Response to review by J.-T. Fischer

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:

ANSWER: We thank Dr. Fischer for his review and very constructive and helpful comments. We envision many changes to the paper to address his comments. Please find our response to the issues raised by him below.

ANSWER: The first version of the paper did not contain a description of the model. We abandoned this version because we could not rationally describe the simulation results and statistical analysis without refering to model input and output. Section 2 serves to define what the model input is, and what the model produces. It is central in understanding the model chain. We stress the goal of the paper is not to make a model comparison, or to present a method of statistical analysis. The goal of the paper is to identify what boundary conditions MUST be ACCURATELY specified in order to produce reliable simulation results. We found that SNOWPACK can be used, however, there are difficulties. Our approach is to keep the avalanche dynamics parameters (more-or-less) constant, but specify the initial and boundary conditions based on SNOWPACK simulations. The model description serves to help the reader distinguish between material parameters (for wet snow) and initial and boundary conditions. Without the model description we found that it was impossible to clearly separate the two. For this reason we need a description of the model. We perfectly understand the comments of the reviewer. It is not our purpose to write a long paper; but without the model description it is impossible for the reader to judge the results of the simulations, and therefore the model chain. Our goal is to present the entire model chain, from snowpack simulations (which we don't describe in detail), through avalanche dynamics simulations and statistical analysis. In order to appease the reviewer, we have restructured the model description, allowing the interested reader to read only those parts of interest.

- *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?*

ANSWER: The goal of the paper is NOT a new simulation evaluation approach. The goal of the paper is NOT to test a flow model. Our goal is to pinpoint the primary difficulty of modelling wet snow avalanches. We come to the somewhat surprising conclusion: "Reliable estimates of avalanche mass (height and density) in the release and erosion zones is identified to be more important than an exact specification of temperature and water content." Moreover:

we come to the conclusion that snowcover models must be able to identify where meltwater accumulates (this defines the amount of release mass.) This is the result that we want to bring forward. We clearly state this goal and result in the abstract. The evaluation approach, which we do not consider new, is used to support this claim.

- *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?*

ANSWER: We agree with Dr. Fischer that it is not appropriate to compare 20cm flow depth with measured depositions of 1m or 2m. Our philosophy is to adopt a practical approach: as a first step we comparted the measured and simulated inundation areas, independent of the flow heights. This is how the simulation models would be applied in practice. We admit that the measured and simulated flow heights WITHIN the deposition area might differ (the reviewer is CORRECT), but suggest the first necessary step is to compare the measured and calculated inundation areas. The models are simply NOT that accurate (yet) to make deposition height comparisons, which are often a function of very, very local conditions. This is why we restrict the paper to the inundated areas. Regarding the comment on runout distance: runout distances provide an intuitive measure. It is also a variable that is used in widely used avalanche classification systems. For this reason, we are motivated to show our results for runout distance. We do not fully comprehend the remark that they are not equally treated. All results are interpreted in terms of inundated areas and runout distances.

- *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)?*

ANSWER: No, they do not deliver the same general message. We consider all FOUR statistical scores to be relevant and necessary. Again, this has to do with the use of inundation areas to describe model performance. We simply want to know when the simulated model results are correct (hits), or when they predict inundated areas where they were not (false alarms). The HKS and ETS are summarizing statistics, giving an overall statistical score, but they don't allow for much interpretation. A low HKS may result from low probability of detection or high false alarm rate. Note that it is important not to use only probability of detection, as it is easy to cheat this score: just make the avalanche as large as possible and you'll optimize the probability of detection. For this reason, most studies using contingency table analysis show multiple statistical scores, to allow for interpretation of the scores.

- *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?)?*

ANSWER: We asked ourselves this question as well. We agree with the reviewer that the interested reader wants to know the difference between a snow property (e.g. density and temperature) and a model parameter (e.g. Coulomb friction). First, snow properties are supplied by the SNOWPACK model. It would be really nice to have measure snow properties everywhere, but this is simply an impossibility. Second, we apply ONLY ONE set of friction parameters (those for wet snow avalanches). We change snow properties (initial and boundary conditions) but do NOT change model parameters. This is one reason why we describe the model in section 2 to distinguish between the model parameters and the snow properties. The reader should obtain this information by reading sections 2 and 3. This is why we want to keep the model description in order to clearly identify what the difference snow properties and model paramters.

- *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)?*

ANSWER: The reviewer asks a good question: Is section 2, the model description needed? A first draft of the paper did not include the model description. We found, however, that when discussing the results that physical knowledge of wet snow avalanche modelling is needed. For example, we cannot talk about LWC without defining how LWC is included in the avalanche dynamics model. These initial conditions (based on physical modelling) need to transferred to the avalanche dynamics model. Again, the purpose of the paper is to highlight what we regard to be an important PHYSICAL result: we need to know where meltwater accumulates with the snowcover (e.g. base or interior layer) to establish the initial conditions of the simulation. This problem, which is immense, might exclude the application of avalanche dynamics models to perform "real time hazard mapping" in future. Because our results QUESTION the application of models, we believe the model and model performance should be presented in the same paper.

- *How does the snow cover simulation perform in comparision to field data (e.g. field observations on fracture depths, densities, . . .)?*

ANSWER: For a few case studies, additional information is available for fracture depths. We will provide these in the revised manuscript.

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?)?).

ANSWER: Thank you for the suggestions and we apologize for the omission of a complete figure description in the supplement. In Supplement B, the color bar denotes deposits height (i.e., flow height in last time step $>$ 20 cm) of the simulation with the initial conditions corresponding to the event. The outlines of the simulated deposits (i.e., flow height in last time step $>$ 20 cm) for each of the other 11 different initial conditions are shown in varying degrees of rosa to red color. We will modify the figures and the supplement according with your suggestions.

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?*

ANSWER: We emphasize the main result of the paper: if you want to mix snowcover models with avalanche dynamics simulations you must be very certain that the snowcover model is predicting the right fracture depth. Otherwise, simple model (VS) or more complicated models (wet snow) will provide the wrong results. That is, it is NOT about the flow model; good results can be obtained by both simple and more complicated procedures. This is our primary result.

- *Why does it make sense to use a mix of modeled snow cover parameters (depth) and guideline parameters?*

ANSWER: In a "real" hazard mitigation analysis, the release depth is set by meteorological extremes. A specific avalanche is not modelled, but an event with a specific return period. Here we have a problem, which the reviewer has correctly identified: Our avalanche data base contains events with different return periods, most of them smaller than an extreme event. Frankly, we don't know the return period of our avalanches. To circumvent this problem we adopt the following approach: We take the modelled snow cover parameters (height, density) for the release, coupled with the entrained snow amounts, and use extreme friction parameters. We demonstrate that this approach CAN lead to good results. In fact, considering the procedure contains ONLY TWO friction parameters, and does not require detailed snowcover conditions, the results are surprisingly good! We apply this approach for all avalanches to make it general. Here we want the reader to come to the conclusion that maybe the guideline model is superiour to the more detailed wet snow avalanche calculation. Thus, our results indicate that good results, at small cost, can be achieved by mixing modeled snow cover parameters with guideline procedures.

- *The growth indices should depend on the choice of flow and the entrainment model/parameters, so they are a result of the model chain?*

ANSWER: Yes and no. Yes, the growth indices are clearly a result of the snowpack simulations which predict the snow distribution. They are therefore a result of the "model chain". The entrainment parameters do not vary strongly, but are limited to a small range. The growth indices are thus largely independent of the model parameters.

- *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?*

ANSWER: The reviewer is correct. The classical model will provide the "same" results with or without the entrained mass. By including the entrain mass, however, we can argue that we have more "extreme" like events, and therefore can use extreme friction parameters which we apply. Again we find the procedure provides solid results at very low cost (i.e. less detail).

- *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?*

ANSWER: Because we restrict ourselves to wet snow avalanches. For wet snow avalanches the friction paramters are (more-or-less) independent of size. This is a procedure often applied in practice. The reviewer is correct: we could not use the approach in general, especially for dry snow avalanches.

- *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?*

ANSWER: They are very small variations. The influence on the results is neglible. In fact, we argue differently: a range of values provides very similar results and we cannot distinguish between specific values (e.g. simulations with entrainment coefficient 0.6 provided the same results as entrainment coefficient 0.8). Note that all other parameters remain constant for wet snow. Of course the exact quality of snow and terrain will vary, and thus the parameters.

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?*

ANSWER: We think that the original manuscript failed to clearly describe our goal of the sensitivity study. One of the novel approaches we present is to use a physics-based snowcover model to determine the initial and boundary conditions for avalanche dynamics calculations. An important role of the sensitivity study is to determine if this approach add information in the simulation process. Here, we consider that all 12 cases represents a variety of wet snow avalanche cases, and the 12 simulations provide realistic, self-consistent initial conditions. Therefore, we decided to interchange the simulated initial conditions, instead to perturb the simulated values. For example, one could vary temperature over a range of -20 to 0 °C separately from LWC, but in this case, a well below freezing snowcover with a noticeable amount of liquid water is provided to the avalanche dynamics model. This is, however, not a realistic scenario. Therefore, we decided to interchange the sets of snowpack conditions from the SNOWPACK model. We will revise Section 3.5 to better introduce the motivation for the sensitivity study. See also a later comment.

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)?

ANSWER: The measured DEM is resampled to the computational resolution. This is the standard procedure.

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.

ANSWER: The other reviewer also mentioned this point. We can consider removing the section. However, we wanted to put the variation in our simulations that arise from interchanging the initial conditions in perspective. We show that the information added by using the SNOWPACK model is noticeable compared to changing calculation grid size resolution. Precisely because previous studies addressed already the influence of grid cell size on avalanche dynamics simulations, and researchers are aware of it, we consider it a good benchmark for the effect of initial conditions. We will revise the manuscript to make this more clear.

Minor comments

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).*

ANSWER: Valid point. It is about wet snow avalanche runout – but we find that it depends on the depth of the meltwater accumulation. I wonder if an alternative title could be, "The role of meltwater accumulation depth in wet snow avalanche modelling" or "Including snowpack properties in release areas for wet snow avalanche modelling". Is this a more appropriate title?

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

ANSWER: No, this is an inconsistency from our side. We will check for consistency when revising the manuscript. Because we denote the "depth" with "h" we will use height throughout.

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

ANSWER: Changed.

- *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.*

ANSWER: Again our goal is not to "pursue" a flow model comparision. Both models appear to work well – if the snowcover model accurately predicts the meltwater accumulation zone. We hope that some readers will conclude that the VS approach is not too bad. Perhaps both models could be applied.

- *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).*

ANSWER: Yes we agree: different statistical methods could be used to evaluate the simulation results. We argue that some statistical procedure is necessary to compare the numerical results and therefore be included in the "model chain". We recognize that different models exist. However, we couple a three-dimensional avalanche dynamics model with a three-diemsional method to calculate statistical scores. This common component led us to include the statistical method in the modelling chain. We include the references.

- *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.*

ANSWER: We would like to get rid of this section too; however, the results are so dependent on model parameters, we believe that we have to include it to describe the differences bewee

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

ANSWER: Changed.

- *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?*

ANSWER: We agree that the wording was not precise, we will formulate it as: "We argue that the elevation differences between the release zones or deposits zones and the weather stations (see Table 1) are sufficiently small to provide representative snowcover simulations to estimate the initial and boundary conditions of the case studies."

- *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).*

ANSWER: We consider only wet snow avalanches. The procedure we apply cannot be used generally, that is, for dry avalanches. Because lubrication is the frictional mechanism driving wet snow avalanches, they are less dependent on size. (Unlike fluidization, which depends strongly on avalanche size, etc.)

- *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)?*

ANSWER: In many ways we regard the methods of Fischer and Rauter to be superiour to the procedure adopted here. At least these methods consider other avalanche flow properties such as velocity. We simply can't apply these methods (and therefore make comparisons) because we are working on a set of documented case studies of wet snow avalanches. Thus, our method is simpler, but reduced. This does not mean that it should not be applied. Perhaps a hybrid method could be developed? But this is out of the scope of the paper.

- *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)?).*

ANSWER: Yes, the reviewer is correct. The two advantages of this approach are that (1) it is simple and (2) it is independent of the inundation area analysis. The disadvantage of this approach is that it does not consider the distribution of mass in the runout area (20cm cutoff etc.). The runout approach was designed to supplement the statistics of the inundation area hit, miss, false alarms. This statistical data can be misleading – therefore we think the combination of the two methods is appropriate. We will include information concerning the flow-depths and stopping criteria in the revised paper. All simulations stoped when 95 percent of the total mass stopped moving.

- *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.*

ANSWER: See also an earlier comment: the idea of exchanging event parameters is to maintain a consistent set of simulated snow covers. In the paper, we want to demonstrate that snowcover conditions that are required to drive the RAMMS-Extended model can be successfully derived from physics based snow cover models. In this regard, we were actually surprised that the connection between simulated snow cover conditions and the avalanche situation was so tight. One can argue that interchanging "true" initial conditions leads to the unsurprising result that the model performance reduces, but we consider it quite significant that this also holds for "simulated" initial conditions. In any case, we want to show that the simulations for an event indeed add information about the specific event. This motivated the exchange of snow cover conditions on an event basis. We think that our study showed that the snowpack model indeed contributes with accurate information about the snowcover conditions in the release area. In our opinion, a sensitivity study, as proposed by the reviewer, would address a different question, namely, purely focusing on the effect of single parameters. However, this approach would not guarantee consistent snowpack conditions. For example, varying the temperature while maintaining the liquid water content constant could lead to an non realistic condition of wet snow at temperatures well below freezing. So it may be considered a trivial result that event based snowpack conditions contribute to good model performance, but it is generally difficult to know the exact snowpack conditions of the release. Often it is dangerous to access

release areas and particularly in wet snow avalanches, changes in the snow cover state can be very rapid, such that manual observations often miss the interesting period. We will more clearly stress this point in the revised manuscript by revising Section 3.5.

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

ANSWER: Changed.

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

ANSWER: It is clearly of interest. But it is also clear speculation.

- *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)?*

ANSWER: If you have the data then one MUST use them (speed, entrainment). However, this is not the usual case. We would like to turn the problem around: how can you best use a massive amount of data (inundation areas) to the greatest possible advanatage. We understand that the analysis is not complete. But we are considering a sub-class of avalanches (wet snow avalanches) where even velocity data is sparse. Data obtained from test sites is likewiselimited because it contains only one terrain geometry, or overlapping avalanche events.

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

ANSWER: Thanks for the suggestion, we will modify the figures accordingly.

- *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?*

ANSWER: There have been many papers written validating the SNOWPACK model. We will select a few citations to address this question directly.