https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

ISeeSnow v1.0 - a pilot study for snow avalanche model intercomparison of thickness-integrated shallow flow approaches and beyond

Anna Wirbel[1], Felix Oesterle[1], Guillaume Chambon[2], Thierry Faug[2], Johan Gaume[3, 4, 5], Julia Glaus[3, 4, 5], Stefan Hergarten[6], Dieter Issler[7], Yoichi Ito[8], Marco Martini[9], Martin Mergili[10], Matthias Rauter[11], Jörg Robl[12], Giorgio Rosatti[13], Kae Tsunematsu[14], Christian Tollinger[15], Hervé Vicari[3,4,5], Daniel Zugliani[13], and Jan-Thomas Fischer[1]

[1]Austrian Research Centre for Forests, Innsbruck, Austria
[2]Univ. Grenoble Alpes, INRAE, CNRS, IRD, G-INP, Institute for Environmental Geosciences IGE, Grenoble, France
[3]WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland
[4]Institute for Geotechnical Engineering, ETH Zürich, Zürich, Switzerland
[5]Climate Change, Extremes, and Natural Hazards in Alpine Regions Research Center CERC, Davos Dorf, Switzerland
[6]University of Freiburg, Institute of Earth and Environmental Sciences, Freiburg, Germany
[7]Norwegian Geotechnical Institute, Natural Hazards Division, Oslo, Norway
[8]Snow and Ice Research Center, National Research Institute for Earth Science and Disaster Resilience (NIED), Japan
[9]Department of Land, Environment, Agriculture and Forestry, University of Padova, Legnaro, Italy
[10]Cascade | The mountain processes and mountain hazards group, Department of Geography and Regional Science, University of Graz
[11]Department of Civil Engineering and Natural Hazards, University of Natural Resources and Life Sciences, Vienna, Austria
[12]University of Salzburg, Department of Environment & Biodiversity, Salzburg, Austria
[13]Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy
[14]Faculty of Science, Yamagata University, Japan
[15]Austrian Avalanche and Torrent Service (WLV), Innsbruck, Austria

Correspondence: Anna Wirbel (anna.wirbel@bfw.gv.at)

Abstract. We present a pilot study for intercomparison of snow avalanche flow simulation tools. Thirteen different groups participated in the study, whose models are categorized into a *core group* using thickness/depth-integrated shallow flow approaches, and an *extended group* including one with a differing shallow water approach, and one depth-resolved 3D approach. The intercomparison is performed for three simple test cases representative of typical applications: snow avalanche flow over an idealized and a real topography with a release area of constant thickness based on a Voellmy and pure Coulomb friction relation and prescribed values for the friction parameters. The aim of this pilot study is to analyse the spread in simulation results and discuss potential sources of the observed differences. A quantitative assessment of the variability is based on the distribution of scalar measures like runout length, runout angle and maximum values of flow thickness and flow velocity. Within the *core group* of thickness/depth-integrated shallow flow approaches, simulation results for the Voellmy test cases (idealized and real topography), excluding outliers, show a spread of roughly 55 m in derived runout length with an interquartile range of about 30/33 m, referring to total runout lengths of 2310/2082 m (median values for idealized and real topography). Runout length is for all three test cases constrained by an abrupt change in slope angle. Maximum peak flow thickness and velocity show a

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

generally larger spread for both, the idealized and real topography test cases. Excluding outliers, differences between the simulation results are more pronounced for the real topography test case compared to the idealized topography case. The largest differences arise if only Coulomb friction is considered, as indicated by an interquartile range of 92 m in runout length. The Coulomb test case also shows a considerably larger number of outliers. This is partly due to the pronounced effect of curvature effects in this case, which is not accounted for by all participating simulations tools. Focusing on the *core group*, this analysis serves as a first assessment of the uncertainty introduced by the different implementation workflows (e.g., numerical schemes, ad-hoc treatments, geo-data handling, curvature treatment, etc.). However, actual attribution of variability to individual sources is beyond the scope of this preliminary study and will necessitate further testing. Performing this pilot study allowed us to identify common issues, gather information on respective requirements regarding input data and problem definition, which will help to optimize the design of a future, more comprehensive model intercomparison study.

1 Introduction

Tools for the simulation of snow avalanche flow are an integral part of performing hazard assessments in mountainous regions. Historically, these assessments were limited to areas affecting settlements and important infrastructure, with a focus on very large to extremely large avalanche events, of 100+ years return periods. Besides detailed case studies, simulation tools are now also increasingly applied as a central building block for the production of large-scale avalanche hazard indication maps (e.g. Harvey et al., 2018; Bühler et al., 2022; Issler et al., 2023; Toft et al., 2024), used for avalanche risk management in the administrative, professional and recreational sectors. Alongside experimental research, development of simulation tools and underlying flow models is ongoing, and related numerical experiments can help to advance our understanding of how sensitive avalanche movement is to specific processes or parameters.

Various simulation tools have been developed and applied to simulate avalanche dynamics at both slope and regional scale. These simulation tools can be broadly categorized into i) physical, process-based models and ii) conceptual, data-driven approaches, the latter including geometrical, statistical, and empirical methods. In this study, we focus on process-based models, which describe avalanche dynamics in terms of partial differential equations based on the conservation of mass and momentum. These models vary widely in terms of complexity, constitutive models, numerical implementation, programming environments, as well as pre- and post-processing routines.

In order to ensure the functionality and applicability of models, testing is a crucial step in the model development process. Testing involves i) model verification, i.e. comparing simulation results to an exact solution to check whether the mathematical model and chosen numerical methods are correctly implemented and ii) model validation, i.e. comparing simulation results to observations for practically relevant cases (e.g. Bueler et al., 2005; Roache, 1997). Beyond this, model intercomparison, where simulation results of different models are compared to each other in a standardized framework, provides information on

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

the individual model performance, model applicability, as well as the uncertainty introduced by the various implementations (e.g. Pattyn et al., 2008; Gueugneau et al., 2021). In the case of snow avalanche flow, model verification tests only exist for simplified problems that can be solved (semi-)analytically and often focus on the early stages of the flow. Examples are the well known dam-break (Riemann) problem (Faccanoni and Mangeney, 2012), the similarity solution for Savage–Hutter type models (Hutter et al., 1993), the uniformly accelerated flow (Zugliani and Rosatti, 2021) and a test based on the conservation of potential and kinetic energy, considering frictional dissipation represented by dry Coulomb friction (Tonnel et al., 2023). However, these tests rely on simple topographic conditions as well as simplified friction relations that may be insufficient to fully represent real-world avalanches. On the other hand, model validation, i.e. reproducing observations from real-world events, often depends heavily on the calibration and optimization of constitutive/rheological parameters as well as the availability of corresponding data. This is because existing simulation tools often make assumptions to simplify the underlying physical processes and their mathematical formulation, and hence require calibration and optimization of certain parameters through back-calculation. Previous studies showed that back-calculating observed events using preceding parameter calibration and optimization tailored to the specific simulation tool can yield satisfactory results (e.g. Gauer et al., 2009; Naaim et al., 2013). Recently, calibrated values of flow material parameters have been tabulated (e.g., Bartelt et al., 2017), for example as functions of flow volume and return period. Yet, these values may depend on the software in which they were calibrated, raising the question of how strongly simulation outcomes diverge between different numerical implementations when the same parameters are applied.

Both model verification and validation are part of the model development process and, hence, usually included in model presentation and documentation. However, so far, there are few studies showing a comparison of different avalanche flow models. These are limited to selected case studies including only a few models (e.g. Barbolini et al., 2000; Issler et al., 2005; Schmidtner et al., 2018; Gauer et al., 2023), or to validation tests when presenting a newly developed simulation tool or application (e.g. Hergarten and Robl, 2015; Rauter et al., 2018; Zugliani and Rosatti, 2021; Martini et al., 2023). Thus, to date, no extended and coordinated intercomparison of simulation tools for snow avalanche flow has been performed.

For this reason, the *ISeeSnow* project has been initiated as a pilot study on model intercomparison. Given the limitations of model verification and validation tests, such as topographical constraints and necessity of model-specific parameter calibration and optimization, we focus on the comparability of simulation tools when applied to problems representative of real-world applications, but without introducing the effect of parameter calibration and optimization. More specifically, this pilot study aims to assess how different software used in engineering and research projects affects simulation results relevant to, for example, hazard mapping (e.g. runout lengths, peak flow thickness and velocity). By comparing simulation tools that are based on the same basic equations and assumptions, the influence of their implementation workflow can be analysed and the corresponding uncertainty in the simulation results quantified. Given the requirements of real-world applications as opposed to those of model verification tests, differences in simulation tools are not restricted to the numerical implementation, but also other components of the simulation workflow come into play, as for example pre-processing of input data. For this reason, rather than referring purely to numerical implementation, we define the implementation workflow more broadly as the combination of i) the choice

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

of numerical methods for solving the set of equations, ii) their implementation as computer code, iii) any ad-hoc treatments to account for omitted terms or to achieve numerical stability, iv) enforcement of initial and boundary conditions, and v) geodata handling (i.e. any pre-processing of input data and post-processing of simulation results to provide required outputs).

To establish a basis for exchange and discussion among the different groups and as a starting point to identify requirements and common issues, all process-based models were invited to participate in *ISeeSnow*. Subsequently, the participating groups were divided into two groups: a *core group* and an *extended group*. The *core group* gathers the thickness/depth-integrated shallow flow models that are most widely used in an operational context. These models solve the equations either using a local coordinate system, with thickness of the fluid measured normal to the slope, or a Cartesian global coordinate system where depth is measured in the direction of gravity. All other tools are assigned to the *extended group*, which consists of one depth-resolved 3D model and one flux-routing model. For the *extended group*, only a qualitative comparison is performed since significant deviations from the *core group*'s results are anticipated due to differences in model formulations.

In order to account for the different requirements of individual simulation tools regarding model-specific setups and configurations, later referred to as implementation-specific parameters, the approach we take in this pilot study is to prescribe input and model parameter values following the idea of the smallest common denominator. We chose them so as to be representative of a standard application for most participating simulation tools. The study consists of three test cases featuring two different topographies and two friction relations, with prescribed release areas with constant release thickness. By also prescribing fixed values of the friction parameters (chosen within the range of reasonable values presented in the literature), we exclude the effect of the calibration and optimization workflow. As we are interested in the comparability of simulation tools, no observed event is included as a reference solution.

By comparing simulation results for the provided test cases, this pilot study aims at gaining information on the actual spread in simulation results among different models under different topographical conditions and applied friction relations, and to point towards potential sources of the observed differences. Furthermore, the study is intended to be a starting point to learn what problems and issues may arise in avalanche simulation intercomparisons, i.e., what kind of test cases are required and how a precise problem definition needs to look like.

The manuscript is structured as follows: the study design and test cases are described in Sect. 2. The participating simulation tools and their key differences (also regarding implementation-specific parameters) are briefly characterized in Sect. 3, with more detailed descriptions provided in Appendix A. In Sect. 4, we present the results of the intercomparison, followed by a discussion of the differences (Sect. 5.1) and their potential causes (Sect. 5.2). Concluding remarks and an outlook are provided in Sect. 6.

2 Study design

The study consists of the following steps: i) comparing the submitted simulation results and analysing the spread in output variables for the different test cases, ii) gathering feedback on input data needs, differences in approaches, common issues, iii)

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

identify and discuss potential sources of found differences and iv) highlighting important issues to consider in the design and realization of a comprehensive simulation tool intercomparison. A full description of the study, the input data sets for the test cases and the simulation results submitted by the participants can be found at the official online repository of the *ISeeSnow* project: https://github.com/avaframe/ISeeSnow (last access: 1 December 2025).

2.1 Call for contribution

An invitation to contribute to the *ISeeSnow* project was sent out to groups that were known to be developing and/or applying tools for the simulation of snow avalanche flow. We also reached out at the first public presentation of *ISeeSnow* at the 2024 General Assembly of the European Geosciences Union (Wirbel et al., 2024), extending the invitation to participate.

2.2 Test cases

As one aim of this pilot study is to establish a platform for a discussion that includes all process-based approaches, we chose simple test cases, model configurations and requested output data sets to gather many different groups and keep their workload manageable. Secondly, to not exclude groups that do not have access to the source code, model verification tests were not included in this pilot study. Such tests rely on simplified friction relations and/or specific initial and boundary conditions that deviate from standard applications, and hence often require modifications of the source code.

2.2.1 Problem definition

The test cases are based on a standard application of incompressible, thickness/depth-averaged snow avalanche flow simulation tools: simulating the flow thickness and velocity evolution for a given topography and release area with constant release thickness using a Voellmy or pure Coulomb friction relation (see e.g. Pudasaini and Hutter, 2007). To exclude the effect of parameter optimization, values for the Voellmy parameters μ and ξ in the expression for the basal shear stress τ_b,

$$\tau_b = \mu\sigma_b + \frac{g}{\xi}\rho\bar{u}^2 = \mu\sigma_b + k\rho\bar{u}^2, \tag{1}$$

are prescribed, where σ_b is the normal stress at the bottom, g the gravity acceleration, ρ the avalanche density and $\bar{u}$ the thickness-averaged flow velocity. The less frequently used formulation based on k instead of ξ has the advantage that the Coulomb friction law (i.e., only the first term on the right-hand side of Eq. 1) results from setting $k = 0$, whereas ξ would need to be set to infinity (see Table 3 for values applied in this study by the respective tools). This is relevant because a test case based on a pure Coulomb rheology is included to allow for a qualitative comparison between 2D and 3D approaches. 3D models require constitutive relations to describe the material's rheology. To facilitate a comparison to the 2D approaches, the non-associative Drucker–Prager yield criterion employed in the participating 3D approach can be coaxed to reduce to pure Coulomb friction. Still, results from the 3D and 2D approaches are not directly comparable, as the latter rely on conservation equations that are integrated over the flow thickness. This thickness-integration inherently simplifies the dynamics, typically assuming that flow velocities are parallel to the surface and constant across the flow thickness. As a result, the 2D model resolves only thickness-averaged velocities. Often, additional assumptions are made, such as constant flow density and the

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

absence of slope-normal accelerations. Nevertheless, a qualitative comparison can still provide valuable insights into the influence of these assumptions and the integration process on the simulated flow behaviour.

No further constraints were imposed on the model configuration. Beyond the prescribed rheological and material parameters, performing a simulation often requires setting parameters that are specific to a model, i.e. implementation-specific parameters. Besides parameters related to the numerics (e.g. time step duration, stabilization techniques, etc.), these include earth-pressure coefficients, settings related to accounting for terrain curvature, stopping mechanisms and stopping criteria, and the pre- and post-processing steps. It is indeed difficult to prescribe values for these implementation-specific parameters, as they are specific to the simulation tool and are not required or available among all the participating tools. Without reducing the problem complexity, this makes it impractical or almost impossible to design a test case that provides values for all these parameters without introducing differences in the simulation procedure between the models. For this reason, we consider the setting of these parameters to be part of the implementation workflow (see Sect. 1 for a definition). For models that are applied operationally, it was suggested to use the respective default setup. Otherwise, the groups were encouraged to define a useful technical setup according to their experience. Allowing for this technically useful freedom, however, comes at the price of additional differences in the simulation results. More information regarding these implementation-specific parameters follows in Sect. 3 and a detailed discussion of related issues can be found in Sect. 5.2.

2.2.2 Input data

For all three test cases, named *VoellmyIdealized*, *VoellmyReal* and *CoulombIdealized*, the following input data and model parameter values are provided:

- a digital elevation model (DEM) as an ASCII raster file with $\Delta x = \Delta y = 5\,\mathrm{m}$,
- a release area polygon as a shapefile (.shp) or as an ASCII raster file with $\Delta x = \Delta y = 5\,\mathrm{m}$,
- constant release thickness (measured normal to the surface),
- fixed, spatially constant values for the friction model parameters μ and ξ.

The test cases are based on two different topographies, an idealized, generic topography (test cases *VoellmyIdealized* and *CoulombIdealized*) and a real topography (test case *VoellmyReal*), shown in Fig. 1. The idealized topography was generated using the *in3Utils* module of AvaFrame (Oesterle et al., 2024). It consists of an inclined plane with a ravine of variable width. After a vertical drop of 1500 m, the slope transitions to a flat foreland. The avalanche release volume is $75\,000\,\mathrm{m}^3$. For the real topography, the release volume is approximately $248\,000\,\mathrm{m}^3$. If the release volume is computed using the ASCII raster file, volumes are roughly 4 % (real topography) and 6.6 % (idealized topography) larger compared to using the polygon from the shapefile (see Sect. 5.2 for more details). The initial conditions and the prescribed values of the friction parameters for the test cases are listed in Table 1.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

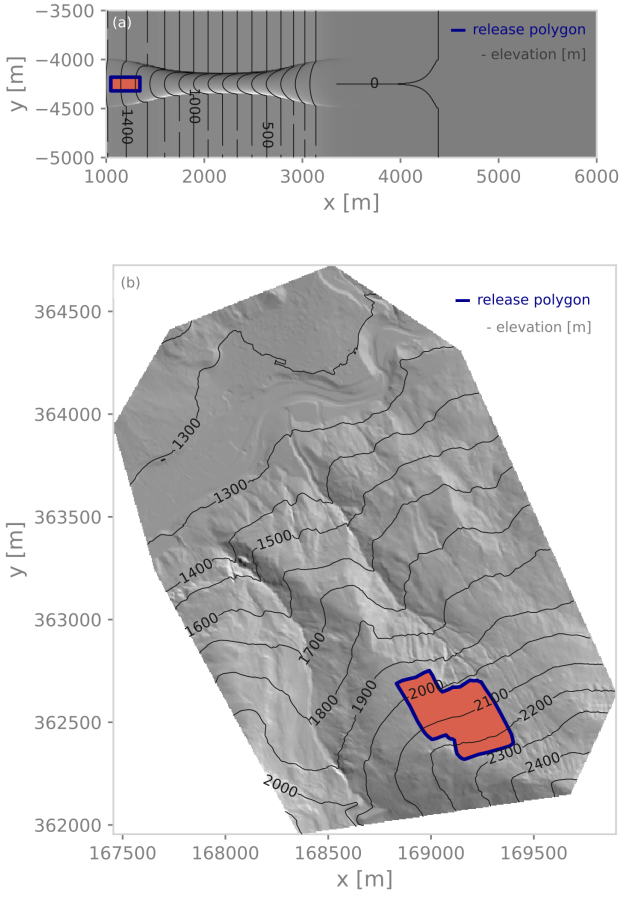

Figure 1. Top view of the computational domain for the test cases *VoellmyIdealized* and *CoulombIdealized* in panel (a) and for the *VoellmyReal* test case in panel (b). The release area is shown in red and the polygon is delineated in blue; the release thickness of $1.5\,\mathrm{m}$ is constant throughout the release area. Elevation contour lines are shown in dark gray.

Table 1. Prescribed model configuration regarding the release thickness and the friction parameters μ and ξ.

Test case	Release thickness [m]	μ [–]	ξ [$\mathrm{m\,s^{-2}}$]
VoellmyIdealized	1.5	0.4	2000
VoellmyReal	1.5	0.2	2000
CoulombIdealized	1.5	0.4	–

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

2.2.3 Required result files

Participating groups were requested to submit the following files:

- peak flow thickness and peak flow velocity fields as ASCII raster files with $\Delta x = \Delta y = 5\,\mathrm{m}$ and the same extent as the input DEM ('peak' refers to the maximum value over the entire duration of the simulation in the given cell),
- values of initial and final volume, the spatial resolution used for the computation, total simulation and computation duration,
- a text file describing the configuration of the simulation tool, the used setup and basic information on the model, including a main reference.

The input data is provided in the form of raster fields with a regular grid spacing of $5\,\mathrm{m}$, and result fields of the peak variables are required to be submitted in the same format. Besides this requirement, the spatial resolution used for the computations was not prescribed. If employed numerical schemes are based on particle methods, irregular grids, the use of adaptive mesh refinement, or simulations were run using a different spatial resolution, participants were asked to interpolate the result to the defined regular mesh with $\Delta x = \Delta y = 5\,\mathrm{m}$. Similarly, if computations are performed based on depth instead of thickness, participants were asked to convert their simulation results to thickness before submission. As a result of a discussion about the settings of any implementation-specific (and therefore non-prescribed) parameters, further information regarding specific model characteristics was added to the list of required result files in the form of a *model information table*.

3 Simulation tools

Thirteen different groups participated in this study with thirteen different simulation tools. Table 2 lists the simulation tool name, the corresponding participants and the group the simulation tool has been assigned to. The last three columns indicate whether or not the participating group has submitted simulation results for the respective test case. Additional information on the simulation tools can be found in the brief descriptions provided by the participants in Appendix A.

For all except *FLO-2D*, *RAMMS::Avalanche* and *TITAN2D*, the participating group is also part of the development team of the respective simulation tool or of the respective extension of an existing numerical solver for computational fluid dynamics. Except *FLO-2D*, all participating groups had access to the source code of the respective model. *FLO-2D* is based on a different friction relation but has previously been used to simulate snow avalanche flow (Martini et al., 2023). As that calibration has formed the basis for the participation in this study, results have only been submitted for the two Voellmy test cases (see Sect. A2).

Model groups and differences in implementations

The participating simulation tools are broadly categorized into two groups: *core group* and *extended group*. The *core group* gathers simulation tools that are based on the same basic equations and assumptions, and all other approaches are assigned

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Table 2. List of participants: name of simulation tool, corresponding participants and respective group, where *core group* refers to the class of thickness/depth-integrated shallow flow models. The *extended group* includes all other approaches. The last three columns provide information on submission of results for the respective test cases. Superscripts refer to the affiliations provided in the list of authors.

Simulation tool	Corresponding participants	Group	*VoellmyIdealized*	*VoellmyReal*	*CoulombIdealized*
com1DFA	Anna Wirbel[1], Felix Oesterle[1], Jan-Thomas Fischer[1]	core	✓	✓	✓
FLO-2D	Marco Martini[9]	extended	✓	✓	-
faSavageHutterFoam	Matthias Rauter[11]	core	✓	✓	✓
Gerris	Jörg Robl[12]	core	✓	✓	✓
minVoellmyv2	Stefan Hergarten[6]	core	✓	✓	✓
MPM	Johan Gaume[3,4,5], Hervé Vicari[3,4,5]	extended	-	-	✓
MoT-Voellmy	Dieter Issler[7]	core	✓	✓	✓
RAMMS::Avalanche	Julia Glaus[3,4,5]	core	✓	✓	✓
r.avaflow	Martin Mergili[10]	core	✓	✓	✓
samosAT	Christian Tollinger[15], Matthias Granig[15]	core	✓	✓	✓
SaVal-2D	Guillaume Chambon[2], Thierry Faug[2]	core	✓	✓	✓
TITAN2D	Yoichi Ito[8], Kae Tsunematsu[14]	core	✓	-	✓
TRENT2D[*]	Daniel Zugliani[13], Girogio Rosatti[13]	core	✓	✓	✓

to the *extended group*. However, the implementation of the equations varies also within the *core group* of thickness/depth-integrated shallow flow approaches (see Sect. 1 for a definition of implementation workflow as is used in this study). For the reasons outlined in Sect. 2.2.1, settings for implementation-specific parameters have not been prescribed. It is important to note that, whilst differences in the setting of numerical parameters should not significantly affect the simulation results, varying settings of other implementation-specific parameters, for example regarding curvature treatment or the implementation of stopping, can lead to, in that respect, 'expected' differences in the simulation results. This introduces additional complexity when assessing the variability in simulations results, as therefore, the differences do not solely reflect the effect of employing different numerical methods. In order to provide an overview of key differences in model implementations, Tables 3 and 4 list the modelling challenges of highest impact and the corresponding approaches chosen by the different models, including settings for the most important implementation-specific parameters. While the provided information refers to the setup used for performing the presented simulations, also other options might be available for the individual tools.

There are two simulation tools in the *extended group*. The 3D approach of *MPM* is based on a Drucker–Prager model matching the Mohr–Coulomb failure criterion in triaxial compression for internal friction (which is set equal to the basal friction). The other member of this category, *FLO-2D*, initializes the release area as a hydrograph along the lower edge of the release area and models friction through Manning's equation; for more information see Sect. A2. Where applicable, information on the tools of the *extended group* is also provided in Tables 3 and 4.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Table 3. List of modelling challenges and the corresponding approaches chosen by the participating models (as provided by participating groups). Information reflects the setup used for performing simulations in this study. This list is not comprehensive, but points towards steps in the simulation workflow where differences in implementation were identified as potential sources of the variability in simulation results. More detailed information can be found in the simulation tool descriptions and related references provided in Appendix A.

Characteristic	Realizations
Curvature effects	Curvature accounted for in 'effective' gravity acceleration in Coulomb friction term (*com1DFA, minVoellmyv2, r.avaflow, RAMMS::Avalanche, TITAN2D, MoT-Voellmy*)
	Not accounted for (*FLO-2D, Gerris, SaVal-2D*)
	First-order accurate due to global coordinates (*faSavageHutterFoam, TRENT2D**)
	Treatment of curvature and dispersion follow naturally from 3D conservation equations (*MPM*)
Stopping mechanism, yield criterion	None (*com1DFA, Gerris, r.avaflow, TITAN2D, RAMMS::Avalanche*)
	Local yield criterion based on Voellmy friction term (*SaVal-2D*, *TRENT2D**)
	Local stopping if Coulomb friction (and pressure gradient) can consume the actual momentum (*minVoellmyv2, MoT-Voellmy*)
	Naturally captured through constitutive model (*MPM, faSavageHutterFoam*)
	Surface detention threshold (*FLO-2D*)
End-of-simulation criterion	Threshold of momentum, velocity or kinetic energy (*com1DFA, Gerris, SaVal-2D, MoT-Voellmy, RAMMS::Avalanche, r.avaflow*)
	User-defined ending time (*TRENT2D***, MoT-Voellmy, faSavageHutterFoam, FLO-2D*)
	Drag based stopping criteria (for Coulomb friction model only) or user-defined ending time/number of time steps (*TITAN2D*)
	None required (*MPM, minVoellmyv2*)
Ad-hoc treatments (non-exhaustive list)	Velocity vector cannot change direction by more than ± 90 °within a single time step (*MoT-Voellmy*)
	Artificial viscosity (*com1DFA*)
	No detachment from surface, enforced through reprojection (*com1DFA*)
	Removing undershoots or dry cells ($h \leq 0$ or $h > 10^{-5}$) (*faSavageHutterFoam*)
	Repeating time step with reduced Δt if undershoot ($h \leq 0$) occurs (*MoT-Voellmy*)
	Effective bed-normal gravity limited to positive values (*faSavageHutterFoam*, *com1DFA*, *MoT-Voellmy*, *minVoellmyv2*)
	Release volume defined as solid–liquid hydrograph and assigned to inflow cells defined by the user (*FLO-2D*)
Coulomb friction relation	Coulomb friction law (*com1DFA, faSavageHutterFoam, TITAN2D, FLO-2D*)
	Voellmy friction law with ξ set to a very high value (*SaVal-2D, minVoellmyv2, Gerris TRENT2D**)
	Voellmy friction law with k set to zero (*MoT-Voellmy, RAMMS::Avalanche (source code adjusted)*)
	Internal friction also accounted for in Coulomb friction model (*TITAN2D*)
	Coulomb basal friction and Drucker–Prager criterion matching Mohr–Coulomb criterion in triaxial compression for internal friction (*MPM*)

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Table 4. List of modelling challenges and the corresponding approaches chosen by the participating models (as provided by participating groups) (continuation of Table 3). Information reflects the setup used for performing simulations in this study. This list is not comprehensive, but points towards steps in the simulation workflow where differences in implementation were identified as potential sources of the variability in simulation results. More detailed information can be found in the simulation tool descriptions and related references provided in Appendix A.

Characteristic	Realizations
Velocity derived from momentum and flow thickness, with minimum thickness/depth threshold	0 m (*minVoellmyv2, RAMMS::Avalanche*)
	0.001 m (*r.avaflow, TITAN2D, SaVal-2D, Gerris*)
	User-selectable, default 0.05 m (*MoT-Voellmy*)
	Velocity directly computed (*com1DFA, MPM, FLO-2D*)
	10^{-5} m (*TRENT2D**)
	User-selectable, 10^{-5} m chosen (*faSavageHutterFoam*)
Minimum thickness/depth threshold used elsewhere	None (*minVoellmyv2*)
	10^{-5}m (*TRENT2D**)
	0.001 m (*r.avaflow, SaVal-2D, Gerris*)
	User-selectable (*MoT-Voellmy, TITAN2D*)
	User-selectable, 10^{-5} m chosen (*faSavageHutterFoam*)
Deposition modelled explicitly	No (*com1DFA, minVoellmyv2, RAMMS::Avalanche, r.avaflow, MoT-Voellmy, TITAN2D, SaVal-2D, TRENT2D**, Gerris)
	Naturally emerging (*MPM*)
	Possible, not active here (*faSavageHutterModel*)
Earth-pressure coefficient	Set to 1 or not implemented equalling setting to 1 (*com1DFA*, *SaVal-2D*, *TRENT2D**)
	User-selectable, active/passive not distinguished (*MoT-Voellmy*)
	Set automatically using parameter values of basal/internal friction angle (*TITAN2D*)
	Not required (*MPM*, as lateral stress effects are resolved naturally through the constitutive model rather than approximated by an earth-pressure coefficient, with the internal friction in the Drucker–Prager failure criterion set equal to the basal friction)
	Not taken into account (*minVoellmyv2*, *Gerris*, *faSavageHutterFoam*)

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

4 Comparison of simulation results

The comparison is performed for the submitted simulation result fields of peak flow thickness (measured normal to the topography) and flow velocity, where 'peak' refers to the maximum value over the entire simulation duration in a given cell of the mesh. Peak flow thickness/depth is a direct result variable when solving the equations for all participating tools (except FLO-2D of the *extended group*), whereas flow velocity is partly derived in a post-processing step from the direct result variables (momentum and thickness/depth; see Table 4). For the visualization of the peak fields, values below a threshold of $0.01\,\mathrm{m}$ are masked. This is required to facilitate a comparison of the general spatial extent of the individual simulation results, as some models produce small non-zero values in large parts of the domain. From these, fields of the average values and the spread are computed for the simulation results of the *core group*, where the spread is defined as the difference between maximum and minimum field values and for the relative spread, this is divided by the average field values. In addition, scalar measures such as maximum peak variable values, runout length and runout angle are derived for all simulation results. For the computation of runout length and runout angle, the data sets are first transformed into a thalweg-following coordinate system (see Appendix B). The thalweg was defined manually as a two-dimensional representation of the main flow direction starting at a point located uphill of the highest point of the release area but close to it so that the entire flow is captured. For a detailed definition, see https://docs.avaframe.org/en/latest/glossary.html (last access: 1 December 2025) and the work that the analysis tools are based on Fischer (2013). The runout length is measured in the thalweg-following coordinate system, starting at the beginning of the thalweg to the furthest point along the thalweg where the chosen peak variable still exceeds the specified threshold. The runout angle is derived from the elevation drop and the length of the thalweg projected onto a horizontal plane, measured from the first to the last point where the chosen threshold of the peak variable is exceeded. The following comparison focuses on peak flow thickness with a threshold of $0.5\,\mathrm{m}$. The same analysis is also performed for the peak flow velocity fields with a threshold of $1\,\mathrm{m\,s^{-1}}$; the corresponding figures can be found in Appendix C. The choice of these thresholds is based on values relevant for practical purposes such as hazard zoning.

4.1 Test case *VoellmyIdealized*

The spatial distribution of peak flow thickness (PFT) of all submitted simulation results is provided in Fig. 2. All models predict an increase in PFT where the channel becomes narrower and within the deposition area. Lateral spreading is observed at the transition into the flat foreland, roughly around $x = 3300\,\mathrm{m}$. However, the pattern of PFT in the runout area varies between the different models. Some results show a more pronounced lateral spreading, which is also reflected in the $0.5\,\mathrm{m}$ contour lines shown in Fig. 3. *MoT-Voellmy's* PFT field shows a more 'elongated' shape at the tip of the runout area with maximum values more than twice as high as maximum values of the other simulations. These high values are not constant within the runout area, but are the maximum values of cross-flow oscillations in PFT. The magnitude of these oscillations increases along the flow direction due to developing numerical instability (visible in Fig. 2; more information on the formation of these oscillations is provided in Sect. A7). Both in terms of general spatial extent of the avalanche flow and the magnitude of PFT values, the largest differences between the simulation results are observed in the runout area rather than in the channel.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Simulation results are in good agreement until the flow reaches the lower part of the channel where lateral spreading starts; until this point the spread is smaller than minimum simulated PFT values (not shown). The overall spread and the standard deviation in PFT results are largest in the runout area, where the spread is partly higher than the mean value. In this part of the domain, this is mainly caused by the high PFT values of *MoT-Voellmy*. Derived runout points, based on a PFT threshold of 0.5 m, are shown in Fig.4a. The majority of the derived runout lengths are in a quite narrow range from 2290 m to 2345 m. However, *MoT-Voellmy* of the *core group* and *FLO-2D* of the *extended group* simulate a larger runout distance of up to 2420 m (*MoT-Voellmy*). In the case of *MoT-Voellmy*, the long runout (as well as the excessive PFT, see Fig.4b) is due to spike-like oscillations along the flow axis; if they are filtered out, the predicted runout is reduced to about 2300 m and PFT to about 10 m.

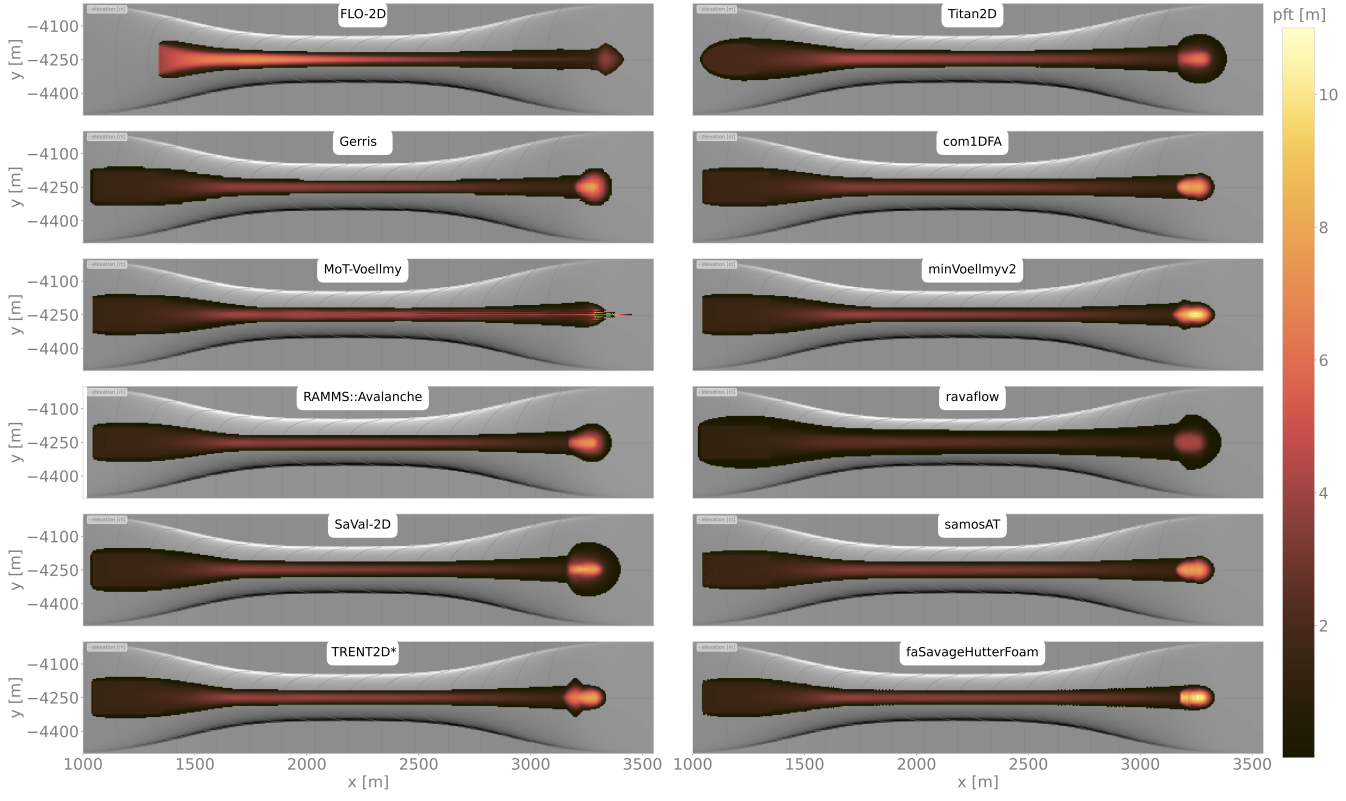

Figure 2. Peak flow thickness (pft) fields for the test case *VoellmyIdealized*. Flow thickness values are capped at 11 m and values below 0.01 m masked. *MoT-Voellmy* values (maximum 27.06 m) exceeding the colour bar limit are indicated in green. Note that, in case of *FLO-2D*, the release area is defined by hydrographs at inflow cells, arranged at a line bordering the downstream part of the release area, hence the upper part of the release area is not represented in the simulation results.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

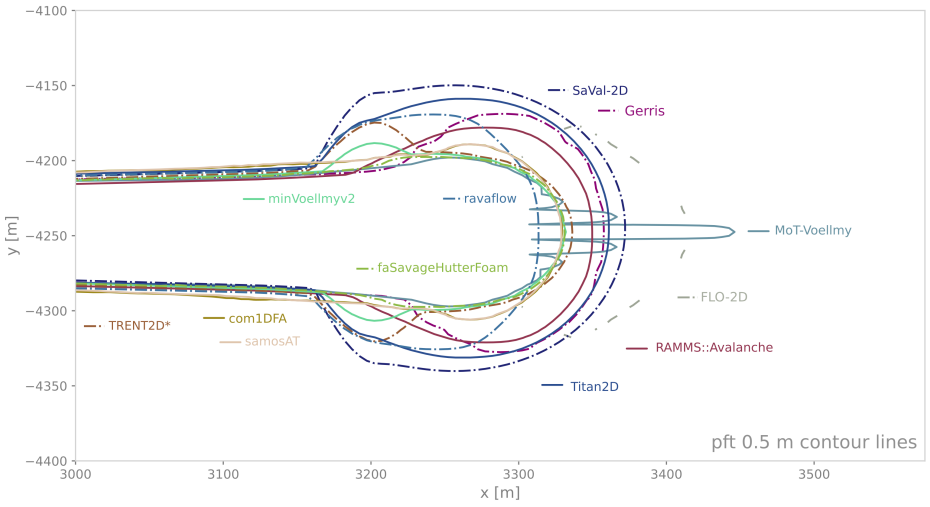

Figure 3. Contour lines of peak flow thickness 0.5 m for the test case *VoellmyIdealized*.

4.2 Test case *VoellmyReal*

In Fig. 5, the submitted peak flow thickness (PFT) fields are shown. The largest differences are found in the runout area and in the channelized region further upstream; simulation results show variations in how much material merges into the channelized part of the topography. The steep, channelized topography shows an abrupt change in inclination when reaching the valley bottom. Along the channelized part of the topography, *r.avaflow*, *MoT-Voellmy* and *SaVal-2D* show larger transverse spreading, with two regions of main flow channelization, (i) a distinct gully in the upper part feeding into the main channel and (ii) where the main channel becomes more narrow further downstream. In contrast, the other simulations show higher PFT values along the entire channelized area and less transverse spreading, indicative of a more pronounced channelling of the flow into the gully. This is reflected in the transect of PFT values shown in the panel in the lower-right corner of Fig. 5. The field of average PFT values is shown in Fig. 6a, and the relative spread (spread / average) in Fig. 6b. Along the channelized area and in the runout area, the maximum values of spread are twice as high as the average PFT value at the respective locations. The differences are largest at the edges of the flow, again indicating that simulations differ in the degree of channelling the flow. In Fig. 6b, the locations of the maximum PFT value are indicated for all simulations. The distribution of maximum PFT cross-profile values along the thalweg (Fig. 6c) also shows lower values for *r.avaflow*, *MoT-Voellmy* and *SaVal-2D*, and an increase where the main channel becomes narrower. The differences in the runout area are more pronounced with respect to the amount of lateral spreading and local PFT values than for the total runout lengths. This is also evident, looking at the relatively small spread in runout lengths measured along the thalweg. The derived runout points are shown in Fig. 6a. The spread in runout lengths is 73 m, ranging form 2051 m to 2124 m. In the simulations, a minor branch develops on the orographically left side, with PFT

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

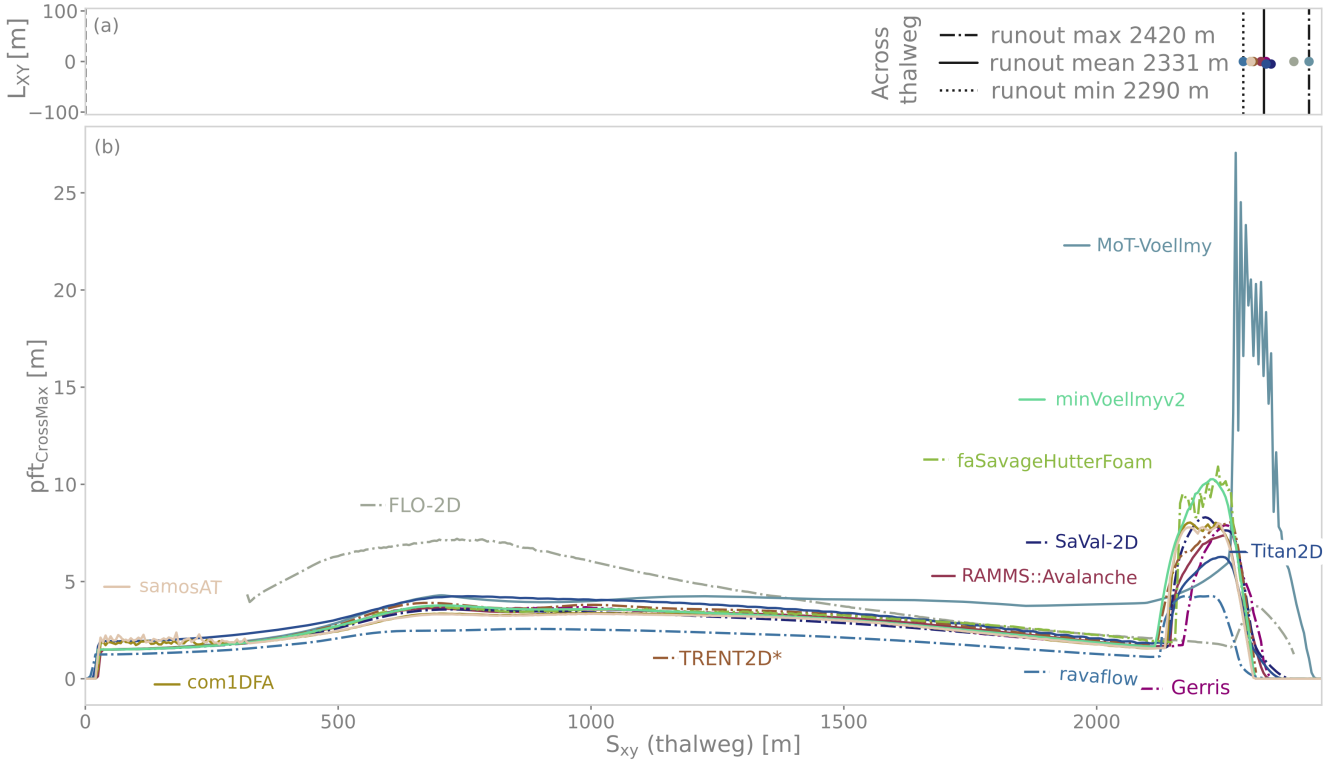

Figure 4. Test case *VoellmyIdealized*: Runout points derived using a threshold of 0.5 m peak flow thickness in panel (a). L_{XY} is the across-thalweg coordinate and S_{XY} the along-thalweg coordinate. Maximum values of peak flow thickness across the flow ($\mathrm{pft}_{\mathrm{CrossMax}}$) along the thalweg are shown in panel (b).

values up to 3 m. This branch is almost absent in the results from *FLO-2D* and *faSavageHutterFoam*; however, for the latter, the simulation domain was cut on the orographic left side as indicated in Fig. 5.

285

4.3 Test case *CoulombIdealized*

An overview of the spatial distribution of peak flow thickness (PFT) and peak flow velocity (PFV) is shown in Fig. 7. The results vary in total runout length as well as the general shape of the runout area, ranging from mostly elliptic to more conical shapes. This is also visible in the 0.5 m PFT as well as the 1 $\mathrm{m\,s^{-1}}$ PFV contour lines in Fig. 8. Looking at the runout length,
290 the simulation results can be separated into two classes, stopping at roughly 4200 m or 4700 m, respectively. This separation can be clearly attributed to whether or not curvature correction terms are accounted for in the model formulation, see Table 3 and Sect. 5.2. As in the pure Coulomb friction test case, no velocity-dependent friction term is taken into account, velocities are

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

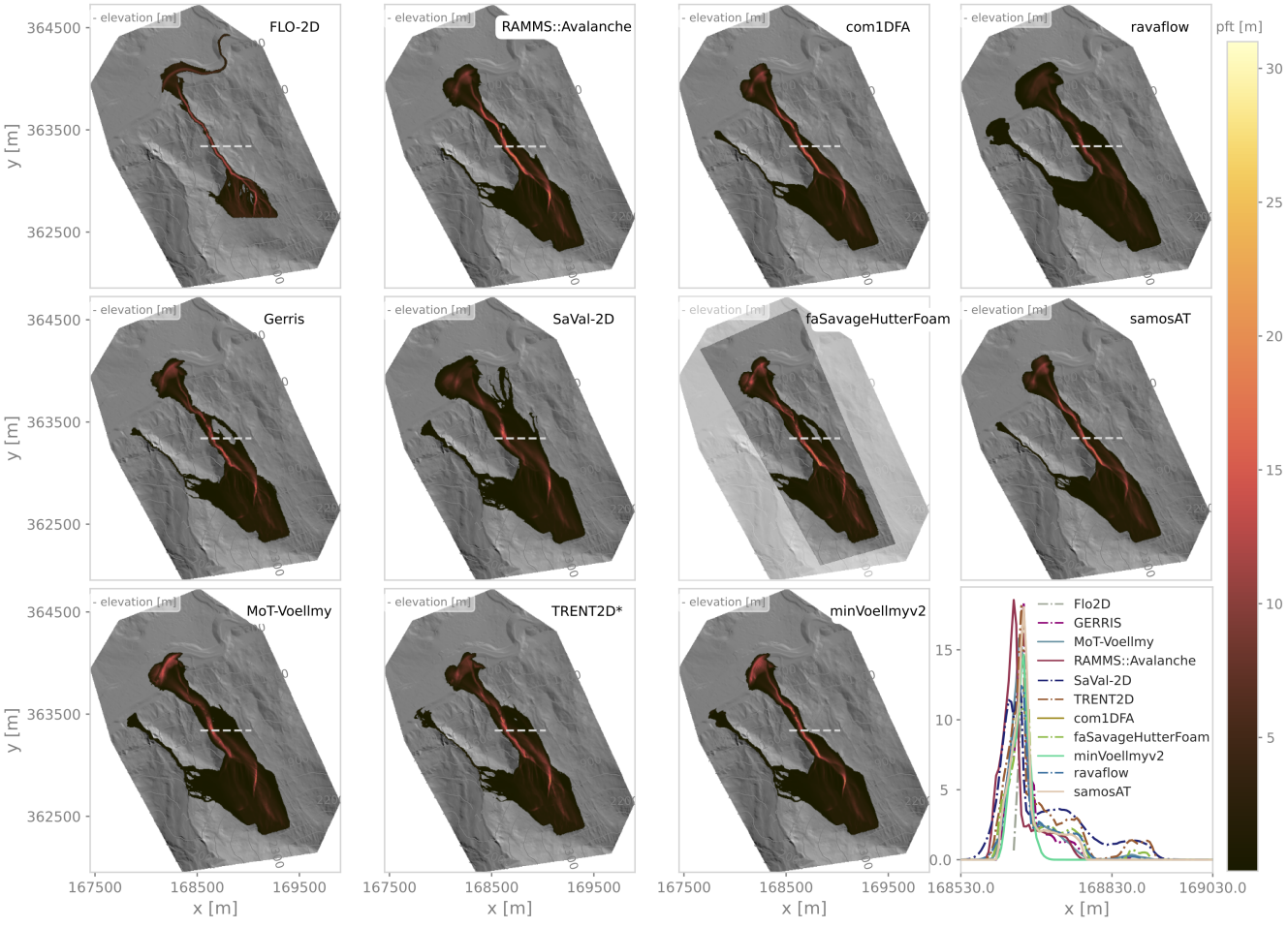

Figure 5. Peak flow thickness (pft) fields for test case *VoellmyReal*. Flow thickness values below 0.01 m are masked. For *faSavageHutterFoam* parts of the computational domain that show no data values are indicated by the transparent area. The panel in the lower-right corner shows a cross-sectional profile of PFT values, the location of the cross section is indicated with a light gray dashed lined in the other panels.

much higher and thus the effect of curvature-induced friction is much more pronounced. For *SaVal-2D* and *Gerris*, simulations were performed not taking into account curvature correction terms and hence, larger runout lengths are expected.

Besides this expected difference, variations in the thickness distribution within the runout area are apparent. With the exception of *TITAN2D* and *MoT-Voellmy*, all simulations show a sharp increase and subsequent decrease of PFT where the flow comes to a rest, i.e. a narrow band of high thickness bounding the runout area (see Figs.7 and 9c). Derived runout lengths range from 2915 m to 3633 m, comprising a spread of 718 m (see Fig. 9a). If based on a threshold of 1 $\mathrm{m\,s^{-1}}$ PFV, the spread in runout lengths is similar (see Fig. C6). However, excluding the results without curvature correction terms (*SaVal-2D* and *Gerris*), derived runout lengths range from 2915 m to 3338 m, which results in a spread of 424 m. Also excluding the result

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

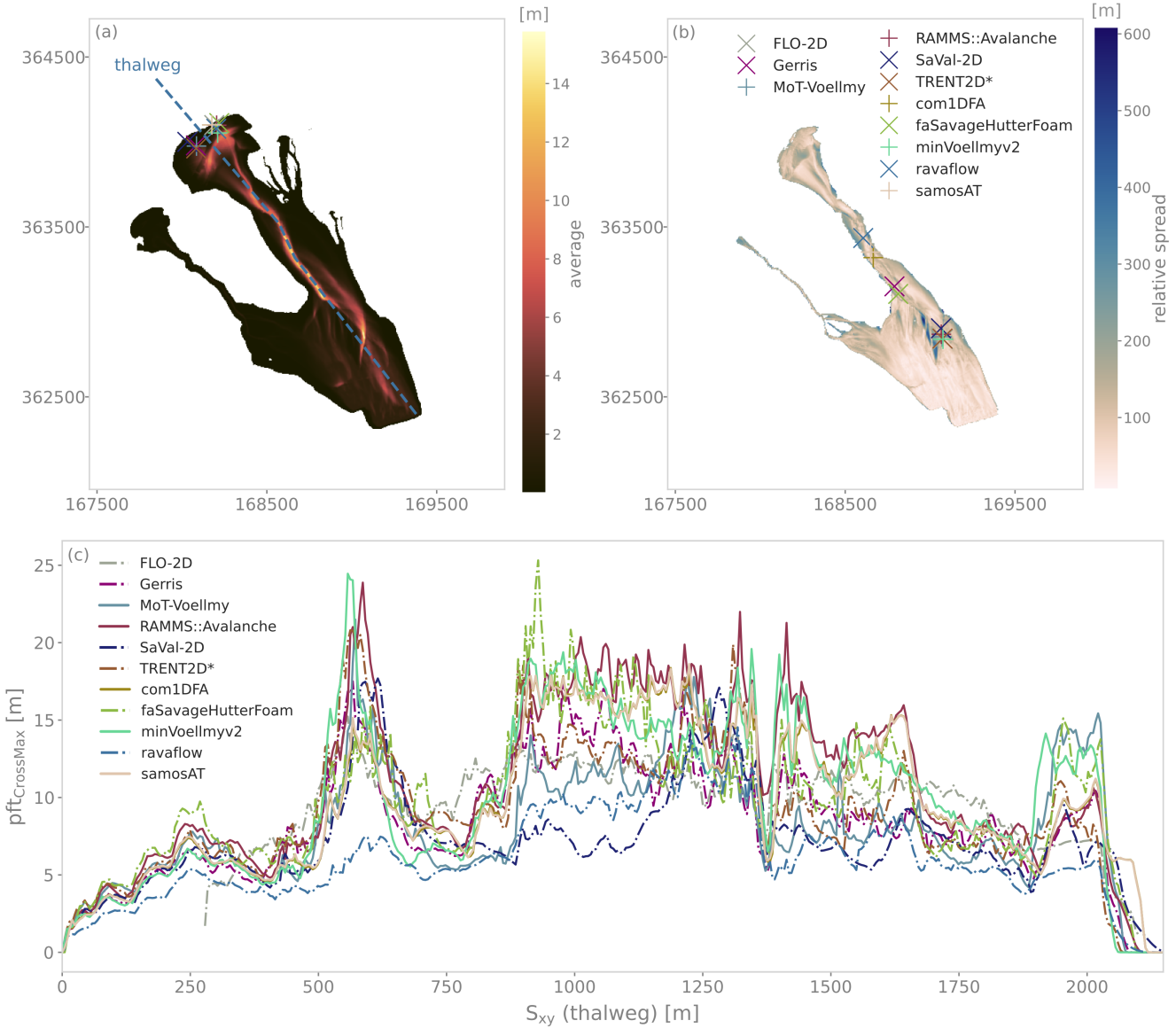

Figure 6. Test case *VoellmyReal*: Average peak flow thickness (PFT) field for all simulations of the *core group* with derived runout points (based on 0.5 m PFT) and the thalweg in panel (a). Relative spread in simulation results derived as the difference of maximum and minimum PFT values divided by the average for each raster cell, with markers indicating the location of the maximum PFT value in panel (b). Cross-sectional maximum peak flow thickness ($pft_{CrossMax}$) along the thalweg for all simulation results in panel (c). For the spatial fields of average and relative spread only results from the *core group* are considered.

of *MoT-Voellmy*, where the derived runout length is strongly influenced by the spike-like oscillations (see Sect. A7), yields a spread of 329 m. Still, this is a significantly larger spread compared to the *VoellmyIdealized* test case. However, as velocities

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

are not constrained by a velocity-dependent friction term in this test case, velocities are much larger, affecting total runout lengths and the observed spread in simulation results. Compared to the *VoellmyIdealized* test case, also larger differences in PFT values are present along the entire avalanche path. Looking at maximum PFT cross-profile values in Fig. 9b, all results show a distinct increase when the channel becomes narrower, and some results indicate a second increase before the channel widens again. However, the maximum PFT values at those locations vary almost by a factor of two, and even more significant differences in PFT values are observed in the runout area. For *MoT-Voellmy*, the high PFT cross profile values in Fig. 9b, are caused by oscillations along the coordinate directions, more information on the formation of these oscillations is provided in Sect. A7. Furthermore, the PFV fields in Fig. 7 reveal distinct symmetric features for *minVoellmyv2*. *r.avaflow* results show very high PFV values (up to $322\,\mathrm{m\,s^{-1}}$) on the inclined part of the topography and an abrupt decrease to values similar to the other simulation results when the topography transitions into the flat foreland. Sect. 5.1 provides further details on the origin of these anomalies. Most of the other simulation results show a steady but much less pronounced increase in PFV followed by a steeper decrease in PFV at the transition into the flat foreland. In the PFV maximum cross-profile values (Fig. C6), this steeper decrease is not present for *SaVal-2D*, where curvature effects are not taken into account.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

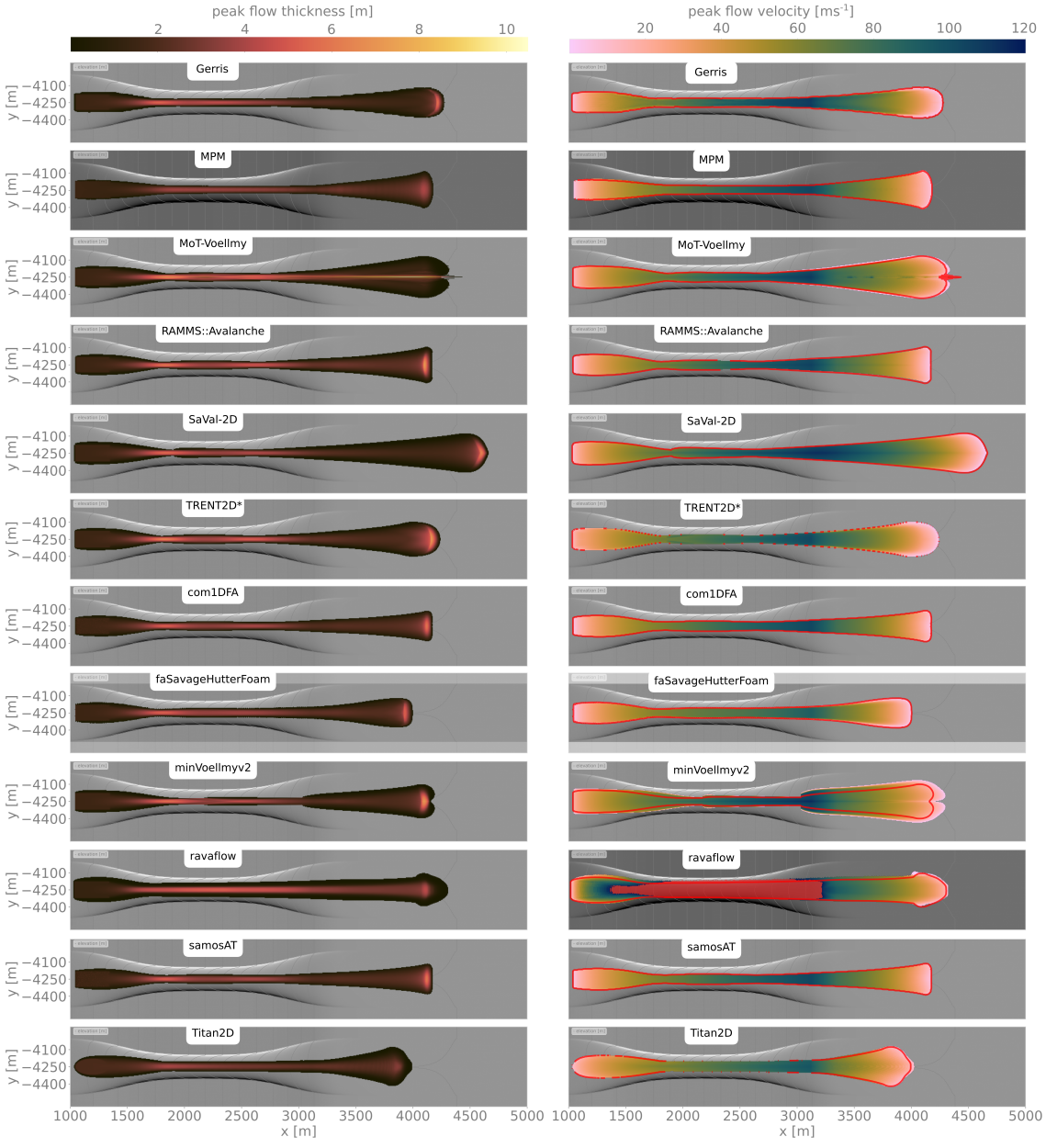

Figure 7. Peak flow thickness and peak flow velocity fields for test case *CoulombIdealized*. Flow thickness values are capped at 10.5 m, whereas flow velocity values are capped at 120 $\mathrm{m\,s^{-1}}$. For both fields, values below 0.01 m or 0.01 $\mathrm{m\,s^{-1}}$ respectively, are masked. Where the values from *r.avaflow* (maximum 322 $\mathrm{m\,s^{-1}}$) exceed this range, they are indicated in red. The red lines show the 0.01 m peak flow thickness contour line.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

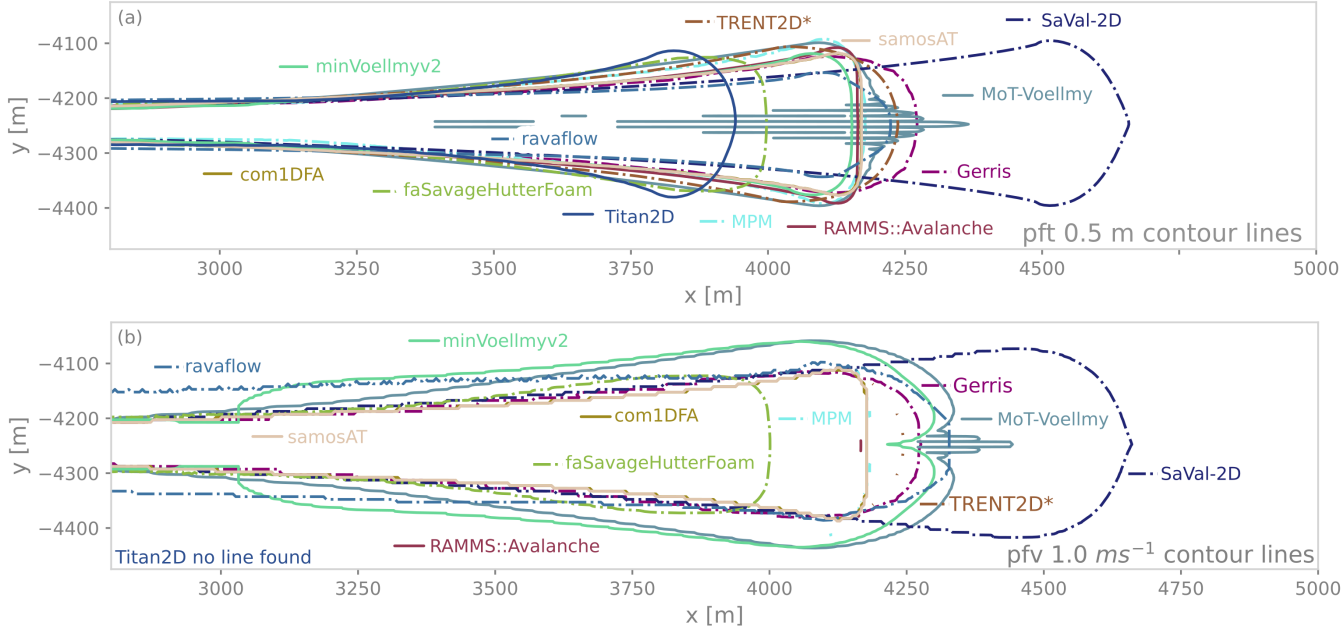

Figure 8. Contour lines for test case *CoulombIdealized* of 0.5 m peak flow thickness (PFT) in panel (a) and 1 m s^{-1} peak flow velocity (PFV) in panel (b). No 1 m s^{-1} PFV contour line was found for the *TITAN2D* simulation, as the minimum PFV value found is 1.3 m s^{-1} before transitioning into no-data values.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

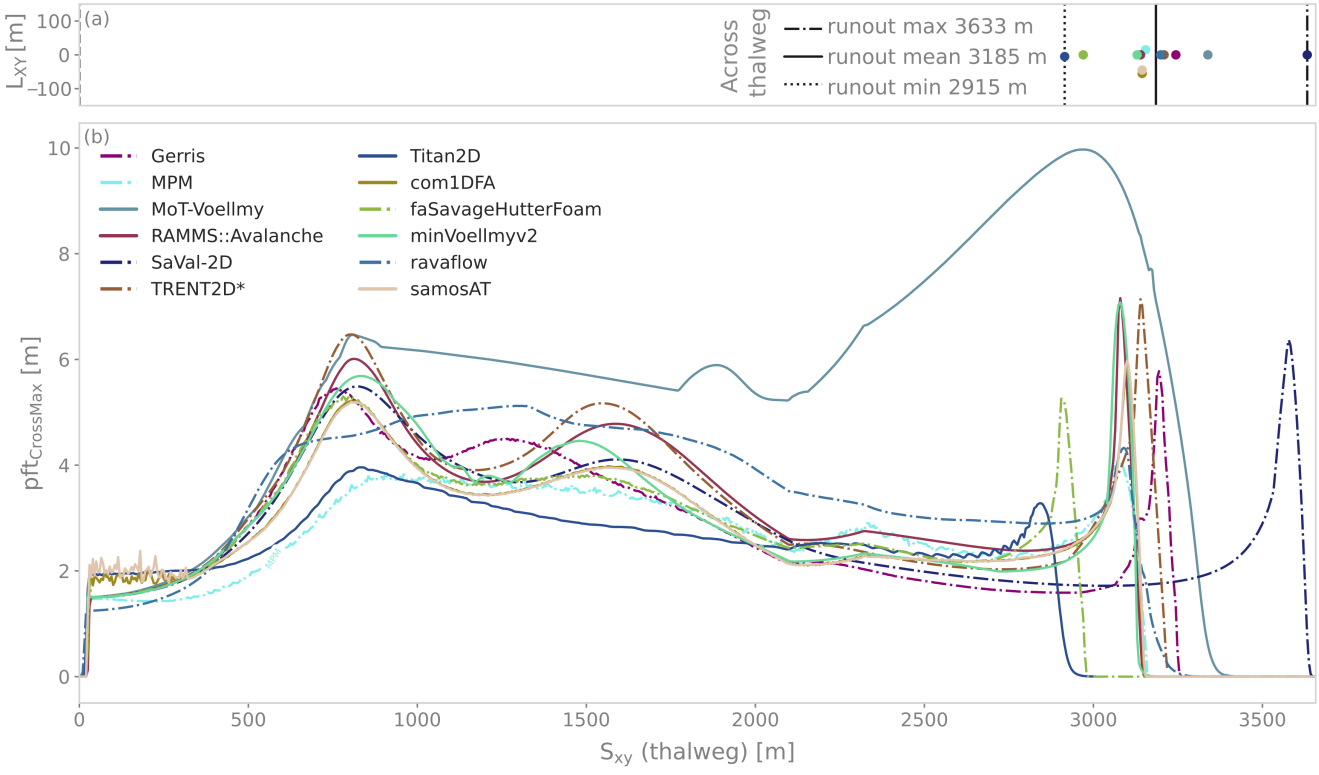

Figure 9. Test case *CoulombIdealized*: Runout points derived using a threshold of 0.5 m peak flow thickness in panel (a). L_{XY} is the across-thalweg coordinate and S_{XY} the along-thalweg coordinate. Maximum values of peak flow thickness across the flow ($\mathrm{pft}_{\mathrm{CrossMax}}$) along the thalweg for all simulation results are shown in panel (b).

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

5 Discussion

In this section, the distribution of scalar measures derived from the simulation results is discussed and compared among the different test cases, representative of different topographical conditions and applied friction laws. This is followed by a discussion of potential sources for the found discrepancies, examining differences in modelling approaches and configurations. Furthermore, common issues and lessons learned with respect to performing an intercomparison study are highlighted.

5.1 Quantitative assessment of differences as an estimate of uncertainty from implementation workflow

Deriving scalar measures like maximum peak flow variables or runout lengths and runout angles allows us to compare the spread in simulation results for the different simulation tools and among the different test cases. This may provide further insight into where the largest differences occur, and whether they are related to, for example, topographical conditions. In addition, comparing the results for tools of the *core group*, which rely on (nearly) the same basic assumptions and set of equations, offers the possibility to investigate the uncertainty introduced by the different implementation workflows. Statistical measures of the distribution of the *core group's* results, like median and interquartile range for example, may serve as a rough estimate for this.

Runout lengths are derived based on a peak flow thickness threshold of $0.5\,\mathrm{m}$ (see definition of how runout lengths are derived in Sect. 4). In Fig. 10, the distribution (solely computed for the *core group*) of runout lengths, angles and maximum peak variable value is presented in the form of box plots. The box extends from the lowermost quartile Q_1 (25%) to the uppermost quartile Q_3 (75%) indicating the interquartile range ($\mathrm{IQR} = Q_3 - Q_1$), the whiskers extend to the highest and lowest value found between $Q_1 - 1.5\,\mathrm{IQR}$ and $Q_3 + 1.5\,\mathrm{IQR}$, all values exceeding this range are classified as outliers. In addition, the individual values for all participating simulation tools are provided alongside the box plots for the respective test cases. The same analysis based on a peak flow velocity threshold of $1\,\mathrm{m\,s^{-1}}$ is shown in Fig. C1. The respective values for the interquartile range for each scalar measure are provided in brackets next to the box plots. The IQRs of the scalar measures runout length and runout angle constitute 1–3 % of the median value for all three test cases. However, as the runout lengths are measured from the start of the thalweg, an IQR of 1–3 % of the median values can still represent a significant deviation, i.e. 30, 33 and $92\,\mathrm{m}$ for the *VoellmyIdealized*, *VoellmyReal* and *CoulombIdealized* test cases, respectively. To make the relative spread more informative, one could measure the runout length from a location further down the valley. However, as this would require a rather arbitrary choice of where to set that location, we decided to look at the entire avalanche track and focus more on the absolute deviations. The maximum peak variable values show a much larger relative spread: the IQR of the median value of the maximum peak flow thickness, MPFT, is 20–27 % and the IQR of the median value of the maximum peak flow velocity, MPFV, is 7–14 %. In addition to the IQR, it is important to also consider whether there are outliers, and if so, what are potential sources of the corresponding deviations. A further detailed description of the distribution of the derived scalar measures is provided for each test case separately in the following sections.

It is important to note that this comparison does not provide a measure of modelling accuracy. Simulation results were neither compared to an exact solution (as these would require a reduced model setup) nor to a reference solution from an

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

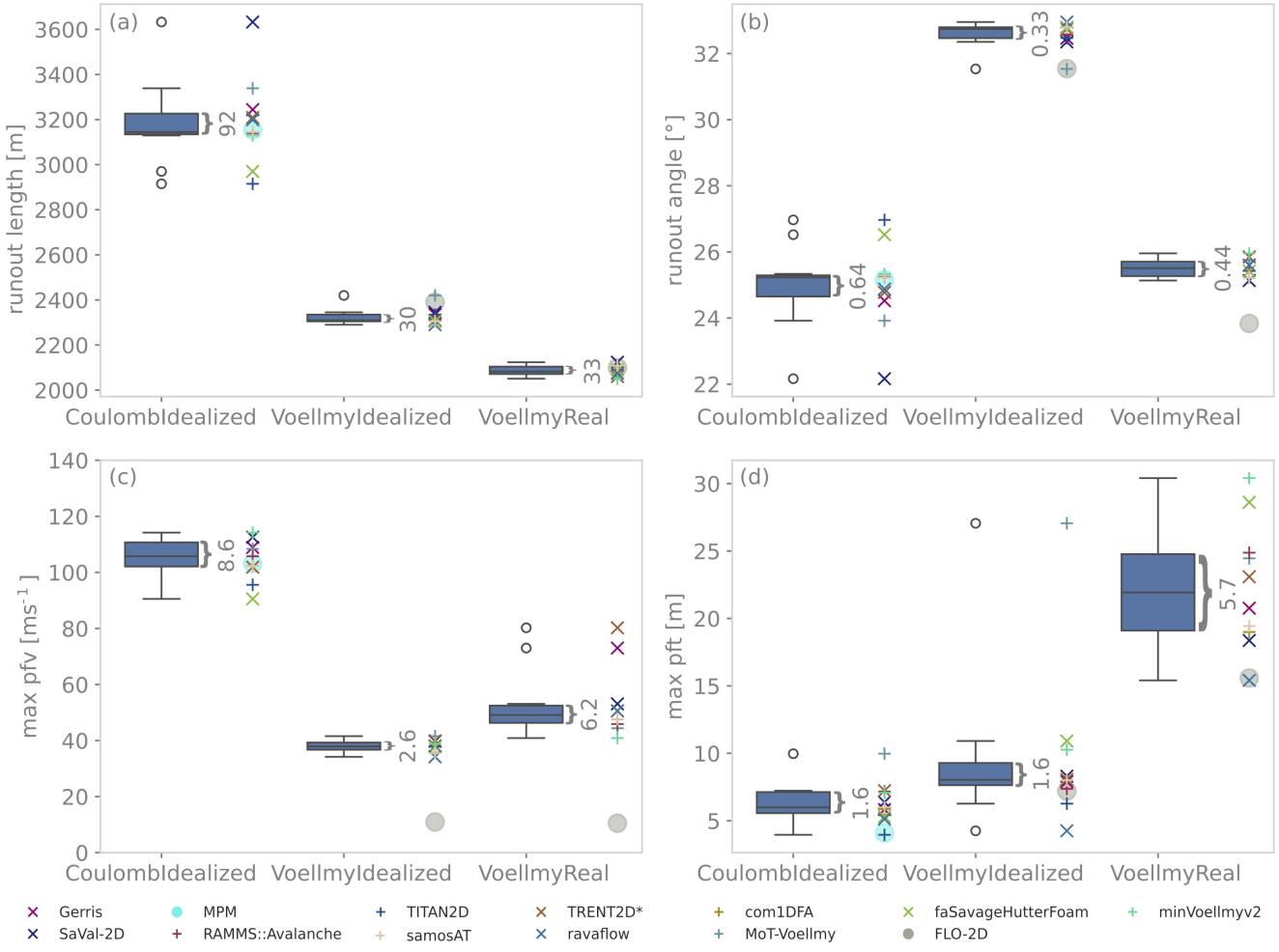

Figure 10. Distribution of scalar measures derived from the simulation results (for runout lengths and angles the analysis is performed along the thalweg and based on a peak flow thickness threshold of 0.5 m). (a) Runout length, distance from start of thalweg to runout point measured along the thalweg. (b) Runout angle, measured from release to runout point. (c) Maximum value of peak flow velocity with the y-axis limited to 140 $\mathrm{m\,s^{-1}}$. There is one value exceeding this limit (*r.avaflow*; 322.6 $\mathrm{m\,s^{-1}}$; *CoulombIdealized*)). (d) Maximum value of peak flow thickness. The box plots show the quartiles of the dataset (only *core group*) and categorized outliers as circles, maximum extent of whiskers is based on quartiles minus (plus) 1.5 times the interquartile range. The interquartile range is added as text with a brace for each box plot. Additionally, the individual data points are shown using coloured × or + markers for the *core group*, and filled ○ for the *extended group*. Only the models belonging to the *core group* are included in the box plot distributions.

observed event (in order to exclude any effects of parameter calibration and optimization). The actual modelling accuracy of an individual simulation tool when simulating a specific case can be higher or lower than the spread in simulation results presented in this study.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

355

Test case *VoellmyIdealized*

For this test case, the spread in runout length, runout angle and MPFV is the smallest of all three test cases. Only *MoT-Voellmy* is classified as outlier for runout length and runout angle, whereas for MPFT, also *r.avaflow* is classified as outlier. The very high value of MPFT of 27.06 m for *MoT-Voellmy*, representing +236 % of the median value (8.04 m), is the result of strong
360 oscillations along the coordinate directions within the runout area (see Fig. 2 and further details on the cause of these oscillation in Sect. A7). For all simulations, MPFT is found in the runout area where the flow comes to rest, except for *FLO-2D* (*extended group*). Whereas *r.avaflow* is the only result classified as outlier, also *faSavageHutterFoam* and *minVoellmyv2* show distinctly higher MPFT values (+35 % and 27 % of the median value in the runout area). The longitudinal PFT patterns (Fig. 4) of these two models are overall very similar. The results from *MoT-Voellmy* and *r.avaflow* also show significant deviations over larger
365 stretches along the thalweg compared to the other simulation tools.

The only simulation result of the *extended group* available here, *FLO-2D*, shows a longer runout length (see filled gray circle in Figs. 10 and C1). The smaller runout angle is partly caused by the difference in how the release area is initialized. *FLO-2D* prescribes a hydrograph at each cell that is classified as inflow cell, which for this test case, are located in a straight line following the front of the downstream border of the release area. *FLO-2D* showed computational instability when simulating
370 inflow cells draining into each another, ultimately causing the simulation to fail. Therefore, the inflow cells were located downstream the release area in order to immediately begin the flow propagation. Compared to the majority of simulation results, higher peak flow thickness is simulated over most of the channel, and in the runout area, the magnitude of the peak flow thickness is approximately half of the average peak flow thickness values (see Fig. 4d). Regarding the peak velocity fields, differences are very apparent. However, *FLO-2D* relies on a different friction relation and was tuned to reproduce the runout
375 length and the shape of the deposition. Hence flow velocity has not been considered in the initial attempt of applying this tool to simulate snow avalanche flow, for further information see Sect. A2.

Except for these outliers, the spread in simulation results of the other models is relatively small for both scalar measures (Fig. 10) as well as for the spatial distribution of peak flow thickness and velocity fields (Figs. 2 and C2).

Test case *VoellmyReal*

380 Regarding the *VoellmyReal* test case, the IQR values are slightly larger compared to the *VoellmyIdealized* test case. This is partly due to the larger overall spread. In addition, the results are more evenly distributed and hence, no result is classified as outlier for runout length, runout angle and MPFT. The IQR for runout length and runout angle is 1.6 % and 1.7 % of the median value, respectively, corresponding to 33 m and 0.44°. For the real-world topography, the total runout length is constrained by the abrupt flattening of the terrain at the valley bottom and the counter slope the flow hits after crossing the riverbed, causing
385 the spread in runout lengths to be small. The spread in MPFT and MPFV (IQR of 26 % and 13 % of the median value, respectively) is, however, considerably larger than in the *VoellmyIdealized* test case. Also the location of MPFT varies between the simulations. While *SaVal-2D*, *RAMMS::Avalanche*, *TRENT2D**, *MoT-Voellmy*, *minVoellmyv2* and *FLO-2D* show the MPFT

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

in the distinct gully feeding into the main channel in the upper part, *Gerris*, *r.avaflow*, *com1DFA*, *faSavageHutterFoam* and *samosAT* simulate MPFT values in the main channel before it becomes narrower (see MPFT markers in Fig. 6b). The different locations of MPFT values as well as the larger spread between the simulation results may be related to differences in the degree of flow concentration into the main channel. Simulations where less lateral spreading is observed show higher MPFT values (see cross profiles in Fig. 5 lower right panel). For MPFV, *TRENT2D** and *Gerris* are classified as outliers, showing values of +63 % and +49 % of the median value, respectively.

In the *FLO-2D* result (only tool from the *extended group* for this test case), the flow is highly concentrated in the main channel, without any significant lateral spreading. Within the runout area, the flow spreads laterally and partly follows the riverbed to the orographic right side. The MPFT is located in the upper part of the flow and its magnitude (15 m) is at the lower end of the distribution (gray filled circle in Fig. 10d).

Compared to the simple, single-channel topography of the idealized test case, the complex topography of the real-world scenario offers more degrees of freedom for the avalanche to move. This might explain the larger spread in the results of the *VoellmyReal* compared to *VoellmyIdealized* test case. Nonetheless, the entire comparison is based on the peak fields, i.e. the maximum values over the entire simulation duration at each cell. These are a result of the spatio-temporal evolution of the result variables, where small differences in the flow direction can accumulate over time and then show a strong signal in the respective peak fields.

Test case *CoulombIdealized*

Among all three test cases, the largest differences are observed for the *CoulombIdealized* test case. These differences are prominent in the distribution of runout lengths and angles, as shown in Fig. 10. However, differences are also visible in the overview plot of peak flow thickness and velocity fields (Fig. 7), as well as in the cross-sectional MPFT values along the thalweg in Fig. 9c. Regarding the distribution of runout lengths and angles, the IQR is up to 3 % of the median values, corresponding to 92 m and 0.64°, respectively. Also, more results are classified as outliers. This is the case for *SaVal-2D*, *faSavageHutterFoam* and *TITAN2D* with respect to runout length and runout angle.

In the case of *SaVal-2D*, simulations were performed without curvature correction terms, hence longer runout lengths are expected. To illustrate this: a mass point that is located at the centre of the release area, assuming pure Coulomb friction, would come to a rest at roughly 4615 m. Despite not taking effects of curvature into account, *Gerris* yields a runout not much longer than that predicted by the models that include effects of curvature. *Gerris* does not preserve the absolute value of the velocity on a curved surface, but only the horizontal component. As a consequence, *Gerris* loses velocity at the transition to the runout zone.

For MPFT, only *MoT-Voellmy* is classified as outlier. Similarly to the *VoellmyIdealized* test case, this high MPFT value is caused by strong oscillations of the peak flow thickness along the coordinate directions (see Fig. 7). Further details on these oscillations are provided in Sect. A7.

Regarding MPFV, only *r.avaflow* is classified as outlier and showing values of up to 322 $\mathrm{m\,s^{-1}}$, which is more than a factor two higher than the median. In general, in *r.avaflow*, velocities tend to be overestimated in areas of low flow thickness: this

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

is a combined effect of spatial redistribution of momenta and the calculation of velocities through division of momentum by flow thickness. It can be mitigated by increasing momentum decay for low flow thicknesses, e.g., through turbulent friction. Therefore, velocities in *r.avaflow* are much more realistic using a Voellmy-type friction compared to a pure Coulomb friction relation. However, further tests indicate that in this test case, high velocities are not only restricted to coinciding low flow thickness values. This issue is currently being studied in more detail. The observed symmetric features in the peak flow velocity field of *minVoellmyv2* (Fig. 7) arise from neglecting lateral stresses in combination with a symmetric discretization scheme. If the topography is symmetric with a flow line exactly in the middle, this central flow line does not interact with the neighbouring flow lines if its surface elevation is higher. In contrast to the general problem with flow parallel to the coordinate axes discussed by Hergarten (2024b) it is irrelevant in real-world applications since it is related to the axial symmetry.

For this test case, a 3D approach was included as part of the *extended group*. *MPM* simulation results are within the IQR of the *core group* regarding runout length, runout angle and MPFV, while a lower value is observed for MPFT (4.1 m). The distribution of peak flow thickness in the runout area is of rather conical shape and the thickness of the band bordering the runout area is less pronounced compared to the bulk of simulation results (Fig. 7). Looking at the cross-sectional MPFT values (Fig. 9), neither of the first two peaks in PFT (first and more distinct peak around $S_{xy} = 750\,\mathrm{m}$, second and broader peak around $S_{xy} = 1500\,\mathrm{m}$) is present in the *MPM* PFT field. After the initial increase, there is a steady decrease followed by a small increase when the inclined slope starts to transition into the flat foreland. A final increase where the flow comes to a rest can be seen in Fig. 9. However, cross-sectional MPFV values show a very similar behaviour and magnitude as the bulk of simulations. Further tests are required to study how the simulation results differ for more complex topographical conditions.

In order to identify all the sources of differences and their relative contribution, more tests and also additional result fields (for example, flow thickness and flow velocity fields for individual time steps) would be required. However, some potential sources of differences can be inferred on the basis of the submitted results, as discussed in Sect. 5.2.

Spatial variability

The maps in Fig. 11 provide further insight into the uncertainty stemming from the differences in implementation workflows for tools of the *core* group. Here, the fraction of simulations exceeding a threshold of 0.5 m peak flow thickness (and $1\,\mathrm{m\,s^{-1}}$ peak flow velocity) is shown for the entire computational domain. The white contour line indicates areas where 95 % of all simulations exceed these thresholds. This type of visualization provides an overview of the spatial variability of the simulation results. Areas with a strong gradient in the fraction indicate better agreement between the individual simulations, whereas a weaker gradient points to areas where the differences in implementation or chosen configuration lead to bigger variability. This might provide further information regarding the source of the differences, for example if these zones exhibit certain features, such as strong terrain curvature, flattening of the terrain or obstacles.

For example, in Fig. 11e the orographically right part of the runout zone in *VoellmyReal* test case, shows less variation between the individual simulations compared to the left part. At the left side, the distance to the counter slope is shorter, which causes a stronger deceleration of the flow. Further upstream, higher variability is present on the orographically right side. This

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

reflects the differences in the amount of transverse spreading observed in Fig. 5. In this region, also the greatest differences between maps derived for 0.5 m PFT and 1 $\mathrm{m\,s^{-1}}$ PFV occur. However, these differences of course depend on the chosen thresholds of the peak variables; the choice here was based on values relevant for practical purposes such as hazard zoning.

In summary, this map of model results variations is a helpful means for investigating the spatial variance of simulation results and provides an estimate of uncertainty introduced by the different implementation workflows. While model uncertainty for a single model can be assessed through methods like parameter variation, implementation uncertainty can be evaluated by aggregating simulation results from several models. These maps could become an interesting tool also in operational applications: combining results of several tools into a single map, similar to an ensemble forecast.

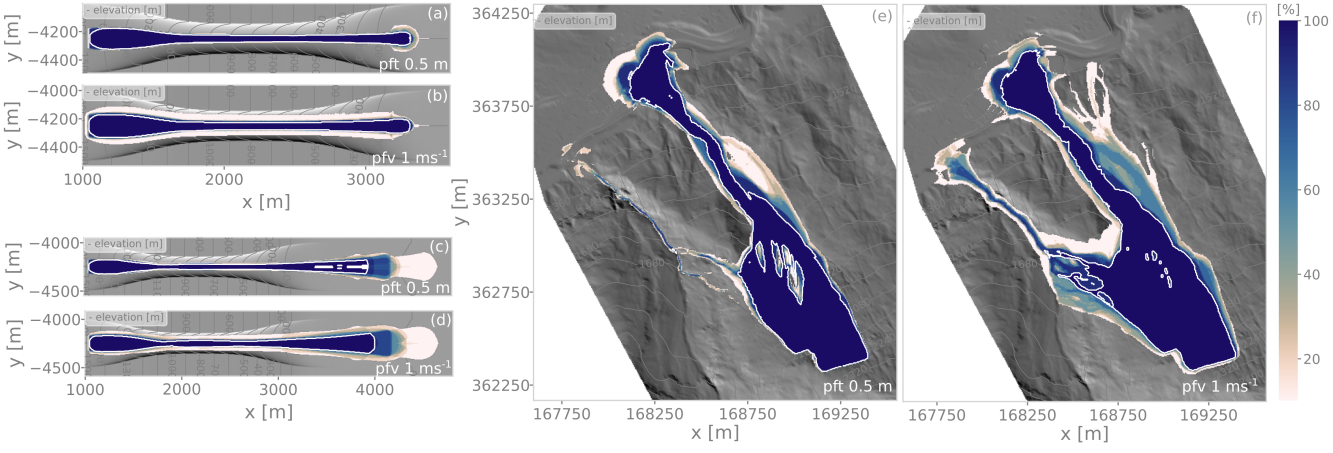

Figure 11. Fraction (in percent) of simulation results exceeding a threshold of 0.5 m peak flow thickness for the test cases *VoellmyIdealized* (a), *CoulombIdealized* (c) and *VoellmyReal* (e); and for a threshold of 1 $\mathrm{m\,s^{-1}}$ peak flow velocity for the test cases *VoellmyIdealized* (b), *CoulombIdealized* (d) and *VoellmyReal* (f). The contour line indicates the area where 95 % (white line) of simulations exceed the threshold of 0.5 m (1 $\mathrm{m\,s^{-1}}$). Only simulations of the *core group* are included.

5.1.1 Correlation across test cases

For practical considerations, it would be helpful to know whether there are trends across the different test cases, for example, whether some models tend to predict a rather long runout or rather high flow velocities under all conditions.

As described in Sect. 4.3, the *CoulombIdealized* test case yields a runout of several hundred meters into the flat area and also a spread of several hundred meters among the different models. In turn, the two scenarios using a Voellmy friction relation yield a much shorter runout and also a smaller spread, mainly owing to the limitation of the velocity by the velocity-dependent friction. Figure 12 compares the runout lengths of the models among the different test cases.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

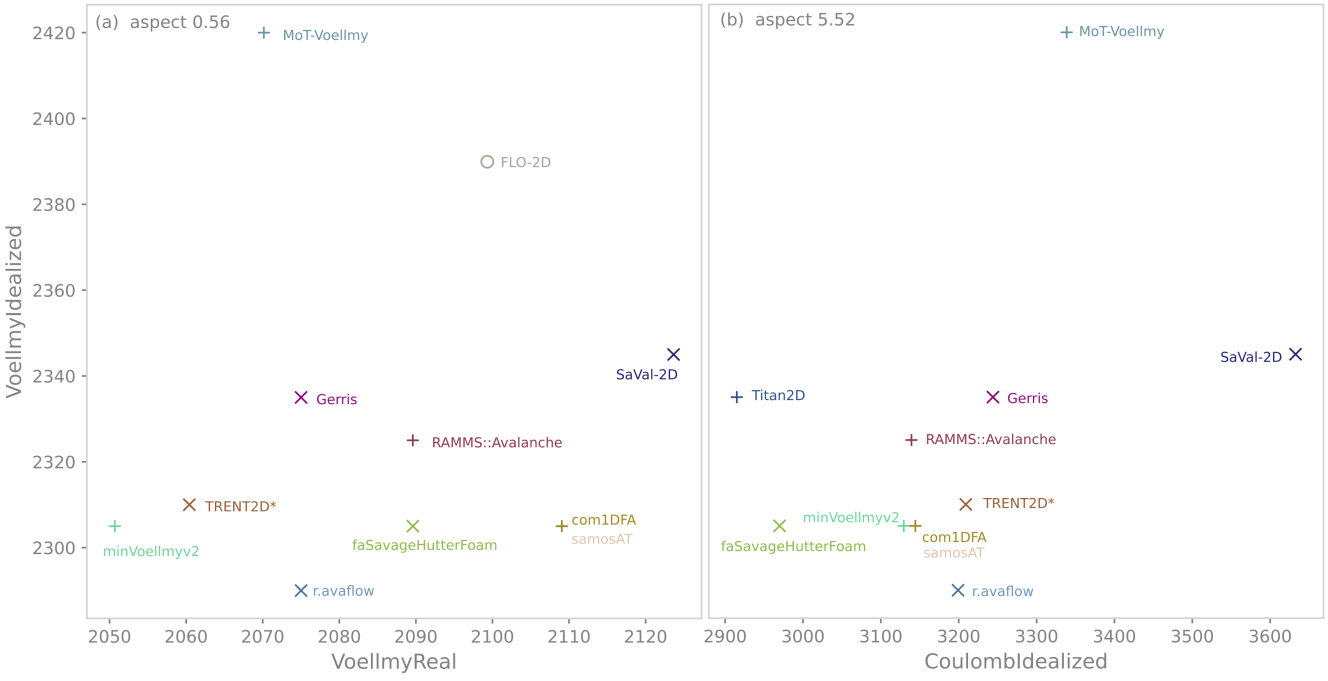

Figure 12. Runout lengths predicted by the models for the three considered test cases: (a) *VoellmyIdealized* versus *VoellmyReal* and in (b) *VoellmyIdealized* versus *CoulombIdealized*. Note the different axes aspect ratios, as indicated in the upper left corner of either panel.

Both diagrams suggest a positive correlation at first, which means that models with a long runout for the idealized topography with Voellmy friction tend to predict a long runout for the respective real-world topography (Fig. 12a), as well as for the idealized topography with Coulomb friction (Fig. 12b). This trend is, however, mainly supported by outliers. Among the models that were evaluated for all three test cases, *SaVal-2D* and *MoT-Voellmy* predict the longest runout for all test cases. This is at least partly caused by neglecting effects of curvature in the case of *SaVal-2D* and the oscillations along the coordinate axis for *MoT-Voellmy*. If we exclude these two models and *FLO-2D* (which is based on a different friction relation), the correlations in runout between the two Voellmy scenarios (Fig. 12a) is lost immediately, and that between the two idealized topographies (Fig. 12b) becomes less clear. If we take into account that *Gerris* should already yield a longer runout because it does not take into account the effect of curvature on friction, the correlation among the remaining models is also lost entirely. In summary, there is a rather large group of models with a moderate variation in predicted runout length, but this variation is not systematic among the considered scenarios.

The maximum peak flow velocities also reveal no clear trend. For the *VoellmyIdealized* test case, the variability of maximum peak flow velocities is rather small (see Fig. 10c). Here, the inclined part of the topography is long enough for the avalanche to reach the terminal velocity with the prescribed friction parameters. Variations are larger for the *VoellmyReal* test case, where *TRENT2D** and *Gerris* predict higher peak flow velocities, and for the *CoulombIdealized* test case, where *r.avaflow* shows

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

exceptionally high peak flow velocity. These outliers, however, do not support a trend across all scenarios. The moderate variation among the remaining models is also not systematic.

5.1.2 Correlation between runout length and peak flow velocity

While variations in runout length and peak flow velocities do not show a strong correlation across the considered scenarios (see Sect. 5.1.1), a positive correlation between flow velocity and runout length is expected. Figure 13 provides an analysis of the relation between runout length and maximum peak flow velocity (PFV). As a guide to the eye, the dashed lines show the simplest model for this relation, based on the idea that a point-like mass enters a flat runout zone at the position L_0 along the track with a velocity v_0. Assuming that the deceleration is dominated by Coulomb friction, the respective runout length is given by Eq. 2:

$$L = L_0 + \frac{v_0^2}{2\mu g}. \tag{2}$$

The dashed lines reflect the hypothesis that v_0 is the maximum PFV.

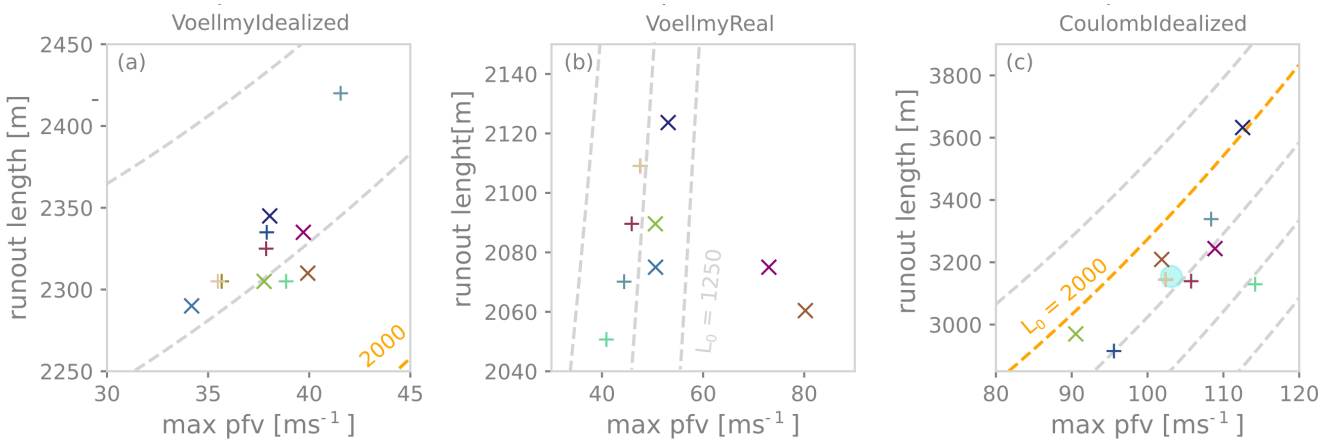

Figure 13. Runout length versus maximum peak flow velocity (max pfv) for the test cases (a) *VoellmyIdealized*, (b) *VoellmyReal* and (c) *CoulombIdealized*. Data points with very high velocities (*r.avaflow* in the *CoulombIdealized* test case) or very low velocities (*FLO-2D*) were omitted for clarity. The dashed lines correspond to Eq. (2) with different values of L_0.

The data for the *VoellmyIdealized* test case (Fig. 13a) align well with the simple prediction, although the scatter is high. *MoT-Voellmy* might be considered an outlier, but as is the case for the *CoulombIdealized* scenario, the numerical instabilities increased the runout length by an estimated 100–120 m (see Sect.A7).

For the *VoellmyReal* test case, there also seems to be a positive correlation between runout length and maximum PFV if the high maximum PFV values of *TRENT2D** and *Gerris* are disregarded. However, the dashed lines of the simple model

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

suggest that the observed variations in maximum PFV should result in a larger variation in runout length than found here. This finding suggests that the maximum PFV is not a useful predictor for other properties. For this test case, the spatial distribution of the PFV is very inhomogeneous and the maximum value occurs at different locations for different models (see Fig. 6b). Furthermore, in addition to the abrupt change in slope angle when entering the valley bottom, the runout is practically limited by the channel that crosses the runout zone. Therefore, the flow dynamics in the runout zone have a strong effect and it cannot be expected that the simple model explains the observed variations in runout.

The simple model in Eq. 2 should be most representative for the *CoulombIdealized* test case; not only since it assumes only Coulomb friction, but also because the maximum PFV is typically reached at the transition to the runout area. Figure 13c indeed reveals a correlation between runout length and maximum PFV. The trend is roughly consistent with the simple model, although the scatter in the runout lengths is still high. The long runout length predicted by *SaVal-2D* can at least partly be explained by the high maximum PFV. The same would apply to *MoT-Voellmy* if the runout length was not affected by the developed numerical instabilities (see Sect.A7). In turn, the runout predicted by *minVoellmyv2* is quite short in relation to the maximum PFV, although the individual values would not define it as an outlier.

5.2 Potential sources of variability: differences in model formulations and configurations

The problem definition of the test cases does not cover all setup choices of the participating simulation tools, i.e., input, model and numerical parameters (see Sect. 2.2.1). One reason is that simulation tools encompass different *model types*, namely depth-averaged vs. depth-resolved models or approaches relying on a different rheology and way of representing the release mass (*FLO-2D*). However, also within the *core group*, model-specific extra parameters must be chosen. These range from the spatial and temporal resolution to methods for suppressing instabilities and the treatment of curvature effects or pre- and postprocessing steps. These parameters are often specific and not required or available among all participating tools. The three test cases were designed to represent the smallest common denominator with respect to a standard application: to simulate snow avalanche flow by prescribing the terrain (digital elevation model), the release area and initial snow thickness (a polygon with a constant thickness value) as well as a Voellmy friction relation with prescribed constant friction parameter values.

The effects of a specific numerical method and its implementation can only be isolated in simplified problems that allow for an exact solution. However, such tests are an integral part of model development and are not the focus of this intercomparison. Instead, we compare simulation results for more complex or real-world scenarios, for which no analytic solutions exist. In such cases, the results are also affected by other steps of the simulation workflow (input data handling, interpolation steps, exporting, etc.) and the chosen settings for any model-specific extra parameters (summarized as *implementation workflow*).

Comparison of only the peak flow velocity and peak flow thickness fields is not sufficient to attribute the observed differences in the simulation results to individual contributions of the potential sources. Nonetheless, beyond gaining an initial insight into the variability in simulation outcomes for a typical application of these tools, this intercomparison serves as a foundation for pinpointing the challenges inherent in conducting such tests and offers guidance on enhancing test designs for a more thorough model intercomparison analysis.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

We conjecture that the following aspects of the implementation workflow can cause major differences in the modelling results: i) initialization of release volume, ii) (non-)implementation of curvature correction terms, iii) ad-hoc treatments (stabilization methods), iv) stopping mechanism and/or stopping criterion, v) the implementation of the friction model, and (vi) the treatment of the longitudinal pressure gradient and the related earth-pressure coefficient.

A factor that we did not analyse is the spatial (and also temporal) resolution used for the computations and also how interpolation to the required spatial resolution might affect the simulation results. Testing for convergence, i.e. ensuring that results converge to a limit value as the time step and grid spacing tend to zero, is part of model development. So we assume that different spatial resolutions between the participating tools do not significantly influence the simulation results. Similarly, interpolating or exporting the result variables to the required 5 m grids for the result file submission is assumed to have a small effect.

Initial volume

As the participating simulation tools require different data formats for initializing the release volume, the release areas were provided in two different formats: i) as shapefiles with constant thickness and ii) as ASCII raster files with constant values of release thickness. As the release area is a closed polygon of arbitrary shape, it is not strictly aligned with the grid cells. When creating the ASCII raster files, all cells intersected by the polygon outline were attributed to the release area. The release volume from the ASCII raster file is therefore 6.6% larger than the one from the polygon in the idealized test cases and about 4% larger for the real topography test. Hence, depending on the file type used, differences in initial volume can be introduced. However, this also represents a common use case, i.e. initialization from polygons to rasters, and hence this step could be seen as part of the implementation workflow.

In Fig. 14, the initial volumes reported by all participating groups are categorized into three groups based on the file type used to initialize the simulation: a) shapefile (light blue), b) ASCII raster file (dark blue) and c) release area is represented differently, i.e. ellipses (*TITAN2D*) or hydrograph (*FLO-2D*) (red). Additionally, a hatched pattern indicates that the provided thickness, measured normal to the topography, has been projected to a vertical depth for the subsequent computations. In this figure, test cases *VoellmyIdealized* and *CoulombIdealized* are summarized into the class *Idealized topography* as both test cases are based on the same topography, release area and thickness (Fig. 14a).

Consistent with the above mentioned estimate in release volume based on the shapefile or ASCII raster file, relative differences are larger for the idealized topography compared to the real topography (Fig. 14b). Besides the variations caused by initializing the simulations using either the provided shapefile or ASCII raster file, further differences can be directly attributed to participating simulation tools employing i) different functions for geodata processing and ii) different numerical schemes. For example, discretizing the release area into either numerical particles or grids, will have an impact on defining the initial volume. This is, for instance, the case for MPM, where the initial volume was calculated as the material bounded between top and bottom material points. As the release area is concave, the theoretical volume obtained by multiplying each projected cell area by the prescribed constant flow depth is larger than the effective 3D volume of the release generated in the simulation. Also, where the release area is not aligned with the computational grid, cells that are only partially covered might be treated

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

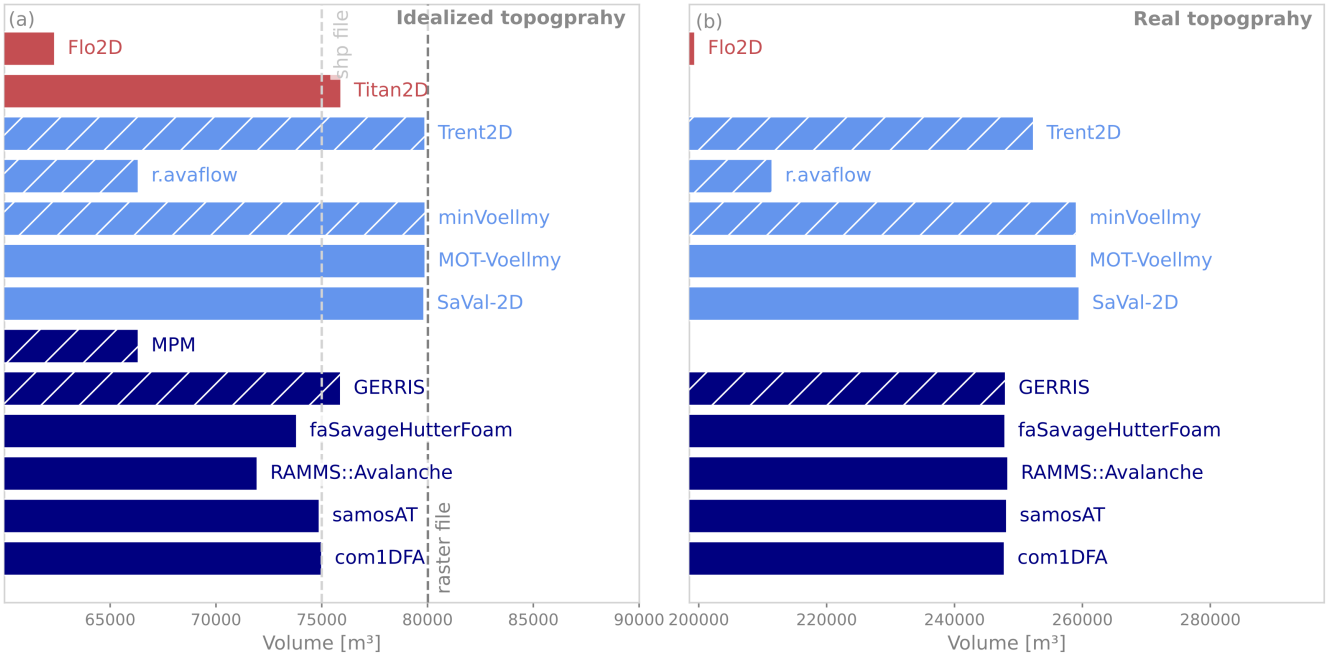

Figure 14. Distribution of initial volume as reported from the individual groups. The participating simulation tools are categorized into three groups, based on the file type used for initialization: shapefile (dark blue), ASCII raster file (light blue) and other initialization methods (red). The light grey line indicates the initial volume derived from shapefile, while the dark grey line the value based on the ASCII raster file, only available for the idealized topography. The hatched pattern indicates that slope normal thickness was projected into vertical depth. Values for the idealized topography are shown in panel (a), and for the real topography in panel (b). The x-axis stretches from -20% to + 20% of the median value in both panels.

differently by different approaches. Furthermore, only the spatial resolution of the input and result fields was prescribed, but not the spatial resolution used for the computation itself.

However, when analysing the differences in final simulation results, it is interesting to look at the total differences in initial volume, irrespective of which file type was used for initialization. If initial volumes are compared to the median of all participating simulation tools, there is a considerable spread in initial volumes among the simulations. For both scenarios, the standard deviation of the initial volume is roughly 8%, but maximum differences reach up to 20% of the median value, where the median values of initial volume are 75 007 m^3 for the *idealized* test cases and 248 181 m^3 for the *VoellmyReal* test case including all values (based on shapefile, ASCII raster file and other methods).

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Curvature treatment

If approaches based on the Saint-Venant model for shallow water or Savage–Hutter type models are applied to complex topography, extra terms have to be taken into account. If these terms are formulated in Cartesian coordinates and integrated along the vertical direction, non-hydrostatic terms in the depth-averaged equations arise (e.g. Iverson and Denlinger, 2001). On the other hand, if using a terrain-following local coordinate system, the effect of curvature changes the effective slope-normal acceleration (e.g. Fischer et al., 2012). This affects the bed shear stress, the pressure gradient and the body force (Peruzzetto et al., 2021).

Participating simulation tools differ in the way curvature has been accounted for in the mathematical formulations as well as how terrain curvature is actually computed. Basic information on the respective treatments is provided in Table 3. Among the three test cases, the effect of curvature in the model formulation is most significant for the *CoulombIdealized* test case because of the high velocity attained without a velocity-dependent friction term. This becomes evident in the case of *SaVal-2D*, which does not account for curvature and thus simulates a longer runout than the other models. In the *VoellmyReal* test case as well, *SaVal-2D* shows less confluence into the ravine, and more lateral spreading also across the more confined parts of the topography.

In the discussion of the results, we distinguish between simulations with and without curvature corrections, but we do not consider differences in how the models implement them. Doing so would require more detailed information on the different implementations as well as more specific tests.

Ad-hoc treatments

Thickness/depth-integrated shallow-flow models make a number of assumptions to simplify the equations and neglect several higher-order terms. However, some of these simplifications can cause problems in simulations and often ad-hoc treatments are employed to increase the plausibility of the simulation results and/or to ensure numerical stability.

As an example, lateral shear stress is typically neglected for the Voellmy friction law, but is mimicked in *com1DFA* by an artificial viscosity term (see Table 3) (Tonnel et al., 2023). Furthermore, the spatial and temporal discretization inherent in the numerical schemes introduce discretization errors of different magnitude. These are often remedied by ad-hoc treatments, such as reprojection of velocity onto the tangent plane in models that operate in a local, slope-parallel coordinate system. On a potentially larger spatial scale, certain processes might not be considered in the formulation. For instance, convex terrain can cause real avalanches to temporarily lose contact with the surface whereas thickness/depth-integrated models force the flow to stay on the surface. Models that neglect terrain curvature ignore the concomitant effect on friction entirely while others approximate it by reducing the slope-normal gravitational component to zero. This approximation, however, only applies to strongly convex topography.

Such ad-hoc treatments are often only poorly described in the model documentation, and hence it is difficult to discern their impact on simulation results. We consider the choice of these treatments to be a part of the implementation procedure, but we

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

provide information on a few key issues in Tables 3 and 4, to highlight differences between the participating tools that may help to explain the differences in simulation results.

Stopping: yield criterion and end-of-simulation criterion

In most thickness-integrated shallow-flow models, the constitutive relations or friction laws are defined for flowing material only (see e.g. Mangeney-Castelnau et al., 2003). To capture the onset of rest, certain models introduce what is termed a yield criterion: the shear stress transmitted at the base, calculated from gravitational traction and the longitudinal pressure gradient, is required to exceed the Voellmy friction plus, in some models, cohesion; otherwise, the flow is assumed to stop. However, this use of the term "yield criterion" deviates from the definition used in classical plasticity theory. These stopping mechanisms are only partly implemented, with *SaVal-2D*, *MoT-Voellmy*, *TRENT2D**, and *minVoellmyv2* including yield-like checks to various degrees (see Table 3). Alternative formulations incorporate deposition directly into the mass conservation equations, whereby some mass of the depositing flow layer is transferred to the bed (Vicari and Issler, 2024). However, except for *MPM* (*extended* group), none of the participating tools explicitly models deposition (see Table 4). *MPM* solves the full 3D equations of motion, with the mechanical behaviour and transitions from fluid- to solid-like during stopping are dictated by the constitutive model, in which a yield surface is embedded rather than prescribed as an external stopping condition.

In addition to physically simulating stopping, artificial end-of-simulation thresholds are also often used to terminate the simulation. This is especially the case if no yield criterion is implemented. These artificial thresholds are mostly based on an absolute value of momentum, velocity, or kinetic energy or a percentage of the respective quantity's maximum value. Furthermore, a maximum simulation time or number of time steps is often employed. Sect. 3 provides information on how the participating models implement stopping and termination of the simulation.

Differences in the implementation of a yield criterion and/or the use of different stopping thresholds will impact final peak flow variable fields and the extent of the inundation area. However, it is challenging to prescribe a standardized treatment in a test setup due to the diverse range of criteria and implementations. Since most practical applications will be impacted by this decision, it is interesting to examine the variations that result from it.

(Coulomb) friction model

The Voellmy friction relation is widely used for the simulation of snow avalanche flow in thickness/depth-integrated shallow flow approaches. The velocity-dependent friction term limits the maximum velocity, and thereby can help to stabilize simulations. For the idealized topography, the slope is steep and long enough for the moving mass to (essentially) reach the terminal velocity in the case of the Voellmy friction law. With Coulomb friction only, much higher flow velocities are attained. Differences can accumulate and are not attenuated by reaching a terminal velocity, hence differences in the simulated runout lengths between models are larger.

In addition, a pure Coulomb rheology is not natively available in all participating simulation tools. In those cases, participants used a "very high" value of ξ (see details on the chosen values of ξ for the respective simulation tool in Appendix A) to perform simulations for the *CoulombIdealized* test case. In contrast, models using k instead of ξ in Eq. 1, allow using a pure Coulomb

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

friction model by setting k to zero. Another aspect to consider is that, if a model is not developed with a view to this limit case, the ad-hoc treatments and stabilization schemes might not be designed to handle conditions that arise when using a pure Coulomb friction law. As an example, *minVoellmyv2* does not contain any specific numerical scheme to avoid numerical diffusion, but relies on keeping the avalanche front sharp with the help of the dependence of effective friction on thickness, which is a specific property of the Voellmy friction law (Hergarten, 2024b).

Longitudinal pressure gradient and earth-pressure coefficient

Depth-integration of the 3D stress divergence term yields, in addition to the basal shear stress, a longitudinal pressure gradient. In practice, this term is often closed using geotechnical earth-pressure coefficients, by which the vertical or slope-normal stress is scaled. The hydrostatic assumption sets this coefficient to one, which was adopted in most simulation tools considered here, see Table 4 for detailed information. Alternative approaches, however, adapt the coefficient to active (dilatant) or passive (compressive) stress states, depending on the sign of the velocity gradient. Differences in both the assumed coefficient and the numerical discretization of the longitudinal pressure gradient term may therefore contribute to the discrepancies observed between the models.

5.3 Result file submission

In this pilot study, a resubmission of simulation results was possible under certain circumstances. This decision was a consequence of the test design, as values were not prescribed for all implementation-specific parameters (see Sect. 2.2.1).

For example, for the *core* group, prescribing whether curvature effects have to be accounted for and in which manner was not feasible because some of the models do not allow choosing to include or disregard curvature effects. In order to streamline this as much as possible, a resubmission applying a treatment for curvature effects was possible. Furthermore, participants could resubmit if a bug was found in the chosen settings or in the code.

For the sake of transparency and reproducibility, any changes in applied settings or employed code compared to the initial submission had to be clearly stated. This information is added in Appendix A for the respective simulation tool. Furthermore, both the initially submitted files and the resubmitted version are available in the GitHub repository.

A benefit from allowing for resubmission of files was that bugs were not just identified but also resolved and documented. Also a comparison to the updated version is therefore already available.

6 Conclusions, lessons learned and outlook

Model intercomparison is an important complement to model verification and validation. In contrast to verification tests, model intercomparison facilitates testing model performance for more complex problem setups, which are relevant for real-world scenarios. Also, by prescribing values for the rheological parameters, potential sources of differences are not concealed by calibrating for a certain outcome, as is done in most validation tests.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

The results of this pilot study show that simulations based on the same input data and prescribed values for rheological parameters generally show good agreement among most participating models with regard to total runout length. However, higher variability is observed in lateral spreading and, in general, differences become larger in the case of more complex topographies or when using simple friction relations that do not consider turbulent friction.

The study highlights the importance of providing detailed information on the entire simulation workflow that is followed, i.e., numerical implementation including any ad-hoc treatments, stopping criteria, curvature correction terms, minimum thickness/depth thresholds, as well as pre- and postprocessing steps such as initialization of input data or interpolation of result variables. This information is not always readily available in model documentations but proves to be crucial in order to understand differences in simulation results among different models. As a first step, in this pilot study we provide basic information on the various aspects summarized in the simulation workflow and indicate where the participating tools vary in their approaches and how this might impact the simulation results.

To draw rigorous conclusions about the origin of the differences, more stringent test designs are required. Based on the findings of this pilot study, this should involve tests prescribing whether or not curvature effects are taken into account, providing release areas where polygons are/are not aligned with raster cells, prescribing minimum thickness/depth thresholds for computing velocity from momentum. Furthermore, test topographies with abrupt changes in slope cause a very strong braking effect, which leads to a sudden deceleration of the flow, potential differences may either become masked in the runout zone, or they may be augmented due to differences in the numerical treatment of shocks. This highlights that tests need to be representative of a range of topographical features, also including more gently sloping runout areas. Strongly idealized topographies prove to be helpful in revealing problems of numerical schemes related to axial symmetry. Furthermore, analysis of the temporal evolution of result variables would provide further insight, for example on the development of internal surges, treatment of shocks and also on the effect of implemented stopping mechanisms or applied end-of-simulation criteria.

The significant variability in the simulation results, even among models based on the same basic equations and assumptions, highlights the critical importance of accounting for uncertainty in simulations. In general, uncertainty and/or errors introduced by applying a numerical method should be very small compared to the uncertainty related to input data or flow model parameters. However, as this study shows, the uncertainty coming from the wider category of implementation-specific parameters must also be considered. In this respect, ensembles of different models could present a way forward.

Finally, performing this pilot study gave insight into the specific requirements regarding input data and also shows that runout length is not a good single measure when comparing simulation results. Building on the experience gained in this collaborative effort, it could be interesting to further streamline input/output data formats and establish a basic workflow for comparing simulation results between different models and also to measurement data.

Code and data availability. The official online repository of the *ISeeSnow* project can be found at https://github.com/avaframe/ISeeSnow (last access: 1 December 2025) and https://zenodo.org/records/17668272 (last access: 1 December, Wirbel (2025)). There, a full description of the test cases including the input data files as well as the result files submitted by the participants are available. We performed the analysis

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

of the simulation results using the *in3Utils*, *ana4Stats* and *ana3AIMEC* modules of AvaFrame (https://github.com/avaframe/AvaFrame, last access: 1 December 2025, Oesterle et al. (2024)).

Appendix A: Additional information on simulation tools

In the following, participating simulation tools are listed alphabetically. If there is a publication describing the model equations and respective implementation, the relevant reference is provided. Besides a brief description, only information on modifications and/or specific settings that have been used to perform the simulations for this study are described.

A1 com1DFA

com1DFA is a computational module within the open-source framework *AvaFrame*. Snow avalanche flow is simulated using thickness-integrated mass and momentum equations and a shallow-flow assumption, discretized with a particle-grid approach. A detailed description and testing of the approach can be found in Tonnel et al. (2023), the online documentation at https://docs.avaframe.org/en/latest/ (last access: 1 December 2025) or the source code at https://github.com/OpenNHM/AvaFrame (last access: 1 December 2025) and Oesterle et al. (2025b). The simulations presented in this study were performed with release *1.8 beta 3* using the default configuration of *com1DFA*, except for the prescribed values of release thickness, the prescribed Voellmy friction relation and values for the friction parameters μ and ξ.

Main reference: Tonnel et al. (2023), Oesterle et al. (2025a)

A2 FLO-2D

Before describing the parameters used to simulate the proposed case study, it is important to note that *FLO-2D* does not utilize the rheology of Voellmy and Salm. In a previous study, the rheology employed by *FLO-2D* for simulating mud/debris flows was adjusted for snow avalanches and a back analysis of the runout and the deposition shape of observed events carried out (Martini et al., 2023).

Instead of the friction parameters μ and ξ, *FLO-2D* considers the yield stress (τ), a turbulent term involving the Manning's coefficient (n) and the square of the velocity, and a viscous term accounting for dynamic viscosity. Presently, there is no calibration of the frictional parameter considered by *FLO-2D* to establish a direct relationship with the Voellmy–Salm friction law. Consequently, the choice of the frictional parameters in this intercomparison relies solely on limited experience.

Furthermore, in *FLO-2D* it is not possible to isolate the Coulomb friction. With the Manning's coefficient set to 0, the model failed to execute.

For both analysed case studies, a snow density of $300\,\mathrm{kg\,m^{-3}}$, a volumetric concentration (C_v) of 0.55, and a laminar flow coefficient $K = 2000$ were used (the viscous frictional term was turned off, however). In the real topography case study, a Manning's coefficient of $0.4\,\mathrm{s\,m^{-(1/3)}}$ and a yield stress of $1000\,\mathrm{Pa}$ were used. In the idealized topography case study, the corresponding values were $0.2\,\mathrm{s\,m^{-(1/3)}}$ and $540\,\mathrm{Pa}$, respectively.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

The methodology applied for determining the rheological parameters is described in Martini et al. (2023) and Barbolini and Savi (2014).

Main reference: Martini et al. (2023); O'Brien and Julien (1985)

A3 faSavageHutterFoam

faSavageHutterFoam is an implementation of the Savage–Hutter model within the *OpenFOAM* framework, making use of its finite area method. It is based on a depth-integrated formulation but retaining a three-dimensional velocity vector to account for topographic effects such as curvature.

The finite area method is a surface-based variation of the finite volume method that enables the solution of this mathematical system. Numerical schemes follow the standard OpenFOAM approach, supporting first- or second-order time-implicit integration. To avoid dry areas, a minimum flow thickness is enforced across the entire simulation domain. Apart from this correction, the model is mass-conserving. Process models, such as friction, are fully customizable by the user. The active and passive earth-pressure terms of the original Savage–Hutter model are not applied.

The simulations were conducted on hexagon-dominated meshes, generated by *pMesh*, with an approximate cell size of $5\,\mathrm{m}$. The minimum flow thickness was limited to $10^{-5}\,\mathrm{m}$ to avoid dry areas. No numerical viscosity was used in the simulations with Voellmy friction model, which resulted in velocities of up to $0.07\,\mathrm{m\,s^{-1}}$ in steep regions, slightly above the cap of $0.01\,\mathrm{m\,s^{-1}}$. For stability reasons, the cases employing Coulomb friction were run with a numerical viscosity of the form $\frac{g}{\chi}\rho\frac{\bar{u}^2}{h^2}$ where $\chi = 10^7\,m^{-1}\,s^{-1}$. This dynamic-viscosity formulation suppresses motion in dry areas while remaining small in the regions of primary interest. Simulations were conducted with both first- and second-order methods to validate the choice of mesh size and other numerical parameters. The second-order results were submitted for this comparison.

The computations were conducted on an AMD Ryzen 7 PRO 5850U CPU. The results required for the comparison were calculated using function objects that were loaded at runtime. The respective fields of peak flow thickness and velocity were mapped to a Cartesian grid and written to an ASCII file with another function object. To obtain a single raster file without post-processing, the solver had to be run with a single process (serial mode).

Main reference: Rauter and Tukovic (2018); Rauter et al. (2018); Rauter and Kowalski (2024)

A4 Gerris

The model proposed by Hergarten and Robl (2015) is based on the shallow water part of the *Gerris* fluid dynamics framework (Popinet, 2009). The approach is, in principle, not tied to *Gerris*, but could be integrated in any solver of the shallow water equations that allows for defining custom friction terms. Owing to the reduction to the shallow water equations, the concept neither uses slope-parallel velocity components nor 3D velocity vectors, but only the horizontal velocity components. While the resulting deviations arising from the inclination of the surface are captured by specific friction terms, effects of curvature are neglected. Due to the consideration of horizontal velocities, this results in an error in the balance of momentum. Velocities are overestimated at convex transitions and vice versa. While it was shown that the resulting error is not crucial for Voellmy friction (Hergarten and Robl, 2015), larger errors are to be expected for Coulomb friction. For this study, the *Gerris* framework

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

was used with the *2D libgfs* version 1.3.2. (131206-155120). For technical reasons, the simulations were performed on a regular grid with 4 m spacing and interpolated from and to the original 5 m grid.

Main reference: Hergarten and Robl (2015)

A5 minVoellmyv2

minVoellmyv2 is a lightweight implementation in Cartesian coordinates. It was originally developed for a modified version of Voellmy's rheology (Hergarten, 2024a). It uses Cartesian coordinates, with the thickness of the flow measured vertically and vertically averaged velocities. This treatment requires a modified expression for the fluid pressure (Hergarten, 2024b). In order to keep the implementation simple, an upstream scheme is used for the advection terms without any measures against numerical diffusion. The applicability of the simple upstream scheme relies on an increase of friction with decreasing flow thickness so that it is not well-suited for situations in which the Coulomb friction term dominates. Furthermore, the upstream scheme introduces an artificial longitudinal diffusion of momentum. *minVoellmyv2* is freely available under the GNU General Public License and consists of MATLAB and Python classes. Data handling is left to the user.

Main reference: Hergarten (2024a), Hergarten (2024b)

A6 MPM

The *MPM* is a hybrid Lagrangian–Eulerian numerical technique in which the Lagrangian particles track history-dependent variables such as position, velocity and deformation gradient, and the Eulerian grid enables computation of the spatial gradients of these quantities. The transfer of information between the grid and the particles is then handled by the so-called APIC-RPIC interpolation scheme. The time is discretized using a symplectic Euler time integrator. Details of the applied algorithm are available in Gaume et al. (2018). The material model used in this study is a non-cohesive non-associative (zero dilatancy) Drucker–Prager model, which is described in more details in Blatny et al. (2023); Cicoira et al. (2022). The depth-resolved simulation outputs are exported in a depth-averaged form following Vicari et al. (2025).

Main reference: Gaume et al. (2018)

A7 MoT-Voellmy

This ISO-C code, developed since 2011 and made open source in 2025, uses the same quasi-Cellular Automaton approach (with continuous state variables rather than discrete ones) as the C++ code DIMAS by Adamska-Szatko (2012). In its essence, it is a first-order upwind scheme that also includes fluxes to diagonal neighbour cells.

The terrain-following mesh is a square grid when projected vertically onto a horizontal plane. Pure Coulomb behaviour is obtained by setting the dimensionless drag coefficient k to zero. The user can choose whether or not to include curvature effects on Coulomb friction. In the present context, the grid spacing was set to 5 m and curvature effects were switched on.

Several stopping criteria are imposed: (i) Cells are considered stopped if the thickness drops below $h_{\min}$ (set to 0.01 m here). (ii) The flow direction in a cell (i, j) cannot change by more than 90° within a single time step. (iii) A cell at rest can begin

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

moving only if the combined downslope weight and pressure gradient exceed the maximum mobilizable Coulomb friction. (iv) The simulation ends if the simulated time exceeds the user-set maximum $t_{\max}$ (set to 75 s for the case *CoulombIdealized*, 90 s for *VoellmyIdealized* and 150 s for *VoellmyReal*) or if the mass-weighted average speed drops below a threshold velocity set to $0.18\,\mathrm{m\,s^{-1}}$ on the idealized topography and $0.32\,\mathrm{m\,s^{-1}}$ on the real topography. (v) The simulation is also stopped if the variable time step (determined by the CFL condition) drops below a user-set limit—here chosen as 0.001 s.

The non-physical oscillations in the peak fields, e.g. in Fig. 2 and Fig. 7, presumably result from chequerboard oscillations along the coordinate directions with a wavelength equal to twice the grid spacing, where very high speeds at tiny flow depths alternate with excessive flow depths at near-zero speed. The exact cause of this instability remains to be elucidated; it is most pronounced in highly regular topographies and symmetric configurations.

An updated version of the simulation results was used; updates to the initially submitted files were required due to a programming error in the curvature-induced friction term. At the same time, the stopping criteria (minimum average speed and maximum simulated time) were set to values used in real-world applications ($t_{\max} = 150\,\mathrm{s}$ for *VoellmyReal*, 90 s for *VoellmyIdealized*, and 75 s for *CoulombIdealized*) to avoid non-physical creeping.

Main reference: Issler (2025)

A8 RAMMS::Avalanche

RAMMS::Avalanche (Rapid Mass Movement Simulation) is a depth-averaged numerical model for simulating dense-flow avalanches, widely used in snow engineering. Developed by the WSL Institute for Snow and Avalanche Research SLF, it solves governing equations with second-order numerical schemes. The model supports multiple release zones in three-dimensional terrain, and here, the Voellmy–Salm (VS) approach is used as rheology. The model uses μ (Coulomb friction), ξ (turbulent drag) and the average release height as input parameters. For the Coulomb-only use case, the friction term $1/\xi$ is set to 0 in the source code. *RAMMS::Avalanche* models additionally curvature effects by splitting the velocity-dependent friction into a terrain-dependent term that amplifies Coulomb friction along curved paths and a terrain-independent Voellmy-like drag (Fischer et al., 2012).

An updated version of the simulation results was used. Updates to the initially submitted files were: description of rheologies was corrected and curvature effects taken into account.

Main references: Christen et al. (2010b, a)

A9 r.avaflow

r.avaflow (Mergili and Pudasaini, 2023) is an open-source tool that offers GIS integration, for the simulation of gravitational mass flows, first introduced by Mergili et al. (2017) and since then continuously improved and equipped with additional functionalities. For extremely rapid flows, the current version *r.avaflow* (Mergili et al., 2025), offers a Voellmy-based mixture model (with basal friction angle = internal friction angle, turbulent friction) and a multiphase model, based on Pudasaini and Mergili (2019). For the Coulomb-only model, the latter is applied with one solid phase, for which a simplified Coulomb rheology is assumed (with basal friction angle = internal friction angle). For both models, the differential mass and momentum

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

balance equations are solved on a regular Cartesian grid with a shock-capturing TVD-NOC numerical scheme (Tai et al., 2002; Wang et al., 2004). The length of each numerical time step is dynamically determined to fulfil the CFL criterion. *r.avaflow* is available in two releases, the more powerful of which is coupled to the GRASS GIS software. It is therefore provided under the GNU GPL License and can be downloaded from Mergili and Pudasaini (2023). A developer version is used for the present work, which is available as a digital supplement along with the start script for the simulations performed.

Main reference: Mergili and Pudasaini (2023); Mergili et al. (2025)

A10 samosAT

samos-AT is a simulation tool developed by the Austrian government and AVL List GmbH. While allowing for powder snow and dense flow avalanches, the code for the latter type served as the basis for the *AvaFrame::com1DFA* module. So it uses the same thickness-integrated governing equations and the smoothed particle hydrodynamics (SPH) method for numerical calculations.

Main reference: Sampl (2007)

A11 SaVal-2D

The *SaVal-2D* code solves the standard thickness-integrated mass and momentum conservation equations using an explicit Godunov-type finite-volume scheme. The model is formulated in a topography-following reference frame. Spatial discretization is based on a structured square mesh. The DEM and release area thickness should be provided as input asc files. An input txt file specifying the model parameters (friction parameters, time step, threshold thickness, etc.) is also required.

Calculations for the *ISeeSnow* benchmark were performed with code version `v5_sc_max`. This version implements an enriched Voellmy law with a cohesion term (set to 0 in this case). Topography curvature correction terms are not accounted for. For the Coulomb test case, the ξ parameter has been set to a value of $10^9\ \mathrm{m\,s^{-2}}$. The threshold thickness was set to $0.001\ \mathrm{m}$, meaning that thickness values smaller than this threshold in the outputs should be considered as null.

Although the model implements a physically-based yield criterion, an additional stopping criterion had to be used due to numerical diffusion. We used here a criterion based on total flow momentum: flow was considered to be arrested when the total flow momentum became less than 4% of the maximum momentum reached during the simulation. Further decrease of momentum below this threshold seems to be mainly controlled by numerical diffusion, both for the *VoellmyReal* and *VoellmyIdealized* cases.

The code is parallelized using OpenMP. Computations were run on 4 cores with Intel Core i7@2.5GHz processors. The indicated CPU time corresponds to the total time used by the process.

The *SaVal-2D* model is currently being integrated in a web-based platform that will be released during the year 2025.

Main reference: Naaim et al. (2004)

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

A12 TITAN2D

TITAN2D is a computer program developed by the Geophysical Mass Flow Group at the State University of New York at Buffalo, for the purpose of simulating granular avalanches over digital elevation models of natural terrain. It is designed to simulate various types of geological mass flows, including volcanic flows, debris avalanches and landslides, as well as snow avalanches. *TITAN2D* is based on a thickness-averaged shallow-water model for an incompressible continuum. The conservation equations of mass and momentum are solved using a finite-volume scheme. In this study, the latest version (v4.2.0) was used.

TITAN2D automatically generates a peak flow thickness (PFT) file (pileheightrecord), however, it is not feasible to directly obtain a peak flow velocity (PFV) file. The calculation of PFV was achieved by extracting the flow height (h), the x-direction momentum (hV_x) and the y-direction momentum (hV_y) for each mesh from the output data file at arbitrary time intervals (1 s in this study). Flow velocity was calculated as $\sqrt{(hV_x/h)^2 + (hV_y/h)^2}$, then the results were superimposed, and the maximum velocity was extracted. It should be noted that, in *TITAN2D*, there is an option to refine the mesh only around the flow instead of the entire DEM (Adaptive mesh refinement) in order to reduce computational cost. However, as PFV calculations require output data from the entire mesh, this option was deactivated for the computations in this study, resulting in a longer computation time.

Before commencing calculations, *TITAN2D* modifies the mesh size slightly from the original. In this study, a mesh with 5 m intervals was employed, producing a new mesh with intervals ranging from 4.7149 to 5.1297 m. Consequently, the PFT and PFV data were created by interpolating to 5 m intervals.

The shape of the avalanche release area can only be set in elliptic cylinders in *TITAN2D*. In this study, the initial thickness was also varied to achieve the same specified initial volume. In the cases of *VoellmyIdealized* and *CoulombIdealized*, an elliptic cylinder with a major radius of 150 m, a minor radius of 70 m, and a height of 1.9 m was positioned at the centre coordinates (1195, -4250). In the *VoellmyReal* case, elliptic cylinders were set: one with major and minor radii of 170 and 110 m, respectively, at centre coordinates of (169230, 362520); and the other with major and minor radii of 160 and 85 m, respectively, at centre coordinates of (168990, 362580). The height of both cylinders was 2.0843 m.

In the Coulomb friction model of *TITAN2D*, internal friction is required in addition to basal friction. Here results for the internal friction angle of 30° are shown.

The computation duration was measured in the non-multiprocessing mode.

An updated version of the simulation results was used; updates to the initially submitted files were: in the resubmission *TITAN2D* $v4.2.0$ was used (initial submission was performed with $v4.0.0$), no adaptive mesh refinement, only consider results where flow thickness is greater than $1\ \mathrm{mm}$

Main reference: Patra et al. (2005)

A13 TRENT2D*

*TRENT2D** (Zugliani and Rosatti, 2021) describes the motion of dense snow avalanches by means of the depth-averaged shallow flow equations over fixed bed written with a global reference system (i.e., with the x-y axes describing a horizontal

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

plane and the z axis vertical). Thanks to the global reference system used, curvature effects are automatically taken into account to a first-order approximation. The friction law employed in the model is the Voellmy law and, for this reason, a pure Coulombian friction law can be achieved by using very high ξ values. The model presents an uppermost limit for the ξ parameter equal to $10^8\,\mathrm{m\,s^{-2}}$.

The equations are solved numerically by means of a finite volume scheme with Godunov fluxes evaluated with a specific Riemann solver (Zugliani and Rosatti, 2024) and the accuracy is second order in space and time. The spatial discretization is based on a structured square mesh. The model is written in Fortran and is parallelized using OpenMP.

The avalanche motion is stopped thanks to the yield criterion of the Voellmy law implemented in the model.

As initial conditions, the DEM and the release depth (measured vertically) must be provided in ASCII format.

The model is available for practitioners and local authorities inside the WEEZARD system (Rosatti et al., 2018) (a web-service, GIS environment, based on the cloud computational capabilities) at the following link: https://tool.weezard.eu/ (last access: 1 December 2025).

An updated version of the simulation results was used. Updates to the initially submitted files were: solely the results for the *CoulombIdealized* test case were resubmitted using $\xi = 10^8\,\mathrm{m\,s^{-2}}$ instead of a higher value for $\xi = 10^{10}\,\mathrm{m\,s^{-2}}$.

Main reference: Zugliani and Rosatti (2021, 2024); Rosatti et al. (2018)

Appendix B: Thalweg-following coordinate transformation

To make scalar measures like runout length objective and comparable between different simulation results, the result fields are transformed into a thalweg-following coordinate system. Here, thalweg refers to: "The thalweg is defined as the line representing the main flow direction of all potential avalanche events within a specific avalanche path..." from: https://docs.avaframe.org/en/latest/glossary.html#term-thalweg (last access: 1 December 2025). This is done using the *ana3AIMEC* module of *AvaFrame*. Figure B1 shows the transformation of the peak flow thickness field of *com1DFA* for the test cases *VoellmyReal* and *VoellmyIdealized*. In the case of the idealized topographies used in this study, the thalweg is parallel to the x-axis, which makes the transformation trivial.

Appendix C: Additional analysis figures

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

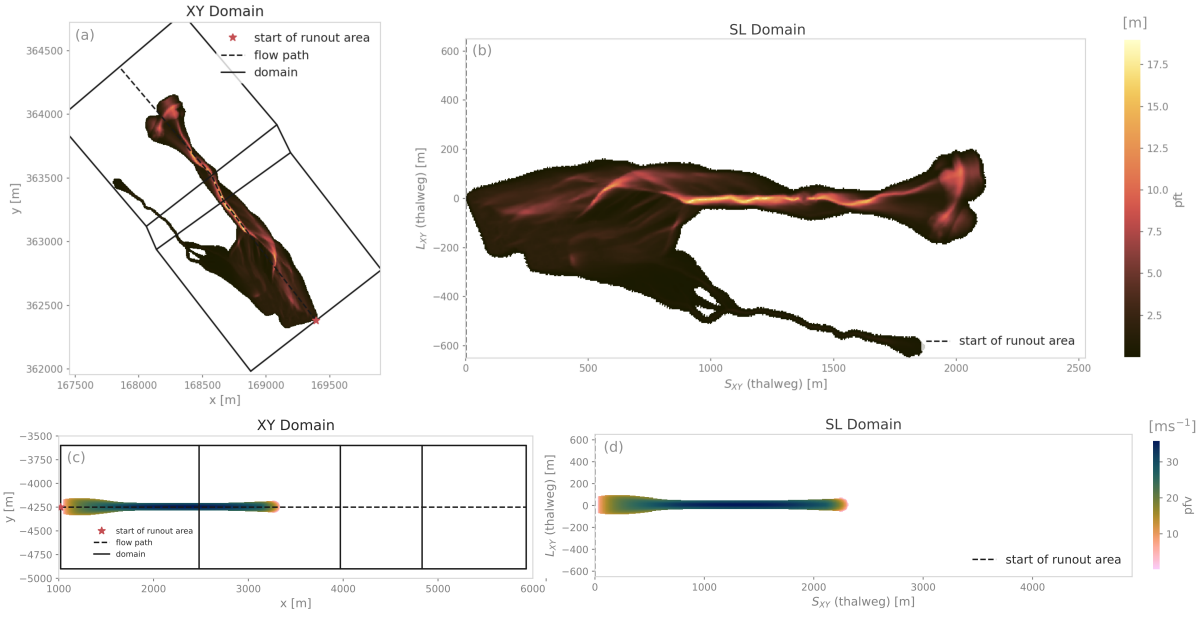

Figure B1. Peak field in Cartesian coordinate system (XY domain) in panels (a) and (c) showing the thalweg and domain boundaries for the coordinate transformation. Peak fields transformed into the thalweg-following coordinate system (SL domain) with the along thalweg coordinate S_{xy} and across thalweg coordinate L_{xy} in panels (b) and (d), for the real and idealized topography respectively.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

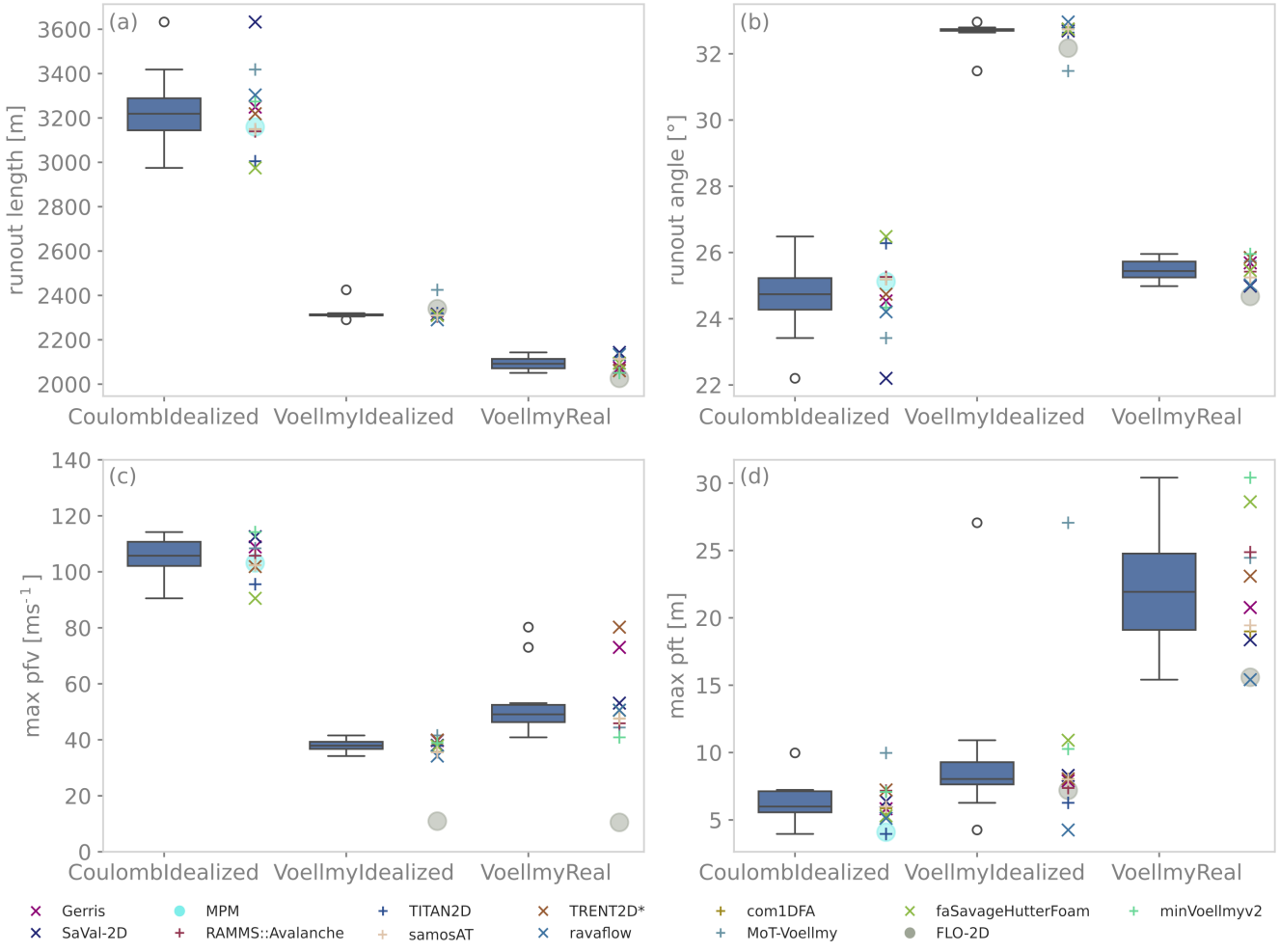

Figure C1. Distribution of scalar measures derived from the simulation results (for runout lengths and angles the analysis is performed along the thalweg and based on a peak flow velocity threshold of $1\,\mathrm{m\,s^{-1}}$). (a) Runout length, distance from start of thalweg to runout point measured along the thalweg. (b) Runout angle, measured from release to runout point. (c) Maximum value of peak flow velocity (the y axis is limited to $140\,\mathrm{m\,s^{-1}}$, outlier is *r.avalflow* ($322.6\,\mathrm{m\,s^{-1}}$). (d) Maximum value of peak flow thickness. The box plots show the quartiles of the dataset (only *core group*) and categorized outliers as circles, maximum extent of whiskers is based on quartiles minus (plus) 1.5 times the interquartile range. The interquartile range is added as text with a bracket for each box plot. Additionally, the individual data points are shown using coloured × or + markers for the *core group*, and filled ○ for the *extended group*. Only the models belonging to the *core group* are included in the box plot distributions.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

C1 Test case *VoellmyIdealized*

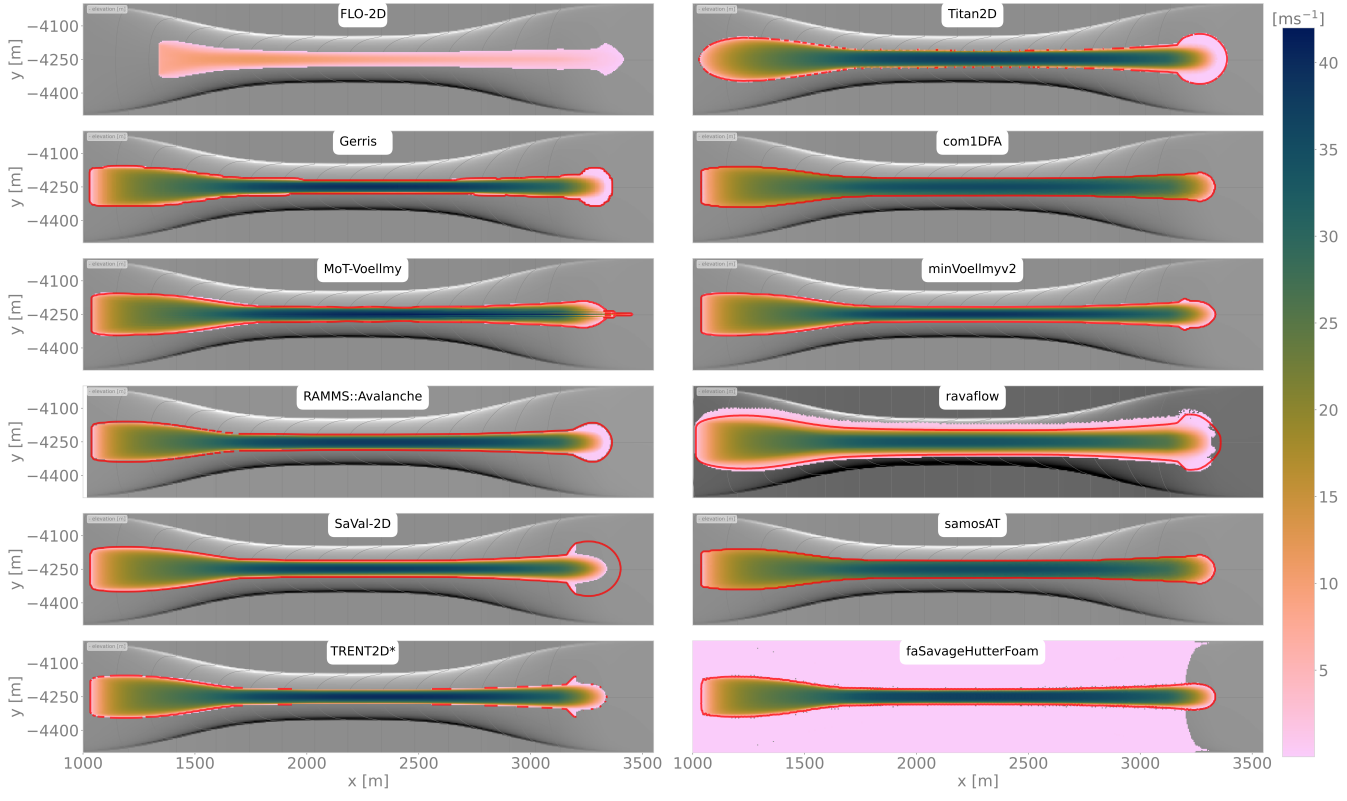

Figure C2. Peak flow velocity fields for test case *VoellmyIdealized*; simulation tool name is added within each panel. Peak flow velocity values are capped at 42 $\mathrm{m\,s^{-1}}$ and values below 0.01 $\mathrm{m\,s^{-1}}$ masked. Red lines indicate the 0.01 m contour lines derived from the corresponding peak flow thickness fields. Note that, in case of *FLO-2D*, the release area is defined by a hydrograph at a line, hence the upper part of the release area is not represented in the simulation results.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

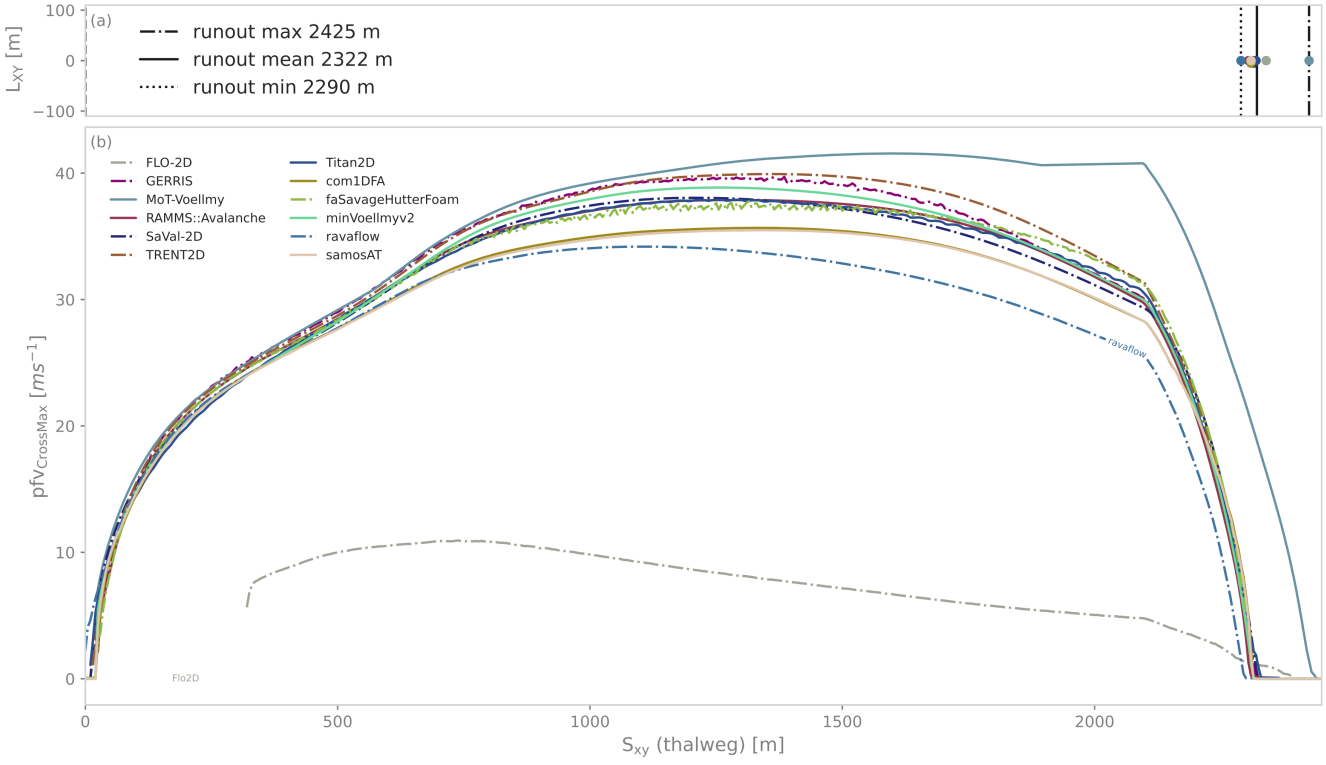

Figure C3. Runout points for test case *VoellmyIdealized* based on 1 $\mathrm{m\,s^{-1}}$ peak flow velocity in panel (a). Maximum across flow peak flow velocity ($pfv_{CrossMax}$) along thalweg for all simulation results in panel (b).

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

C2 Test case *VoellmyReal*

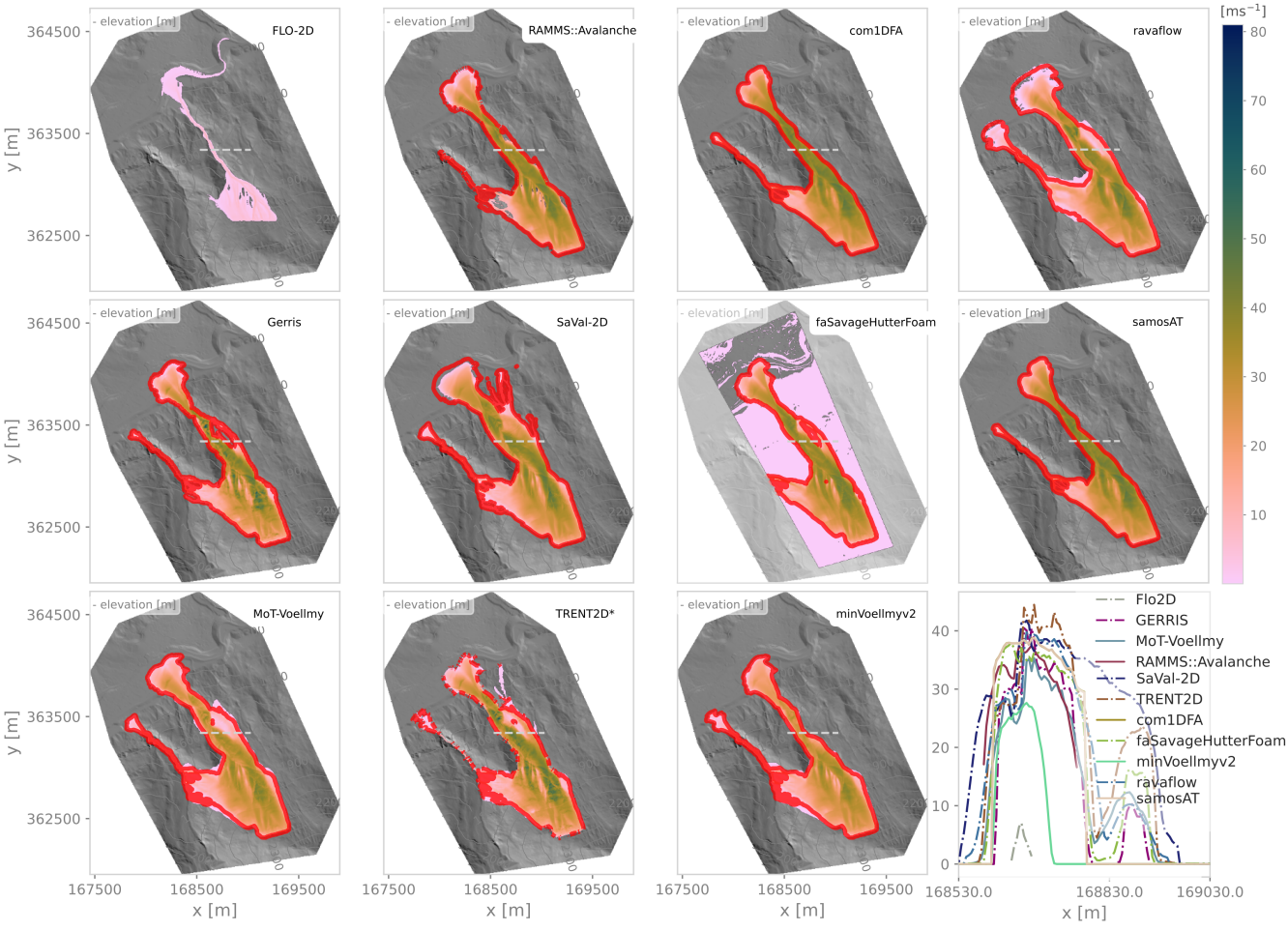

Figure C4. Peak flow velocity fields for test case *VoellmyReal*; simulation tool name is added within each panel. Peak flow velocity values are capped at 81 $\mathrm{m\,s^{-1}}$ and values below 0.01 $\mathrm{m\,s^{-1}}$ masked. Red lines indicate the 0.01 m contour lines derived from the corresponding peak flow thickness fields. For *faSavageHutterFoam* parts of the computational domain show no data values and are indicated by the transparent area. The panel in the lower right corner shows a cross profile of peak flow velocity values, the location of the cross profile is indicated in the other panels with a light gray dashed line. Note that, the in case of *FLO-2D*, the release area is defined by a hydrograph at a line, hence the upper part of the release area is not represented in the simulation results.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

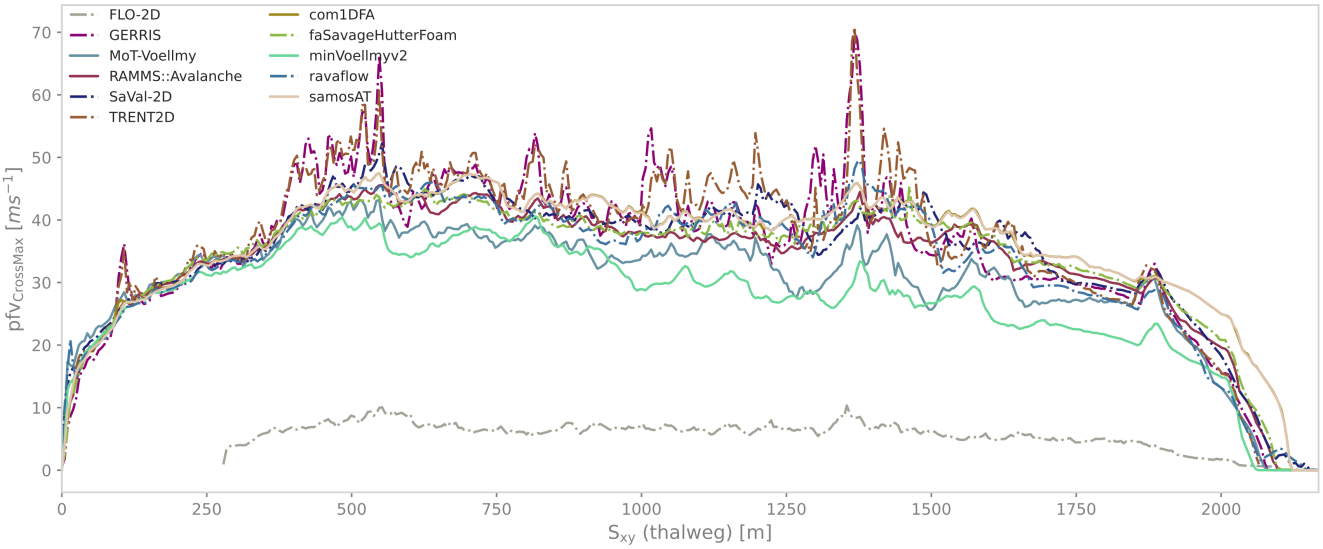

Figure C5. Test case *VoellmyReal*: Maximum across flow peak flow velocity ($\mathrm{pfv}_{\mathrm{CrossMax}}$) values along the thalweg for all simulation results.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

C3 Test case *CoulombIdealized*

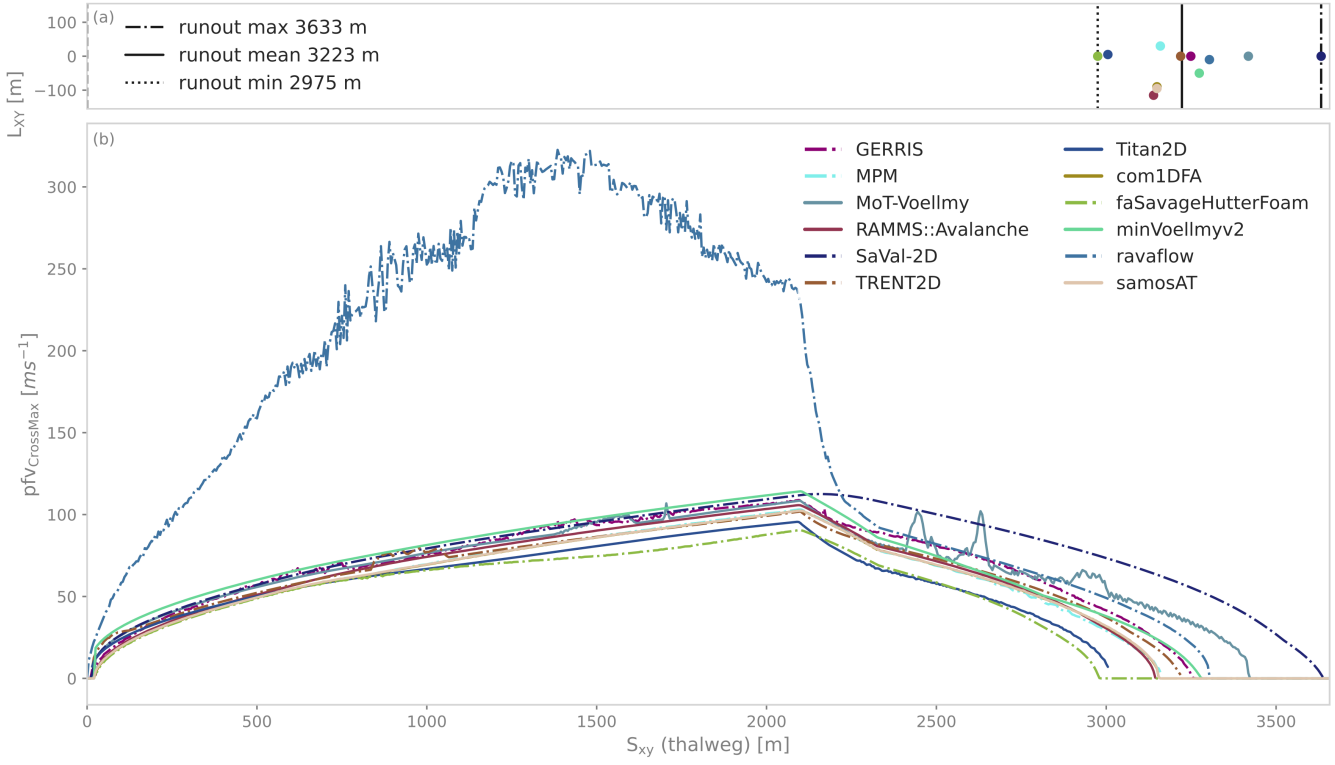

Figure C6. Test case *CoulombIdealized*. Runout points based on $1\,\mathrm{m\,s^{-1}}$ peak flow velocity in panel (a). Maximum across flow peak flow velocity ($pfv_{CrossMax}$) values along the thalweg for all simulation results in panel (b).

C4 Statistical analysis table

In Table C1, the values shown in Fig. 10 are listed.

Author contributions. AW, JT, FSO initiated and designed the pilot study. AW performed the main analysis of submitted simulation results and led the writing of this manuscript. All authors performed simulations and contributed information regarding the respective simulation tools, as well as issues encountered performing the simulations for the test cases and ideas on potential sources of differences and on how to improve the test design and also gave helpful feedback on the manuscript text. All authors helped to shape the final form of the manuscript.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Table C1. Scalar measures derived from simulation results as shown in Fig. 10; RL: runout length, RA: runout angle, MPFV: maximum peak flow velocity, MPFT: maximum peak flow thickness. Runout length and runout angle are derived using peak flow thickness (PFT) and a threshold of 0.5 m. The runout length is measured from the start of the thalweg.

	CoulombIdealized				VoellmyIdealized				VoellmyReal			
model name	RL [m]	RA [°]	MPFV [ms^{-1}]	MPFT [m]	RL [m]	RA [°]	MPFV [ms^{-1}]	MPFT [m]	RL [m]	RA [°]	MPFV [ms^{-1}]	MPFT [m]
com1DFA	3144.30	25.22	102.35	5.85	2305	32.75	35.67	8.037	2109.06	25.22	47.57	18.99
FLO-2D					2390	31.54	10.93	7.2	2099.33	23.83	10.49	15.6
faSavageHutterFoam	2969.91	26.52	90.52	5.29	2305	32.8	37.77	10.92	2089.60	25.42	50.55	28.61
Gerris	3243.95	24.52	108.89	5.82	2335	32.46	39.71	7.9	2075.01	25.6	73.0	20.76
minVoellmyv2	3129.35	25.33	114.18	7.08	2305	32.8	38.86	10.27	2050.69	25.95	40.87	30.40
MPM	3154.27	25.15	103.14	4.09								
MoT-Voellmy	3338.63	23.91	108.385	9.97	2420	31.53	41.55	27.06	2070.15	25.74	44.37	24.46
samosAT	3144.30	25.22	102.33	5.99	2305	32.8	35.48	8.0	2109.05	25.22	47.62	19.44
SaVal-2D	3632.60	22.16	112.541	6.37	2345	32.35	38.05	8.3	2123.65	25.14	53.09	18.37
RAMMS::Avalanche	3139.32	25.24	105.745	7.17	2325	32.57	37.87	7.37	2089.60	25.42	45.88	24.87
r.avaflow	3199.11	24.87	322.59	5.12	2290	32.95	34.18	4.25	2075.01	25.6	50.59	15.41
TITAN2D	2915	26.96	95.56	3.95	2335	32.47	37.9	6.27				
TRENT2D[*]	3209.08	24.77	101.865	7.22	2310	32.74	39.93	8.05	2060.41	25.85	80.23	23.09

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank Tómas Johanesson and Alexander Jarosch for helpful comments. *ISeeSnow* is supported by *AvaFrame*, the open Avalanche Framework (https://www.avaframe.org/, last access: 1 December 2025), a cooperation between the Austrian Research Centre for Forests (Bundesforschungszentrum für Wald; BFW) and the Austrian Avalanche and Torrent Service (Wildbach- und Lawinenverbauung; WLV). Additional support has been provided by the international cooperation project "AvaRange - Particle tracking in Snow Avalanches", supported by the the Austrian Science Fund (FWF, grant no. I 4274-N29) and the German Research Foundation (DFG, grant no. 421446512).

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

References

Adamska-Szatko, M.: Modelowanie matematyczne powierzchniowych zjawisk dynamicznych z zastosowaniem automatów koórkowych [Mathematical Modeling of Surface Dynamic Processes with the Use of Cellular Automata], PhD thesis, Akademia Górniczo-Hutnicza (AGH University of Science and Technology), Dept. of Geoinformatics and Applied Computer Science, Cracow, Poland, 2012.

Barbolini, M. and Savi, F.: Snow avalanche dynamics simulation and hazard mapping using FLO-2D, pp. 26–27, Interpraevent 2008, Dornbirn, Austria, 2014.

Barbolini, M., Gruber, U., Keylock, C., Naaim, M., and Savi, F.: Application of statistical and hydraulic-continuum dense-snow avalanche models to five real European sites, Cold Regions Science and Technology, 31, 133–149, https://doi.org/10.1016/S0165-232X(00)00008-2, 2000.

Bartelt, P., Bieler, C., Bühler, Y., Christen, M., Deubelbeiss, Y., Graf, C., McArdell, B. W., Salz, M., and Schneider, M.: Manual RAMMS - Rapid Mass Movements Simulation, 2017.

Blatny, L., Löwe, H., and Gaume, J.: Microstructural controls on the plastic consolidation of porous brittle solids, Acta Materialia, 250, 118 861, https://doi.org/10.1016/j.actamat.2023.118861, 2023.

Bueler, E., Lingle, C. S., Kallen-Brown, J. A., Covey, D. N., and Bowman, L. N.: Exact solutions and verification of numerical models for isothermal ice sheets, Journal of Glaciology, 51, 291–306, https://doi.org/10.3189/172756505781829449, 2005.

Bühler, Y., Bebi, P., Christen, M., Margreth, S., Stoffel, L., Stoffel, A., Marty, C., Schmucki, G., Caviezel, A., Kühne, R., Wohlwend, S., and Bartelt, P.: Automated avalanche hazard indication mapping on a statewide scale, Natural Hazards and Earth System Sciences, 22, 1825–1843, https://doi.org/10.5194/nhess-22-1825-2022, 2022.

Christen, M., Bartelt, P., and Kowalski, J.: Back calculation of the In den Arelen avalanche with RAMMS: interpretation of model results, Annals of Glaciology, 51, 161–168, https://doi.org/10.3189/172756410791386553, 2010a.

Christen, M., Kowalski, J., and Bartelt, P.: RAMMS: Numerical simulation of dense snow avalanches in three-dimensional terrain, Cold Regions Science and Technology, 63, 1–14, https://doi.org/10.1016/j.coldregions.2010.04.005, 2010b.

Cicoira, A., Blatny, L., Li, X., Trottet, B., and Gaume, J.: Towards a predictive multi-phase model for alpine mass movements and process cascades, Engineering Geology, 310, 106 866, https://doi.org/10.1016/j.enggeo.2022.106866, 2022.

Faccanoni, G. and Mangeney, A.: Exact solution for granular flows, International Journal for Numerical and Analytical Methods in Geomechanics, 37, 1408–1433, 2012.

Fischer, J.-T.: A novel approach to evaluate and compare computational snow avalanche simulation, Natural Hazards and Earth System Sciences, 13, 1655–1667, https://doi.org/10.5194/nhess-13-1655-2013, 2013.

Fischer, J.-T., Kowalski, J., and Pudasaini, S. P.: Topographic curvature effects in applied avalanche modeling, Cold Regions Science and Technology, 74-75, 21–30, https://doi.org/10.1016/j.coldregions.2012.01.005, 2012.

Gauer, P., Medina-Cetina, Z., Lied, K., and Kristensen, K.: Optimization and probabilistic calibration of avalanche block models, Cold Regions Science and Technology, 59, 251–258, https://doi.org/10.1016/j.coldregions.2009.02.002, international Snow Science Workshop (ISSW) 2008, 2009.

Gauer, P., Body, N. S., and Aalerud, A. H.: What avalanche observations tell us about the performance of numerical models, in: Proceedings of the International Snow Science Workshop, Oregon, 2023.

Gaume, J., Gast, T., Teran, J., van Herwijnen, A., and Jiang, C.: Dynamic anticrack propagation in snow, Nature Communications, 9, 3047, https://doi.org/10.1038/s41467-018-05181-w, 2018.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Gueugneau, V., Charbonnier, S., Esposti Ongaro, T., de'Michieli Vitturi, M., Peruzzetto, M., Mangeney, A., Bouchut, F., Patra, A., and Kelfoun, K.: Synthetic benchmarking of concentrated pyroclastic current models, Bulletin of Volcanology, 83, 75, 2021.

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

Hergarten, S.: Scaling between volume and runout of rock avalanches explained by a modified Voellmy rheology, Earth Surface Dynamics, 12, 219–229, https://doi.org/10.5194/esurf-12-219-2024, 2024a.

Hergarten, S.: MinVoellmy v1: a lightweight model for simulating rapid mass movements based on a modified Voellmy rheology, Geoscientific Model Development, 17, 781–794, https://doi.org/10.5194/gmd-17-781-2024, 2024b.

Hergarten, S. and Robl, J.: Modelling rapid mass movements using the shallow water equations in Cartesian coordinates, Natural Hazards and Earth System Sciences, 15, 671–685, https://doi.org/10.5194/nhess-15-671-2015, 2015.

Hutter, K., Siegel, M., Savage, S. B., and Nohguchi, Y.: Two-dimensional spreading of a granular avalanche down an inclined plane. Part I. Theory, Acta Mechanica, 100, 37–68, https://doi.org/10.1007/bf01176861, 1993.

Issler, D.: Basic equations and numerical methods in MoT-Voellmy, NGI Technical Note 20230100-06-TN, Norwegian Geotechnical Institute, Oslo, Norway, 2025.

Issler, D., Harbitz, C. B., Kristensen, K., Lied, K., Moe, A. S., Barbolini, M., De Blasio, F. V., Khazaradse, G., McElwaine, J. N., Mears, A. I., Naaim, M., and Sailer, R.: Comparison of avalanche models as applied to dry-snow avalanches observed in the full-scale test site Ryggfonn, Norway, in: Landslides and Avalanches: ICFL 2005 Norway. Proc. 11th Intl. Conference and Field Trip on Landslides, Norway, September 2005, edited by Senneset, K., Flaate, K., and Larsen, J. O., pp. 173–179, CRC Press, Taylor & Francis, London, U.K., ISBN 978-0415386784, 2005.

Issler, D., Gisnås, K. G., Gauer, P., Glimsdal, S., Domaas, U., and Sverdrup-Thygeson, K.: NAKSIN -– a new approach to snow avalanche hazard indication mapping in Norway, https://doi.org/10.2139/ssrn.4530311, preprint, 2023.

Iverson, R. M. and Denlinger, R. P.: Flow of variably fluidized granular masses across three-dimensional terrain: 1. Coulomb mixture theory, Journal of Geophysical Research: Solid Earth, 106, 537–552, 2001.

Mangeney-Castelnau, A., Vilotte, J.-P., Bristeau, M.-O., Perthame, B., Bouchut, F., Simeoni, C., and Yerneni, S.: Numerical modeling of avalanches based on Saint Venant equations using a kinetic scheme, Journal of Geophysical Research: Solid Earth, 108, 2003.

Martini, M., Baggio, T., and D'Agostino, V.: Comparison of two 2-D numerical models for snow avalanche simulation, Science of the Total Environment, 896, 165 221, https://doi.org/10.1016/j.scitotenv.2023.165221, 2023.

Mergili, M. and Pudasaini, S.: r.avalfow – The mass flow simulation tool., https://www.avaflow.org, 2023.

Mergili, M., Fischer, J.-T., Krenn, J., and Pudasaini, S. P.: r.avaflow v1, an advanced open-source computational framework for the propagation and interaction of two-phase mass flows, Geoscientific Model Development, 10, 553–569, https://doi.org/10.5194/gmd-10-553-2017, 2017.

Mergili, M., Pfeffer, H., Kellerer-Pirklbauer, A., Zangerl, C., and Pudasaini, S. P.: r.avaflow v4, a multi-purpose landslide simulation framework, EGUsphere, 2025, 1–31, https://doi.org/10.5194/egusphere-2025-213, 2025.

Naaim, M., Naaim-Bouvet, F., Faug, T., and Bouchet, A.: Dense snow avalanche modeling: flow, erosion, deposition and obstacle effects, Cold Regions Science and Technology, 39, 193–204, https://doi.org/10.1016/j.coldregions.2004.07.001, snow and Avalanches: Papers Presented at the European Geophysical Union Conference, Nice, April 2003. Dedicated to the Avalanche Dynamics Pioneer Dr. B. Salm., 2004.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Naaim, M., Durand, Y., Eckert, N., and Chambon, G.: Dense avalanche friction coefficients: influence of physical properties of snow, Journal of Glaciology, 59, 771–782, 2013.

Oesterle, F., Wirbel, A., and Fischer, J.-T.: avaframe/AvaFrame: 1.8.3, https://doi.org/10.5281/zenodo.11234874, 2024.

Oesterle, F., Wirbel, A., Fischer, J.-T., Huber, A., and Spannring, P.: OpenNHM/AvaFrame: Concept (all versions), https://doi.org/10.5281/zenodo.4721446, 2025a.

Oesterle, F., Wirbel, A., Fischer, J.-T., Huber, A., and Spannring, P.: OpenNHM/AvaFrame: 2.0b2, https://doi.org/10.5281/zenodo.17713591, 2025b.

O'Brien, J. S. and Julien, P. Y.: Physical properties and mechanics of hyperconcentrated sediment flows, Proc. ASCE HD Delineation of landslides, flash flood and debris flow Hazards, 1985.

Patra, A. K., Bauer, A. C., Nichita, C. C., Pitman, E. B., Sheridan, M. F., Bursik, M., Rupp, B., Webber, A., Stinton, A. J., Namikawa, L. M., and Renschler, C.: Parallel adaptive numerical simulation of dry avalanches over natural terrain, Journal of Volcanology and Geothermal Research, 139, 1–21, 2005.

Pattyn, F., Perichon, L., Aschwanden, A., Breuer, B., de Smedt, B., Gagliardini, O., Gudmundsson, G. H., Hindmarsh, R. C. A., Hubbard, A., Johnson, J. V., Kleiner, T., Konovalov, Y., Martin, C., Payne, A. J., Pollard, D., Price, S., Rückamp, M., Saito, F., Souček, O., Sugiyama, S., and Zwinger, T.: Benchmark experiments for higher-order and full-Stokes ice sheet models (ISMIP–HOM), The Cryosphere, 2, 95–108, https://doi.org/10.5194/tc-2-95-2008, 2008.

Peruzzetto, M., Mangeney, A., Bouchut, F., Grandjean, G., Levy, C., Thiery, Y., and Lucas, A.: Topography Curvature Effects in Thin-Layer Models for Gravity-Driven Flows Without Bed Erosion, Journal of Geophysical Research: Earth Surface, 126, e2020JF005 657, https://doi.org/10.1029/2020JF005657, 2021.

Popinet, S.: An accurate adaptive solver for surface-tension-driven interfacial flows, Journal of Computational Physics, 228, 5838–5866, https://doi.org/10.1016/j.jcp.2009.04.042, 2009.

Pudasaini, S. P. and Hutter, K.: Avalanche dynamics: dynamics of rapid flows of dense granular avalanches, Springer Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-32687-8, 2007.

Pudasaini, S. P. and Mergili, M.: A multi-phase mass flow model, Journal of Geophysical Research: Earth Surface, 124, 2920–2942, 2019.

Rauter, M. and Kowalski, J.: OpenFOAM-avalanche 2312: depth-integrated models beyond dense-flow avalanches, Geoscientific Model Development, 17, 6545–6569, https://doi.org/10.5194/gmd-17-6545-2024, 2024.

Rauter, M. and Tukovic, Z.: A finite area scheme for shallow granular flows on three-dimensional surfaces, Computers & Fluids, 166, 184–199, https://doi.org/10.1016/j.compfluid.2018.02.017, 2018.

Rauter, M., Kofler, A., Huber, A., and Fellin, W.: faSavageHutterFOAM 1.0: depth-integrated simulation of dense snow avalanches on natural terrain with OpenFOAM, Geoscientific Model Development, 11, 2923–2939, https://doi.org/10.5194/gmd-11-2923-2018, 2018.

Roache, P. J.: Quantification of uncertainty in computational fluid dynamics, Annual review of fluid Mechanics, 29, 123–160, 1997.

Rosatti, G., Zorzi, N., Zugliani, D., Piffer, S., and Rizzi, A.: A Web Service ecosystem for high-quality, cost-effective debris-flow hazard assessment, Environmental modelling & software, 100, 33–47, 2018.

Sampl, P.: SamosAT - Modelltheorie und Numerik, AVL List GMBH, https://doi.org/10.5281/zenodo.8007681, 2007.

Schmidtner, K., Bartelt, P., Fischer, J.-T., Sailer, R., Granig, M., Sampl, P., Fellin, W., Stoffel, L., Christen, M., and Bühler, Y.: Comparison of powder snow avalanche simulation sodels (RAMMS and SamosAT) based on reference events in Switzerland, in: International Snow Science Workshop Proceedings 2018, Innsbruck, Austria, pp. 740–745, 2018.

https://doi.org/10.5194/egusphere-2025-6053
Preprint. Discussion started: 2 February 2026

Tai, Y.-C., Noelle, S., Gray, J. M. N. T., and Hutter, K.: Shock-capturing and front-tracking methods for granular avalanches, Journal of Computational Physics, 175, 269–301, https://doi.org/10.1006/jcph.2001.6946, 2002.

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

Tonnel, M., Wirbel, A., Oesterle, F., and Fischer, J.-T.: AvaFrame *com1DFA* (v1.3): a thickness-integrated computational avalanche module – theory, numerics, and testing, Geoscientific Model Development, 16, 7013–7035, https://doi.org/10.5194/gmd-16-7013-2023, 2023.

Vicari, H. and Issler, D.: MoT-PSA: a two-layer depth-averaged model for simulation of powder snow avalanches on three-dimensional terrain, Annals of Glaciology, 65, e16, https://doi.org/10.1017/aog.2024.10, 2024.

Vicari, H., Kyburz, M. L., and Gaume, J.: Brief Communication: Depth-averaging of 3D Depth-Resolved MPM Simulation Results of Geophysical Flows for GIS Visualization, Natural Hazards and Earth System Sciences, 25, 3897–3904, https://doi.org/10.5194/nhess-25-3897-2025, 2025.

Wang, Y., Hutter, K., and Pudasaini, S. P.: The Savage–Hutter theory: a system of partial differential equations for avalanche flows of snow, debris, and mud, ZAMM – Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik, 84, 507–527, 2004.

Wirbel, A.: avaframe/ISeeSnow: 1.0, https://doi.org/10.5281/zenodo.17668272, 2025.

Wirbel, A., Oesterle, F., Fischer, J.-T., Chambon, G., Faug, T., Gaume, J., Glaus, J., Hergarten, S., Issler, D., Jarosch, A., Jóhannesson, T., Martini, M., Mergili, M., Rauter, M., Robl, J., Rosatti, G., Spannring, P., Tollinger, C., Vicari, H., and Zugliani, D.: ISeeSnow – initiating an avalanche simulation tool intercomparison, https://doi.org/10.5194/egusphere-egu24-17750, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-17750, 2024.

Zugliani, D. and Rosatti, G.: TRENT2D*: An accurate numerical approach to the simulation of two-dimensional dense snow avalanches in global coordinate systems, Cold Regions Science and Technology, 190, 103 343, https://doi.org/10.1016/j.coldregions.2021.103343, 2021.

Zugliani, D. and Rosatti, G.: DOT-type schemes for hybrid hyperbolic problems arising from free-surface, mobile-bed, shallow-flow models, Journal of Computational Physics, 507, 112 975, https://doi.org/10.1016/j.jcp.2024.112975, 2024.