Reply to Referee 2#’s comments regarding the article *“A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover“.*

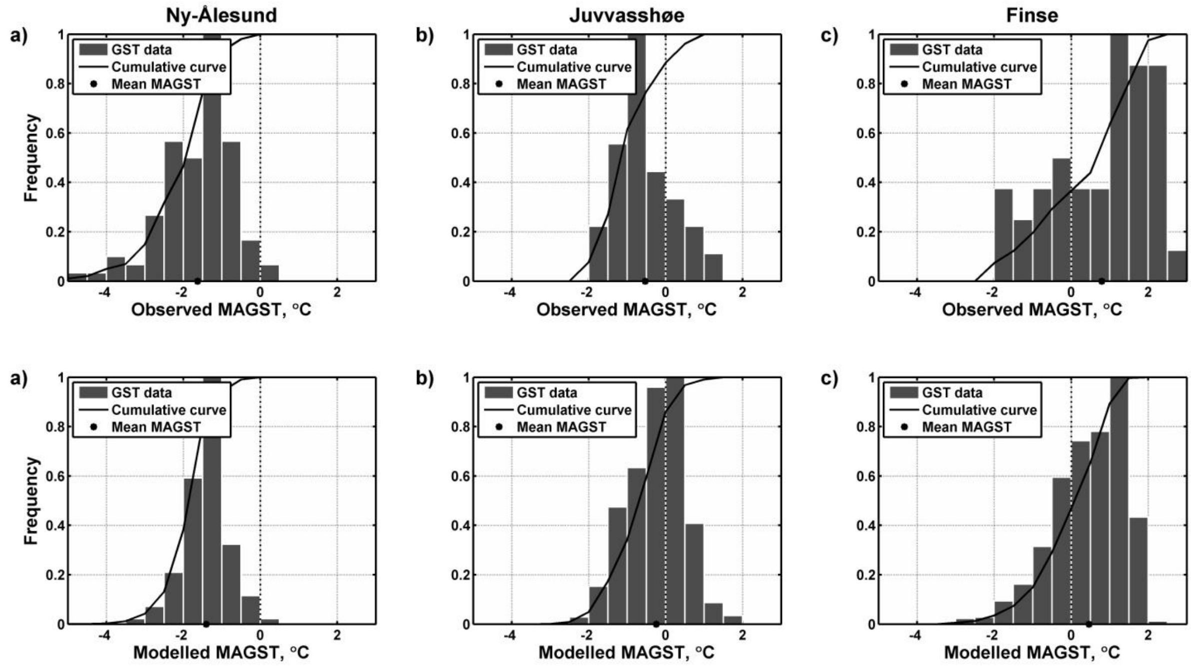
Referee comments are in bold, our answers are without formatting, and *changes to the initial manuscript are in Italics*.

Common points raised by all reviewers were:

A) The n-factor relations are not fully independent from the dataset used for the calibration.

We have changed the model approach with a nF-factor relation based on an independent dataset of 15 stations distributed in 3 different mountain areas in southern Norway. The dataset contains observations of air and ground surface temperatures as well as maximum height of snow over the period 2009-2012. This is the same dataset that makes the basis of the nF-factor relation used in Gisnås et al (2013), except one more year that is now included. The nF is now given as: nF = -0.187 * ln(HS) + 0.399 where HS is maximum height of snow in meters. We have cut the snow-dependent relation of nT-factors, and use a constant nT value of 1, following the value for the surface class “barren ground” in Gisnås et al. (2013).

The new distributions are as follows (measured in first row, modelled in second row):



Changes:

Page 517, line 20: Changed into:

*where nF = -0.187*ln(HS) + 0.399, and nT has a constant value of 1. This relation is based on independent observations of air and ground surface temperatures as well as snow height at 15 stations in southern Norway over the period 2009-2012, published in Gisnås et al. (2013).*

Table 2 is cut and changed into a result table (see comment at Pg 518 from referee #1).

Page 519, line 20 – 27 + page 520, line 6 – 10: Modelling results are updated.

Fig. 5 is updated.

B) Why are other surface characteristics, such as aspect, slope, solar radiation, sediment and vegetation type not investigated to show that snow is a dominating factor?

Surface characteristics including sediment type, vegetation cover, aspect, slope and wetness have been recorded for 107 logger locations, in addition to maximum snow height and days of snow cover. From regression analysis of all factors it was clear that maximum snow height to a large degree explains the small scale variation at our sites. This supported by Fig.3 where the largest spatial variation in mean monthly GST clearly is found during mid winter (Dec – March), and is also strongly indicated by the fact that including height of snow in a simple model strongly improves the modelling result. The importance of snow on ground temperatures in similar areas have been highlighted in several previous publications (Westermann et al., 2013; Gisnås et al., 2013; Farbrot et al., 2011; Isaksen et al., 2002; Isaksen et al., 2011). We agree that a detailed statistical study of these data would be interesting; however, this would extend the scope of this paper and lengthen the manuscript significantly. The focus of this paper is that the distribution of ground temperatures to a large degree can be reproduced using a simple approach only including one parameter. We have therefore chosen not to include the full statistics of all surface characteristics for this manuscript, but could of course include it after an editor decision.

Below we present a point-by-point response to all individual referee comments:

General Comments:

1. What do we learn from scaling factors like nF and nT? These scaling factors do not enhance our process understanding and are in general questionable in this context. The scaling factors were mainly introduced by Lunardini (1978) with a main focus on engineering applications. Therefore, I would expect for a paper like it is presented here, a much more process-based approach, which would fit more into a geoscientific journal like 'The Cryosphere', aiming at a better process understanding. However, I completely agree that these scaling factors are sometimes useful for the application of calibrated permafrost models simulating the current or maybe also the past permafrost distribution in areas like the Arctic and Scandinavia but in alpine to high-alpine environments, they are not applicable, because for example the freezing n-factor, nF, has a very large seasonal variability, which cannot be captured with annually calibrated factors.

The reviewer has brought up an interesting point: are sophisticated process-oriented models "better" than more simple empirical models? While the reviewer's answer to this question appears to be more of a "yes", our point of view is more to: "maybe, but we don't know yet". Process-based approaches are sensitive to a large number of model parameters, which can vary considerably in space. They can be well calibrated at the point scale to fit measured data, but in spatially distributed applications, their performance may not be better (or actually worse!) than that of robust and more simple approaches with few empirical parameters. Ultimately, this question can only be answered by case studies validated with spatially distributed in-situ data sets on permafrost conditions. And there are still only very few spatially distributed modeling studies where validation is based on more than a few in-situ point measurements. For Norway, for instance, we consider it very likely that a simple TTOP approach WITH subgrid representation of snow depths, as outlined in this manuscript, is a better description of nature than a more process-based one (such as in Westermann et al., 2013) WITHOUT such subgrid representation. With the further advance of process-based models, the TTOP approach may have become redundant in 10 years time, but currently we still consider it an important part of the tool set in permafrost modeling.

2. In addition, if n-factors are used to perform future scenario modelling based on a TTOP approach several problems will arise. Temporal extrapolations with statistical derived n-factor models are not straightforward, because these models do not contain any non-linear feedback mechanisms such as the influence of non-conductive processes such as water transport in the ground or air within coarse grained ground and neglecting any transient changes of the n-factors itself.

This relates to the previous comment. We agree with the reviewer that the empirical parameters may (and probably will) change in the future. But at least in data-sparse regions, where it is hard to impossible to constrain the parameters of process-based models (i.e. in any application over very large spatial domains), the performance of simple empirical approaches may not be worse than that of more sophisticated approaches even on a 100 year timescale. Again, only case studies validated by a solid basis of in-situ observations can decide this issue.

3. The model validation with an independent dataset is in my opinion very important. I believe that the used dataset is not completely independent although the authors used different methods of snow height measurements like manual snow probing and GPR measurements. I would suggest that the model should have been applied at an independent place where the measured data was not used for the establishment of the transfer functions.

See point A) in the introduction.

4. The authors calibrate their model with only one factor (snow height) knowing that although they get statistically seen very good results, it does by no mean reflect the variability of the 'real' nature, where snow height is an important factor among many others like aspect, slope, ground characteristics (vegetation, bedrock, fine or coarse material). As the authors correctly state at the end of their discussion on page 523, this would require a more sophisticated statistical approach. Therefore immediately the question arises 'why you did not include other parameters like the ones mentioned before?'

See point B) in the introduction.

5. Fiddes and Gruber (2012, 2013) and Gubler et al. (2011, 2013) for example show approaches, being more process-based but with the same objective like this paper to overcome the problem of sub-grid variability. The question arises here 'why not using a more sophisticated approach to overcome the sub-grid variability?'

See Points 1 and 2 above. Here we may ask the question what would perform better for modeling the ground thermal regime in a country like Norway: TopoSCALE/SUB/etc. with parameter values from the Alps, and driven by potentially biased ERA reanalysis data, or CryoGrid 1 (as in Gisnås et al, 2013) with a subgrid representation of snow depths? We think that there is no clear answer, and we really don't know, at least not at present! It is at least possible that the second scheme is as good or even better, at a fraction of the computational costs.

More sophisticated approaches demands normally more input data which not always are available, and certainly not always at the sites included here. We demonstrate here a more simple approach, capable to later be applied on larger areas. As long as similar results can be achieved with a more simple approach, we think our approach is justified.

6. The authors talk in their paper always from ground surface temperature. However, the sensors are NOT at the surface itself, they are some cm in the ground, therefore they should better use the expression ground temperatures. If a surface temperature is measured you will of course receive completely different results because of radiation effects.

Yes, the loggers are buried a couple of centimeters to avoid the effect of direct radiation. Therefore, strictly speaking we do not measure surface or skin temperatures but very near-surface temperatures. We add a clarifying sentence in the method paragraph. But for the understanding for the reader we keep the term "surface temperatures" We change the sentence at page 515, line 1 into the following: *"A total of 171 temperature data loggers recorded GST <u>(2 cm below the ground surface)</u> with 2 and 4h temporal resolution…."*

Specific Comments:

Page 512, line 15: please use always the same units in meter instead of centimeters.

Done

2. Page 513, line 1: suggestion: better use maritime instead of marine. Done

3. Page 513, line 12: maybe better to talk about 'active permafrost' in the sense that active layer refreezes during winter. At your site probably relict permafrost will still exist in larger depth for hundreds of even thousands of years based on the fact that the heat flow is strongly reduced and low gradients are observed as mentioned on line 8 above.

We agree. However, at the sites in Jotunheimen, we have shallow sediment cover over bedrock, which is quite coarse grained. This implies a relatively good coupling between atmosphere and ground. At the study sites, permafrost was hardly very thick, even during the the Little Ice Age (e.g. Hipp et al. 2011), and probably disappeared during the Holocene Thermal Maximum (Lilleøren et al. 2011). But to avoid confusion we change the sentence into:

"… and during the last decade have undergone warming and possible degradation"

4. Page 515, line 5: missing an before accuracy. Done

5. Page 515, line 19: missing a before probe. Done

6. Page 517, line 15: missing the before number. Done

7. Page 517, line 25 and 26: I do not believe that the estimate of this temperatures can easily be transferred to the ground temperature without taking into account the highly non-linear processes within the ground itself. I would probably agree at a site like Juvasshøe, where you have a quite direct relation between the atmosphere and the

ground (mainly bedrock). However, at sites with complex ground characteristics it is by far not so straight forward.

We agree that this is difficult, but the TTOP model has been successfully applied in a number of studies (e.g. Gisnås et al., 2013; Juliussen and Humlum, 2007; Smith and Riseborough, 2002), with the particular focus on large-scale spatial application. As already mentioned in the replies to Points 1, 2, and 5, it is ultimately validation with in-situ data which decides. So far, we only claim that the scheme gives promising results concerning MAGST at the three study sites (which is thoroughly validated), and refer to the above mentioned publications that the combined TTOP scheme gives promising results on large-scale applications.

8. Page 518, line 18: The authors write: 'The median at Finse is close to 0°C during the entire winter, showing that GST is decoupled from the air temperature in large parts of the field area.' I interpret this sentence probably different as the authors do, therefore this sentence needs clarification. I would interpret the fact that the temperatures at Finse are close to 0_C not only as a 'decoupling' effect but much more an effect that there is mainly no permafrost at the investigation site and therefore the ground temperatures stay at zero degree during the whole winter time. You have a decoupling also at the permafrost sites, but then the permafrost influences your measured ground temperature strongly. We clarified the sentence as follows: „*and that permafrost is not present in most of the area.*"

9. Page 520, line 9-11: This is quite an 'academic' exercise as in reality as already mentioned above other factors like ground characteristics have sometimes even a larger influence on the ground temperatures than the snow cover. The entire paper is based on the assumption that snow is the main contributor to the variation of ground surface temperatures at this scale (< 1x1km). We are of course aware that sub-surface characteristics sometimes are more important, such like organic surface cover, different types of sediment covers, block fields etc. However, at the scale we operate here in this manuscript we think that it is not wrong to suggest snow being the most influential parameter. See further comment B) in the introduction.

10. Page 521 & 522, line26 and 1-3: I fully agree with this sentence. However, it also shows how dangerous such statistical models can be. If you calibrate your snow distribution from the past with a statistical approach like you have it chosen here, this approach could fail for future conditions when other effects would change your snow redistribution for example by different future synoptic conditions.

We agree. However, pretty much ALL future predictions of permafrost conditions use a very simple assumption on the snow distribution: the snow depth is considered constant for a generally large grid cell. We know that this assumption is wrong in most (mountain) permafrost areas, already for the present conditions! This work aims to improve on this major deficiency at least for the present. We are extremely confident that future conditions are better reproduced assuming the present distribution is unchanged, than with a constant snow depth which is grossly wrong already for the present.

11. Page 522, line 15: I agree that for Scandinavian conditions, the chosen approach is feasible but for other mountain areas it is not easily transferable.

We agree with that comment and we therefore state in line 13 that this is valid for Scandinavian conditions. We believe this is clear and have not further elaborated this in the text.

12. Page 522, line 20: It is feasible but you loose process understanding and it is likely not applicable for predictions! There are always trade-offs, we find that our approach is feasible.

13. Page 522 and 523, line 27, 28 and 1-4: I do not agree with this statement because of the already mentioned reasons in my general comments. Statistical relations are not useful for simulations of the future behavior of permafrost because of many non-linear effects. See comment 2 above.

References:

Gisnås, K., Etzelmuller, B., Farbrot, H., Schuler, T. V., and Westermann, S.: CryoGRID 1.0: Permafrost Distribution in Norway estimated by a Spatial Numerical Model, Permafrost and Periglacial Processes, 24, 2-19, 10.1002/ppp.1765, 2013.

Juliussen, H., and Humlum, O.: Towards a TTOP ground temperature model for mountainous terrain in central-eastern Norway, Permafrost and Periglacial Processes, 18, 161-184, 2007.

Smith, M. W., and Riseborough, D. W.: Climate and the limits of permafrost: a zonal analysis, Permafrost and Periglacial Processes, 13, 1-15, 10.1002/ppp.410, 2002.