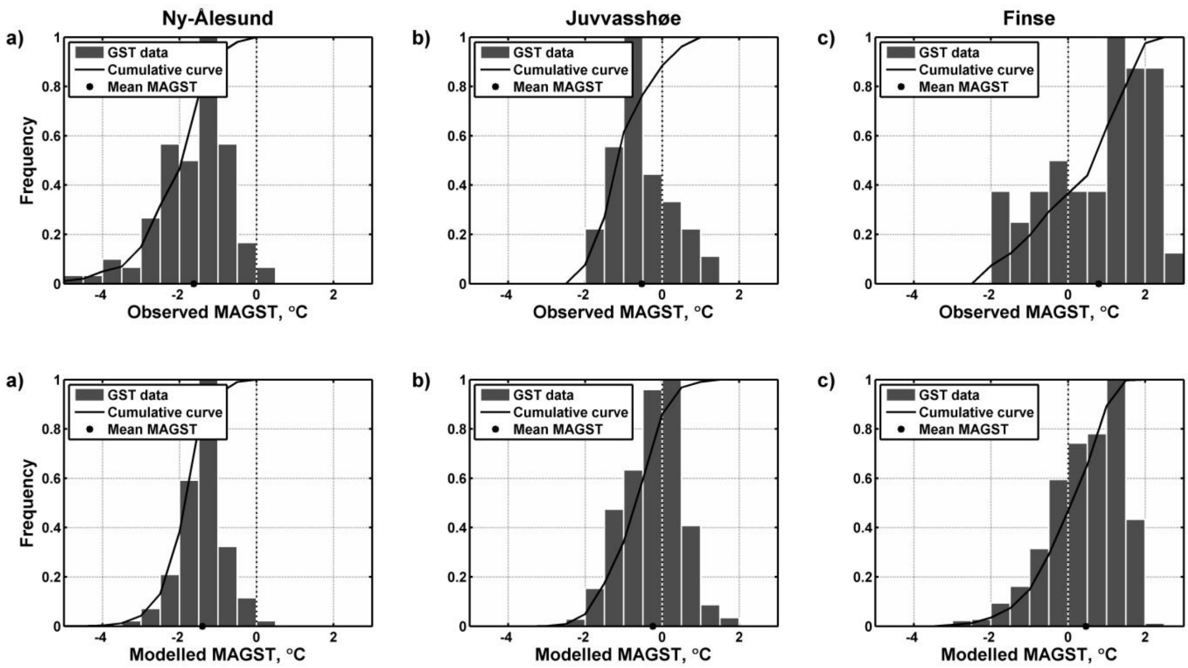
Reply to Joel Fiddes' comments regarding the article *"A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover"*.

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. In northern climates, such as the focus of this study, solar radiation plays a much reduced effect compared to more southerly (and steeper) terrain (although it is difficult to generalise as a large range of latitudes 61-79 deg N are covered). However, I am not convinced that snow height can be stated so unequivocally as the single most important variable governing ground temperatures. Aspect and ground type can be significant even at very fine scales (e.g. Gubler et al. 2011). For example, a snow pack of insulating depth may fail to de-couple the ground from the atmosphere when the surface is composed of large-block material. Additionally, choice of sampling site has a large influence on this studies results and interpretation of significant processes. For example, if the Ny-Ålesund field site was located some 2 km south-east on the steep N facing slopes - other processes related to aspect or slope (e.g. avalanche deposition requiring a different modelling approach to that of wind-blown snow) may become significant. The sites in general appear to be reasonably homogeneous so it is difficult to assess the importance of other variables on MAGST variability in these climatic zones. Along the same lines, a thorough description of variability of other variables (topography, surface, subsurface) that influence MAGST and how these were sampled would be useful to interpret the results (e.g. Fig. 3).

The presented model approach which takes only snow depth into account, is to a large extent capable of reproducing measured distributions of MAGST at three sites across a significant climatic gradient. In the revised version, the published nf-snow depth relationship of Gisnås et al., 2013, which has been applied to entire mainland Norway, is employed, and the agreement with the measured distributions is still satisfactory. It is quite possible, that other factors play a significant role at the point scale, i.e. for each logger, but this does not seem to play a major role for the MAGST distribution at the sites. The important point is that a largely empirical model approach succeeds in reproducing in-situ data for the range of environmental conditions described in the manuscript, which are typical for a significant part of the permafrost areas in Scandinavia. By definition, an empirical model approach can only be justified through comparison with in-situ data, which we have done for three sites. Therefore, the results are strictly speaking only valid for these sites. However, installing arrays of 100 loggers for a representative number of sites to cover a representative cross-section of environmental conditions (which would then contain the mentioned steep N-facing slope) is clearly a prohibitive effort. If TTOP with a statistical formulation for snow is to be applied on large scales, e.g. entire Norway, it would again have to be validated with a range of in-situ observations, e.g. boreholes, BTS measurements, or geomorphological observations (as in Gisnås et al., 2013, Westermann et al., 2013). But at this point, we are confident that the results obtained at the three sites are a clear indication that the statistical scheme is a major improvement compared to modeling schemes with a constant snow depth, as they have been applied for Norway in the past.

2. Wind-blown snow modules (such as Alpine3D) which could be used to compute a snow height distribution can be expensive to run on a fine subgrid, particularly over large areas as they require a fully distributed simulation. It would be good to provide some more details on the costs of running such modules and impact on efficiency of the proposed subgrid scheme (how many years would you run etc.). In addition a quick comment on the ability of these schemes to re-create wind-drift patterns at various spatial scales (assuming we are interested in the fine scale at which topographically modified wind patterns can be very different from the larger scale forcing) and uncertainties related to processes such as sublimation, would be useful. Especially considering such a method is fundamental to the implementation of the proposed scheme over larger areas.

Sentence included at page 522, line 20: *Because of computational expenses and input-data requirements, Alpine-3D and Snowmodel are suited for local approaches while the Winstral terrain parameterization is applicable over larger regions.*

3. Do you think the snowmobile surveys would have a sampling bias (aside from snow/no snow that is mentioned) due to the terrain it is possible to cross (because of difficult terrain, steep slopes) and therefore choice of field site? How does this influence the whole experimental design and wider application of results?

The representativeness of the GPR-tracks with respect to curvature, slope and aspect is assessed by Litherland (2013), that found that the tracks in general are representative.

4. Study sites do not represent really complex topography, especially in Svalbard – how transferable is this method to more heterogeneous environments? For example, how would the approach be expected to perform if the study footprint (or coarse grid) lay across a North/South mountain ridge where other processes could be important in driving MAGST? This point relates to how this scheme upscales as a regional modeling approach as the footprint of a coarse model grid unit (e.g. 1 km x 1 km CryoGRID), cannot be assumed to be as homogeneous as the field sites appear to be.

We agree that other factors, such as solar radiation might contribute more to the small scale variation of MAGST in other permafrost areas. In that case a different parameterization should be used. However, some of theses factors would co-vary with snow height and hence our approach might even have a wider applicability than just Scandinavia.

5. How transferable are the calibrated N-factors to other years? It would be good to include some comments on how this would upscale temporally/ spatially. Would the model be recalibrated every year? What is the spatial resolution over which a given calibration would be considered valid?

The n-factor relation to maximum height of snow is general, and would not be calibrated for every year. However, the amount of snow in the grid cell would vary between years; hence the n-factor distribution would also vary between years.

In the revised version, we employ nF-factors from Gisnås et al., 2013, which represent a space- and time average for entire Norway. Still, the representation of the MAGST distributions is satisfactory for three different sites and one random study year. This suggests that the statistical approach is quite robust.

6. In general, the assertion that wind-blown snow is the most dominant process at subgrid scales (while possibly true at the field sites in this study) is not proven by the data as other variables which govern MAGST are not tested. In addition, the conclusion that GST variability is small during summer and early winter (p524 l.6) really depends on the heterogeneity that exists within the footprint tested.

See comment 1 above, and comment B) in the introduction.

7. The authors state that this approach enables a simple equilibrium permafrost model to reproduce observed ground temperature distributions. I think this statement needs to be backed up a little more strongly than simply eye-balling Fig. 2+5.

We have now included a table (see referee #1) including mean, standard deviation, minimum, maximum and skewness of both observed and modeled distributions. In addition we have included the following text at page 519, line 24:
The model results (Fig. 5) are in good agreement with the observed distributions (Fig. 2), *with r^2 between the distributions of observed and modelled MAGST being 0.9 (Ny-Ålesund), 0.6 (Juvvasshøe) and 0.4 (Finse), using 0.5°C bin width.*

8. It seems that some of the same datasets/ sites were used to calibrate as well as evaluate the model which calls into question independence of results.

See comment A) in the introduction.

SPECIFIC COMMENTS:

1. p.511 l.23: Perhaps change “implemented” to “established”. Done

2. Sect. 3: How was the random distribution of loggers achieved? Random number generator in MATLAB. Sentence included: “*...,with coordinates generated by a random number generator”.*
3. Sect. 3: What was the basis of each field-site footprint selection? The Juvvass site is an already established field site (Isaksen et al., 2002; Isaksen et al., 2003; Isaksen et al., 2011), and is chosen to represent a typical Scandinavian mountain topography, with moraine/block field cover. The Ny-Ålesund site was chosen to represent a high Artic type of setting, and is also surrounding an already established surface energy balance station. The site at Finse is chosen to represent the typical topography (rugged but not really alpine surface roughness)

and ground cover (bedrock to bouldery terrain). The three sites/foot-prints represent different topographical settings, as well as the continuous, discontinuous and sporadic permafrost zones.

4. p.515 l.19: I think this sentence needs to be re-phrased. Changed into: *"The height of snow (HS) at each data logger was measured manually with a probe at maximum height of snow. Daily height of snow is measured at one location within each field area, and the observation date matches well with maximum height of snow at all three sites."*

5. A definition of "snow maximum" would be helpful, e.g. p.515 l.19. Acknowledgement of uncertainty with the selection of this date would be good. Daily height of snow is measured within each field area, and the field observations are made at snow maximum at all three sites. To make the paper more compact we did not include a figure to show the development of the snow pack including the date of maximum snow cover. However, to clarify, we included the sentence in the previous comment (see point B.4).

6. p.515 l.21: "lacks 13 days to an entire year" - do you really mean a range here (13 days–1year)? No, it contains 352 days of measurements. Changed into: *„The data series from Juvvasshøe lacks 13 days to contain a full year of data".*

7. p.515 l.21: Is surface temperature assumed to equal air temperature in gap filling? Yes, in this particular case it is. The data gap is in the last weeks of July, without snow cover and after the ground has dried. We have 12 years of air and ground surface data at 6 of the loggers, covering the variety of ground cover types along the transect. All of the loggers had a marginal offset to air temperatures during July, after the melt-out. We therefore found gap-filling from air temperatures the most sufficient method in this case. We changed the sentence into: *"The data series from Juvvasshøe lacks 13 days <u>in the end of July</u> to contain a full year of data"*

8. p.515 l.25-26: Re-phrase to make what you mean clearer. Rephrased into: *"Snow surveys using ground penetrating radar (GPR) were carried out around the time of maximum snow heights, the same date as the manual probe measurements at the data loggers. Maximum snow height was derived from snow depth sensors at each of the three sites.*

9. p.517 l.1: What are the likely implications of under-sampling shallow snowpacks? For this study the main implication is that there will be a shift between the distributions of snow heights measured at the logger sites and with the GPR survey, consequently the modelled MAGST will be slightly too warm.

10. p.517 l.18: How are the degree days extrapolated to the field-sites? How would this be done without measurement stations near by i.e. away from established experimental sites? The degree days are not extrapolated but measured within each field area. We assume that air temperatures do not vary within the 1x1 km field area. Gridded daily air temperatures at 1km resolution are available for entire Norway from the Norwegian Meteorological Office, and degree days are calculated based on this in Gisnås et al. 2013.

11. p.517 l.22: "surface vegetation type" - did you also consider non-vegetated surfaces?

In this study we only consider sparsely vegetated to non-vegetated surfaces (since we here study alpine to arctic settings). In Gisnås et al 2013 we also consider vegetated surfaces and forest, but had only one nT-factor for barren ground. We found that within this barren ground –surface class nT is related to snow cover as well. However, to simplify we have changed back to a constant nT-factor of 1 in accordance with Gisnås et al (2013) in the revision of the paper, and this sentence is deleted.

12. p.518 l.15: How exactly is the snow cover duration calculated? Do you only consider a thermally-insulating snowpack (decoupled surface and atmosphere)? Schmid et al. (2012) discuss this topic and how a melt-date can be robustly estimated from ground temperature measurements. We are aware of the publication by Schmid et al. (2012). The number of snow covered days in this study is derived from deviation in daily variance between air and ground surface temperatures, following Hipp (2012). This method has also proved to be robust, and the melt-out dates in Ny-Ålesund is validated with hourly photos from an automatic camera.

13. p.518 l.20: Perhaps describe this phenomenon as the zero-curtain which is related to thermal inertia due to phase change (Outcalt et al. 1990). Included sentence: *"and the zero-curtain effect resulting from the phase-change when the ground thaws."*

14. p. 521 l.10: Exactly 1 m? always? Perhaps modify to "approximately 1 m". This is changed as suggested.

15. p.521 l.13: "is" > "are" This is changed.

16. p.521 l.20-23: Make sure this isn't misunderstood as a generally applicable conclusion. It's applicable for this type of topography; included sentence: *„in high alpine environments such as in the Scandinavian mountain range."*

17. p.522 l.20: Agreed, but as stated above a wind-model still needs to be run on the fine-grid with suitable wind field and there are significant uncertainties with processes such as sublimation. These associated costs and uncertainties should be mentioned. It takes about 1min for running the *Winstral terrain parameterization* at 10 m resolution over 1x1 km (10 000 cells). To run one season of *Snowmodel* (Liston and Eldar) over 10 000 cells takes around 10 hours on a normal computer. To run only the wind distribution scheme (*ARBS*) for *Alpine-3D* takes a few weeks for the same area, on a normal computer. However, there are many different possibilities and combinations of this, and it also depends on how you set up your model (how you parallelize, how many cores you have available etc.) and we therefore believe this is not within the scope of this paper. We still included a sentence at page 522, line 20: *Because of computational expenses and input-data requirements, Alpine-3D and Snowmodel are suited for local approaches while the Winstral terrain parameterization is applicable over larger regions.*

18. p522 l.27-28: I don't think you can confidently say this as the statistical technique applied here may not be valid under future conditions. 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 or even 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.

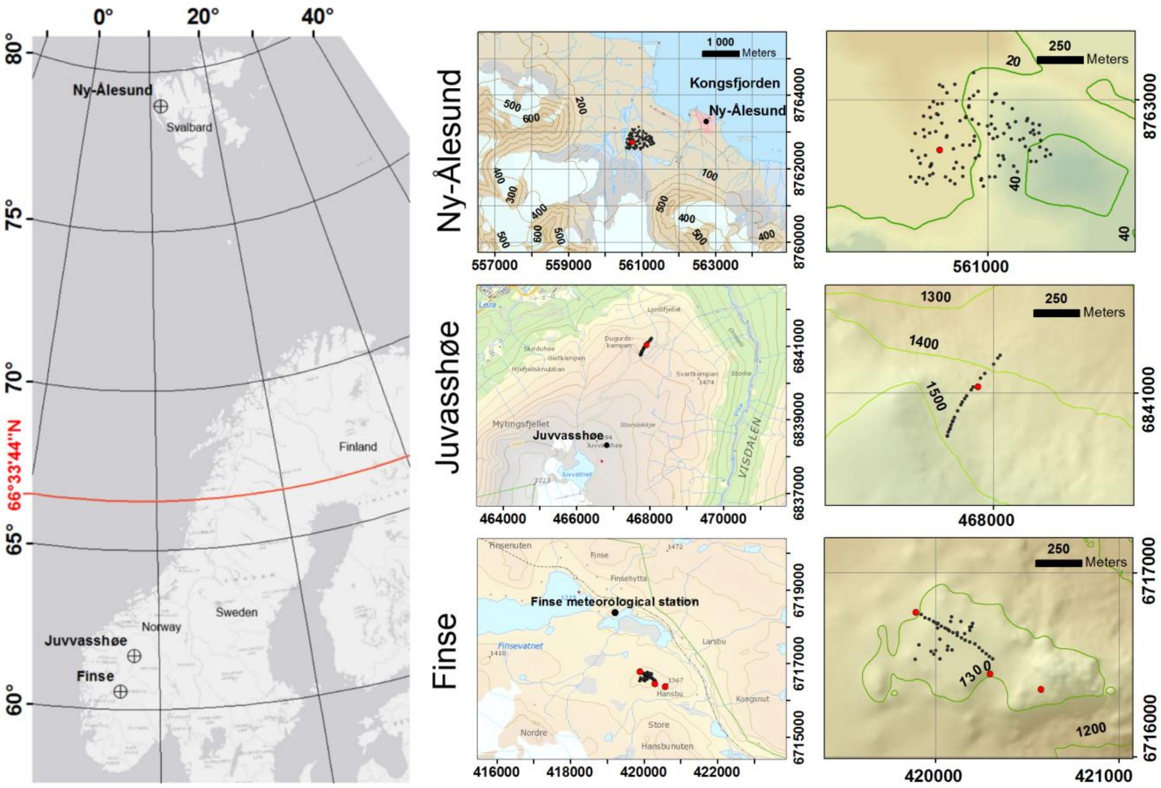
19. p523 l.16: I would suggest this is a strong statement considering that aspect and slope, ground etc. were not sampled. How representative is the study footprint from the broader study area?

We changed the sentence into: *"In the study areas, the variability of ground temperatures can to a large degree be described by the variability snow depth, which in turn is depending on …."*

20. Qualifying statement about solar radiation you have on p.523 l.18 should already be made much earlier to allow the reader to know you are not generalising your results to regions where other variables are certainly significant. We hypothesize that snow is the most important factor for these areas - alpine to arctic areas in the Nordic countries.

21. Fig. 1 needs to be larger.

We made a new layout of Figure 1 that will fit better with the TCD format:



References:

Farbrot, H., Hipp, T. F., Etzelmüller, B., Isaksen, K., Ødegård, R. S., Schuler, T. V., and Humlum, O.: Air and Ground Temperature Variations Observed along Elevation and Continentality Gradients in Southern Norway, Permafrost and Periglacial Processes, 22, 343-360, 10.1002/ppp.733, 2011.

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.

Hipp, T.: Mountain Permafrost in Southern Norway. Distribution, Spatial Variability and Impacts of Climate Change., PhD, Faculty of Mathematics and Natural Sciences, Department of Geosciences, University of Oslo, 166 pp., 2012.

Isaksen, K., Hauck, C., Gudevang, E., Ødegård, R. S., and Sollid, J. L.: Mountain permafrost distribution in Dovrefjell and Jotunheimen, southern Norway, based on BTS and DC resistivity tomography data, Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 56, 122-136, 10.1080/002919502760056459, 2002.

Isaksen, K., Heggem, E., Bakkehøi, S., Ødegård, R., Eiken, T., Etzelmüller, B., and Sollid, J.: Mountain permafrost and energy balance on Juvvasshøe, southern Norway, Eight International Conference on Permafrost, Zurich, Switzerland, 2003, 467-472,

Isaksen, K., Ødegård, R. S., Etzelmüller, B., Hilbich, C., Hauck, C., Farbrot, H., Eiken, T., Hygen, H. O., and Hipp, T. F.: Degrading Mountain Permafrost in Southern Norway: Spatial and Temporal Variability of Mean Ground Temperatures, 1999–2009, Permafrost and Periglacial Processes, 22, 361-377, 10.1002/ppp.728, 2011.

Litherland, T.: Snow Redistribution Modelling in Alpine Norway : Validation of SnowModel for a wet, high mountain climate, Master degree, Department of Geosciences, Faculty of Mathematics and Natural Sciences, University of Oslo, DUO, Digital publications at UiO, 101 pp., 2013.

Westermann, S., Schuler, T. V., Gisnås, K., and Etzelmüller, B.: Transient thermal modeling of permafrost conditions in Southern Norway, The Cryosphere, 7, 719-739, 10.5194/tc-7-719-2013, 2013.