powder snow avalanches on 3-D terrain', *Annals of Glaciology*, vol. 65, 2024, doi: 10.1017/aog.2024.10.

[34] R. G. Newcombe, 'Improved confidence intervals for the difference between binomial proportions based on paired data', *Statistics in Medicine*, vol. 17, no. 22, pp. 2635–2650, 1998.

[35] A. Wirbel *et al.*, 'ISeeSnow v1.0 – a pilot study for snow avalanche model intercomparison of thickness-integrated shallow flow approaches and beyond'. 2026. doi: 10.5194/egusphere-2025-6053.

[36] E. D. Hafner, F. Techel, R. C. Daudt, J. D. Wegner, K. Schindler, and Y. Bühler, 'Avalanche size estimation and avalanche outline determination by experts: reliability and implications for practice', *Natural Hazards and Earth System Sciences*, vol. 23, pp. 2895–2914, 2023, doi: 10.5194/nhess-23-2895-2023.

[37] G. M. Machado, M. M. Oliveira, and L. A. F. Fernandes, 'A Physiologically-based Model for Simulation of Color Vision Deficiency', *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1291–1298, 2009, doi: 10.1109/TVCG.2009.113.

[38] P. Bartelt, Y. Bühler, O. Buser, M. Christen, and L. Meier, 'Modeling mass-dependent flow regime transitions to predict the stopping and depositional behavior of snow avalanches', *Journal of Geophysical Research*, vol. 117, p. F1015, 2012, doi: 10.1029/2010JF001957.

[39] M. Christen, P. Bartelt, and J. Kowalski, 'Back calculation of the In den Arelen avalanche with RAMMS: interpretation of model results', *Annals of Glaciology*, vol. 51, no. 54, pp. 161–168, 2010, doi: 10.3189/172756410791386553.