Reviewer 2:

We are grateful for this detailed review which contains many thoughtful suggestions on how the manuscript can be improved. Before answering each of the raised points in detail, we would like to give an overview over the main points of criticism raised by the reviewer (as summarized by the last part of the review) and the corresponding changes to the manuscript:

The high-resolution model should allow for site specific questions to be addressed, demonstrating that the downscaling approach has the ability to simulate spatial variation of actual field measurements. If it can do that, then consequently, we would have a much improved understanding of future permafrost conditions and their variation along environmental gradients.

However, only active-layer conditions were considered from this perspective, but in a limited way. As a result, the paper is not convincing. This modeling is next-level, modeling a high degree of spatial variation, and therefore requires a great amount of field data to support the increased spatial resolution.

Considering the comments above and the model's high spatial resolution, there are at least three options available for a revised manuscript, and the discussion in any case should explicitly treat the spatial dynamic associated with the distinct vegetation classes and range of snow depths.

1. Focus on the active layer. Much of the permafrost related work in this study area treats active-layer conditions and relation to greenhouse gas dynamics, and the active-layer field data appear to be the most extensive in space and time. With more intensive activelayer data collected from grid cells along the transect this coming field season, there would be strong support for any conclusions and the benefit from downscaling would be clear.

2. Rather than positioning the paper as predictive, explicitly treat the paper as a sensitivity study, and in addition to future air temperature forcing under multiple predicted trajectories, explore the implications of changing snow depth regimes that may arise from long-term increasing or decreasing future winter precipitation and/or vegetation change.

3. Maintain stated objectives, and collect more ground temperature data in addition to active-layer data.

We comment on three main directions of criticism and explain, how these were improved in the revised version:

- **Lack of validation data for thaw depth along Zero-Line**: ZERO-line is the reference transect for surveying changes of a variety of different environmental factors, in particular vegetation, in the Zackenberg valley. This, however, implies that the area around ZERO-line should receive as little disturbance as possible. A regular survey of thaw depths, as they are conducted at the CALM sites, is not conducted to avoid such disturbances. However, we now present such validation for one point in time in the revised version (new Fig. 6). In August 2013, a survey of thaw depths has been performed at 100 points along the modeled part of ZERO-line, and we compare the resulting distributions of thaw depths. This comparison also demonstrates that thaw depths exceeding 1m depth occur for parts of ZERO-line, which the presented modeling approach delivers for the Fell class.

- **Lack of validation data for ground temperatures:** This is a highly important point raised by the reviewer. In particular regarding the validity of future predictions, we agree that it is highly important that the modeling can reproduce BOTH thaw depth and the ground temperatures from in-situ measured data for the present or past. In transient modeling approaches, as the one presented here, it is possible to fit measured thaw depths both by adjusting the summer temperature forcing (e.g. in our case by the NDVI-nt-relationship) and by adjusting the stratigraphy of ground properties (i.e in our case through stratigraphy assigned to the soil classes). The climate sensitivity of such different solutions would be different, though, and modeled future ALT may be strongly biased. However, these solutions would generally feature different annual average ground temperatures, and therefore can be distinguished by comparing to in-situ measurements. For this reason, we did not follow "Alternative 1" for improvement of the manuscript suggested by the Reviewer (i.e. "1. Focus on the active layer. Much of the permafrost related work in this study area treats active-layer conditions and relation to greenhouse gas dynamics, and the active-layer field data appear to be the most extensive in space and time. With more intensive activelayer data collected from grid cells along the transect this coming field season, there would be strong support for any conclusions and the benefit from downscaling would be clear.")
  We agree with the reviewer that the data basis for validation of ground temperatures was rather weak in the original version of the manuscript. In the revised version, we have employed annual average ground temperatures from 17 sites measured at at depths between 0.15 and 1.0 m. Although these measurements to not constitute a statistically representative sample of ZERO-line, they do show that the modeling is widely capable of reproducing the ground thermal regime within the Zackenberg valley (new Fig. 7). Thus, we present field data in the revised version that allow to assess the model performance both with respect to thaw depth and annual average ground temperature.

- **The "study design":** In our opinion, there exist very few published modeling studies on permafrost which can employ validation data sets from measurements that were specifically performed to provide a statistically sound validation of the modeling. This is particularly true for ground temperatures, which requires significant installations in order to obtain a sound record, e.g. annual average temperatures. Recently, Gisnås et al. (2014) showed for comparatively simple and homogeneous permafrost landscapes in Svalbard and Norway that on the order of 100 randomly distributed measurement points are required to obtain a representative sample of the distribution of ground temperatures. In Zackenberg (as well as in the large majority of high-arctic field sites), such a high number of ground temperature measurement sites does not exist (and cannot be set up in the future, given the strict environmental rules at the site). In the revised manuscript we use all available in-situ measurements in the Zackenberg valley to compare annual average ground temperatures, in total 17 sites. These sites do not at all follow a random sampling design and were set up in a variety of different locations for a variety of different purposes. For this reason (and also since the total number is too low to allow for representative sample, see above), it is not possible to "validate" the modeled distribution of ground temperatures. Furthermore, for these measurement sites, information on the snow conditions is generally not available so that it is also not possible to investigate the influence of the snow cover and the soil class on the ground thermal regime. However, the comparison clearly shows that variations in annual average ground temperatures are well in the range of the model results. As demonstrated by the newly added sensitivity analysis concerning snow depth and NDVI (Sect. 4.2), such variability is only reproduced in the model scheme, since spatially variable snow depths from MicroMet/SnowModel were used. In the revised version, the confidence in the presented model

scheme is based on: a) soil stratigraphies largely based on field measurements (Table 1), b) snow depth and melt-out data validated by field observations (Sect. 3.3, new Fig. 4), c) thaw depths validated for three of the soil classes with a multi-year time series (Fig. 5), d) comparison of a snap-shot of the thaw depth distribution for ZERO-line (new Fig. 6), e) comparison of 47 data points of annual average ground temperatures from in total 17 sites to the modeled range of ground temperatures, e) a sensitivity analysis showing that the variability of thaw depths is largely caused by ground and surface properties (see point c), while the variability of annual average ground temperatures is strongly related to snow depths (see point e).
(Reference: Gisnås, K., Westermann, S., Schuler, T. V., Litherland, T., Isakson, K., Boike, J., and Etzelmüller, B.: A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover, The Cryosphere, 8, 2063-2074, doi:10.5194/tc-8-2063-2014, 2014.)

General Comments

This paper sets out to his paper sets out to predict future permafrost conditions for a 4-km long transect of elevation and vegetation change through the Zackenberg valley along the ZERO-line. Considering the depth of research associated with Zackenberg Ecological Research Operations, with 9 or so related papers published on permafrost – carbon dynamics, this is a timely paper. From a methodological perspective, the stepwise downscaling approach taken in this paper enabling 10 m grid cells is intriguing, the inclusion of snow redistribution modeling is important, an consequently the overall approach is of great interest. Model parameterization approach is thoughtfully undertaken, clear, and essential physical elements that contribute to local ground temperature variation are considered appropriately. For example, the elegant use of the peak summer NDVI to compute nt. The overall structure of the paper is fine as long as it is aligned with the stated purpose which is to present simulations for the transect. Whereas simulations along the ice-edge to sea transect are not tied into the purpose and objectives, and modeling for that transect does not include the degree of downscaling used along the ZERO-line, any related sections should be removed and the title adjusted. There are, however, considerable problems with this paper that relate to field data, specifically how they were used to set model parameters, and the manner in which they are used to verify modeled contemporary conditions. As this downscaling approach predicts permafrost thermal evolution at a local scale, adequate field data are essential; however active-layer data are limited in extent with respect to the possible variation along the ZERO-line transect, no data exist to validate modeled 1-m ground temperatures, and shallow borehole data are not used convincingly to validate modeling. Rectification of these problems, which are outlined in Specific Comments, requires major revisions and additional field data, and as a result the paper should be rejected as it is. However, I strongly suggest that the authors submit a revised paper, as this work will make an important contribution with additional effort.

Specific Comments

Abstract

Clear abstract, provides a concise summary of the ZERO-line related work as presented in the text, but does not mention the ice edge to sea transect.

As suggested, the ice edge to sea transect is removed in the revised version

P 3908, L114-15: "permafrost remains thermally stable until 2100 in most model grid cells" should read "permafrost remains thermally stable until 2100 in all model grid cells". Fig4 shows10-m temperatures all remain below 0°C until 2100 and Figure 5 shows that active-layer thickness does not exceed 1.2 m (though fell is not included). Therefore permafrost as a condition remains stable, it is either there or is isn't. It is best not to confound increased activelayer thickness with "permafrost stability".

We have changed the statement, now only referring to the modeled temperatures (see next comment).

P 3908, L15: "Thaw threshold is exceeded" could mean a few different things. I think what you mean to day is that annual mean ground temperatures at 1-m depth were greater than 0°C.

We have changed the statement, now only referring to the modeled temperatures: "While ground temperatures at 10m depth remain below 0°C until 2100 in all model grid cells, positive annual average temperatures are modeled at 1m depth for a few years and model grid cells at the end of this century."

However, the paper should point out that active layer depth exceeded 1-m decades earlier in some vegetation classes when the annual mean 1-m temperature is still below 0°C. "Thaw threshold" is used in the introduction, results and discussion, each with a different implication. More appropriate terminology should be adopted.

We have removed all qualitative terminology from the manuscript and now only refer to the model results.

1. Introduction

Fine for the most part, except the ice edge to sea transect is not mentioned.

As suggested, the ice edge to sea transect is removed in the revised version

2. The Zackenberg site

Though there are already 7 figures, this paper requires a figure of the study area indicating the location of the model grid cells or at least the ZERO-line, in addition to topography, NDVI/vegetation classification, and locations for CALM grids and boreholes. It is difficult to assess the appropriateness of the field data without this critical study design information.

A map of the study area has been included, showing the location of Zackenberg in Greenland, as well as the NDVI map of the area around the modeled part of Zero-line, the location of the ZeroCalm sites and the location of ground temperature measurements employed for validation.

P391, L18-20: "This increasing thickness of the active layer represents only a fraction of the permafrost degradation taking the high ice contents (40–80%) into account." Is this % excess ice, gravimetric moisture content, or volumetric moisture content? Table 1 indicates most sites have a sand substrate with a 40% volumetric fraction of ice, which is saturation, so it appears that most

sites to not have excess ice. Consequently, increasing active layer thickness would be equivalent to the amount of permafrost degradation. In other words there would be little thaw strain. L18-20 would be true if there were excess ice and significant thaw strain.

The statement L18-20 has been removed.

P391, L21 to P3912, L11: Includes some repetition that can be tightened up. Write a subsection that details the field study design, including where field data was collected and why so that the reader can better assess their utility and appropriateness for model parameterization and validation. Specifically:

- **Why was a transect used?**
- **What are the site conditions around the boreholes, where are they located, and what depth**
- **intervals are temperatures measured at, what is the snow depth?**
- **Were any 1-m ground temperatures measured, and at what locations?**
- **What is the range of measured snow depths within each vegetation classification?**
- **What is the range of active-layer depths along the ZERO-line?**

The study site description has been modified and a new figure, which includes an NDVI map of the area around ZERO-line, the positions of the CALM sites and the positions of ground temperature measurements, has been inserted.

Clearly, the ZERO-line transect was used because data were available, but that reason alone is not adequate. The paper must demonstrate that use of a transect was a part of the study design because it allows testing of specific hypotheses related to differences in elevation and vegetation. Otherwise, one might argue that if elevation change was not important, it might have been better to choose completely random sites to parameterize and test the model. Is variation between vegetation classes greater than variation due to elevation change alone? In addition, active-layer soil moisture, an important control on ground temperatures, varies substantially along the ZEROline gradient, but its importance to variation of the ground thermal conditions in space and through time is not addressed. Very generally, ground temperatures are influenced by the surface conditions in a radius approximately equivalent to 3 times the depth the temperature is measured at. Thus with a 30-m radius, nearly a third of a CALM grid would influence the 10-m ground temperature. It would be very good to know if the boreholes were located in homogeneous settings and represented two distinct vegetation classes, or whether the boreholes represent composite classes. This also relates to snow conditions, as one site has "more regular snow" and the other is "at a site with a snowdrift". If the snow drift is not extensive, it may not have a substantial influence on the 10-m temperature. Further, as you rightly mention on P3924,L19-20, deeper ground temperatures are influenced by longer-term temperature forcing, and consequently there is a depth of zero annual amplitude (DZAA) above which ground temperatures reflect short-term trends. Temperature change below the DZZA is related to long-term climate forcing. Based upon the cold permafrost temperatures, it

would not be surprising if the DZAA was below 10-m, but the DZAA is not stated. Other studies that model permafrost thermal evolution, get around time lag effects by comparing a measured annual mean temperature profile with a predicted profile (e.g., Burn and Zhang, 2009, doi:10.1029/2008JF001087). As the model run begins in 1960, this kind of validation approach should be possible. Such validation would give much more strength to the model parameterization, rather than simply stating that the 10-m borehole temperatures are within the range predicted across all vegetation types. With such a high resolution model we should be able to see specific outcomes related to vegetation type and snow conditions, rather than generalized outcomes that one might expect from a non-downscaled modeling scheme.

We agree with the reviewer. However, since only two deep boreholes are available, any more detailed analysis is, in a statistical sense, not satisfactory, since coincidence cannot be ruled out both in case the model results agree and in case they do not. In the revised version, we present a more extensive in-situ data set of annual average ground temperatures at 1m depth instead.

1-m ground temperatures are modeled, so it seems necessary, or at least quite beneficial, that some near-surface ground temperature measurements should be compared against modelled results. The active-layer is treated this in this manner, and it should be done with 1-m ground temperatures. The 1-m data should represent all vegetation classes modelled. It would be even better if this shallow-ground temperature dataset also included a thin snow site and deep snow site. Then specific questions can be addressed such as: Is the model better at predicting sites next to sea level, or at 1040 m a.s.l.? Is elevation more important to ground temperature variation than snow depth overall?

In the revised version, we use the available in-situ data on annual average ground temperatures in the active layer to compare to the model results of ground temperature. Note that the modeling focuses only on on the lower part of ZERO-line, to an altitude of 200 m a.s.l. (Sects. 2 and 3).

Much of the novelty in the modeling approach comes from incorporation of MicroMet/SnowModel, and the agreement between predicted and observed snow depth observations is not quantified during validation. It seems that manual snow depths were determined within the CALM grids, and this implies that snow depth at fell was not assessed. In addition, as many CALM grids are established on relatively level topography, thus the effects larger snow drifts related to topography may not be adequately treated in the paper. Is there any significant difference in snow depth along the transect, or between ecosystem types?

In the revised version, we provide more quantitative validation for the performance of MicroMet/SnowModel, both for a point and for ZERO-line. ZeroCalm 2 actually features quite some topography, with the part classified as Dryas located near the crest of a hill, while the part classified as wetland is located at the foot of this hill.

Active-layer data are critical to this paper and they are derived solely from the CALM grids. This is problematic because the Fell classification, which may ultimately thaw more deeply, is not included, and because the within grid variation in a given year may or may not be representative of the variation along the transect. 4-km is not a great distance to cover, and it would be possible to determine a range of active-layer depths within several of the model grid cells for each of the four vegetation classes. The CALM data are used to validate the predicted active-layer increase over

time (as is done in the paper), but the additional sites would be used to test model's ability to predict the spatial component of active layer variation., i.e., are there any patterns associated with change along the transect? This test seems necessary due to the model's fine resolution and incorporation of processes that vary at the local scale.

Regular surveys of thaw depth are not available from ZERO-line (see introductory statement). In the revised version, we have included a comparison of modeled and measured distributions of thaw depth for August 2013, where a systematic survey was conducted.

3. Modeling tools

P3913, L15: Without a map of the study area, it is difficult to assess what part of the ZERO-line was modeled.

A new figure including a map of the study site has been inserted.

3.1 The permafrost model CryoGrid 2

P 3915, L15 to P3916,L6: Tighten up this paragraph. Figure 2, mentioned at the bottom of the paragraph shows the relation described by Equation 3 at the top, but what is the degree of explanation (the R2 , the p-value)?

The text has been formulated more concisely, and coefficients of correlation are provided.

P3917: The wetland class includes grassland and fen, but as fens are in wet depressions with a saturated active layer, they probably have a distinctly different climate-permafrost relation than the grasslands that are likely more well drained (based upon the study area descriptions given on P3912 this seems true). Unless shown otherwise, the fen is likely an end member of the spectrum of climate-permafrost relations and it should probably be treated separately. Similarly, Salix snow-bed delineates a class that is likely thermally distinct from others because of its relation to snow rather than soil moisture, so merging it with wetland is a problem. This uncertainty is glossed over to some degree, and the implications on spatial variation of ground thermal conditions are not discussed at the end of the paper. These merging problems arise, in part because the classes have to be differentiated according to NDVI values that are subsequently related to nt, but NDVI is also influenced by soil moisture. Soil moisture adjusted vegetation indices such as the SAVI (Huete, 1988, doi:10.1016/0034-4257(88)90106-X) or the new MAVI (Zhu et al., 2014, doi:10.1371/journal.pone.0102560) are probably better able to differentiate between vegetation classes, especially moist versus saturated soils, and it is likely that a strong relation can be established between either index and nt . These alternative indices may be worth exploring.

In the preparation of the manuscript, we have tried to use other indices, including SAVI, but the performance was worse than NDVI. While we fully agree with the reviewer's assessment of the situation concerning Salix and Fen/wetland, the presented modeling approach is at the moment not capable of reproducing such phenomena. As an example, both fens and grasslands are generally water-saturated near the surface. However, in some extreme situations, e.g. in dry summers following winters with little snow, the grasslands can become dry at the surface, while the fens areas remain saturated also in such situations. As a result, the sites feature different plant communities with slightly different NDVI values. However, our modeling employs constant soil water/ice contents over time, and typical values must be prescribed to achieve satisfactory results. These typical values are the same for both classes (though the extremes are not), which is why we do not distinguish between fens and grasslands.

Furthermore, we note that the functional relationship between NDVI and nt is at least partly caused by surface soil moisture content which causes different evapotranspiration and thus different surface temperatures (higher soil moisture leads to higher NDVI, but also to lower surface temperatures and thus lower nt).

P3918,L13-14: I assume depth to bedrock is not known from either borehole, otherwise it could be used here.

Correct.

3.3 Modelling snow distribution by MicroMet/SnowModel

As discussed above, it was surprising not to see a quantitative comparison of modeled and measured snow depths. At bare minimum, since the area is probably wind swept, give the shd for each vegetation class since the snow depths will be close to these except where drifting is related to topography. The reader has no idea what the normal range of snow depth is in this region.

Quantitative comparisons have been added to the revised version, both using point data and spatially distributed data on the snowmelt date along Zero-line.

How well was snow depth replicated outside of the CALM grids such as at topographic drifts like the one above one of the boreholes?

Systematic snow depth measurements do not exist from these sites, so a comparison cannot be performed. However, the comparison of the melt-out dates along ZERO-line which was added to the revised version of the manuscript provides insight in the spatial performance of SnowModel.

3.5 Permafrost simulations along the ice edge to sea transect

This section comes as a surprise as it is not mentioned in the introduction. There are no data to validate any results, and the simulation does not include downscaling.

We have removed this section from the revised version.

4 Model results

4.1 Comparison to field data

Given the scope of the modeling, the field data are largely inadequate.

1. **There seems to be no relevant field data for the fell classification. Consequently the validity of the total range of AL thicknesses and permafrost temperatures is questionable as the**

contribution by fell is unclear. Fell should be treated with caution and probably removed from analysis, validation, and prediction.

We agree, a major problem is that manual measurements of thaw depth are difficult and thus generally not available for fell due to large thaw depths and generally rocky/sandy ground. In the new comparison of thaw depths along ZeroLine, we note the presence of areas with thaw depths larger than 1m, which in the modeling are reproduced only by the fell class.

2. **Active-layer thickness; P3924, L7-18; Figure 3:**

- **It seems very counterintuitive that active-layer thickness, a function of summer conditions, is presented and discussed in relation to snow. If there is a clear physical link between antecedent snow depths and active-layer thickness in this study area then it should be established and discussed. This relation is not commonly observed, but it can occur.**

We agree, this is most likely a misunderstanding due to a misleading formulation in our manuscript. We have removed "and a realistic snow distribution can be assumed" from this section. We did not mean to make a causal connection between snow depths and thaw depth.

- **Figure 3 shows modeled and measured thaw depths are reasonably consistent over time within the CALM grid. An unanswered question is whether or not the downscaling approach was able to reproduce thaw depth variation along elevation and moisture gradients associated with the ZONE-line, controlling as best as is possible for vegetation/snow depth.**

Validation for thaw depth along Zero-line is presented in the revised version.

3. **Permafrost temperatures; P3924,L19 to P3925,L2:**

- **What kind of surface/vegetation/snow setting does each borehole represent? Is 10-m depth below the DZAA?**
- **Compare modelled results with mean ground temperature profiles from the deep boreholes. To say that the two point measurements are within a range of modeled temperatures is reasonable for coarse resolution models, but is not a strong argument for a model designed for high spatial resolution predictions. It would be better to be able to say that modeled output closely fit measured ground temperature data. How good is the agreement under specific conditions?**
- **Evolution of highly variable 1-m ground temperatures is presented in the next section, but no field data are presented for validation.**

In the revised version, we present much more extensive validation using near-surface ground temperatures. A more detailed characterization of the temperature profiles in the boreholes will be left to another study which will also include boreholes in other parts of Greenland.

4.2 Evolution of active layer thickness and ground temperatures

P3925,L9-11: A “weak spot” is not a description that adds to our understanding of the evolution of ground temperatures. See earlier comments about permafrost “stability” and increased active layer thickness and exceeding the “thawing threshold”. What are the vegetation/snow conditions associated with the “weak spots”? This should be answerable due to the nature of the model.

The formulation “weak spot” has been removed, we now only refer to the results of the modeled ground temperatures. An additional statement emphasizing the role of snow depth for such weak spots has been included: “These sites are characterized by above-average snow depths in the long-term average, which suggests that talik formation may be initiated at snowdrift sites.”

P3925,L20-22: No, it cannot be due to high ground ice content at the top of permafrost because all vegetation types were assigned the same volumetric ground ice content at the top of permafrost (see P3918,L10-13). It should be due to the saturated wetland active layer, that year over year has a much higher water content than the next wettest vegetation type Cassiope.

We agree and have changed this statement accordingly.

P3925,L24-25: Have you any field data to support these modeled results?

4.3 Ice edge to sea transect

Probably remove this section.

Done

5. Discussion

5.1 Model uncertainty

A big caveat is presented, that the results should be considered a first order approximation. Considering the light amount the field data used to parameterize and validate model results, this caveat should be emphasised in the abstract, and conclusions.

We have added a statement emphasizing the considerable uncertainty to the Conclusion: “Despite of the complex model approach, the projections of the future ground thermal regime are associated with considerable uncertainties related a variety of environmental factors which exemplifies the need for intensified process studies on permafrost environments.”

5.2 From model results to permafrost landscape development

Readers want to know if the scaling strategies from GCM to plot scale enable improved understanding about the spatial variation of thaw depth and ground temperatures, and their coevolution over time? Rather than producing overall ranges of modeled results, such as from a sensitivity study or from sub-grid estimates based upon a coarse model, readers are interested in knowing if the downscaling enables more specific questions to be addressed, such as:

- **How good was the temperature fit between snow drifts and bare-blown locations?**
- **Was the ground thermal regime in each vegetation class reproduced satisfactorily?**

- **Are there particular vegetation/snow conditions associated with the highest/lowest increase in active-layer thaw or ground temperature increase?**
- **Was snow depth the greatest driver of spatial variation of ground temperature or soil moisture, and at what scale?**

Unfortunately, the field study design prevents many of these questions from being tested, making it harder to demonstrate the added benefit of undertaking this downscaling approach.

In the revised version, we have added a sensitivity analysis (Sect. 4.2) for the two factors NDVI (which determines summer ground surface temperatures and the soil class) and snow depth, which sheds more light in these questions. The following text has been inserted: "In order to investigate the sources for this spatial variability, a sensitivity analysis was performed by running CryoGrid 2 for ZERO-line with a uniform ground stratigraphy and associated characteristic NDVI value (Sect. 3.1) for each of the four stratigraphic classes Fell, Dryas, Cassiope and wetland. This analysis suggests that snow depth has the largest effect on 1m ground temperatures, with a variability 3-5 times larger than the variability caused by ground and surface properties. On the other hand, modeled maximum thaw depths are much more influenced by ground and surface properties than by snow depths which only lead to differences on the order of 0.1 to 0.2m, compared to differences of more than 0.5m for different stratigraphic classes/NDVI values. A statistically significant correlation between NDVI values (and thus stratigraphic classes) and snow depths modeled by SnowModel/MicroMet does not exist in the employed data set."

P3930,L5-8: No evidence of excess ground ice has been presented, so this text is probably not relevant.

This paragraph addresses permafrost modeling in general, not only in Zackenberg. The resolution of a model approach is a highly relevant issue for phenomena like thermokarst which are related to thawing of excess ground ice. Therefore, we did not change this paragraph.

P3930,L14-15: "the spatial variability of ground temperatures to a large extent caused by spatially variable snow depths". No analyses are presented in a figure or statistically to support this statement, and there is no line in Model results connecting ground temperature variation with snow depth. However, Figure 3 suggests that soil moisture is probably important to thaw depth and it is therefore probably important to overall ground temperature variation.

A sensitivity analysis has been added, both for ground temperatures and thaw depth, which answers these questions.

P3931: The model approach should also capture within ecosystem class differences related to spatial variation along the ZERO-line.Simulated ground temperatures are not adequately compared with borehole data.

See above

As discussed earlier, '"Weak" spots' is not informative. "Onset of permafrost thaw" is not indicated by when the annual mean temperature at 1 m is above 0°C, it happens earlier when the increasing active layer thickness includes that depth.

We agree and only state the findings related to the model in the revised version.

The final conclusion is not clear, and perhaps this is due to wording. Why is it important that the range of temperatures due to spatial change is near to the amount of increased ground temperature change? Do you mean to say something like: The projected increase of ground temperatures until the end of this century varies spatially within a few kilometers according to ecosystem class, ranging overall from about 3 to 5°C. This conclusion regarding ground temperatures does not have strong support because of the validation approach which does not include fell, and the lack of clarity of the representativeness of the borehole sites. Also, is this 1-m or 10-m depth? Rather than K, °C should be used as it is throughout the paper.

We have added a clarifying statement: "Therefore, both modeling and in-situ monitoring of the ground thermal regime may provide an incomplete assessment of present and future permafrost thaw if they are restricted to one or a few points within an area."

This model predicted a variable ground thermal regime, but it is not clear how that spatial variation is divided amongst the various ecosystem classes.

In the revised version, a sensitivity analysis has been added that can answer this question at least from the point of view of modeling.

Technical Corrections

P3908,L20: Change "discussed in the" to "discussed regarding the".

Done

L22: Delete "as well as in the light of".

Done

L24 to P3909,Ll1: Clarify. Near-surface permafrost degrades because of significant deepening of the active layer, so provide threshold depth that should be referred to here to qualify a decrease in extent?

"near-surface" has been removed – in GCM projections, "deep" permafrost generally does not exist due to the limited depth of the employed soil domain.

P3909,L5: change "assessment for" to "assessment of".

Done

L8: Delete"e.g."

We provide examples of such studies and other studies using RCM output exist, so that the wording is correct.

P3910,L11: Change "Zhang et al., 2013" to "Zhang et al., 2012, 2013".

Done

P3912,L20: Change "are dominating" to "dominate".

Done

P3913,L4: Change "sample" to "determine".

Done

L15-16: Change to "The different parts of the scheme and their interplay are described as follows".

Done

P3914,L13: Change "The shape" to" The characteristic curve".

Replaced by "This curve…"

P3916,L7: change "accumulated" to "determined".

Done

L9-13: Change to "Whereas the acquisition date is close to the annual maximum NDVI values, it represents a single point in the time and there is strong seasonal and interannual variability in plant growth and consequent evolution of NDVI values (Tamstorf et al., 2007). While this error source is hard to quantify"

Done

P3918,L1 change "oriented at" to "derived from".

Done

L8-9: change to "The volumetric organic material contents are low in all classes (5 % or less)".

Done

P3919,L27: Change "too high albedo values" to "albedo values that are too high".

Done

P3924,L24: Change "the meteorological" to "at the meteorological".

Done

P3927,L12: Change "thus e.g. reproduce" to "thus reproducing".

Done

P3928,L11: Change "it must remain unclear" to "it remains unclear".

Done

P3930,L11: Change “could capture” to “captured”.

Done

Figure 4: Delete the number 38 on the figure.

This was due to the page layout of the Discussion journal, will be removed in the final version

Figure 5: Heath is classified as Cassiope in the rest of the text.

changed

Figure 6: Delete the number 40 on the figure.

This was due to the page layout of the Discussion journal, will be removed in the final version

Figure 7: Probably do not need this figure.

Removed

Thank you very much for your effort!

On behalf of the authors,

Sebastian Westermann