Reply to Referee #2

GENERAL COMMENTS
The authors of this paper derive a new index for snowpack stability that can be derived directly from snow micro-penetrometer (SMP) measurements. This new index relates the strength of the weak layer (σ_m, micro-structural strength directly measured with SMP) with a the stress exerted on the weak layer by an idealized compression test (tapping with a falling forearm; 0.036 Nm of absorbed energy), which is represented by a SMP derived penetration depth (ps; inversly proportional to stress). The capabilities of the new stability index are then explored with respect to compression test results at the point and slope scale. Finally, the stability index is used to explore spatial patterns and relate it to slope stability.
The study uses an existing dataset, which is well suited for the objectives of this study. The index is neatly derived and the methods for the subsequent analyses are generally well described. The authors clearly highlight the limitations of their approach. Even though the relationship between the stability index and slope stability was poor and the authors were unable to identify significant spatial patterns, I believe that this manuscript is an significant contribution as it provides an important step towards a more effective examination of spatial patterns of stability at the slope scale. I recommend this manuscript to be accepted for publication in NHESS after MINOR REVISIONS.

MAJOR COMMENTS
1. Data - Description of datasets
For somebody reading the manuscript for the first time, keeping the various datasets apart is somewhat challenging. I am wondering whether it would be useful to explicitly label the datasets (e.g., Dataset A, Dataset B) and possibly even include a flow chart that highlights how the datasets relate to each other (How much overlap is between the datasets?) and how they are used for the various analyses. Currently, the order of how the datasets are presented seems out of sync with how they are actually used in the analysis.

> *Though we have already tried to structure the data into (a), (b), and (c), we will improve the description, change the sequence and explicitly refer to (a), (b), and (c) in the revised manuscript.*

2. Methods - Derivation of modelled penetration depth (page 4690)
I am currently unable to completely follow your explanation of how you derived the threshold of dissipated energy (ea). I assume that you are referring to dataset B (19 concurrent manual snow profiles and SMP profiles), but it is not completely clear. It is further unclear to me where or how the observed penetration depth (PS) was observed. I assume this information is coming from the manual profiles, but I am not sure.

> *Your assumptions are correct, but of course we need to describe this explicitly and will do so in the revised manuscript.*

3. Methods - Description of statistical methods
There are numerous elements of the statistical analysis that are not described in the methods section. I believe the manuscript would be easier to read if the statistical methods were described more completely. Examples of omissions include:

3a. Page 4691 - line 26:
You seem to typically assess correlations with the Spearman rank order coefficient, which I believe you refer to with rs (e.g., first paragraph in results section). Later in your manuscript,

you describe relationships between observations with rc (e.g., last paragraph of section 4.1). Does rc also refer to the Spearman rank order coefficient or is it actually a different correlation coefficient. Please clarify.

> *The correlation coefficient r_c is the Pearson correlation coefficient that we missed to introduce in the Methods section, but we will do so in the revised manuscript.*

3b. Page 4692 - line 8 & 9
You are using both Kruskal-Wallis H and pairwise U tests when assessing differences in the distribution of your stability index with respect to the three compression test result classes. Please mention both of these tests in your methods section. Please also describe how you adjusted the p-values of your pairwise comparisons to address the issue of multiple comparisons (e.g., Bonferroni correction).

> *We will better describe this as suggested in the revised manuscript.*

3c. Page 4692 - line 10 & 11
I am a bit puzzled by the sentence "Split values between two categories were determined with the classification tree method (Breiman et al. 1998)." It is unclear to me what 'split levels' are and I do not see this method referred to in the results section at all.

> *We use the classification tree method to define the split between the categories as shown in Figure 4.*

3d. Page 4692 - line 12
It would be useful to include the fact that you used a 10-fold cross-validation approach in the sentence that describes how you assessed classification accuracy.

> *We presently specify this in the Results section but can certainly move this to the Methods section.*

4. Results - Non-spatial analysis: analysis perspectives
You currently present the non-spatial analysis in two stages. First, you examine the relationship between concurrent SMP measurements and compression tests and second, you examine all SMP measurements and compression test results of 11 grids.
The results of these two analyses are quite similar and I am a bit unclear on the purpose of the two perspectives. I think it would help the reader if you stated the objectives of the two analyses more explicitly.

> *The first analysis is presented to derive the model. Second, we apply the model to all grids. We will better explain the objectives in the revised manuscript.*

5. Discussion
In its current format, the discussion section primarily discusses the limitations of your study. It would be nice to start this section with a short description of how all of your results fit together to provide an overall interpretation before you launch into the limitations.

> *We certainly welcome this suggestion. We have not done so as we already summarize the results in the Abstract and the Conclusions.*

MINOR COMMENTS

1. Page 4691 - Equation 4

I believe that the stress term in the denominator should be delta-sigma$_g$ as described on the previous page. If the term in Equation 4 is currently correct, please define it in the text.

The term is correct, but there is a typo on page 4690, line 19: There should be no subscript g.

2. Page 4691 - line 5

You point out that in comparison to Bellaire and Schweizer (2011), you did not include the profile classification as part of your slope stability assessment. Can you elaborate on why you made this choice.

We wanted to keep the classification as objective as possible and moreover primarily consider slope scale information (e.g. slope median CT score) rather than single point scale observations (profile classification).

3. Page 4692 - line 26 & 27

The p-value of the Kruskal-Wallis H test supporting the statement of this sentence is missing.

We agree and will add the p-value in the revised manuscript.

4. Page 4694 - line 11

Typo: Delete second 'in' at the end of the line.

We will change as suggested.

5. Page 4694 - lines 5-9

This section, which starts with "As shown in Fig. 6, ..." is a bit confusing or hard to visualize as the mentioned grids (0708_1, 0708_3, 0708_5 and 0708_9) are actually not shown in Fig. 6.

We agree that the interpretation is not clear. In fact, "as shown in Fig. 6" only refers to the statement that stability depends on the slab as well as the weak layer properties. We will reword these sentences to make it clearer.

6. Page 4694 - line 10

The sentence "In most grids, stability values were either rated rather "stable" or "unstable"." is a bit confusing as other studies in this general research area have used stability categories called "rather stable" and "rather unstable". I would suggest the following wording to avoid this confusion: "In all grids, the distribution of stability indices showed clear tendencies towards either primarily stable and primarily unstable value."

We agree and will change the wording as suggested.

7. Page 4694 - line 13

The term 'mixed results' seems a bit odd. I suggest a wording like "More even distributions of stability index values, ... , were not observed."

> *We agree and will change the wording as suggested.*

8. Page 4694 - line 18
I assume that the term 'range' refers to the interquartile range (IQR), which you describe in the methods section. To avoid confusion, I would like to suggest that you use the abbreviation IQR throughout the manuscript.

> *In fact, in this context, the range refers to the variogram analysis. We will improve the description accordingly in the revised manuscript.*

9. Various locations
It seems to me that you frequently use 'stability' when you actually refer to 'stability index' (e.g., page 4692 - line 24 & 25). I think it would benefit the clarity of the manuscript if you consistently used 'stability index' whenever you refer to the index.

> *We will improve consistency and use stability index wherever appropriate in the revised manuscript.*

10. Table 1
Please describe the abbreviations for fracture character in the table caption.

> *We will include the explanations of the fracture character in the revised manuscript.*

11. Table 3
I think it would be useful to add columns for the number of compression tests and number of SMP profiles to this table. Furthermore, it would be helpful for the reader if grids that are included in Fig. 5, 6 and 7 were somehow flagged or highlighted.

> *We will consider improving Table 3, but in general prefer to provide the information directly in the Figures so there is no need to refer to the Table.*

12. Fig. 5
I think it would be informative for the reader if the number of data points would explicitly be added to the individual charts in this figure similar to Fig. 6. It would also be nice if the grid labels were consistent in the two figures.

> *We can certainly explicitly add the number of data points in each graph instead of mentioning it in the caption only.*

13. Fig. 6
To be consistent with the presentation of the median values in Fig. 6 and the presentation in Fig. 7, the units for the strength axis should be kPa instead of Pa.

> *We will change the units for the strength axis as suggested in the revised manuscript.*

14. Fig. 6 and 7

It seems to me that the grids presented in Fig. 6 and 7 were chosen to illustrate the characteristics of more stable and more unstable slopes. I think it would be useful for the reader if this could be made more obvious by, e.g., adding labels or some kind of an axis (more unstable <-> more stable) to these figures.

We have selected these grids to illustrate that the stability depends on the slab as well as the weak layer properties. We will better describe the motivation for the choice in the revised manuscript.