Review of 'Modeling the influence of snowcover temperature and water content on wet snow avalanche runout' by C. VERA VALERO et al.'

In this paper, the authors present a model chain for the back calculation of twelve well documented (mostly wet) snow avalanches. The snowcover simulation model "SNOWPACK" is used to derive snow cover properties as input data (release and model parameters) for avalanche simulations. Avalanche simulations are performed with the toolbox "RAMMS", employing a classical flow model with Voellmy friction relation and an extended thermomechanical flow model. Different statistical scores are introduced to evaluate the simulation performance regarding the comparison of simulated flow depths and documented deposition patterns. With these statistical scores and runout estimates the simulation sensitivity is investigated with respect to different kinds of input sources (simulation input, model parameters, grid resolution). Topic and content of the paper fit well to the audience of NHESS. However the reader may be confused because important links and a central theme seems to be missing. A possible solution to finalize this paper could be to either concentrate on one of the three main subjects or to somehow relate them in a consistent way.

The presented model chain consists of the two components: a *snow cover model*, which runs on measured meteo data and *avalanche simulations*, which use the snow cover properties provided by the snow cover model. Statistical scores and runout comparison appear as very useful tool to objectively evaluate the avalanche simulation, i.e. the last part of the model chain - variations of snow cover model performance and variability are not presented. The analysis can be divided in three main (somehow mixed but independent!) contributions: (i) model chain performance check and cross comparison to the classical approach, (ii) sensitivity analysis of the thermomechanical avalanche simulations with respect to avalanche path location (model input parameters), (iii) avalanche simulation sensitivity analysis with respect to computational/terrain resolution. Although the presented approaches appear to be highly interesting and promising some parts are incomplete or at least not well structured/distinguished. Throughout the paper there is a need to clarify what (and why) the authors exactly do:

general questions:

- What is the main goal of the analysis? A new simulation evaluation approach? Introducing or testing a new flow model? Sensitivity study with respect to grid resolution?
- What exactly is deposition in terms of simulation results (deposition is not directly modeled in RAMMS, hence 20cm flow depth are compared to observed deposition, but when does an avalanche simulation stop? why is this an appropriate choice?). Why is the runout analysis separated from the statistical scores and not equally treated?
- What is the advantage of four different statistical scores, when they are based on two independent measures that could deliver the same general message (variation of simulation results)?
- How are simulation input, model parameters, boundary and initial conditions distinguished (e.g. density is a snow cover property in terms of snowpack simulation, describing the release mass and also a flow model parameter in terms of avalanche simulations?)?
- Is section 2 needed or would it be more beneficial to discuss the evaluation approach in more detail and simply refer to Valero et al. (2016)?
- How does the snow cover simulation perform in comparision to field data (e.g. field observations on fracture depths, densities, ...)?

Overall the manuscript is well written and the derived figures 4-9 appear useful to interprete the statistical outcome of the sensitivity analysis. However, for better comparability, the figure axis should have the same limits (e.g. HKS of figures 6-7). Same holds for the figures in supplemental material (e.g.

supp. A, figure S8 a-b). Generally it should be stated what exactly is shown in the supplement figures (A+B) (deposit depth is not a direct simulation result - is it flow depth at a certain (which?) time step? What is depicted by the red outlines (which are very hard to distinguish) in supplement B (20 cm flow depth outlines?)?).

specific questions:

In section 3.2. ((i) model chain performance check and cross comparison to the classical approach) the authors outline their performance evaluation strategy (guideline parameters with classical flow model vs. modeled snow pack properties + ad hoc parameter assumptions for the new thermomechanical flow model). It appears that some crucial questions remain unclear:

- Can the simulation approaches really be compared like this? Is this a comparison of simulation strategies/procedures or of flow models?
- Why does it make sense to use a mix of modeled snow cover parameters (depth) and guideline parameters?
- The growth indices should depend on the choice of flow and the entrainment model/parameters, so they are a result of the model chain?
- With the thermomechanically modelled growth indices, the initial mass of the classical simulations are set. But since classical VS parameters where *calibrated* implicitly including entrainment (field observations that include entrainment) this should not be necessary?
- Why is it appropriate to choose model parameters for "small, frequent" avalanches for all events, when release volume of e.g "Gatschiefer" is up to 330.000 m^3?
- How can the ad hoc model parameter choices for the thermomechanical model be justified (that vary for different avalanche paths, e.g. Entrainment coefficient (0.6-0.8) and α parameter)? Or does the choice not matter because the result influence is negligible?

In section 3.5 ((ii) sensitivity analysis of the extended avalanche simulations with respect to avalanche path location) three different approaches to study the sensitivity of the thermomechanical model are performed (interchanging all or combinations of the model parameters that are related to the snow cover model - fracture and erosion depths, density, snow temperature and LWC). The sensitivity analysis is evaluated on a qualitative level, e.g. no single parameters ranges are investigated (varied with respect to their absolute values) with a quantification of the output variability (which would actually be the advantage of the introduced statistical measures). Open questions are:

- The snow cover model parameters are permuted by event location. With this no quantitative evaluation is possible with respect to the absolute variation of avalanche simulation input, which are (as depicted in table 2) ≈26% for release depth, ≈16% for densities and ≈46% for LWC and ≈151% for temperatures (compared to the respective mean value). Considering these differences (in magnitudes) a direct, systematic comparison and sensitivity analysis is hardly possible - how can we finally conclude which parameters are more important if the are not *equally* treated?

In section 4.3 ((iii) avalanche simulation sensitivity analysis with respect to computational/terrain resolution) the sensitivity with respect to grid size is evaluated. Main questions are:

- As i understand it - this analysis treats the computational grid resolution. How is the DEM resolution treated (resampled to the computational resolution)?
- The main result is that the presented method (statistical scores) can show that parameter values are bound to certain spatial resolutions. Since this has been observed before (e.g. by Bühler et al., 2011, as stated by the authors) this section could maybe be moved to the appendix to smooth the entire manuscript.

Please find some more detailed line-by-line comments/questions below:

- *title* The title of the paper **Modeling the influence of snowcover temperature and water content on wet snow avalanche runout** could focus more on the main contributions (simulation evaluation/sensitivity) and results of the paper (as stated in the abstract *Reliable estimates of avalanche mass (depth and density) in the release and erosion zones is identified to be more important than an exact specification of temperature and water content.* - which slightly contradicts the title).

- *abstract* Do height and depth have different meanings? Is it consistent throughout the paper?

- *3, 66, ... deposits area ...* deposition area prediction

- *3, 67, Instead of parameter optimization, ...* This is a crucial point. If you pursue a flow model comparison, both models should be equally treated, i.e. performing a full optimization and comparing the result performance, not to compare apples and oranges (c.f. Rauter et al., 2016, where a extended flow model is also compared to a Voellmy friction relation with different measures). If you pursue a comparison of simulation approaches/strategies, guidelines should not be mixed with model chain results.

- *3, 74, 3. ..., Fig. 1* To me it appears that the "model chain" is the combination of snowpack and avalanche simulations. The statistical scores/analysis is valid tool to evaluate the results (jointly with the runout estimates) but not a part of the chain. Similar evaluations have been performed for operational avalanche simulations Naaim et al. (2013) (snow properties and simulated avalanche runout) or Fischer et al. (2015) and recently for other mass flows Mergili et al. (2017) (introducing statistical scores to evaluate model performance).

- *Section 2: Wet snow avalanche modeling* In this section the underlying avalanche flow model is described. Since it corresponds to Valero et al. (2016) it could be omitted or transferred to the avalanche dynamic modeling (section 3.2 or appendix) part, as it distracts from the main topic of this paper.

- *9, 219, ... apply a three-dimensional avalanche dynamics model* Maybe better: Two dimensional model operating in three dimensional terrain.

- *9, 221-223, The small elevation difference between the release zones and the weather stations ... provides the sufficient conditions to ...* What do you mean with "sufficient conditions", i.e. sufficiently small?

- *13, 294, class Small avalanches.* Same class for all release volumes from 4.000 m^3 up to 330.000 m^3 - is this in correspondence to (Salm et al., 1990)?. There are also reasons to assume that no mass/volume dependency is necessary and that parameters cannot be interchanged between locations (especially regarding non extreme events, c.f. Issler et al., 2005; Gauer et al., 2010).

- *13, 298, Section 3.3 Contingency table analysis for deposition area* How do these scores compare to similar approaches evaluating snow avalanche simulations (Fischer et al., 2015; Rauter et al., 2016) and other mass flow simulations Mergili et al. (2017)? Would it also be possible to show the result variability with only two of the scores (since they are based on two independent measures)?

- *14, 316, section 3.4 Avalanche runout* This is an interesting definition of runout in a simulation framework - what are the advantages and disadvantages of this definition (are there limitations for multipath effects?, c.f. Fischer, 2013)? Some more details on how the final time steps or simulation patterns are determined would be interesting (dependence on numerical parameters, e.g. cut off for flow depths? what are the stopping criteria/simulation times?, c.f. Teich et al. (2014)?).

- *14, 323, section 3.5 Influence of initial conditions on avalanche runout: sensitivity study* **and** *17, 403, section 4.2 Sensitivity analysis* The intention of an objective sensitivity analysis seems promising, but a systematic approach, which leads to clear and quantifiable results regarding the influence of single parameter/input variables is missing (see general questions above). The general result, that interchanging model parameters from one event to another, reduces the simulation performance is not surprising.

- *16, 372,* the guideline-VS model. The

- *25, 540, ...such as speed, dynamic flow depths ....* Is it possible to give an estimate on the magnitude of their variability?

- *25, 540-542, ... avalanche risks.* What would be the benefit of using further modeling results? Why is it not necessary to consider them (compare, e.g. for avalanche velocities Sailer et al. (2002); Ancey and Meunier (2004); Gauer (2014) or Sovilla et al. (2007); Fischer et al. (2015) for growth indices/mass balance)?

- *26, figure 11* For better comparability the same scaling of the y-axis of the single figures (a), (b) and (c) would be desirable.

- *28, 635, ... depth and spatial extent of the avalanche release area was known.* How does the SNOWPACK model perform regarding the documented release depths - are there any measurements available?

References

Ancey, C. and Meunier, M. (2004). Estimating bulk rheological properties of flowing snow avalanches from field data. *Journal of geophysical research*, 109(F01004):doi:10.1029/2003JF000036.

Bühler, Y., Christen, M., Kowalski, J., and Bartelt, P. (2011). Sensitivity of snow avalanche simulations to digital elevation model quality and resolution. *Annals of Glaciology*, 52:72–80, doi:10.3189/172756411797252121.

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.

Gauer, P. (2014). Comparison of avalanche front velocity measurements and implications for avalanche models. *Cold Regions Science and Technology*, 97(0):132 – 150.

Gauer, P., Kronholm, K., Lied, K., Kristensen, K., and Bakkehøi, S. (2010). Can we learn more from the data underlying the statistical $\alpha - \beta$ model with respect to the dynamical behavior of avalanches? *Cold Regions Science and Technology*, 62:42–54.

Issler, D., Harbitz, C., Kristensen, K., Lied, K., Moe, A., Barbolini, M., De Blasio, F., Khazaradze, G., McElwaine, J., Mears, A., Naaim, M., and Sailer, R. (2005). A comparison of avalanche models with data from dry-snow avalanches at Ryggfonn, Norwaykav. *Proc. 11th Intl. Conference and Field Trip on Landslides, Norway, 110 September, 2005. Netherlands, A. A. Balkema, Taylor & Francis Group*, pages 173–179.

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.

Naaim, M., Durand, Y., Eckert, N., and Chambon, G. (2013). Dense avalanche friction coefficients: influence of physical properties of snow. *Journal of Glaciology*, 59(216):771–782.

Rauter, M., Fischer, J.-T., Fellin, W., and Kofler, A. (2016). Snow avalanche friction relation based on extended kinetic theory. *Natural Hazards and Earth System Sciences*, 16(11):2325–2345.

Sailer, R., Rammer, L., and Sampl, P. (2002). Recalculation of an artificially released avalanche with SAMOS and validation with measurements from a pulsed Doppler radar. *Natural Hazards and Earth System Sciences*, 2:211–216.

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.

Sovilla, B., Margreth, S., and Bartelt, P. (2007). On snow entrainment in avalanche dynamics calculations. *Cold Regions Science and Technology*, 47(1-2):69–79.

Teich, M., Fischer, J.-T., Feistl, T., Bebi, P., Christen, M., and Grêt-Regamey, A. (2014). Computational snow avalanche simulation in forested terrain. *Natural Hazards and Earth System Science*, 14(8):2233–2248.

Valero, C. V., Wever, N., Bühler, Y., Stoffel, L., Margreth, S., and Bartelt, P. (2016). Modelling wet snow avalanche runout to assess road safety at a high-altitude mine in the central andes. *Natural Hazards and Earth System Sciences*, 16(11):2303.