A BAYESIAN APPROACH TO CONSIDER UNCERTAINTIES IN AVALANCHE SIMULATION

Andreas Kofler[1,*], Jan-Thomas Fischer[1], Andreas Huber[2], Martin Mergili[3,4], Wolfgang Fellin[2], Michael Oberguggenberger[2]

[1] *Department of Natural Hazards, Austrian Research Centre for Forests (BFW), Innsbruck, Austria*
[2] *Department of Engineering Science, University of Innsbruck (UIBK), Innsbruck, Austria*
[3] *Institute of Applied Geology, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria*
[4] *Department of Geography and Regional Research, University of Vienna, Vienna, Austria*

ABSTRACT: Over the last decades, an increasing number of software tools for modelling rapid mass flows (e.g. avalanches, debris flows) has been developed, tested and applied in scientific and practical studies. But the accurate description of the involved processes still remains a challenge and assumptions are necessary for a simplified description of the natural process. Due to these assumptions, model parameters (e.g. friction) may not present physical properties and thus are commonly back-calculated to fit observed data, which also involve a degree of uncertainty.

We present a Bayesian approach to perform a parameter optimization for the mass flow model r.avaflow, based on documented avalanche events, where uncertainties arising from model simplifications and imprecise observations are explicitly considered. To compare simulation results and documentation data, multiple avalanche characteristics (e.g. run-out lengths, deposition patterns or maximum velocities) are investigated. To derive a *posterior distribution* for the parameters of the basal friction relation, the Metropolis-Hastings algorithm is applied.

The *posterior distribution* is used to perform (i) a probabilistic forward simulation of the same avalanche event and (ii) a probabilistic prediction for a 'theoretical unknown' avalanche track. The dynamic peak pressure results of multiple model runs are evaluated in terms of probability maps. These display the probabilities, that an avalanche hits a certain region of the respective avalanche track, conditional on the used optimization data and considered uncertainties. Observations allow an assessment of the correspondence between theoretically predicted and real events. The outcome illustrates that including uncertainties in both the optimization and prediction process helps to asses the reliability of simulation results for future avalanche events.

Keywords: Bayes' theorem, Metropolis-Hastings algorithm, parameter estimation, probabilistic simulation, posterior distribution, back calculation, prediction

1. INTRODUCTION

Avalanche simulation software packages rely on process models, which describe the physical process. Due to a lack of process understanding and also computational capabilities, the models have to be simplified. This leads to process model parameters, which may not represent physical properties and therefore have to be optimized. The parameter estimation is done by solving an inverse problem (Ancey, 2004), where the arising uncertainties have to be considered in the optimization process.

In the past, several probabilistic approaches have been proposed, which incorporate uncertainties in the calibration or optimization of process parameters of simplified block models (Eckert et al., 2007a,b, 2010) or in the assessment of avalanche risk (Favier et al., 2016), often limited to single optimization variables, mainly the run-out length. (Straub and Grêt-Regamey, 2006) apply a probabilistic simulation set up to AVAL-2D, combining a probabilistic avalanche release and deterministic model parameter scenarios for the mass flow model.

In this work, we show the application of known mathematical concepts and theories from the field of Bayesian statistics to the optimization of process model parameters for the two-dimensional avalanche simulation tool r.avaflow (Mergili et al., 2017). We propose a technical work flow to incorporate uncertainties in the optimization process and derive parameter distributions, which can be used for probabilistic forward simulation and prediction. Finally we introduce two-dimensional probability maps. These show conditional probabilities, that an avalanche simulation hits a certain region, i.e. the dynamic peak pressure results exceeds 1 kPa, given the considered optimization data. The probability maps allow for intuitive interpretation of pre-

*Corresponding author address:
Andreas Kofler
Austrian Research Centre for Forests (BFW), Department of Natural Hazards, Rennweg 1, 6020 Innsbruck, Austria
Tel.: +43-512 573933 5175
Email: andreas.kofler@bfw.gv.at

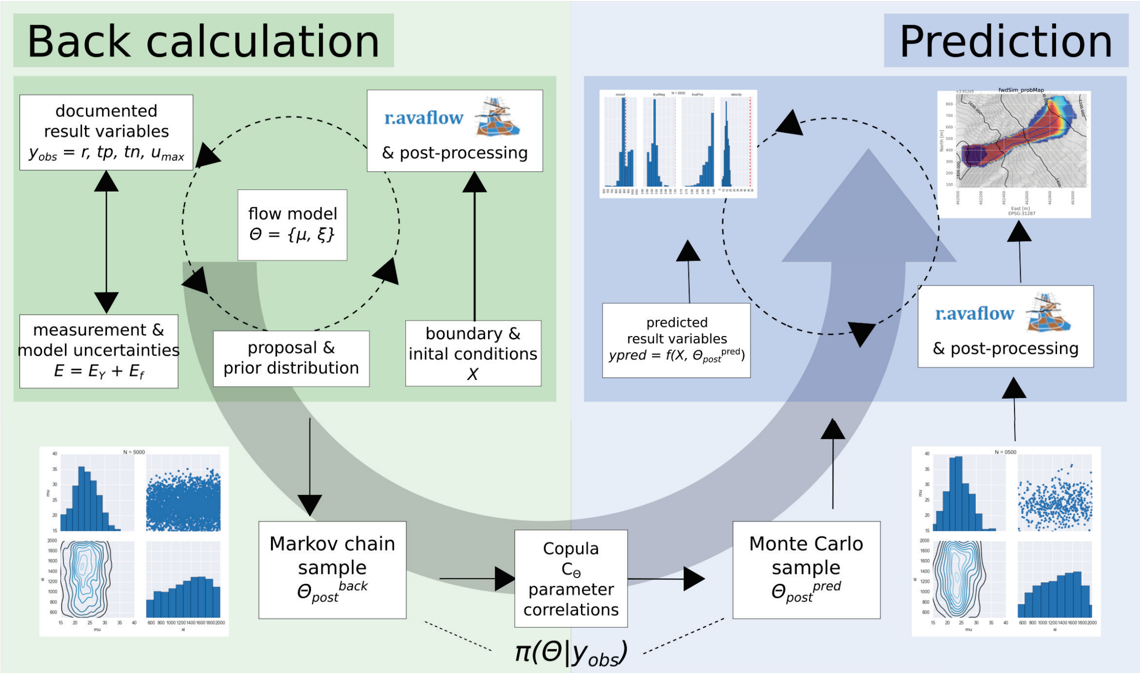

Figure 1: Overview of optimization framework (adapted from Hellweger et al., 2016).

dictive avalanche simulation results and associated uncertainties. The statistical framework for back calculation (green) and prediction (blue) is depicted in figure 1.

2. THE INPUT-OUTPUT MODEL

The used input-output model $y = f(x, \theta)$ combines the mass flow model r.avaflow Mergili et al. (2017) and a postprocessing of simulation results (AIMEC, Fischer, 2013; Fischer et al., 2015). r.avaflow includes different flow models: (i) a general two phase model and (ii) a one phase model with a classical Voellmy friction (Voellmy, 1955; Salm et al., 1990) or a modified frictional relation, accounting for random kinetic energy (Buser and Bartelt, 2009). In this work, we make use of the one phase flow model with the process parameters $\theta = \{\delta_0, \epsilon\}$ of the Voellmy friction relation (Voellmy, 1955; Salm et al., 1990) for the basal shear stress $\tau^{(b)} = \sigma^{(b)} \tan \delta_0 + \frac{g}{\epsilon} \bar{\mathbf{u}}^2$.

The boundary and initial conditions x (e.g. release volume, V_{rel}) are derived from documentation data. Together with a set of process parameters θ, a simulation run is performed and the results are transformed in a flow path relative coordinate system. In this coordinate system, the scalar result variables

- r ...the run-out,
- tp ...the true positive area,
- tn ...the true negative area,
- u_{max} ...the maximal velocity

are derived and collected in the output vector $y = \{r, tp, fp, u_{\mathrm{max}}\}$. The maximal velocity is a direct simulation result, whereas the other three result variables are derived from the peak pressure result and represent the 1 kPa reach (run-out) and 1 kPa outline.

3. TEST CASES

We make use of two documented avalanche events from Austria, the Kerngraben avalanche (Salzburg) and the Wolfsgruben avalanche (Tyrol). Facts about the events can be found in table 1.

The Kerngraben avalanche event is used for the back calculation task, including the respective optimization variables and arising uncertainties. The result of the back calculation are optimized parameter distributions for the process model parameters, covering the associated uncertainties.

Using the parameter distributions, obtained from back calculation of the Kerngraben avalanche, we then perform probabilistic Monte Carlo simulations evaluate the simulation results in terms of probability maps for the following scenarios

- back calculation - forward simulation for the Kerngraben avalanche,
- back calculation - prediction for the Wolfsgruben avalanche.

	V_{rel}	fall height	r	$u_{\max}$	tp	tn
Kerngraben avalanche	65,000 m^3	780 m	1,741 m	55 m/s	241,272 m^2	1,648,774 m^2
Wolfsgruben avalanche	275,000 m^3	980 m	2,103 m	58 m/s	550,992 m^2	1,747,317 m^2

Table 1: Documented data for investigated avalanches.

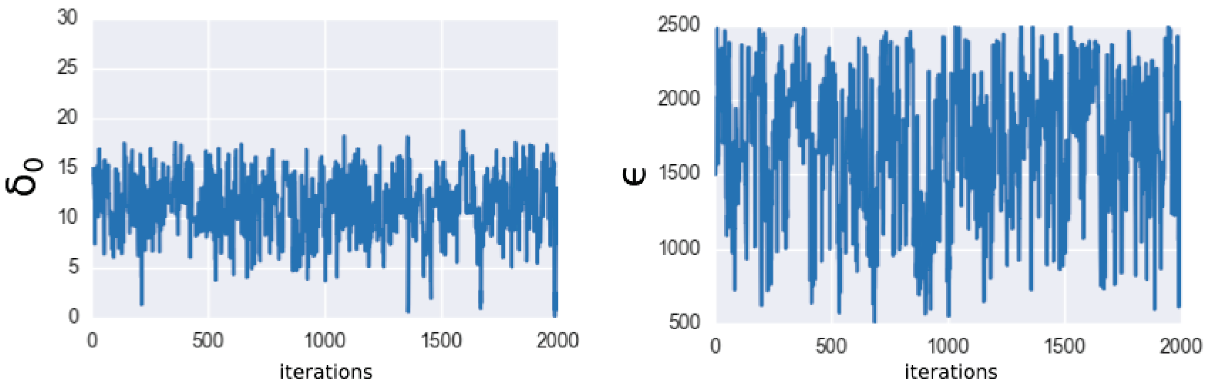

Figure 2: Trace plots: shown are all proposed candidates θ^*.

4. BACK CALCULATION

The aim of the back calculation task is to translate arising uncertainties (documentation and input-output model) in a probabilistic distribution of process parameters. In here, we perform a back calculation for the Kerngraben avalanche (for details, see table 1).

For each simulation run with given input (boundary and initial conditions x, process parameters θ) the result variables $y = \{r, tp, fp, u_{\max}\}$ are calculated. These are compared to the documentation variables y_{obs} (see table 1), using a likelihood function, which expresses the probability of the observed data given the parameter combination θ

$$\pi(y_{\text{obs}}|\theta) = \pi_{err}(y_{\text{obs}} - f(x, \theta))\,.$$

The error term models arising uncertainties

- in the measured data (run-out, velocity, affected area and deposit volume),
- in the model and its numerical implementation respectively.

To apply the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970), we further define the probability density $\pi_{\text{pr}}(\theta)$, called the *prior density*, which in our case simply represents uniform distributions within the intervals $\delta_0 \in [0°, 30°](\mu \in [0, 0.577])$, $\epsilon \in [500\ \frac{\text{m}}{\text{s}^2}, 2500\ \frac{\text{m}}{\text{s}^2}]$. Now we specify deterministic starting values $\theta^{(0)} = \{15°, 1500\ \frac{\text{m}}{\text{s}^2}\}$. A *proposal distribution* $q(\theta^*, \theta_t)$ serves to suggest possible candidates of parameters θ^* within the parameter space for each iteration step t. The likelihood of the respective simulation results is assessed and used to decide, which parameter combination is added to a Markov chain. If the candidate is accepted, the parameter combination θ^* is set as new state ($\theta^* \mapsto \theta_{t+1}$); if the candidate is rejected, the actual state of the chain is kept for the next iteration step ($\theta_t \mapsto \theta_{t+1}$). This way we generate a Markov chain of parameter combinations, asymptotically following the *posterior distribution* $\pi_{\text{post}}(\theta)$. The *proposal distribution* is chosen as a two-dimensional Gaussian distribution centered at the actual state θ_t of the Markov chain, with mean zero and covariance matrix Σ_{prop}

$$\Sigma_{\text{prop}} = \begin{bmatrix} 5^2 & 0 \\ 0 & 500^2 \end{bmatrix}.$$

In figure 2, trace plots of all proposed candidates after 2000 iterations are displayed. They show, how the parameter space for the respective parameters has been explored. The candidates for δ_0 are suggested in the range from $\delta_0 \in [0°, 18°]$, whereas the distribution of possible candidates of ϵ is wider and covers the whole parameter space of the *prior distribution*.

From the candidates, 968 different parameter combinations were accepted, thus resulting in an acceptance rate of 0.48. The mean values of the resulting distributions are $\bar{\delta}_0 = 11.3°$ and $\bar{\epsilon} = 1714\ \frac{\text{m}}{\text{s}^2}$ (see figure 3). The calculation time is about $18\ h$ on a single core CPU with 3.40 GHz.

5. PREDICTION

In this section we utilise the derived parameter distributions in a predictive simulation set up. Monte Carlo simulations with a sample of the optimized distributions allow us to estimate the variability of simulation results with respect to the considered uncertainties. With the given initial and boundary con-

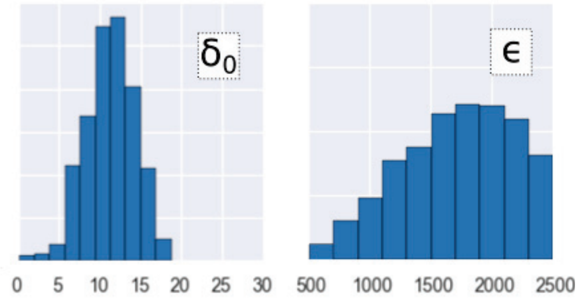

Figure 3: Histograms of the derived Markov chain for θ from the back calculation.

ditions x for the respective avalanche events (see table 1) we perform

- a probabilistic forward simulation for the Kerngraben avalanche,
- and a probabilistic prediction for the Wolfsgruben avalanche.

We distinguish between forward simulation and prediction, because of the considered data in the back calculation process.

The Monte Carlo sample consists of $N = 500$ parameter combinations; Gaussian copulas ensure that the marginal distributions of the sample follow the *posterior distribution* θ, conserving the correlations between the parameters. This leads to 500 simulation results which can be further processed to 500 avalanche characteristics y_{pred} and evaluated statistically (e.g. distributions of run-outs r).

In this work we focus on the visualisation of predictive avalanche simulation results in terms of probability maps. We counted the relative hitting frequency

$$I_j = \begin{cases} 1, & \text{location } j \text{ gets hit} \\ 0, & \text{location } j \text{ does not get hit} \end{cases}$$

for an arbitrary location j of the simulation raster. A hit means that a threshold, i.e. 1 kPa of the dynamic pressure result, is exceeded. Hence the conditional probability that an area j is hit by a single simulation I_j^i, $i = 1, \ldots, N$ can be evaluated as

$$P(I_j = 1|y_{\text{obs}}) = \frac{1}{N}\sum_{i=1}^{N} I_j^i,$$

given the observed data y_{obs} included in the respective optimized parameter distribution. The displayed results differ from approaches in e.g. Straub and Grêt-Regamey (2006), where the annual probability distribution of the run-out distance is conditional on deterministic model parameters. The impact indicator scores in Mergili et al. (2018) are not associated to a probabilistic derivation of process parameters.

In figure 4, the results of the predictive simulations are shown for the Kerngraben avalanche (left) and Wolfgruben avalanche (right). Therein the boundary of the documented affected area is highlighted with a blue line. The colormap indicates the probability, that a respective area of the simulation raster is affected by an avalanche, given the considered data. In the case of the Kerngraben avalanche, the boundary served to determine the run-out length and affected areas for the back calculation and is covered by nearly all predictive simulations. Just at the border of the orographic lower left branch, the probability drops to $\approx 50\ \%$. It can also be seen, that the process model predicts a high probability in the orographic right part of the deposit area, which has not been covered by the optimization data y_{obs}.

The prediction for the Wolfsgruben avalanche shows an overall good agreement to the observation in the upper part of the avalanche track. In the deposition zone, the simulations tend to be a bit shorter than the documentation, thus showing a low probability ($\leq 10\ \%$) in the orographic lower right branch of the affected area, where in fact an impact has been documented. This can be explained by the main flow direction of the simulations, which tends to move straight downstream and thus a bit more to the left, compared to the documentation. Furthermore a finger on the right side has been predicted by the simulation, which has not been documented so far. However this finger is related to a low probability. The predicted run-out range in the central flow direction is also more stretched for the Wolfsgruben avalanche, compared to the Kerngraben avalanche.

6. CONCLUSIONS

A probabilistic concept has been applied to the simulation software r.avaflow to (i) derive a *posterior distribution* of the 2-dimensional process model parameter $\theta = \{\mu, \epsilon\}$ and (ii) employ this *posterior distribution* to perform predictive simulations.

We showed the different ingredients required to apply a Metropolis-Hastings algorithm in the back calculation procedure of the Kerngraben avalanche event. After 2000 iterations and a total CPU-time of $\approx$ 18 h, a Markov chain with 968 combinations of the process model parameters could be found.

We then distinguished between (i) forward simulation of an avalanche event (Kerngraben avalanche), which was considered in the back calculation and (ii) prediction of an (theoretically) unknown avalanche event (Wolfsgruben avalanche). Using the *posterior distribution*, Monte Carlo samples (500 parameter combinations) were derived and applied to the two test cases. The probabilistic results showed generally a higher accordance of the documented affected area for the forward simulation than for the prediction of the unknown event.

The used probability maps are a useful tool to evaluate the variability of simulation results, with re-

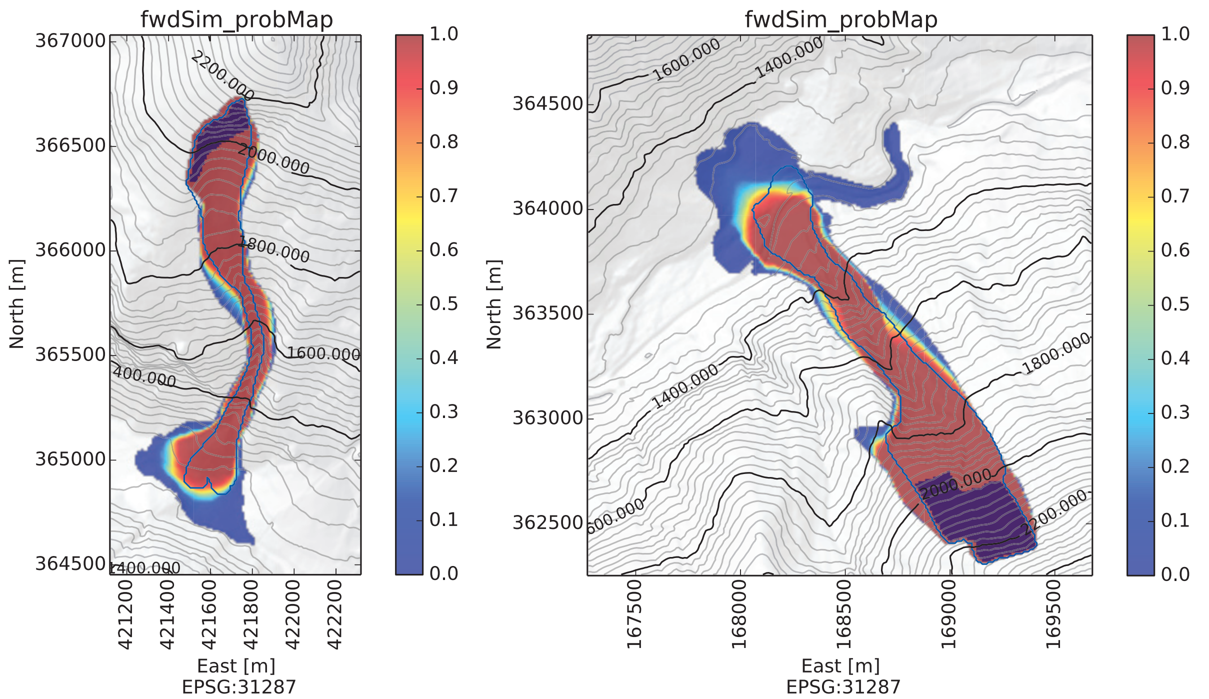

Figure 4: Probability map of $N = 500$ Monte Carlo simulations for the Kerngraben avalanche (left, forward simulation) and the Wolfsgruben avalanche (right, prediction). The colormap indicates the hitting probability from 0 (low) to 1 (high). The blue outline marks the border of the respective documented event.

spect to the considered optimization data. The visualisation allows for easy interpretation and introducing thresholds (e.g. 95 % quantiles according to engineering sciences) and confidence intervals helps to eliminate outliers. Consideration and evaluation of uncertainties associated with avalanche simulations is imperative for researchers and practitioners as rational basis to further employ a risk based hazard assessment.

REFERENCES

Ancey, C. (2004). Powder snow avalanches: Approximation as non-Boussinesq clouds with a Richardson number-dependent entrainment function. *J. Geophys. Res*, 109(F01005):4–5.

Buser, O. and Bartelt, P. (2009). Production and decay of random kinetic energy in granular snow avalanches. *Journal of Glaciology*, 55(189):3–12.

Eckert, N., Naaim, M., and Parent, E. (2010). Long-term avalanche hazard assessment with a Bayesian depth-averaged propagation model. *Journal of Glaciology*, 56(198):563–586.

Eckert, N., Parent, E., Belanger, L., and Garcia, S. (2007a). Hierarchical Bayesian modelling for spatial analysis of the number of avalanche occurrences at the scale of the township. *Cold Regions Science and Technology*, 50(1-3):97–112.

Eckert, N., Parent, E., and Richard, D. (2007b). Revisiting statistical-topographical methods for avalanche predetermination: Bayesian modelling for runout distance predictive distribution. *Cold Regions Science and Technology*, 49(1):88–107.

Favier, P., Eckert, N., Faug, T., Bertrand, D., and Naaim, M. (2016). Avalanche risk evaluation and protective dam optimal design using extreme value statistics. *Journal of Glaciology*, pages 1–25.

Fischer, J.-T. (2013). A novel approach to evaluate and compare computational snow avalanche simulation. *Natural Hazards and Earth System Science*, 13(6):1655–1667.

Fischer, J. T., Kofler, A., Fellin, W., Granig, M., and Kleemayr, K. (2015). Multivariate parameter optimization for computational snow avalanche simulation. *Journal of Glaciology*, 61(229):875–888.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hellweger, V., Fischer, J. T., Kofler, A., Huber, A., Fellin, W., Rauter, M., and Oberguggenberger, M. (2016). Stochastic mehtods in operational avalanche simulation - from back calculation to prediction. In *International Snow Science Workshop, Breckenridge, Colorado, USA - 2016*.

Mergili, M., Emmer, A., Juřicová, A., Cochachin, A., Fischer, J., Huggel, C., and Pudasaini, S. P. (2018). How well can we simulate complex hydrogeomorphic process chains? the 2012 multilake outburst flood in the santa cruz valley (cordillera blanca, perú). *Earth Surface Processes and Landforms*, 43(7):1373–1389.

Mergili, M., Fischer, J.-T., Krenn, J., and Pudasaini, S. P. (2017). r. avaflow v1, an advanced open-source computational framework for the propagation and interaction of two-phase mass flows. *Geoscientific Model Development*, 10(2):553–569.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

Salm, B., Burkhard, A., and Gubler, H. U. (1990). Berechnung von Fliesslawinen: Eine Anleitung fuer Praktiker; mit Beispielen. *Mitteilungen des Eidgenoessischen Instituts fuer Schnee- und Lawinenforschung*, 47:1–37.

Straub, D. and Grêt-Regamey, A. (2006). A Bayesian probabilistic framework for avalanche modelling based on observations. *Cold Regions Science and Technology*, 46(3):192–203.

Voellmy, A. (1955). Über die Zerstörungskraft von Lawinen. *Schweizerische Bauzeitung*, Sonderdruck aus dem 73. Jahrgang(12, 15, 17, 19 und 37):1–25.