Authors Response:

Two reviews for our manuscript are available. Following the suggestions and criticism of the reviewers, we have compiled a revised version and additional Supplementary Material.

In the following, we provide the replies to both reviews (Sect. 1, 2), followed by a summary of the major changes (Sect. 3). We then proceed with the revised version of the manuscript with changes marked in bold (Sect. 4), and the new Supplementary Material (Sect. 5).

On behalf of all co-authors,

Sebastian Westermann

1. Reply to Reviewer 1

We are grateful to the reviewer for the thoughtful comments and suggestions to our manuscript. We have compiled a revised version and in the following provide a point-by-point reply to all issues raised. The reviewer's comments appear in bold font, our replies in normal font, and changes to the manuscript are in italics.

In this study the authors present a remote sensing based scheme for transient modelling of the ground surface regime together with the previously published numerical model CryoGrid2. The scheme is applied over a large area in the Lena River Delta (LRD), Siberia. Forcing datasets at 1km and weekly resolution are derived from MODIS LST, MODIS SCE, GlobSnow SWE plus meteo fields from ERA-Interim reanalysis. Spatially distributed ground properties are based on geomorphological observations and mapping drawing on previous studies in the region. Results are compared to insitu observations of ground temperatures from boreholes, CALM active layer depths and measurements from the Samoylov Island Permafrost observatory. The authors conclude that comparison to in-situ measurements shows that the scheme is capable of estimating the thermal state of permafrost and its time evolution in the LRD.

This paper is a further contribution to the work of using remote sensing data together with numerical models (eg. Westermann 2015) which I think is a very interesting and promising approach to large area and/or operational assessments. The paper is well written with a clear methodology, presentation of results and critical discussion. The authors acknowledge shortcomings of the approach such as dependency upon a well estimated snow density and difficulty in applying in heterogeneous terrain due to coarse scale of the LST data. I have one main comment with respect to the forcing timeseries, other comments are reasonably minor.

Comments:

1. **P8 Section 3.3: In the merged LST /reanalysis product, 2m air temperature and LST are merged. I think it would be helpful to add some discussion of how comparable surface air temperature and LST are and how this is expected to vary under both different atmospheric and surface cover conditions. The most obvious example is when a snowcover is present and air temperature and snow surface temperature can differ strongly. This reference (Gallo et al 2011) would probably be useful: http://dx.doi.org/10.1175/2010JAMC2460.1. This study from Raleigh et al. http://dx.doi.org/10.1002/2013WR013958 suggests that the 2metre dewpoint temp (also**

available from ERA-Interim) is perhaps a better approximation of snow surface temperature than 2m air temperature. What kind of biases can be expected by forcing the upper boundary condition of surface temperature with a 2m air temperature field? Or are these different forcings treated differently by the model?

We have taken up this comment in Sect. 3.3 and in the Discussion, Sect. 5.1.1. In Sect. 3.3, we have explained how we handle situations when positive temperatures of the surface forcing (which can occur as a result of admixing of air temperatures) occur for still snow-covered ground:

"During cloudy skies, differences between air and surface temperatures are strongly reduced compared to clear-sky conditions (Gallo et al., 2011), so that air temperatures can be regarded an adequate proxy when MODIS LST is not available due to cloud cover. For melting snow, surface temperatures are confined to the melting point of ice, while air temperatures can be positive. Positive values of the surface temperature forcing are therefore set to 0°C when a snow cover is present."

In-situ measurements on Samoylov Island indicate that air temperatures are a relatively good proxy for snow surface temperatures in winter, most likely because the ground heat flux from the refreezing active layer and the cooling permafrost is a substantial source of energy to the surface which prevents strong near-surface temperature inversions. In Sect. 5.1.1 we write:

"Based on in-situ measurements, Raleigh et al. (2013) suggest that for snow-covered ground dew point temperatures are a better approximation for surface temperatures compared to air temperatures at standard height. However, observations on Samoylov Island suggest only a small offset between snow surface and air temperatures, with the difference increasing from near zero in early winter to about 1° C in late winter (Table 3, Langer et al., 2011b). The reason for this is most likely that the ground heat flux is a strong heat source especially in early winter (Langer et al., 2011b) which warms the surface and thus prevents formation of a strong near-surface inversion. Therefore, we consider air temperatures an adequate proxy for snow surface temperatures in the LRD, but dew point temperatures should clearly be considered for gap-filling in the snow-covered season in future studies."

2. **P10 l20-22: is this spatial variability due to residual snow patches? Perhaps state the cause here.**

In July, residual snow patches do not occur on Samoylov Island, the snow pack has fully melted by at latest mid June – the spatial variability is caused by different surface cover and soil moisture conditions. Using a thermal camera, Langer et al. (2010) showed that the spatial differences in polygonal tundra can be up to 10K for single scenes, but become much smaller for temporal averages over longer periods. However, a residual net difference between the point measurements used for comparison and the larger-scale MODIS LST values cannot be excluded. Text changed to:

"However, surface temperatures can feature a strong spatial variability during summer due to differences in surface cover and soil moisture conditions…"

3. **P7 l3 + 33 on line 3 you say "extensive set of observations available" whereas on l33 you say "which temporally /spatially distributed sets are not available" - are these statements contradictory?**

An extensive data set is available from Samoylov Island, but there is no data set covering the entire Lena River Delta (which we refer to in the second statement). We have made this clearer in the text, l. 33 now reads:

"Therefore, the snow … is a highly crucial parameter for which spatially or temporally distributed data sets covering the entire LRD are not available. However, an extensive set of measurements from polygonal tundra on Samoylov Island suggests …"

Can you describe the snow density data briefly in Section 2.2, particularly at which times of year these measurements were made.

We have inserted a statement in Sect. 2.2:

"*In addition, a spatially distributed survey of snow depths and densities (216 points in polygonal tundra) was conducted in early spring 2008 (25 April to 2 May) before the onset of snowmelt (Boike et al., 2013*)."

In addition, the range of snow depths obtained from the spatial survey has been added to Fig. 3.

4. **Fig 6: Is there an offset in your measurements as looks like in Fig 6 that zero curtain is occurring 0.5deg or so below the 0degC point.**

Yes, there is a slight drift of the sensor in the later years (visible from 2010), with the zero curtain occurring at about -0.2°C instead of 0°C. A statement has been added to the figure caption.

5. **Fig 6: can you explain why there is no zero curtain at phase transition from ice to water in spring/summer in the wet polygon? Would you expect this?**

The zero curtain effect is a result of a two-sided freezing front which only occurs in fall for permafrost ground (the top freezes in fall and the ground below the active layer is permanently frozen). In this case, the temperatures at both sides of the still unfrozen domain are confined to 0°C due to the freezing of soil water at the freeze front. As a consequence, temperatures inside this domain quickly reach 0°C as well. As the freeze fronts progress slowly due to the considerable amounts of latent heat provided by the phase change of the water, this zero curtain state can last for several weeks and only ends when the two freeze fronts meet each other, which is generally followed by rapid cooling of the then frozen soil. In spring, the ground thaws from top down and only one freeze/thaw front exists. In this case, there is always a temperature gradient both above and below the progressing thaw front, so that temperatures in a certain depth/grid cell are never confined to 0°C for extended periods. For seasonally frozen ground, the situation is opposite and the zero curtain effect occurs in spring, when the seasonally frozen layer thaws from top down and from the bottom up, again resulting in two freeze fronts. In fall, only on freeze front exists, corresponding to the freezing of the ground from top down, so that no zero curtain effect occurs.

Technical issues:

1. **p6 l29: add terms in brackets after items in text so that equation is more easily understood.**

done

2. **P7 l6: ...LRD for which... → ...LRD which..**

   Changed to "for which we define"

3. **P9 l27-29: I think it is more common to use term "layers" when talking about vertical discretisation of model units?**

   Changed to "layers"

4. **P10 l16: Figure 2 seems to lose most bar elements upon printing (not digital form). Perhaps my printer issue - but check this.**

   We have tried printing Fig. 2, but did not encounter any problems.

5. **P10 l24: "well suited as input for ground thermal modelling" - qualify this statement with something like ", at least in homogeneous terrain".**

"at least in homogeneous terrain" added

6. **P11 l7 over a an → over an.**

Done, thanks!

7. **p11 l10 repeated word "cloudiness" → you mean snow?**

Yes, we mean snow cover. Thanks!

8. **P14 l32-33: qualify statement with something like 'in homogeneous terrain'.**

We agree, this statement is too general. In the revised version, we have made clear that it only applies to our study area in the LRD, and only to homogeneous terrain. The sentence now reads:

"*We conclude that surface temperatures synthesized from MODIS LST and ERA-interim reanalysis are an adequate choice for the purpose of ground thermal modeling in the LRD, at least in homogeneous terrain. However, it may introduce a slight cold-bias in modeled ground temperatures*."

9. P17 l11: had → hand.

Done, thanks!

10. p18 l8: ares → area.

Done, thanks!

New reference:

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The surface energy balance of a polygonal tundra site in northern Siberia Part 2: Winter, Cryosphere, 5, 509–524, 2011b.

2. Reply to Reviewer 2

We thank the reviewer for the critical thoughts on the manuscript and for raising important points which have led to changes to the manuscript, including additional Supplementary material.

A main line of criticism is that GlobSnow SWE cannot be a suitable data set for the purpose of ground thermal modeling in the Lena River Delta, based on the characteristics of our study area and the few published GlobSnow validation studies. The critical issue is how the studies mentioned by the reviewer can be related to the conditions in the Lena River Delta. Before we provide a point-by-point reply to the review, we would like to clearly state our assessment of these studies and evaluate their suitability for characterizing the performance of GlobSow for our study area (the Lena River Delta) and study period (2000-2014 with particular emphasis on the period 2002-2014 for which validation data for ground temperatures are available).

a) Luojus et al., 2010 and Takala et al., 2011: Both studies present validation information for Eurasia using the INTAS-SCCONE data set (Kitav et al., 2002). The comparison to this data set represents (to our best knowledge) the only systematic GlobSnow validation that includes our study area in a geographical sense. However, Fig. 2 in Takala et al., 2013, suggests that only a small fraction of the data set is obtained from Arctic tundra sites (with a rough count, we came to <50/1264 sites, i.e. <5%), and even fewer (<15/1264, i.e. <1.5%) are located in Northeast Siberian lowland tundra with climate and landscape characteristics that are at least somewhat similar to the Lena River Delta. The overwhelming majority of the sites, on the other hand, is located in the Boreal Forest and in steppe environments, where completely different snow pack properties must be assumed (Takala et al, 2011, Derksen et al., 2012). We emphasize that the comparison to the INTAS-SCCONE data is a highly meaningful benchmark on a continental scale, at least for SSM/I-based GlobSnow SWE retrievals. However, the study only presents an evaluation of the entire data set und information on subregions is not given. Therefore, we cannot see why the results, especially the absolute values for RMSE and bias presented in Table 2 (Takala et al., 2011), should be valid also for small subregions like the Lena River Delta, in particular when shallow Arctic tundra snow is strongly underrepresented in the data set. Furthermore, it should be noted that the INTAS-SCCONE data set covers the years 1978-2000, for which the GlobSnow data are compiled with older generations of passive microwave sensors compared to our study period.

An important point is that SWE in the Lena River Delta is generally low enough so that saturation effects do not influence SWE retrievals: starting at about 120-150mm average SWE (Fig. 1 in Luojus et al., 2010), the GlobSnow retrievals become biased to too low values, as brightness temperatures are affected by radiation emitted within the snow pack (and not only by scattering of radiation emitted from the ground below). This effect is a consequence of the physics of snow grain - radiation interactions and the results can therefore be expected valid for all microwave remote sensors. For SWE values of 60mm and less (as in the LRD, Fig. 5f, our study), the relationship between measurements and GlobSnow retrievals is on average linear. Therefore, the retrievals must be considered more reliable in this range, although GlobSnow to a certain extent overestimates measured SWE values (Takala et al., 2011). However, these low SWE values in the data set (roughly 30% is SWE<50mm, Fig. 7b, Takala et al., 2011) could in many cases represent data from areas other than Arctic tundra, so that the bias is once again difficult to interpret with respect to our study area.

b) Takala et al., 2011: In addition to the INTAS-SCCONE data, Takala et al., 2011, present validation for Finland and Canada. The data set from Finland is from a region with a high density of ground stations so that an improved performance can be expected compared to regions without a dense station network. For this reason, a comparison is challenging due to the lack of ground stations in the vicinity of the Lena River Delta. However, the saturation effect at higher SWE values is again evident, confirming this feature also for areas with high density of ground stations. Furthermore, a comparison to extensive in-situ data sets from Northern Canada is presented. These data were mainly obtained during a snowmobile traverse in 2007 (Derksen et al., 2009), largely in tundra areas at latitudes between approx. 64 and 68° N which is an environment in many aspects comparable to our study area (located approx 72-74° N). Only average results for the entire data set are presented in Table 3 (which the reviewer explicitly refers to), showing a mean SWE of 120mm, an RMSE of 47mm and a mean negative bias of 36mm. Takala et al., 2011, comment on this: "The relatively high uncertainty over tundra regions (Table 3) is likely driven by three issues: the extremely sparse network of surface climate stations across the Canadian sub-Arctic, the complex microwave emission from lake rich snow covered tundra (see Derksen et al., 2009), and the extremely heterogeneous tundra snow cover which complicates the determination of 'ground-truth' SWE at coarse spatial resolutions."
The first point "sparse station network" is at least partly true also for the Lena River Delta, but we note that the WMO station in Tiksi is in a distance of 50 (eastern part) to 200 km (western part), much closer than the closest station for a large part of the N Canadian traverse (Fig. 5a, Takala et al, 2011). Furthermore, the environmental and climatic conditions at Tiksi are similar to the Lena River Delta, so that the snow pack properties inferred for Tiksi are very likely a good representation for the Lena River Delta, which is favorable for SWE retrievals in our study area (see Derksen et al., 2012 for a discussion of the role of snow pack properties in SWE retrievals). Also the second point, "high water body fraction", is clearly applicable to the Lena River Delta, as pointed out by the reviewer. Using Landsat (Schneider et al., 2009) and MODIS (MODIS water mask) based land cover classifications, we estimate the water fraction in the interior of the Lena River Delta (the part for which the modeling was performed) between 12 and 30% in 25km EASE grid cells, with a single grid cell in the Eastern Delta reaching 37% (of which more than half is estimated to be river arms, see below). Almost three quarters of the grid cells feature water fractions of less than 20%. However, the character of the water bodies is very different to the ones in N Canada: themokarst lakes and river arms dominate in our study area, while the traverse crossed the Canadian Shield where lakes are mostly a result of glacial erosion. The track of the

traverse provided in Derksen et al., 2009, suggests that unforested, ground-ice-rich, flat lowland areas only made up an insignificant portion of the traverse: most of the area E of Great Bear Lake is characterized as "continuous permafrost with low ground ice content and thin overburden or exposed bedrock" in the IPA permafrost and ground ice map. Hereby, an important difference regarding SWE retrievals with passive microwave could be that many of the thermokarst lakes in the Lena River Delta are shallow and can even freeze to the bottom in winter (Schwamborn et al., 2012, Antonova et al., 2016), while the Canadian lakes are in general deeper. With respect to microwave emission, this could be an important difference: microwave emissions become more similar to land areas, although the emission characteristics of fully frozen water bodies are not yet entirely clarified (Gunn et al., 2011).

Furthermore, the Lena River features a very low winter discharge, as much of the catchment is in the continuous permafrost zone. Despite of recent increases, we estimate the winter discharge to be only about 10% of the average summer discharge (Fig. 2 in Yang et al., 2002), and large river areas visible as water in summer-derived satellite imagery (which are hence classified as water in the above mentioned classifications) fall dry in winter, which will decrease the water fraction in particular in the central and eastern part of the delta considerably. Furthermore, shallow river arms will freeze to the bottom, similar as the above mentioned thermokarst lakes, so that we expect the true "open water" fraction relevant for microwave brightness temperatures in winter to be significantly lower than the open water fractions obtained from summer imagery (see above) suggest. As a consequence, we argue, that the high summer water fraction in some parts of the LRD (although it may affect passive microwave retrievals in particular in fall) is not a priori an exclusion criterion for GlobSnow in the LRD, especially since the comparison with in-situ data set in the relatively water-body-rich area around Samoylov Island suggests a satisfactory performance.

Third, the pronounced spatial variability at scales smaller than 25km is certainly an issue also in the Lena River Delta, although the landscape is generally flat (compare the images of the borehole sites in the Supplement to the revised version of the manuscript) in most parts and large snow drifts only occur in localized spots, such as edges of islands or thermokarst gullies, which we do not target with our modeling. For the extensive N Canada data set, spatial variability at small scales will only affect the RMSE and not the bias when comparing GlobSnow retrievals to small-scale in-situ data. Fig. 10 in Takala et al., 2011, is clear evidence that the spatial variability is extreme at least in some parts of the Canadian study area, with a spread from 40 to more than 200mm SWE. Such strong differences at scales of less than 25km could to a large part explain the high RMSE value found in the N Canadian data set. However, we find it encouraging with respect to our study that GlobSnow in the presented cases can indeed capture a SWE value in the center of the SWE distribution (Fig. 10, Takala et al., 2011), which would be a satisfactory representation for the average snow conditions in the grid cell.

Finally, the largest difference between the N Canada data set and the Lena River Delta is the strong difference in the absolute values of SWE, with on average 120mm instead of 40-60mm. For an average of 120mm and values of more than 200mm occurring regularly (Fig. 10), the considerable negative bias of 36 mm is an indication that the GlobSnow SWE retrievals could partly be affected by saturation effects (see above) in this area, which is highly unlikely for the much lower SWE values in the Lena River Delta. Furthermore, we do not see why absolute uncertainties (both bias and RMSE) from the N Canada study could simply be assigned to the Lena River Delta (as suggested by the reviewer), despite the large differences in absolute values for SWE itself. As evident for Fig. 1 in Luojus et al., 2010, higher SWE values are associated with higher absolute uncertainties, while lower SWE values have a lower absolute uncertainty. We therefore argue that, if at all, relative errors from the N Canada data set should be employed when assessing the possible performance of GlobSnow SWE in the Lena River Delta. In the revised version, we show that the uncertainties found for the N Canda data set can then be reconciled with the comparison for Samoylov Island (see below for details).

c) **Derksen et al., 2012:** This study is not a validation of the operational GlobSnow SWE product itself, but investigates an important aspect of the retrieval algorithm: it evaluates the landcover dependence of microwave emission, distinguishing between open tundra areas, forest and lakes. We have carefully evaluated the information provided in the publication with respect to the conditions in the LRD.

The study area of Derksen et al., 2012, is located at 58-59° N near Churchill, Canada. The mean annual air temperature in the Churchill area is approx. -6.5°C (Fig. 3a, Zhang et al., 2012), while it is around -12.5°C for Samoylov Island in the Lena River Delta (Boike et al., 2013), a significant difference, which most certainly causes differences in the freezing behavior of lakes and possibly the snow pack properties. Moreover, the radiation regime during winter is necessarily different due the latitudinal difference, with polar night conditions dominating in the Lena River Delta, while Churchill is located several 100km south of the polar circle. This factor may influence the snow pack properties and lake freezing. No information on the lakes (depth, origin) is provided, which makes it difficult to compare to thermokarst lakes in our study area.

In-situ observations of passive microwave brightness temperatures and snow pack properties were conducted for different landcover types at sites located approx. 5-15km from the coast of Hudson Bay (Fig. 1, Derksen et al., 2012), showing a strong landcover dependence. We note that the 25km EASE grid cell #2 (Fig. 1, Derksen et al., 2012), in which all in-situ measurements were conducted, is located directly at the coast (with even a small ocean fraction included), and the "ocean overspill problem" (see Other remarks by the reviewer, point 1) is not mentioned or taken into account in this study. However, only grid cell #2 facilitates, in our opinion, a direct comparison between in-situ and satellite brightness temperature measurements: the "ground truth information" for the other grid cells is synthesized from a landcover classification, assuming that the in-situ measurements conducted at localized sites in the vicinity of the coast can deliver unique values for "lake snow depth", "open tundra snow depth" and "forest snow depth" (and for all the other snow pack and lake ice properties) that are representative also for sites more than 50km inland. In this procedure, possible spatial differences in precipitation, temperature and wind speed (controlling snow redistribution) between coast and inland are not mentioned or taken into account (although they may well exist, see Fig. 3a/b in Zhang et al., 2012 for temperature and precipitation).

Summarizing our assessment, we conclude that a) the two study areas feature somewhat different characteristics with respect to landcover and climate, and it is not entirely clear in how far this affects the transferability of the findings; b) no in-situ measurements are available from the grid cells other than one coast grid cell. The ground truth information is instead based on land-cover-weighted average of in-situ measurements taken near the coast.

In this light, we further evaluate Table 9 and Fig. 12 in Derksen et al., 2012, which are mentioned by the reviewer. Table 9 and 10 display absolute RMSE and bias for SWE retrievals for all grid cells, i.e. the data set affected by issue b) mentioned above. RMSE values strongly increase if grid cells other than # 2 (where the in-situ measurements were conducted) are evaluated, while there is only little effect of different lake fractions (<25% and <50% are distinguished with grid cell maximum lake fraction 73%, Table 3, note that the overwhelming majority of grid cells in the Lena River Delta would fall in the low lake fraction category). In our opinion, it can at least not be ruled out that the results in Tables 9 and 10 are affected by systematic differences in snow cover between grid cells on a coast-inland gradient that are not captured by the study design. Such differences could at least contrbute to the RMSE increase for grid cells other than #2. In addition, the results are strongly affected by the presence of forest as a third landcover class, featuring completely different snow depths and densities (Fig. 3, Derksen et al., 2012), which is not an issue in the Lena River Delta. This effect might even override the uncertainty caused by water bodies, at least for grid cells with lake fractions <25%. Finally, it is not clear what the results exactly mean for operational GlobSnow SWE retrievals, which are also controlled by the quality of the station-interpolated background fields. We conclude by noting that Derksen et al., 2012, do not share the reviewer's pessimistic view with respect to passive microwave retrievals. They write: "The results in Table 9 are encouraging with respect to passive microwave SWE retrievals, however, this represents an ideal scenario in which snow cover characteristics were thoroughly measured through the complete winter season and available for input to the forward modeling component of the retrieval. In order to test retrieval performance under less idealized circumstances, the retrieval simulations were re-run with only a single land use tile. This better replicates an operational scenario where information from only a single snow survey or weather station would be available. Because these observations are often located in open areas (i.e. adjacent to airports) only the snow measurements from the open site were used as model inputs. Table 10 provides a summary of these retrievals; the accuracy is not influenced appreciably (and actually improved in some cases) compared to the simulations using the full set of snow observations. This suggests that snow information for a single land cover class can still result in useful retrievals. "

Fig. 12 in Derksen et al., 2012, shows the spatial variability of SWE within and between landcover classes (based on the in-situ measurements in grid cell #2), in conjunction with the single value retrieved by the satellite, which we presume corresponds to grid cell #2, containing

62% open, 27% forest and 11% lake (Table 3). When roughly computing the grid cell average with the above landcover fractions, we find that the satellite retrievals seem to match quite well, at maximum overestimating the in-situ value with about 25-30% and thus far from the 100% error mentioned by the reviewer. Moreover, presence of forest significantly complicates the picture compared to the Lena River Delta. Considering the high spatial variability within and between the classes and the fact that systematic sampling over the entire 25km grid cell has not been performed, it is, in our opinion, not entirely clear if this moderate mismatch is due to the scaling of the in-situ data or the satellite retrievals. The obvious decrease in satellite-derived SWE in March (that clearly must be associated with a higher uncertainty) might be explained by formation of ice lenses, as described for March in the paper (an effect that generally does not occur in the colder winter climate of the LRD). As a consequence, we do not agree with the reviewer's statement regarding Fig. 12 ("even more revealing (over 100% error; Figure 12) when examining errors on a monthly basis"), at least when regarding grid cell average SWE. It is clear that mismatches of more than 100% occur for point measurements due to the strong spatial variability of SWE, but this is an inherent issue in coarse-scale products such as GlobSnow SWE, which aim at delivering grid cell averages.

In the following, we provide point-by-point replies to all issues raised by the reviewer. In the end, we summarize the major changes to the manuscript that were inserted following the review. The reviewer's comments are provided in bold font, our replies in normal font and changes to the manuscript in italics:

This paper presents an approach to map the spatial distribution of ground temperatures and thaw depths using a 1D transient ground thermal model (CryoGrid 2). The model uses remote sensing derived surface temperature (MODIS 1 km complemented by 2-m air temperature from atmospheric reanalysis ERA-Interim 0.75 deg. grid spacing) and snow depth obtained from the GlobSnow snow water equivalent (SWE; 25 km grid spacing) product as forcing data. The study builds on the earlier work of Langer et al. (2013), moving the application of the CryoGrid 2 model from the local scale (station on Samoylov Island) to the regional scale by including the entire Lena River Delta (LRD). From Figure 1, one notices that the LRD is covered by many small lakes and branches of the Lena River (i.e. a large freshwater fraction).

We agree. We now provide quantitative information of the water fractions of the GlobSnow grid cells in the Lena River Delta. More than 70% of the employed grid cells have open water fractions of less than 20%, and only a single grid cell has a high fraction of more than 35%, most of which are river arms. Grid cells near the coast with even higher open water fractions do not contribute to our model forcing data set, these areas have been excluded. As detailed in the above assessment of Takala et al., 2011, these open water fraction correspond to the summer state. As the winter discharge of the Lena River is reduced to approx. 10% of the summer flow (Yang et al., 2002), large river areas must be expected to fall dry during winter. Furthermore, many shallow water bodies freeze to the bottom in winter in the Lena River Delta (see Schwamborn et al., 2002, and Anotonova et al., 2016, for field data), so that significantly lower “winter open water fraction” can be expected compared to the summer values, at least after mid winter. Therefore, we do not see a reason why the performance of GlobSnow should be significantly worse than in lake-rich areas in N Canada (Takala et al., 2011).

This is a complex area to study using coarse resolution passive microwave satellite data (or derived products) due to the large sub-grid scale variability within pixels,

Regarding the spatial variability of SWE, our scheme does not aim for directly capturing the small-scale variability of snow depths within 1km grid cells, but a significant scale mismatch remains between the 25km GlobSnow pixels and the 1km model resolution (which we discuss in detail in Sect. 5.2), although it is unlikely to assume that abrupt changes in average SWE occur in the flat landscape of the LRD. Nevertheless, this mismatch cannot be resolved, and is therefore likely to constitute an important source of uncertainty. However, when we compare the model output to point measurements of the ground thermal regime, we find it encouraging that the model results in most cases can fit the measurements, which could be considered an indication that both model and measurements reproduce “average conditions” in the relatively flat tundra landscape of the Lena River Delta. In the revised version, we have added Supplementary material showing images of the borehole sites, which demonstrates that the landscape is indeed rather homogeneous in the vicinity of the boreholes.

notably due to the presence of water bodies, which introduces significant uncertainty in SWE estimates (GlobSnow or other satellite-based SWE products).

We partly agree. It is evident that water bodies have an effect on the brightness temperatures obtained by passive microwave sensors, leading to problems in SWE products based on passive microwave data. However, the GlobSnow SWE retrieval features a data assimilation procedure, which also takes in-situ measurements at WMO stations into account. The station data are interpolated in space to provide a background or “a priori” field for the assimilation of the satellite data. This background field is then weighted against SWE information derived from the passive microwave sensor. Takala et al., 2011, state: “A basic feature of the algorithm is that if the sensitivity of space-borne radiometer observations to SWE is assessed to be close to zero (…), the weight of the radiometer measurements on producing the ‘assimilated SWE’ approaches zero (…). The higher the estimated sensitivity of TB to SWE, the higher the weight given to the radiometer data. Thus, the weight of the radiometer data varies both temporally and spatially in order to provide a maximum likelihood estimate of SWE.” It is beyond the scope of our study to evaluate in detail how GlobSnow retrievals are obtained in the Lena River Delta. However, it is clear that the GlobSnow algorithm is capable of handling situations with reduced reliability of the passive microwave retrievals, other than algorithms entirely based on satellite retrievals. Therefore, studies focusing solely on the effect of lakes on passive microwave brightness temperatures provide an incomplete picture on how the GlobSnow algorithm will handle these situations.

This issue has been recognized by the group who originally developed the GlobSnow product (reported in Takala et al., 2011) and at least one of its members in a latter publication (Derksen et al., 2012). Takala et al. (2011; Table 3) report mean bias errors of -36 mm and RMSE of 47 mm for a tundra area with small water bodies.

Table 3 in Takala et al. (2011) is also clear evidence that the absolute values of SWE in the Canadian data set are considerably higher than in our study area, 120mm vs. 35-60mm. Therefore, we do not find it adequate to assign <u>absolute errors</u> from this data set to the much lower SWE values in the Lena River Delta (see our detailed assessment above). In fact, if one simply did that without further thinking, negative SWE values would occur regularly in the Lena Delta, which is impossible both in reality and in the GlobSnow processing algorithm. Instead, if we do assume that error estimates can be transferred from the N Canada data set to the Lena River Delta, we suggest that <u>relative errors</u> are more appropriate.

On Samoylov, we have an average snow depth of ca. 16 cm at 225kg/m^3 (Fig. 5f) which corresponds to a SWE of about 40mm. Assuming a relative error similar to the Canadian study (RMSE of 47mm at 120mm average SWE, i.e. a 40% relative error), we obtain an absolute error of 15mm SWE (i.e. 6-7cm snow) at Samoylov. Considering our comparison of snow depth model forcing to in-situ data (Fig. 3, our study), this RMSE appears indeed realistic: using only the non-zero in-situ snow depth measurements, we obtain an RMSE of 6 cm for the snow depths, while the average bias of +1.5cm is of opposite sign and considerably smaller than for the N Canadian data set (-36mm at 120mm total SWE corresponds to ca. -5 cm snow depth, if scaled to the average SWE values from Samoylov Island). This underestimation in the N Canada data set could at least partly be due to saturation effects at SWE values >150mm, which is not an issue in the LRD. Although only a single point measurement is available, this comparison is an indication that the performance of GlobSnow-derived SWE in the LRD is at least not worse than in other arctic areas.

Derksen et al. (2012) also show that SWE retrieval errors can be large (see Table 9 and Figure 12 of this publication) from passive microwave data, even more revealing (over 100% error; Figure 12) when examining errors on a monthly basis.

Please see our detailed discussion of Derksen et al. (2012) above. The study showcases the land-cover-dependence of snow depth/properties and the effect on passive microwave emissions, but we do not think that the findings can be regarded as evidence that GlobSnow is entirely inadequate for the LRD. If we re-perform the above scaling exercise for the values given in Table

9 and Fig. 12, we once again arrive at an order of magnitude for the uncertainty that is well comparable to the performance for Samoylov Island (see previous point).

As indicated by Takala et al. (2011): "Additionally, the consideration or compensation for the effect of (frozen) lakes requires further study and algorithm development work."

Our model scheme is designed in a way that improved future SWE products can be directly ingested, which consequently has the potential to improve the modeled ground thermal regime. In the revised version of the manuscript we write in Sect. 5.1: "*In the future, enhanced SWE retrieval algorithms taking waster bodies explicitly into account (e.g. Lemmetyinen et al., 2011) may become available.*"

The authors of the present manuscript state, in Langer et al. (2013), that: "The thermal state of permafrost is reproduced with an uncertainty of about ±2.5 °C with a SWE accuracy of about ±10 mm.

This statement is based on a sensitivity analysis at a single point which applies a constant bias to the entire time series. In this study, however, we provide a characterization of uncertainties based on a comparison to in-situ measurements. As the number of validation sites is limited, we have added the following statement to the Conclusion of the revised version of the manuscript: *"However, due to the relatively small sample of validation sites, this accuracy assessment must be considered preliminary. "*

This is still below the performance that can be reached with a realistic LST accuracy of about ±2 °C. However, a much lower SWE accuracy level (±40 mm) must be considered in regions with sparse weather stations (Luojus et al., 2010) and when field measurements are not available for calibration. Our results show that realistic permafrost simulations with a transient heat transfer model would be almost impossible with such low accuracies in the SWE forcing.

Here, Langer et al., 2013, refers to the accuracy stated in Luojus et al., 2010, which is once again intended as a global benchmark, including many different conditions (see above). It is not intended as a benchmark on the regional scale. In the context of this statement, it represents a worst case that one can possibly encounter if the model scheme of Langer et al., 2013, (with model parameters of Samoylov Island) is applied to a random point in the Northern Hemisphere. However, in the present study we are interested in the performance in the Lena River Delta itself, and the above considerations are clear indication that an absolute error of 40mm is not adequate.

In contrast to the permafrost temperatures, the thaw depths are found to be more or less independent from the SWE accuracy. However, this might be different in regions where the permafrost temperature is already close to the freezing point as observed by Åkerman & Johansson (2008). In any case, the impact of snow on the active layer dynamics can be very complex and dependent on regional factors (Zhang, 2005). The performed sensitivity study

demonstrates that a highly accurate snow cover forcing is crucial for reliable permafrost modeling."

This statement once again refers to application at a random point. From Langer et al., 2013, it is not really clear what "highly accurate snow cover forcing" means, and it is necessary to evaluate this point in further regional studies – this is exactly our intention with the present manuscript, and we will evaluate the scheme further in other case studies in the future.

Given: 1) the above statement by the authors in a previous paper;

As detailed above, the present study is a follow-up of Langer et al., 2013, and in many aspects represents an update. Some deviations in the findings are therefore not surprising, although we argue that the SWE threshold estimates can largely be reconciled with GlobSnow error estimates from Takala et al., 2011, if relative instead of absolute errors are employed. The statements from Langer et al., 2013, as picked by the reviewer, clearly refer to application at a random point anywhere in the world, not to the relatively similar landscape of the Lena River Delta, which even includes the study site of Langer et al., 2013.

2) the known retrieval errors in similar regions reported by the developers of GlobSnow SWE;

As detailed above, we do not consider absolute errors derived from continental-scale studies applicable for the LRD. If they were, the satisfactory agreement at Samoylov Island (Fig. 3, our study) could only be explained by fortuitous coincidence. If we assume relative errors instead, the reported errors from "similar" regions can largely be reconciled with our comparison for Samoylov Island.

and 3) the lack of validation of snow depth (derived from GlobSnow SWE with density values of 200-250 kg m-3) over a larger area (transects) than just the small island of Samoylov (located to the south of the LRD),

This relates to "Other remarks, #4: **How much confidence should we have in the snow depth map of Figure 5"**? We agree, this is a very relevant question, in particular whether the slight increase of snow depths/SWE from E to W is real. In the revised version of the manuscript, we provide Supplementary Material in which we compare our GlobSnow-based forcing data with Canadian Meteorological Centre (CMC) Snow Depth Analysis Data (Brasnett, 1999; Brown & Brasnett, 2015). The CMC product provides SWE values at a spatial resolution of 24km, comparable to GlobSnow SWE, and has been used as a reference product to evaluate global snow retrievals (Frei et al., 2012). The CMC product does not employ satellite-derived passive microwave data for deriving SWE, and is hence completely unaffected by water bodies. Instead, a background field of SWE is calculated from snowfall in an atmospheric circulation model, which is subsequently updated by assimilating in-situ measurements from WMO ground stations. Both data sets use WMO data from Tiksi which could affect the absolute values, but not the spatial patterns, as no other WMO station is located close-by to the W or N of the Lena River Delta. For the comparison, we have used the CMC monthly SWE data set for the period 2004 to 2013, corresponding to the period displayed in Fig. 5 of our manuscript. The result of the spatial comparison is shown in Fig. R1. Both products show a similar spatial pattern and absolute values mostly agree to within 10mm, with CMC generally showing lower values than GlobSnow. When interpolated to 1km scale, a significant correlation between the data sets is found (r^2=0.71). The coarse-scale pattern in the LRD is further backed up by winter precipitation from the ERA-interim reanalysis (Fig. R2). Despite of the coarse resolution and the insufficient representation of the coastline, there is a clear W-E gradient over land areas in the LRD, with lowest values occurring in the SE edge, similar to GlobSnow SWE.

While these comparisons to independent model data sets do not constitute a validation of GlobSnow SWE in a strict sense, they indicate that the large-scale pattern derived from GlobSnow is not an artefact of the GlobSnow SWE retrieval, but related due to regional trends in winter precipitation.

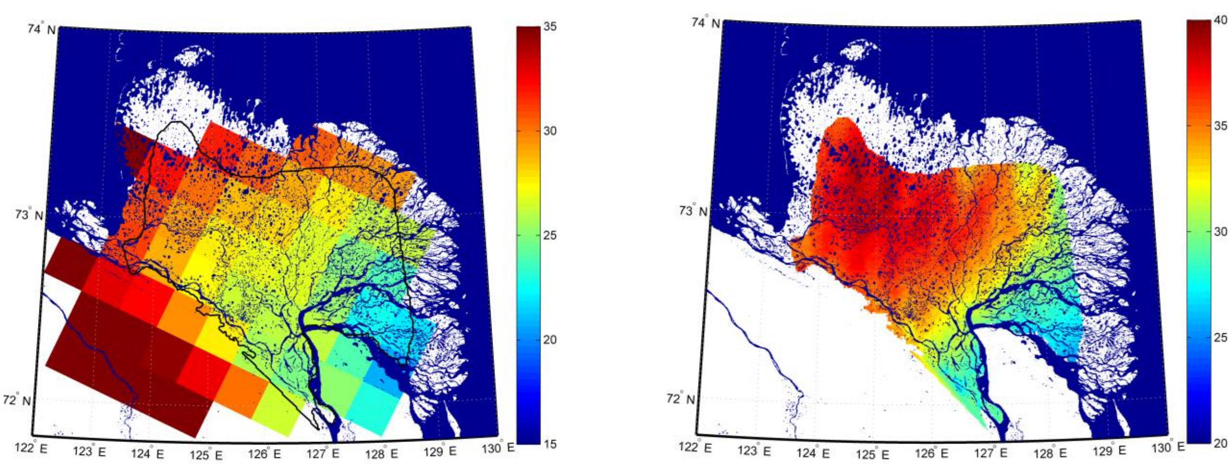

Fig. R1: Average SWE in the months October to June from 2004 to 2013. Left: Canadian Meteorological Centre (CMC) Snow Depth Analysis Data, 24 km resolution; the black line corresponds to the outline of our model domain. Right: CryoGrid 2 forcing data (1km resolution) based on GlobSnow SWE and MODIS snow cover. Note the offset of 5 mm between the color scales.

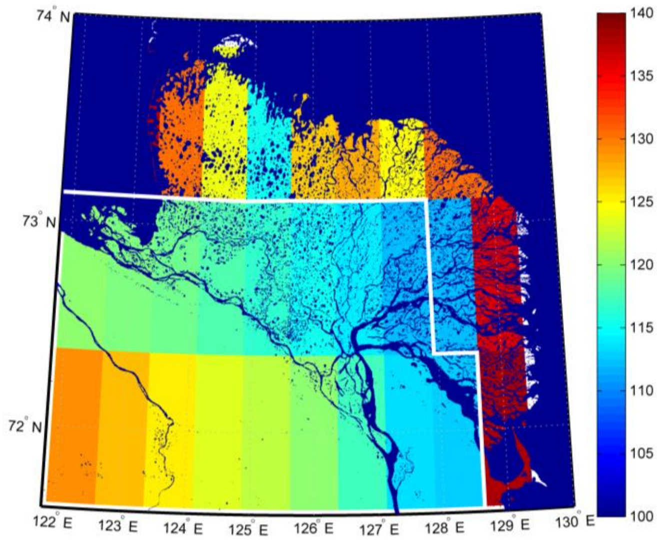

Fig. R2: Annual averages of the total precipitation ([mm]) falling at 2m-air temperatures of less than 0°C for the months October to June for 2004-2013, based on the ERA-interim reanalysis at 0.75° resolution (values represented as grid cells). The land-sea mask is indicated by a white line.

I am afraid to say that the manuscript submitted is not acceptable for publication in The Cryosphere. In fact, I am quite concerned by the fact that the authors missed the publication of Takala et al. (2011) which is the key paper reporting uncertainties of the GlobSnow SWE product. It is important to read and cite others who work in similar areas or at least with similar data sets, and who have reported uncertainties in the forcing variables used by CryoGrid 2.

We have provided a detailed assessment of Takala et al., 2011, and further literature on the GlobSnow algorithm (see Major changes to the manuscript). We point out that Takala et al., 2013, is clearly dedicated to a continental-scale assessment, and caution is warranted when assigning the error estimates to regional scales. We believe that absolute error values taken from this study are inappropriate for the Lena River Delta (see above).

Other remarks:

1. The authors do not seem to be aware that the SSM/I footprints for the 19 GHz and 37 GHz frequency brightness temperature channels are in the order of 70x45 km and 38x30 km, respectively. These brightness temperature measurements are then interpolated into a 25x25 km grid which is then used for SWE retrieval in GlobSnow. Therefore, although the authors masked some areas along the coast, the ocean "overspill" problem within the footprints is a larger problem than reported herein.

We have generally kept a distance of 20-30km from the coast, so that contributions from the ocean are at least not large, considering the footprint sizes. At the western end of the modeled domain, the distance to the coast is less in order to be able to include the validation sites “Olenyoskaya channel mouth” and “Turakh Island”. We have stated in the revised version that a higher uncertainty is likely for these sites.

2. The large fraction of the landscape covered by lakes/river channels represents the largest uncertainty in GlobSnow SWE values. The authors need to read further on this topic in order to better understand the limitations of GlobSnow SWE and, perhaps, search for other products (satellite or reanalysis, including assemble) that could be considered in a new manuscript submission to The Cryosphere or another journal.

Please see our detailed assessment above.

3. Boike et al. (2013) is given as the reference for the snow depth and density values of Samoylov Island. However, I personally browsed this paper to find that there are mismatches between values reported in Table 5 and Figure 6 of that paper and the values reported in Figure 3 (and the text) of the present manuscript. I am not sure how, as a reviewer, I can reconcile the two sources. The range (and maximum) of measured snow depths in Boike et al. (2013) do not always match those of this paper. For example, in winter 2004 (a high snow year), the maximum snow depth found in Table 5 and Figure 6 of Boike et al. (2013) is 56 cm while that plotted in the graph of Figure 3 of this paper is at a value of about 47 cm. This is only one of several examples.

In Fig. 3, we have applied a running average filter with a weekly window, corresponding to the temporal resolution of our forcing data. This is now stated in the figure caption of the revised version, and explains deviations from Table 5, where the absolute maximum recorded is displayed: differences of several centimeters can be easily explained by snowfall during periods of low wind speeds that gets quickly removed or compressed by wind action afterwards. Fig. 6 in Boike et al., 2013, refers to a different data set, i.e. spatially distributed measurements taken at one point in time. We have now added this spatial survey as a data point to Fig. 3 in the revised version manuscript, which indicates that GlobSnow SWE is an adequate representation.

4. How much confidence should we have in the snow depth map of Figure 5 and the ground temperature (1-m depth) map of Figure 11, given that snow density comes from Samoylov Island only and that there is a large degree of uncertainty in GlobSnow SWE retrievals over complex (lake-rich) areas such as the LRD?

We have provided a comparison to an independent SWE data (Figs. R1, R2), confirming the SWE general pattern in the Lena River Delta. Please see our detailed response above.

As shown in Figures 6-8, winter temperatures are significantly underestimated in wintertime by the model (up to 8°C, most frequently by 3-4°C). Of course, taking the

average of all years combined reduces the error reported (1-1.5°C given in the Abstract), but the errors are larger when inspecting each individual year.

We do not agree with the reviewer's statement "winter temperatures are significantly underestimated in wintertime by the model (up to 8°C, most frequently by 3-4°C)". We presume that the reviewer refers to periods when a) thermokarst development at the borehole sites was obvious (Sardagk and Kurunghak), or b) a change in the snow regime due to the building of a new research station had taken place (Samoylov). The timing of these events was described rather qualitatively in the original version of the manuscript. In the revised version, we have now clearly marked the affected periods in Figs. 7 and 8 and stated in the figure captions that they should not be employed for comparison. We also provide images showing the thermokarst development around the boreholes (as well as new buildings around the borehole on Samoylov Island) in the new Supplement

To facilitate a more quantitative evaluation of model results for the annual cycle, we have added a new Fig. 9, which displays a scatter plot for monthly averages for all boreholes. In agreement with the comparison of yearly averages, we find that the model results have a slight cold-bias of -0.9°C and an RMSE of 1.1°C (for a snow density of 225 kg/m^3). In the revised version, we write in Sect. 4.2.1:

"A comparison of monthly averages for all five boreholes is shown in Fig. 9. For a snow density of 225 kg m^{-3}, the model results feature an RMSE of 1.1°C and an average bias of -0.9°C, mainly due to underestimation of measured values during the summer and fall seasons. For a snow density of 200 kg m^{-3}, the model bias is on average positive (+0.8°C), but the RMSE is increased (1.6°C). The model performance is worst for the highest snow density (RMSE 2.1°C, bias -2.1°C). If the Samoylov Island borehole (for which the ground stratigraphy was adjusted, see above) is removed, the model performance for the best-fitting snow density of 225 kg m^{-3} remains largely unchanged (RMSE 1.2°C, bias -0.9°C)."

5. The scaling issue between point (single station measurement(s)) and large satellite pixels should not be ignored throughout the manuscript.

We agree, and we have mentioned the scale mismatch several times in the manuscript. Our in-situ measurements are generally located at points that the installation team deemed to be representative for the larger-scale environment, which may partly explain the satisfactory agreement with the model results. In the revised version, we provide Supplementary material with images of the borehole sites, which show the flat and relatively homogeneous landscape around the borehole sites.

Major Changes to the manuscript in response to the reviewer:

Sect. 3.3 Model forcing data:

"GlobSnow SWE (Daily L3A SWE, level 2.0) data are derived from passive microwave remote sensors which are not affected by clouds, so that a gap-free daily time series is in principle available for entire model period from 2000 to 2014. The GlobSnow processing algorithm is based on a data assimilation procedure, which also takes in-situ measurements at WMO (World Meteorological Organization) stations into account (Takala et al., 2011). For the LRD, the closest station is located at Tiksi, about 50 km to the E, while the closest stations to the W are several hundred kilometers away. The station measurements are interpolated in space to obtain a SWE background field which is then weighted against SWE information derived from the passive microwave sensor by means of forward modeling of snowpack microwave emission using the HUT model (Pulliainen et al., 1999). The SWE values in the LRD are typically below the critical threshold of about 150mm (see Sect. 4.1) above which SWE can no longer reliably derived from passive microwave retrievals (Takala et al., 2011). On the other hand, SWE retrieval is hampered for shallow snow cover and for wet melting snow, so that the start and the end of the snow season is not well covered by GlobSnow. Furthermore, water bodies constitute a major error source (e.g. Derksen et al., 2012) which generally leads to underestimation of SWE, in particular when the ice cover is thin (Lemmetyinen et al., 2011). Due to admixing of microwave radiation emitted from the ocean, the number of SWE retrievals is very small or even zero in the coastal areas of the LRD, so that almost half of the area of the LRD could not be included in the modeling. The boundary of the final model domain was finally chosen so that all validation sites (Fig. 1) are located within. In a few cases (in particular the sites AN, Tu and OM, Fig. 1), the available SWE data had to be extrapolated by about one grid cell or 25 km, which seems adequate considering the smoothness of the remote sensing derived SWE field in the LRD."

Sect. 4.1.1 Comparison to in-situ data:

"(...) At least some of the observed interannual differences are reproduced in the remote sensing-derived snow product, e.g. the above-average snow depths in winter 2003/04 and the below-average snow depths in 2012/13 (the latter was qualitatively noted by the station personnel, pers. comm., N. Bornemann). For values with non-zero snow depth, the model forcing (using a snow density of 225 kg m^{-3}) features an RMSE of about 0.06m, and a slight positive bias of 0.015m. The average snow depth in polygonal tundra (obtained by a spatially distributed survey, Boike et al., 2013) in early spring 2008 is slightly higher than both point measurements from the snow depth sensor and the model forcing. However, the difference is only about 0.05m for the model forcing with snow density 225 kg m^{-3}, well within the observed spatial variability of snow depths (Fig. 3)."

Sect. 5.1 Model forcing

“As demonstrated by Langer et al. (2013), snow depth and snow thermal properties are crucial factors for correctly modeling ground temperatures in the LRD. In this light, the coarsely resolved estimates of GlobSnow SWE must be considered the key source of uncertainty for the thermal modeling.

– The performance of GlobSnow SWE has been evaluated on continental scales by comparison to systematic in-situ data sets (Luojus et al., 2010; Takala et al., 2011). For Eurasia, surveys spanning the entire snow season (Kitaev et al., 2002) were compared from 1980 to 2000. For shallow snow (approx. SWE<60mm), GlobSnow SWE tends to overestimate observed values slightly, but the relationship between measurements and GlobSnow retrievals is on average linear. When SWE exceeds approx. 100mm, the GlobSnow algorithm tends to underestimate measured SWE, and for values larger than 150mm the signal from passive microwave retrievals saturates and SWE can no longer reliably be detected (Takala et al., 2011). For the LRD, both in-situ measurements and GlobSnow values indicate that SWE is generally below this critical threshold so that saturation effects most likely do not play a role for the uncertainty. The Eurasia data set is strongly biased towards sites in steppe environments and the boreal forest zone (where SWE retrieval is strongly affected by the canopy, e.g. Derksen et al., 2012), while northern tundra areas with characteristics similar to the LRD are strongly undersampled. A more representative data set is available from an extensive transect across Northern Canada (Derksen et al., 2009), for which comparison of GlobSnow SWE retrievals yielded an RMSE of 47mm and an average bias of -36mm. The average SWE of 120mm (Takala et al., 2011) was significantly larger than in the LRD, so that it is not meaningful to transfer the absolute uncertainties. When using relative uncertainties, on the other hand, we arrive at a similar RMSE as for the comparison of the time series on Samoylov Island (0.06m, see Sect. 4.1.1): for N Canada, a relative RMSE of around 40% was found, which corresponds to an absolute RMSE of 0.065m in snow depth, when scaled to the average of around 0.16m on Samoylov Island (Fig. 5f). Although the character of the two data sets differs (spatial transect vs. multi-year point measurement), the good agreement is an indication that the GlobSnow performance in the LRD could be similar to N Canada. We emphasize that the RMSE corresponds to undirected fluctuations around the average value which have much less influence on the modeled average gorund thermal regime (Figs. 12, 13) than a systematic bias. For a systematic bias of 10mm SWE (applied uniformly to the entire time series), the sensitivity study by Langer et al. (2013) suggests a deviation of approx. 2.5°C of the modeled average ground temperatures at 2.5m depth. In this study, we find an agreement within 1 to 1.5°C with borehole temperatures for multi-year averages at similar depths (Sect. 4.2.1). Therefore, a SWE bias of more than 10mm seems unlikely for the borehole sites, although modeled ground temperatures are also influenced by the ground stratigraphy (Table 3).

– Water bodies strongly affect microwave emission of the ground, which is known to lead to underestimation of SWE in passive microwave-based retrievals (Rees et al., 2006; Lemmetyinen et al., 2011). For the above mentioned N Canada data set, water bodies might explain the significant bias of 36mm (Takala et al., 2011), but the average values (120mm) are also sufficiently high that saturation effects (Luojus et al., 2010) are likely to contribute to the bias. In the LRD, water bodies are abundant features (Fig. 1), so that GlobSnow retrievals are likely to be affected. Using a Landsat (Schneider et al., 2009) and MODIS (MODIS water mask) based land cover classifications, we estimate the water fraction in the employed 25 km grid cells in the Lena River Delta to be between 12 and 30%, with a single grid cell in the E part reaching 37% (of which more than half is estimated to be river arms, see below). Almost three quarters of the grid cells feature water fractions of less than 20%. However, relatively shallow themokarst lakes dominate in the LRD, which at least partly freeze to the bottom in winter (Schwamborn et al., 2002a; Antonova et al., 2016), so that microwave emission becomes similar to land areas, although in particular the wave-length dependency of the effect may be complex (Gunn et al., 2011). Furthermore, the winter discharge of the Lena River is very low compared to other northern rivers, as the catchment is largely located in the continuous permafrost zone (Yang et al., 2002). We estimate the winter discharge to be only about 10% of summer averages (Fig. 2 in Yang et al., 2002), and large river areas identified as water in summer-derived satellite imagery must fall dry in winter, which decreases the water fraction in the central and eastern part of the delta (where the water fractions are highest) considerably. Furthermore, also shallow river arms and even coast-near areas of the Laptev Sea (Eicken et al., 2005) freeze to the bottom, so that we expect the true "open water" fraction relevant for microwave emission in winter to be significantly lower than the open water fractions obtained from summer imagery (see above) suggest. This is corroborated by the comparison to in-situ measurements for Samoylov Island (Fig. 3) situated in a relatively water-body-rich area where we find a satisfactory performace for GlobSnow. The largest impact on SWE retrievals is most likely during lake freezing and snow cover build-up in fall, when GlobSnow SWE retrievals must be considered highly uncertain. In the future enhanced SWE retrieval algorithms taking water bodies explicity into account (e.g. Lemmetyinen et al., 2011) may become available."

Additional References:

Antonova, S., Duguay, C., Kääb, A., Heim, B., Westermann, S., Langer, M., & Boike, J. (2016). Monitoring bedfast ice in lakes of the Lena River Delta using TerraSAR-X backscatter and coherence time series, Remote Sensing, 8(11), 903; doi:10.3390/rs8110903.

Brasnett, Bruce. 1999. A Global Analysis of Snow Depth for Numerical Weather Prediction. Journal of Applied Meteorology 38: 726–740.

Brown, Ross D. and Bruce Brasnett. 2015, updated annually. Canadian Meteorological Centre (CMC) Daily Snow Depth Analysis Data. © Environment Canada, 2010. Boulder, Colorado USA: National Snow and Ice Data Center.

Derksen, C., Silis, A., Sturm, M., Holmgren, J., Liston, G. E., Huntington, H., & Solie, D. (2009). Northwest Territories and Nunavut snow characteristics from a subarctic traverse: Implications for passive microwave remote sensing. Journal of Hydrometeorology, 10(2), 448-463.

Frei, A., Tedesco, M., Lee, S., Foster, J., Hall, D. K., Kelly, R., & Robinson, D. A. (2012). A review of global satellite-derived snow products. Advances in Space Research, 50(8), 1007-1029.

Gunn, G. E., Duguay, C. R., Derksen, C., Lemmetyinen, J., and Toose, P.: Evaluation of the HUT modified snow emission model over lake ice using airborne passive microwave measurements, Remote sensing of Environment, 115, 233–244, 2011.

Lemmetyinen, J., Kontu, A., Kärnä, J.-P., Vehviläinen, J., Takala, M., and Pulliainen, J.: Correcting for the influence of frozen lakes in satellite microwave radiometer observations through application of a microwave emission model, Remote Sensing of Environment, 115, 3695–3706, 2011.

Yang, D., D. L. Kane, L. Hinzman, X. Zhang, T. Zhang, & H. Ye (2002). Siberian Lena River hydrologic regime and recent change, J. Geophys. Res., 107(D23), 4694, doi:10.1029/2002JD002542.

Zhang, Y., Li, J., Wang, X., Chen, W., Sladen, W., Dyke, L., ... & Kowalchuk, S. (2012). Modelling and mapping permafrost at high spatial resolution in Wapusk National Park, Hudson Bay Lowlands, Canadian Journal of Earth Sciences, 49(8), 925-937.

3. Summary of major changes

a) We now provide more direct information for the periods the borehole data are suitable for comparison to model results. Periods not found suitable are marked in Figs. 7 and 8. For Samoylov, it is evident that the extension of the research station in summer 2012 led to new buildings and structures in the direct vicinity, so that snow depths drastically increased in the following winters. For Kurungnakh and Sardakh, we employ the observed timing of pond formation around the borehole to restrict the time series. Pictures showing the boreholes and thermokarst formation are provided in Sect. 2 of the Supplementary Material.

b) In response to reviewer 2, we have added a new Fig. 9 and a quantitative assessment of modeled monthly average temperatures. Using all boreholes, we find a mean bias of -0.9°C and an RMSE of 1.1°C.

c) Following reviewer 1, we have added information on the relationship between air temperatures and surface temperatures in Sects. 3.3 and 5.1.

d) Following reviewer 2, we have added information and an extended discussion on GlobSnow SWE in Sect. 3.3 and 5.1.

e) In response to reviewer 2, we provide independent snow products based on atmospheric model output for the LRD to back up the distribution of snow depths obtained from GlobSnow. As this study is primarily about permafrost and not a validation study of remote sensing products, we are of the opinion that this detailed assessment is beyond its scope and should not become part of the manuscript. Nevertheless, the reservations on the usability of GlobSnow outlined by reviewer 2 are comprehensible. We have therefore compiled a Supplement to the manuscript, in which the comparison is included. A short summary of the findings is provided in Sect. 4.1.2 of the manuscript.

4. Revised version with changes in bold

Transient modeling of the ground thermal conditions using satellite data in the Lena River Delta, Siberia

Sebastian Westermann[1], Maria Peter[1,2,*], Moritz Langer[3,2], Georg Schwamborn[2], Lutz Schirrmeister[2], Bernd Etzelmüller[1], and Julia Boike[2]

[1]Department of Geosciences, University of Oslo, P.O. Box 1047, Blindern, 0316 Oslo, Norway
[2]Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Telegrafenberg A43, 14473 Potsdam, Germany
[3]Department of Geography, Humboldt-University, Unter den Linden 6, 10099 Berlin, Germany
[*]now at: Université du Québec à Trois-Rivières, 3351 boul. des Forges, C.P. 500, Trois-Rivières G9A 5H7, Canada

Correspondence to: Sebastian Westermann
(sebastian.westermann@geo.uio.no)

Abstract. Permafrost is a sensitive element of the cryosphere, but operational monitoring of the ground thermal conditions on large spatial scales is still lacking. Here, we demonstrate a remote-sensing based scheme that is capable of estimating the transient evolution of ground temperatures and active layer thickness by means of the ground thermal model CryoGrid 2. The scheme is applied to an area of approx. 16 000 km^2 in the Lena River Delta in NE Siberia for a period of 14 years. The forcing data sets at 1 km spatial and weekly temporal resolution are synthesized from satellite products (MODIS Land Surface Temperature, MODIS Snow Extent, GlobSnow Snow Water Equivalent) and fields of meteorological variables from the ERA-interim reanalysis. To assign spatially distributed ground thermal properties, a stratigraphic classification based on geomorphological observations and mapping is constructed which accounts for the large-scale patterns of sediment types, ground ice and surface properties in the Lena River Delta.

A comparison of the model forcing to in-situ measurements on Samoylov Island in the southern part of the study area yields a satisfactory agreement both for surface temperature, snow depth and timing of the onset and termination of the winter snow cover. The model results are compared to observations of ground temperatures and thaw depths at nine sites in in the Lena River Delta which suggests that thaw depths are in most cases reproduced to within 0.1 m or less and multi-year averages of ground temperatures within 1 to 1.5 $^\circ\mathrm{C}$. **Comparison of monthly average temperatures at depths of 2 to 3 m in five boreholes yielded an RMSE of 1.1 $^\circ\mathrm{C}$ and a bias of -0.9 $^\circ\mathrm{C}$ for the model results.** The warmest ground temperatures are calculated for grid cells close to the main river channels in the south, as well as areas with sandy sediments and low organic and ice contents in the central delta, where also the largest thaw depths occur. On the other hand, the coldest temperatures are modeled for the eastern part, an area with low surface temperatures and snow depths. The lowest thaw depths are modeled for Yedoma permafrost featuring very high ground ice and soil organic contents in the southern parts of the delta.

The comparison to in-situ observations indicates that the satellite-based model scheme is generally capable of estimating the thermal state of permafrost and its time evolution in the Lena River Delta. The approach could hence be a first step towards remote detection of ground thermal conditions and active layer thickness in permafrost areas.

1 Introduction

Permafrost is an important element of the terrestrial cryosphere which is likely to undergo major transformations in a warming climate in the 21st century. At present, near-surface permafrost covers about a quarter of the land area of the Northern Hemisphere, but future projections with Earth System Models (ESMs) suggest a reduction between 30 and 70% until 2100, depending on the applied anthropogenic emission scenario (e.g. Lawrence et al., 2012). Observations of the ground thermal state are evidence that the ground is already warming in many permafrost areas (Romanovsky et al., 2010) and near-surface permafrost is in the process of disappearing from peripheral areas (e.g. Borge et al., 2016). In-situ monitoring efforts are coordinated world-wide within the Global Terrestrial Network for Permafrost (GTN-P, www.gtnp.org, Burgess et al., 2000) which is comprised of two components: (1) the Circumpolar Active Layer Monitoring (CALM) with measurements of active layer thickness at about 250 sites, and (2) the Thermal State of Permafrost (TSP) in which ground temperatures are measured in over 1000 boreholes with depths ranging from a few to more than 100 m.

While GTN-P can deliver high-quality direct observations of permafrost state variables, TSP and CALM sites represent point measurements on spatial scales of 100 m and less. Transferring this knowledge to larger regions is hampered by the considerable spatial variability of the ground thermal regime (which limits the representativeness of a measurement) and the strong concentration of TSP and CALM sites in a few regions, while vast permafrost areas are not at all covered (Biskaborn et al., 2015).

A possibility to infer ground temperatures on large spatial scales is the use of grid-based models that use meteorological data as forcing. Spatially distributed permafrost modeling was e.g. demonstrated by Zhang et al. (2013) and Westermann et al. (2013) forced by interpolations of meteorological measurements, or by Jafarov et al. (2012) and Fiddes et al. (2015) by downscaled atmospheric model data. Remote sensing data sets have been extensively used to indirectly infer the ground thermal state through surface observations, e.g. occurrence and evolution of thermokarst features (e.g. Jones et al., 2011), vegetation types characteristic for permafrost (Panda et al., 2014), or change detection of spectral indices (Nitze and Grosse, 2016). As permafrost is a subsurface temperature phenomenon, it is not possible to observe it directly from satellite-borne sensors. However, remotely sensed data sets can be used as input for the above-mentioned permafrost models (Hachem et al., 2009; Westermann et al., 2015).

Langer et al. (2013) demonstrated and evaluated a transient ground temperature modeling scheme forced by remote sensing data for a point in the Lena River Delta. In this work, we **update and** extend this earlier approach to facilitate spatially distributed mapping of the ground thermal regime based on satellite-derived data sets on surface temperature and snow cover. The model results are compared to in-situ observations of ground temperatures and thaw depths, thus facilitating a coarse assessment of the performance of the scheme regarding important permafrost variables.

2 Study area

2.1 The Lena River Delta

The Lena River Delta (LRD) is located in NE Siberia at the coast of the Laptev Sea. It constitutes one of the largest river deltas in the Arctic, covering an area of around 32 000 km^2 between 72 and 74°N. The LRD is dominated by continuous permafrost in a continental climate, with extremely cold winter and relatively warm summer temperatures (Boike et al., 2013). Mean annual ground temperatures are the order of -10 °C, and the frozen ground is estimated to extend to about 400 to 600 m below the surface (Yershov et al., 1991).

With altitudes between 0 and 60 m a.s.l., the LRD can essentially be regarded as "flat", so that medium and low resolution data sets (1 km or coarser) can be employed without the need of topographic corrections. However, the surface and ground properties feature a strong heterogeneity at spatial scales of 1 m to 1 km (with e.g. a large number of small water bodies, Muster et al., 2012, 2013) which is not reflected in medium and low resolution data sets. Despite such small-scale variability, the LRD can be classified in three main geomorphological units (Fig. 1), which have distinctly different characteristics regarding their surface and subsurface properties, such as ground ice contents, thermokarst features and vegetation cover (Morgenstern et al., 2013; Fedorova et al., 2015).

The *first river terrace* covers large parts of the eastern and central delta. It is the youngest and most active part of the delta, shaped by river erosion and sedimentation during the Holocene. Polygonal tundra with mosses, sedges, grass and occasional dwarf shrubs dominates the surface (Schneider et al., 2009; Boike et al., 2013). The subsurface material consists of silty sands and organic matter in alluvial peat layers with thicknesses up to 5 to 6 m (Schwamborn et al., 2002b). Ice wedges of more than 9 m depth have been described on the first terrace (Grigoriev et al., 1996; Schwamborn et al., 2002b). The ice contents in the uppermost few meters reach 60 to 80% in volume, while the mineral and organic contents reach 20-40% and 5-10%, respectively (Kutzbach et al., 2004; Zubrzycki et al., 2012). A considerable fraction of the first terrace is composed of the modern floodplain of the Lena River which is periodically inundated. These floodplain areas feature a different ground stratigraphy, with sandy, generally well-drained soils with low organic contents.

The *second river terrace*, located in the northwestern part of the LRD, was created by fluvial deposits between 30 and 15 kaBP when the sea level was lower than today. These sandy sediments generally feature low ice and organic contents (Schirrmeister et al., 2011). Arga Island is the biggest island of this terrace and the geomorphologic unit is often called Arga complex.

The *third river terrace* is composed of late Pleistocene sediments which have not been eroded by the Lena River during the Holocene. It is distributed in isolated islands in the southern margins of the LRD (Grigoriev, 1993; Zubrzycki et al., 2012). The third terrace is part of the Yedoma region which contains substantial quantities of ground ice and organic carbon down to several tens of meters below the surface (Strauss et al., 2013). The Yedoma was accumulated during the extremely cold climate of the last glacial period between 43 and 14 ka, which created ice wedges of more than 25 m depth (Grigoriev, 1993; Schwamborn et al., 2002b; Schirrmeister et al., 2003). The vegetation consists of thick 0.1 to 0.2 m hummocky grass, sedge and moss cover, and the upper horizon of the soil has a thick organic layer. Holocene permafrost degradation resulted in the current complex thermokarst landscape characterized by thermokarst lakes and drained basins (Morgenstern et al., 2013). The three river terraces occur in clusters of at least a few square kilometers (Fig. 1) so that they can be resolved by grid-based mapping at 1 km scale. A model study by Westermann et al. (2016) suggests that the subsurface stratigraphies of the three river terraces lead to a distinctly different ground thermal regime and susceptibility to future surface warming. Spatially distributed permafrost modeling hence must account for these geomorphological units and their characteristics of subsurface heat transfer.

2.2 Field sites and in-situ observations

2.2.1 The Samoylov Permafrost Observatory

Samoylov Island is an about four square kilometer large island (72°22'N, 126°28'E) located at the southern apex of the LRD, close to where the the Olenyokskaya Channel flows out of the main stem of the Lena River (Fig. 1). It is situated on the first river terrace and dominated by wet polygonal tundra and thermokarst lakes and ponds of various sizes (Boike et al., 2013). A Russian-German research station has been operating on Samoylov Island for more two decades and facilitated scientific studies on energy and carbon cycling (Kutzbach et al., 2007; Wille et al., 2008; Sachs et al., 2010; Abnizova et al., 2012, e.g.), validation of satellite data sets (Langer et al., 2010) and ESM development (e.g Ekici et al., 2014; Yi et al., 2014; Chadburn et al., 2015). Permafrost temperatures have been increasing, and ice-wedge degradation is occurring "subtly" on sub-decadal timescales, but with long term consequences for the hydrologic drainage (Liljedahl et al., 2016). A detailed overview on the climate, permafrost, vegetation, and soil characteristics on Samoylov Island is provided by Boike et al. (2013). On Samoylov Island, a long time series of meteorological and environmental variables is available (Boike et al., 2013) which forms an excellent basis for validation of satellite data sets and ground thermal modeling (Langer et al., 2010, 2013; Westermann et al., 2016). In the following, we briefly describe the in-situ data sets employed in this study (Sects. 4.1.1 and 4.2.1):

Surface temperature: On Samoylov Island, surface (skin) temperature has been measured continuously since 2002 by a downward facing long wave radiation sensor (CG1, Kipp & Zonen, Netherlands). The outgoing long wave radiation is converted to surface temperature using Stefan-Boltzmann law (see Langer et al., 2013, for details).

Snow depth and properties: On the point scale, snow depth measurements have been conducted with an ultra-sonic ranging sensor (SR50, Campbell Scientific, USA; located close to the long wave radiation sensor) since summer 2003, but a few winter seasons are not covered due to sensor failure. **In addition, a spatially distributed survey of snow depths and densities (216 points in polygonal tundra) was conducted in early spring 2008 (25 April to 2 May) before the onset of snowmelt (Boike et al., 2013).** The onset and termination of the snow cover were manually determined from pictures taken by an automated camera system, with dates from 1998 to 2011 provided in Boike et al. (2013).

Ground temperature: In this study, we make use of measurements of active layer temperatures in a low-center polygon established in 2002, and ground temperatures in a 26 m deep borehole since 2006 (Boike et al., 2013). The measurement site of the active layer temperatures can be considered representative for the polygonal tundra of the first river terrace (Boike et al., 2013). The deep borehole is located near the southern bank of the island close to the research station in an area with ground properties that differ from the “typical” stratigraphy of the first terrace: the area around the borehole features sandier soils with low organic contents that are generally well-drained due to the proximity to the river bank. **In the course of an upgrade of the research station, new buildings and structures were erected in the direct vicinity of the borehole in summer 2012 (See Supplementary Material), leading to much higher snow accumulation around the borehole in the following winters (compared to the surrounding terrain on Samoylov Island). Therefore, only borehole data until summer 2012 are used for comparison to model results.**

Thaw depth: Oriented at the measurement protocol for CALM sites (Burgess et al., 2000), thaw depths have been manually mapped on a grid with 150 points in polygonal tundra on Samoylov Island since 2002. According to the land cover classification in Boike et al. (2013), the grid points are located both on dry polygon rims and wet polygon centers. In most years, several surveys are available covering the entire period from the onset of thaw until maximum thaw depths are reached.

2.2.2 In-situ observations in the LRD

Outside of Samoylov Island, only sparse observations on the ground thermal regime are available. In 2009 and 2010, ground temperature measurements at several meters depth were established in four boreholes distributed across the LRD (Fig. 1), all of which are located in a rather homogeneous surroundings (see Supplementary Material for images):

- Olenyokskaya Channel, mouth: located on the third terrace at the W edge close to the Laptev Sea (72°49’20.1” N, 123°30’45.0” E),

- Olenyokskaya Channel, center: located on the first terrace in the SW part of the LRD (72°33’56.9” N, 125°03’52.3” E),

- Kurungnakh Island: located on the third terrace in an *alas* depression on Kurungnakh Island about 10 km SW of Samoylov Island (72°19’12.5” N, 126°11’35.7” E). The installation of the borehole destroyed the surface vegetation which triggered melting of excess ground ice and the formation of a thermokarst pond around the borehole within one year (see Supplementary Material). The ground temperature record must therefore be considered disturbed and most likely features a warm-bias compared to the surrounding undisturbed terrain. We therefore only employ the first three months of data following the drilling of the borehole.

- Sardakh Island: located in the SE part of the LRD near the main channel of the Lena River (72°19’12.6” N, 127°14’29.4” E). Sardakh is generally classified as part of the third terrace due to similar surface cover and height above river level, but the ground is actually comprised of neogene sandstone with a cover of Yedoma deposits (Kryamyarya et al., 2011). At the borehole site, melting of excess ground ice has occurred since the installation of the borehole like in the case of Kurungnakh, which has led to subsidence of the surface and the formation of a pond around the borehole. **This was observed for the first time in summer 2012 (see Supplementary Material) and we therefore exclude the later parts of the borehole record from the comparison to model results.**

For the second terrace, there are no measurements of ground temperatures available.

Systematic measurements of thaw depths according to the CALM protocol have not been conducted outside Samoylov Island.

However, there exist observations of thaw depths for single points in time and space for all three river terraces, which facilitate validation of regional differences in thaw depths:

- First terrace: In addition to the comprehensive record on Samoylov Island, a single measurement near the borehole site "Olenyokskaya Channel, center" is available from the year 2010.
- Second terrace: In summer 2005, thaw depths were recorded at several sites on Turakh Island (72° 56'24.4" N, 123°47'54.9" E) in the southwestern LRD near exposures at the shoreline and at a drill core site (Schirrmeister, 2007; Ulrich et al., 2009). Another manual thaw depth measurement was performed in the northern part of Arga Island (73°29'39.2" N, 124°22'33.1" E) in 2010. These observations are the only available ground truth information for the second terrace in the model period 2000-2014. Two additional observations are available from summer 1998 from the central part of Arga Island (73°20'18.5" N, 124°12'30.5" E) near Lake Nikolay and on Dzhipperies Island (72°51'14" N, 125°50'22" E) near Lake Yugus-Jie-Kuyele (Rachold and Grigoriev, 1999). While these cannot be compared to model output in a strict sense, they confirm the general order of magnitude of thaw depths on the second terrace.
- Third terrace: Thaw depth measurements are available from two distinct areas. At the W edge of the LRD, the thaw depth was recorded near the borehole site "Olenyokskaya Channel, mouth" in summer 2010. At three dates in July and August 2013, thaw depths were recorded at nine locations in the S part of Kurungnakh Island, near so-called "Lucky Lake" (72°17'41.0"N 126°9'34.0" E). The nine locations are spread over an area of several square kilometers which is contained within six 1 km model grid cells.

3 Methods

In this study, we update and extend the satellite data-based transient modeling of the ground thermal regime as outlined in Langer et al. (2013) to an area of approx. 16 000km^2 within the LRD. The general idea is to employ time series of remotely sensed surface temperatures and snow depths to force a transient ground thermal model.

3.1 The CryoGrid 2 ground thermal model

CryoGrid 2 is a transient 1D ground thermal model based on Fourier's Law of heat conduction (Westermann et al., 2013). The model does not account for changing subsurface water contents due to infiltration and evapotranspiration, but instead assigns fixed values for the porosity and saturation of each grid cell. Freezing/thawing of soil water/ice is accounted for by a temperature-dependent apparent heat capacity (e.g. Jury and Horton, 2004) which is determined by the soil freezing characteristic according to the formulation by Dall'Amico et al. (2011). The apparent heat capacity and thermal conductivity of each layer are computed according to the volumetric fractions of water/ice (determined by the temperature), air and sediment matrix material composed of a mineral and an organic component. A more detailed description of the model physics and the numerical solvers is provided in Westermann et al. (2013).

CryoGrid 2 is capable of representing the annual build-up and disappearance of the snow cover with a variable number of snow grid cells, but only allows for constant thermal properties of the snow (both throughout the snow pack and over time). For this study, we assign a functional dependency between snow thermal conductivity k_{snow} and density ρ_{snow} according to Yen (1981):

$$k_{\mathrm{snow}} = k_{\mathrm{ice}} \left(\frac{\rho_{\mathrm{snow}}}{\rho_{\mathrm{water}}} \right)^{1.88}, \tag{1}$$

with k_{ice} and ρ_{water} denoting the thermal conductivity of ice and the density of water, respectively. This parameterization performed well over a wide range of snow densities and types in a dedicated validation study (Calonne et al., 2011). As a result, the thermal properties of the snow pack are described by only a single parameter, the snow density ρ_{snow}, for which an extensive set of in-situ observations is available from Samoylov island (Boike et al., 2013).

3.2 Subsurface properties and additional model parameters

At $1\,\mathrm{km}$ resolution, it is not possible to resolve small-scale differences of surface and subsurface properties. Therefore, we only distinguish the three river terraces as the main geomorphological units within the LRD for which we define "typical" subsurface stratigraphies oriented at the available field observations (Sect. 2.1). The stratigraphies are provided in Table 1, while the boundaries of the terraces (Fig. 1) are based on Morgenstern et al. (2011), which were subsequently gridded to $1\,\mathrm{km}$. For all terraces, a saturated bottom layer with mineral content of 70 vol.% is assumed, corresponding to densified fluvial deposits underlying the modern delta (Schirrmeister et al., 2011; Schwamborn et al., 2002b).

For the first terrace, a $0.15\,\mathrm{m}$ thick upper layer with high porosity and organic content is assigned, which is not entirely saturated with water or ice (Schneider et al., 2009; Langer et al., 2013). Below, the ground is assumed to be saturated, but the porosity remains high, corresponding to the ice-rich sediments. Based on field observations on Samoylov Island (Kutzbach et al., 2004; Zubrzycki et al., 2012), fine-grained silty sediments dominate the matrix material, with organic contents of approx. 5 vol. %. The depth of this layer is set to $9\,\mathrm{m}$, based on observations for the depth of ice wedges in the first terrace (Schwamborn et al., 2002b). Note that these ground properties are also assigned to the active floodplain areas within the first terrace (Sect. 2.1) which cannot be meaningfully delineated at $1\,\mathrm{km}$ scale. In such floodplain areas, the model results must therefore be considered with care. Furthermore, the polygonal tundra landscape features a strong variability in surface soil moisture and vegetation/sediment conditions over distances of a few meters (Boike et al., 2013), which cannot be captured by the single stratigraphy employed for the modeling.

The sandy sediments of the second terrace largely lack an organic upper horizon (Rachold and Grigoriev, 1999; Ulrich et al., 2009; Schneider et al., 2009), so that a uniform upper layer with typical porosity of sand is prescribed (Table 1).

The third terrace is dominated by a relatively dry organic top layer with high porosity (Schneider et al., 2009; Zubrzycki et al., 2012), followed by a thick layer with very high ice contents (and organic contents of 5 vol. %), corresponding to the late Pleistocene Yedoma deposits (Schwamborn et al., 2002b; Schirrmeister et al., 2011). While the mineral fraction of this layer in reality is composed of fine-grained silty sediments, we assign "sand" as sediment type (Table 1) to account for the freezing characteristic of the extremely ice-rich ground which can be expected to resemble that of free water/ice rather than that of saturated silt.

The thermal conductivity of the mineral fraction of the sediment matrix required for the calculation of the soil thermal conductivity (Westermann et al., 2013) is set to $3.0\,\mathrm{W\,m^{-1}K^{-1}}$, as in previous modeling studies on Samoylov Island (Langer et al., 2011a, b, 2013). The sensitivity study by Langer et al. (2013) showed that the snow thermal properties are the most important model parameter controlling the simulated ground thermal regime. Therefore, the snow density (which controls both snow depth, heat capacity and thermal conductivity, Sect. 3.1) is a highly crucial parameter for which spatially or temporally distributed data sets **covering the entire LRD** are not available. However, an extensive set of measurements from polygonal tundra on Samoylov Island suggests snow densities of $(225\pm25)\,\mathrm{kg\,m^{-3}}$ Boike et al. (Fig. 6b, 2013) for polygon centers with well-developed snow cover, so that it is possible to explicitly account for the uncertainty of this important parameter by conducting model runs for a range of snow densities. For comparison to in-situ data (Sects. 4.1.1, 4.2.1), we present model runs with confining values of 200 and $250\,\mathrm{kg\,m^{-3}}$ (thus providing a range of ground temperatures), while the spatially distributed model runs (Sect. 4.2.2) are conducted with an average snow density of $225\,\mathrm{kg\,m^{-3}}$. Note that the confining values represent one standard deviation and that higher and lower snow densities occur regularly (Boike et al., 2013).

3.3 Model forcing data

CryoGrid 2 requires time series of surface temperatures and snow water equivalent as forcing data sets.

Surface temperature: As temperature forcing at the upper model boundary, a product synthesized from clear-sky land surface temperatures (LST) from the “Moderate Resolution Imaging Spectroradiometer” (MODIS) and $2\,\mathrm{m}$ air temperatures from the ERA–interim reanalysis (Dee et al., 2011) was applied. For this purpose, the daily MODIS level 3 LST products MOD11A1/ MYD11A1 in the version 005 were employed, which deliver four LST values per day (Terra and Aqua satellites, day and night time LST each). The merging procedure is similar as described in Westermann et al. (2015) in which spatially distributed data sets of freezing and thawing degree days were generated. In essence, gaps in the MODIS LST record due to cloud cover are filled by the the reanalysis data, which creates a data record with homogeneous data density and has the potential to moderate the cold-bias of temporal averages of surface temperatures computed from clear-sky MODIS LST (Westermann et al., 2012, 2015). **During cloudy skies, differences between air and surface temperatures are strongly reduced compared to clear-sky conditions (e.g. Gallo et al., 2011), so that air temperatures can be regarded an adequate proxy when MODIS LST is not available due to cloud cover. For melting snow, surface temperatures are confined to the melting point of ice, while air temperatures can be positive. Positive values of the surface temperature forcing are therefore set to $0\,^{\circ}\mathrm{C}$ if a snow cover is present (see below).** For this study, we create a time series of weekly averages of surface temperatures to force the CryoGrid 2 model. The reanalysis data which are available at 0.75° resolution are interpolated to the center point of each MODIS LST pixel (in the sinusoidal projection native to MOD11A1/MYD11A1 data). The satellites carrying the MODIS instrument were launched in 2000 (Terra) and 2002 (Aqua), respectively, while ERA–interim reanalysis is available since 1979. The synthesized time series used for model forcing therefore extends from 15 May 2000 to 31 October 2014 and thus covers the period for which remotely sensed LST data from at least one satellite are available. For the first two years, the data density of MODIS LST measurements in the composite product is lower than after summer 2002 when LST measurements from Aqua become available. Spatially, the fraction of the successful MODIS LST retrievals is relatively constant throughout the LRD, varying between 50 and 55%. In summer and fall, retrieval fractions are generally lower (40-50%) than winter and spring (55-70%), indicating more frequent cloudy conditions in summer and fall.

Snow depth: Similar to the procedure outlined in Langer et al. (2013), a weekly snow water equivalent (SWE) product was synthesized from GlobSnow SWE (Pulliainen, 2006) (25 km resolution) and the MODIS level 3 Snow Extent (SE) products MOD10A1/MYD10A1 (0.5 km resolution), which for clear-sky conditions deliver two values of binary flags (1: snow; 0: no snow) per day (one for Terra and Aqua each). The latter products were averaged over the 1 km sinusoidal grid of the MODIS LST data and the two satellites, yielding a number between 0 and 1 for each day with available data, corresponding to the fraction of successful retrievals at the 0.5 km pixel level flagged as "snow". We then applied a "maximum change" detection algorithm to the data set to determine the most likely dates for the start and the end of the snow cover in each 1 km pixel. For this purpose, we compute the fractions of 1 km values with values of 0 and 1, respectively, both within a window of four weeks before and after each date. The snow start date is determined as the date for which the sum of fractions of 0 before and fractions of 1 after is largest. This sum can be up to 2 when there are 100% retrievals flagged as snow-free before and 100% retrievals flagged snow-covered before the date. For the snow end date, the opposite criterion is applied, i.e. the sum of the fractions of 1 before and fractions of 0 after features a maximum. Note that the large window is required as prolonged cloudy periods often occur in the study area, for which no measurements are available. The MODIS SE products cover the same periods as the MODIS LST data (see above).

GlobSnow SWE (Daily L3A SWE, level 2.0) data are derived from passive microwave remote sensors which are not affected by clouds, so that a gap-free daily time series is in principle available for entire model period from 2000 to 2014. **The GlobSnow processing algorithm is based on a data assimilation procedure, which also takes in-situ measurements at WMO (World Meteorological Organization) stations into account (Takala et al., 2011). For the LRD, the closest station is located at Tiksi, about 50 km to the E, while the closest stations to the W are several hundred kilometers away. The station measurements are interpolated in space to obtain a SWE background field which is then weighted against SWE information derived from the passive microwave sensor by means of forward modeling of snowpack microwave emission using the HUT model (Pulliainen et al., 1999).**

The SWE values in the LRD (see Sect. 4.1) are typically below the critical threshold of about 150 mm above which SWE can no longer reliably derived from passive microwave retrievals (Takala et al., 2011). On the other hand, SWE retrieval is hampered for shallow snow cover and for wet melting snow, so that the start and the end of the snow season is not well covered by GlobSnow. Furthermore, water bodies constitute a major error source (e.g. Derksen et al., 2012) which generally leads to underestimation of SWE, in particular when the ice cover is thin (Lemmetyinen et al., 2011). Due to admixing of microwave radiation emitted from the ocean, the number of SWE retrievals is very small or even zero in the coastal areas of the LRD, so that almost half of the area of the LRD could not be included in the modeling. The boundary of the final model domain was finally chosen so that all validation sites (Fig. 1) are located within. In a few cases (in particular the sites AN, Tu and OM, Fig. 1), the available SWE data had to be extrapolated by about one grid cell or 25 km, which seems adequate considering the smoothness of the remote sensing derived SWE field in the

LRD.

As a first step, the daily SWE data were interpolated from the Northern Hemispherical EASE-Grid projection (25 km resolution) to the 1 km sinusoidal grid of the MODIS LST data. We subsequently assign linearly increasing SWE from the date identified as the most likely snow start date (using the MODIS SE product, see above) and the next available GlobSnow SWE measurement. The same procedure is applied for the snow end date. Not that this procedure can result in a step-like increase or decrease of the snow depth, if a valid GlobSnow SWE value is available for the identified start/end date. As a final step, the daily time series is averaged to the same weekly periods as the employed surface temperature forcing (see above) and SWE converted to snow depth with the applied snow density (Sect. 3.2). The use of medium-resolution MODIS SE facilitates correcting the coarse-scale GlobSnow SWE product regarding the start and the end of snow cover period, both of which can crucially influence the modeled ground thermal regime. **Nevertheless, passive microwave-derived SWE is associated with considerable uncertainty in the LRD. We therefore compare the model snow forcing to in-situ measurements from Samoylov Island (Sect. 4.1.1) and to independent spatial SWE data sets (Sect. 4.1.2, Supplementary Material).**

3.4 Model set-up

For each 1 km grid cell, the ground thermal regime was simulated for a specific ground stratigraphy and forcing time series of surface temperatures and snow depths. In the vertical direction, the ground between the surface and 100 m depth is discretized in 163 **layers**, which increase in size from 0.02 m near the surface (until 1.5 m depth so that the active layer is modeled at maximum resolution) to 10 m near the bottom, similar to the set-up in Westermann et al. (2013). Within the snow cover, the minimum layer size of 0.02 m is prescribed. At the lower boundary, a constant geothermal heat flux of 50 $\mathrm{mW m^{-2}}$ is assumed, as estimated from a 600 m deep borehole 140 km east of Samoylov Island (Langer et al., 2013).

To estimate a realistic initial temperature profile, a model spin-up is performed to achieve steady-state conditions for the forcing of the first five model years, using the multi-step procedure outlined in detail in Westermann et al. (2013). In a first step, the model is run to estimate the average temperature at the ground surface (i.e. below the snow cover in winter), for which the steady-state temperature profile in the ground is assigned to all grid cells (considering the geothermal heat flux at the bottom and the thermal conductivity of all grid cells). In a second step, CryoGrid 2 is run twice for the first five model years, so that the annual temperature cycle to the depth of zero annual amplitude is reproduced. The simulations for the entire time series can thus be initialized by a temperature profile that is both adequate for the upper and the lower parts of the model domain. We emphasize that the initialization procedure limits the CryoGrid 2 results to the uppermost few meters of the soil domain since deeper temperatures are still influenced by the surface forcing prior to the model period, for which satellite measurements and thus model forcing data are not available.

4 Results

4.1 Forcing data sets

4.1.1 Comparison to in-situ data

Systematic in-situ observations on surface temperature and snow depths are only available for the Samoylov permafrost observatory, so that a validation of the spatial patterns of the model forcing data within the LRD is not possible.

Surface temperature: We compare the surface temperature forcing synthesized from MODIS LST and ERA reanalysis air temperatures (Sect. 3.3) to measurements of surface (skin) temperature from Samoylov Island from 2002 to 2009 (Boike et al., 2013). The results of the comparison for the $1\,\mathrm{km}$ grid cell in which the observation site is located, are displayed in Fig. 2: while the annual temperature regime is reproduced very well, a systematic cold-bias of on average -0.8$^\circ\mathrm{C}$ remains which is consistent throughout the year. Fig. 2 (bottom) also shows a comparison of monthly averages of all available MODIS LST measurements, i.e. without filling the gaps in the time series with ERA reanalysis air temperatures. Here, a significantly larger cold-bias of up to $3^\circ\mathrm{C}$ is found for all months except July, which is in line with validation studies from Svalbard which demonstrate a similar cold-bias during the winter moths (Westermann et al., 2012; Østby et al., 2014). In July, the average of all MODIS LST measurements is significantly warmer than the observations. However, surface temperatures can feature a strong spatial variability during summer **due to differences in surface cover and soil moisture conditions** (Langer et al., 2010; Westermann et al., 2011b), so that the scale mismatch between the $1\,\mathrm{km}$ remotely sensed LST values and the in-situ point observations may explain at least part of the deviation. In summary, the time series of surface temperatures synthesized from MODIS LST and ERA-interim reanalysis air temperatures facilitates an adequate representation of in-situ observations and thus well suited as input for ground thermal modeling **(at least in homogeneous terrain)**, which supports earlier results from the N Atlantic permafrost region (Westermann et al., 2015). However, the slight, but systematic cold-bias must be taken into account when analyzing the uncertainty of modeled ground temperatures.

Snow cover: As for surface temperatures, only point measurements on Samoylov Island are available for snow depth which are compared to the forcing time series of snow water equivalents synthesized from $25\,\mathrm{km}$ GlobSnow SWE and $0.5\,\mathrm{km}$ MODIS SE (Sect. 3.3). In general, snow depths computed from GlobSnow SWE with snow densities between 200 and $250\,\mathrm{kg\,m^{-3}}$ can reproduce the order of magnitude of the in-situ measurements, with differences generally smaller than $0.1\,\mathrm{m}$ (Fig. 3). At least some of the observed interannual differences are reproduced in the remote sensing-derived snow product, e.g. the above-average snow depths in winter 2003/04 and the below-average snow depths in 2012/13 **(the latter was qualitatively noted by the station personnel, pers. comm., N. Bornemann)**. **For values with non-zero snow depth, the model forcing (using a snow density of 225 $\mathrm{kg\,m^{-3}}$) features an RMSE of about 0.06 m, and a slight positive bias of 0.015 m. The average snow depth in polygonal tundra (obtained by a spatially distributed survey, Boike et al., 2013) in early spring 2008 is slightly higher than both point measurements from the snow depth sensor and the model forcing. However, the difference is only about 0.05 m for the model forcing with snow density 225 $\mathrm{kg\,m^{-3}}$, well within the observed spatial variability of snow depths (Fig. 3).**

Start and end dates of the snow cover are compared to in-situ observations (Fig. 4) based on interpretation of time-lapse imagery from an automatic camera system (Boike et al., 2013). The snow melt date, which is crucial for capturing the onset of soil thawing correctly, is generally well captured, although differences of more than half a month exist for some of the years. We emphasize that the transition from a completely snow covered to a completely snow-free surface occurs over an extended period of time due to spatially variable snow depths, so that a “snow melt date” in a strict sense does not exist. The MODIS SE processing algorithm based on surface reflectances may apply a different threshold for the characterization of a snow-free surface than the subjective interpretation of the in-situ camera images. Furthermore, prolonged periods of cloudiness make remote detection of **snow cover** impossible, so that a considerably reduced accuracy must be expected in such years. The same issues apply to the detection of the snow start date. While deviations of more than 15 days exist in the beginning of the period, the remotely detected snow start date in general follows the in-situ observations very well (Fig. 4). **We conclude that the model forcing can adequately reproduce the general magnitude of snow depth and the timing of the snow-covered season on Samoylov Island. However, due to the considerable uncertainties associated with GlobSnow SWE retrievals (Takala et al., 2011) the snow depth model forcing for the entire LRD must be considered less reliable than the surface temperature forcing.**

4.1.2 Spatial distribution in the LRD

Fig. 5 displays the spatial distribution of yearly average surface temperatures (b), freezing degree days (c), thawing degree days (d), snow-free days (e) and average snow depth (f) for a ten-year period 2004-2013, as well as the classification of subsurface stratigraphies (a, see Sect. 3.2). Average surface temperatures feature only moderate spatial differences in the order of $2^{\circ}\mathrm{C}$, with the warmest areas close to the main river channels in the southern part of the LRD. Similarly, the differences in freezing degree days are only on the order of 10 to 15%, with the largest number of freezing degree days recorded in the central parts of the LRD, which is located furthest away from the coastline and main river channels. On the other hand, thawing degree days feature a pronounced north-south gradient, with values almost twice as large in the southern parts of the LRD compared to the areas at the north coast. A similar pattern is found for the average number of snow-free days which varies between around 100 in the northern areas and around 140 in the southern areas.

Average snow depths are largest in the western areas and decrease towards the southeastern parts of the LRD, although the differences are only small. **This spatial distribution is in coarse agreement with Canadian Meteorological Centre (CMC) Snow Depth Analysis Data (Brasnett, 1999), an independent global snow product at 24 km resolution based on precipitation data from an atmospheric model (see Supplementary Material). As passive microwave data are not employed in the CMC Snow Depth Reanalysis, the match is an indication that the overall snow depth pattern in Fig. 5f is not an artifact of the GlobSnow retrieval algorithm, but rather reflects spatial differences in snowfall. This conclusion is further supported by winter precipitation from the ERA-interim reanalysis which also displays a west-east gradient over the land areas in the LRD (see Supplementary Material). However, we emphasize that the effective spatial resolution of the remotely sensed snow depth data is significantly coarser than for the other variables, so that large biases are likely to**

occur at the model scale of 1 km, at least for single grid cells. Furthermore, the quality of the SWE retrievals is insufficient in coastal areas (Sect. 3.3) which hence are not covered by the ground thermal modeling.

4.2 Modeled ground thermal regime

4.2.1 Comparison to in-situ data

The model results are validated for ground temperatures and thaw depth for nine field sites, Samoylov Island, Olenyokskaya Channel center and mouth, Arga Island north and center, Dzhipperies Island, Turakh Island, Kurungnakh Island and Sardakh Island (Fig. 1, Sect. 2.2). With this data basis, all three stratigraphic classes are covered by two or more in-situ measurement sites. However, for the second terrace only few unsystematic thaw depth measurements are available and observations of ground temperatures are lacking entirely.

Ground temperature: To assess modeled ground temperatures, we use in-situ measurements of active layer temperatures from Samoylov Island (first terrace), as well as measurements of permafrost temperatures at 2-3 m depth in boreholes. At this depth, the temperature regime is dominated by the surface forcing over a couple of square meters surface area which averages over smaller-scale variability of surface and subsurface properties. On the other hand, the modeled temperature field is not strongly dominated by the initial condition, at least after the first years of simulation.

Fig. 6 displays a comparison of modeled and measured active layer temperatures at 0.4 m depth in a wet polygon center on Samoylov Island in the first terrace. In general, the in-situ values are contained within the range of modeled ground temperatures for the two confining snow depths, but some deviations exist during refreezing in fall. In a few years, the length of the so-called "zero-curtain" when temperatures remain in the vicinity of 0°C is underestimated in the simulations. Possible reasons are a too high thermal conductivity of the uppermost, already frozen soil layers, higher than average surface temperatures in the more moist sites during refreezing (compare Langer et al., 2010), or a shallow snow or rime cover at the surface which is not detected by remote sensors.

Although small, a similar effect is visible in several years for the modeled temperatures in shallow boreholes on the first and third terrace (Fig. 7) for which the pronounced cooling in fall occurs too early in the model runs. The consistent occurrence at several locations in the LRD points to a shortcoming of the model scheme rather than local conditions, e.g. caused by spatial variability of the subsurface properties. Despite such problems, the model scheme allows an adequate representation of measured ground temperatures within the range of uncertainty due to the snow density, **except for the periods when thermokarst development around the boreholes was evident (shaded grey in Fig. 7).** The 26 m deep borehole on Samoylov Island (Boike et al., 2013) is located near the south-west edge of the island in a relatively well-drained environment. With the relatively water- and ice-rich stratigraphy used for the first terrace (Table 1), considerably colder ground temperatures are modeled compared to the measurements (Fig. 8 left), particularly during summer and fall. Using the same surface forcing, but a stratigraphy oriented at the true conditions at the borehole (sandy sediments; 0-0.5 m: 30 vol. % water/ice, 10 vol. % air, 60 vol. % mineral; 0.5-9 m: 40 vol. % water/ice, 60 vol. % mineral; deeper layers as for first terrace) significantly improves the match between modeled and measured values, especially during summer (Fig. 8 right).

A comparison of monthly averages for all five boreholes is shown in Fig. 9. For a snow density of 225 $\mathrm{kg\,m^{-3}}$, the model results feature an RMSE of 1.1 $^\circ\mathrm{C}$ and an average bias of -0.9 $^\circ\mathrm{C}$, mainly due to underestimation of measured values during the summer and fall seasons. For a snow density of 200 $\mathrm{kg\,m^{-3}}$, the model bias is on average positive (+0.8 $^\circ\mathrm{C}$), but the RMSE is increased (1.6 $^\circ\mathrm{C}$). The model performance is worst for the highest snow density (RMSE 2.1 $^\circ\mathrm{C}$, bias -2.1 $^\circ\mathrm{C}$). If the Samoylov Island borehole (for which the ground stratigraphy was adjusted, see above) is removed, the model performance for the best-fitting snow density of 225 $\mathrm{kg\,m^{-3}}$ remains largely unchanged (RMSE 1.2 $^\circ\mathrm{C}$, bias -0.9 $^\circ\mathrm{C}$). Fig. 10 displays an inter-site comparison of measured and modeled yearly average ground temperatures for a two-year period for which largely gap-free in-situ records from four sites are available. All measurements are contained in the range of modeled ground temperatures for the confining snow densities of 200 and 250 $\mathrm{kg\,m^{-3}}$, although the in-situ value for Sardakh is located near the upper bound of the modeled temperature range. For the average snow density of 225 $\mathrm{kg\,m^{-3}}$, the measured and modeled values agree within 1 to 1.5 $^\circ\mathrm{C}$, which can serve as a coarse accuracy estimate for the spatially distributed simulations of the ground thermal regime in the LRD (Sect. 4.2.2). **While the model performance is encouraging, we emphasize that it is mainly based on only four sites (the Kurungnakh record comprises only a short period) which are all located in the southern part of the LRD.**

Thaw depth: In the LRD, temporally resolved measurements of thaw depths are only available from Samoylov Island. Fig. 11 compares modeled thaw depths with the average of 150 points for which thaw depths have been measured manually over a period of 13 years (Boike et al., 2013). In general, the model scheme can represent the measured thaw depths very well, with deviations of 0.1 m or less. In particular in the second half of the model period, the agreement is excellent with deviations of 0.05 m or less. Furthermore, the annual dynamics of the thaw progression is adequately resolved. We emphasize that the in-situ measurements are evidence of a considerable spatial variability of thaw depths even, with an average standard deviation of 0.06 m. This variability is not captured by the model runs with different snow densities which only induces differences in modeled thaw depths of a few centimeters Fig. 11. These results are in agreement with the sensitivity analysis of Langer et al. (2013) who showed for Samoylov Island that ground temperatures are most sensitive to snow thermal properties, while the thaw depth is more dependent on ground properties and ice contents which are set constant in the simulations (Table 1).

The comparison of modeled and measured thaw depths for the point measurements in the three stratigraphic units of the LRD is shown is Table 2. The in-situ observations are clear evidence that thaw depths are by far shallowest for the third terrace, while the largest thaw depths occur in the second terrace. The model scheme can reproduce this pattern very well, although deviations between measured and modeled thaw depths of 0.1 m or more can occur. The largest deviations occur for Turakh Island for which the model significantly underestimates the measured thaw depths. However, the measurements were performed near terrain edges and at slopes (Schirrmeister, 2007), so that a reduced match must be expected when comparing to thaw depths obtained for the simplified “model case” of flat homogeneous terrain. All in all, the comparison suggests that the presented model scheme accounts for the main drivers of active layer dynamics and facilitates an overall adequate representation of thaw depths in the LRD.

4.2.2 Spatial distribution in the LRD

Fig. 12 presents average ground temperatures at 1.0 m depth (i.e. well below the active layer, see next section) for the ten-year period 2004-2013. Within each stratigraphic unit, modeled ground temperatures generally decrease from west to east, following the spatial pattern of snow depth in the LRD (Fig. 5), and towards the North, presumably as a result of low summer surface temperatures and shorter snow-free period (Fig. 5). At the same time, the ground stratigraphic units have a pronounced impact on modeled ground temperatures, with lowest temperatures modeled for the third and warmest for the second terrace (compare Fig. 12). This is corroborated by the results of a sensitivity analysis towards the ground stratigraphy for the nine validation sites in the LRD (Table 3). When using the same forcing data, but different ground stratigraphies, the modeled ground temperatures are generally coldest for the third terrace and warmest for the second terrace stratigraphy.

The warmest ground temperatures are modeled for parts of the second terrace in the northwest and for the areas around the Olenyokskaya Channel in the southwest part of the LRD where ground temperatures warmer than -9°C are mapped. Medium cold temperatures of -9 to -11°C are obtained for the center of the delta and thus large parts of the first terrace. In the eastern part of the LRD, the coldest average temperatures with less than -11°C are modeled for parts of the third terrace.

Thaw depth: The spatial distribution of modeled maximum thaw depths (Fig. 13) is mainly related to two factors: the thawing degree days which decrease strongly from south to north (Fig. 5) in the LRD, and the ground stratigraphy. For the third terrace, average maximum thaw depths of less than 0.3 m are modeled, while the second terrace features maximum thaw depths of 0.65 to 0.95 m. In the first terrace, the modeled thaw depths are largest in the southern part (approx. 0.5 m), while the northeastern part feature considerably lower maximum thaw depths that are of similar magnitude as for the third terrace (0.3 m). These results are in agreement with the sensitivity analysis for the validation sites (Table 3), which clearly shows the strong dependence of modeled thaw depths on the ground stratigraphy.

5 Discussion and Outlook

5.1 Model forcing

5.1.1 Surface temperature

Validation studies have revealed a significant cold-bias of long-term averages derived from MODIS LST in Arctic regions (Westermann et al., 2012; Østby et al., 2014) which is attributed to the over-representation of clear-sky situations and deficiencies in the cloud detection during polar night conditions (Liu et al., 2004). The same bias is found for Samoylov Island (Fig. 2) for which averages directly computed from MODIS LST measurements are cold-biased by about 1-2°C for most of the year. In this study, we therefore employ a gap-filling procedure with ERA-interim near-surface air temperatures. During cloudy periods, reanalysis-derived air temperatures may indeed facilitate an adequate representation of surface temperatures, as the near-surface temperature gradient is smaller compared to clear-sky conditions (e.g. Hudson and Brandt, 2005; Gallo et al., 2011; Westermann et al., 2012).

As demonstrated by Westermann et al. (2015) for the N Atlantic region, the composite product features a considerably reduced bias and is significantly better suited as input for permafrost modeling than the original MODIS LST record. However, a small, but consistent cold-bias of about 0.8°C remains. This could be explained by the fact that the gap-filling procedure only applies to gaps due to clouds that are successfully detected, but does not remove strongly cold-biased LST measurements of cloud top temperatures (Langer et al., 2010; Westermann et al., 2011b) that regularly occur when the MODIS cloud detection fails. Here, further improvements seem feasible, e.g. through simple plausibility criteria when comparing the remotely sensed LST against meteorological variables of the ERA-reanalysis data set. However, such methods are most likely sensitive towards a range of factors, such as landcover and exposition (which strongly influence the true surface temperature), so that they should be carefully developed and validated for a range of sites. **Based on in-situ measurements, Raleigh et al. (2013) suggest that for snow-covered ground dew point temperatures are a better approximation for surface temperatures compared to air temperatures at standard height. However, observations on Samoylov Island display only a small offset between snow surface and air temperatures, with the difference increasing from near zero in early winter to about 1°C in late winter (Table 3, Langer et al., 2011b). The reason for this is most likely that the ground heat flux is a strong heat source especially in early winter (Langer et al., 2011b) which warms the surface and thus prevents formation of a strong near-surface inversion. Therefore, we consider air temperatures an adequate proxy for snow surface temperatures in the LRD, but dew point temperatures should clearly be considered for gap-filling in the snow-covered season in future studies.** We conclude that surface temperatures synthesized from MODIS LST and near-surface air temperatures from the ERA-interim reanalysis are an adequate choice for the purpose of ground thermal modeling **in the LRD, at least in homogeneous terrain,** although it may introduce a slight cold-bias in modeled ground temperatures.

5.1.2 Snow

As demonstrated by Langer et al. (2013), snow depth and snow thermal properties are crucial factors for correctly modeling ground temperatures in the LRD. In this light, the coarsely resolved estimates of GlobSnow SWE must be considered the key source of uncertainty for the thermal modeling.

– **The performance of GlobSnow SWE has been evaluated on continental scales by comparison to systematic in-situ data sets (Luojus et al., 2010; Takala et al., 2011). For Eurasia, surveys spanning the entire snow season (Kitaev et al., 2002) were compared from 1979 to 2000. For shallow snow (approx. SWE<60 mm), GlobSnow SWE tends to overestimate observed values slightly, but the relationship between measurements and GlobSnow retrievals is on average linear. When SWE exceeds approx. 100 mm, the GlobSnow algorithm tends to underestimate measured SWE, and for values larger than 150 mm the signal from passive microwave retrievals saturates and SWE can no longer reliably be detected (Takala et al., 2011). For the LRD, both in-situ measurements and GlobSnow values indicate that SWE is generally below this critical threshold so that saturation effects most likely do not play a role for the uncertainty. The Eurasia data set is strongly biased towards sites in steppe environments** and the boreal forest zone (where SWE retrieval is affected by the canopy, e.g. Derksen et al., 2012), while northern tundra areas with characteristics similar to the LRD are strongly undersampled. A more representative data set is available from an extensive transect across Northern Canada (Derksen et al., 2009), for which comparison of GlobSnow SWE retrievals yielded an RMSE of 47 mm and an average bias of -36 mm. The average SWE of 120 mm (Takala et al., 2011) was significantly larger than in the LRD, so that it is not meaningful to transfer the absolute uncertainties. When using relative uncertainties, on the other hand, we arrive at a similar RMSE as for the comparison of the time series on Samoylov Island (0.06 m, see Sect. 4.1.1): for N Canada, a relative RMSE of around 40% was found, which corresponds to an absolute RMSE of 0.065 m in snow depth, when scaled to the average of around 0.16 m on Samoylov Island (Fig. 5f). Although the character of the two data sets differs (spatial transect vs. multi-year point measurement), the good agreement is an indication that the GlobSnow performance in the LRD could be similar to N Canada. We emphasize that the RMSE corresponds to undirected fluctuations around the average value which have much less influence on the modeled average ground thermal regime (Figs. 12, 13) than a systematic bias. For a systematic bias of 10 mm SWE (applied uniformly to the entire time series), the sensitivity study by Langer et al. (2013) suggests a deviation of approx. 2.5 $^{\circ}\mathrm{C}$ of the modeled average ground temperatures at 2.5 m depth. In this study, we find an agreement within 1 to 1.5 $^{\circ}\mathrm{C}$ with borehole temperatures for multi-year averages at similar depths (Sect. 4.2.1). Therefore, a SWE bias of more than 10 mm seems unlikely for the borehole sites, although modeled ground temperatures are also influenced by the ground stratigraphy (Table 3).

– Water bodies strongly affect microwave emission of the ground, which is known to lead to underestimation of SWE in passive microwave-based retrievals (Rees et al., 2006; Lemmetyinen et al., 2011). For the above mentioned N Canada data set, water bodies might explain the significant bias of 36 mm (Takala et al., 2011), but the average values (120 mm) are also sufficiently high that saturation effects (Luojus et al., 2010) are likely to contribute to the bias. In the LRD, water bodies are abundant features (Fig. 1), so that GlobSnow retrievals are likely to be affected. Using a Landsat (Schneider et al., 2009) and MODIS (MODIS water mask) based land cover classifications, we estimate the water fraction in the employed 25 km grid cells in the Lena River Delta to be between 12 and 30%, with a single grid cell in the E part reaching 37% (of which more than half is estimated to be river arms, see below). Almost three quarters of the grid cells feature water fractions of less than 20%. However, relatively shallow themokarst lakes dominate in the LRD, which at least partly freeze to the bottom in winter (Schwamborn et al., 2002a; Antonova et al., 2016), so that microwave emission becomes similar to land areas, although in particular the wave-length dependency of the effect may be complex (Gunn et al., 2011). Furthermore, the winter discharge of the Lena River is very low compared to other northern rivers, as the catchment is largely located in the continuous permafrost zone (Yang et al., 2002). We estimate the winter discharge to be only about 10% of summer averages (Fig. 2 in Yang et al., 2002), and large river areas identified as water in summer-derived satellite imagery must fall dry in winter, which decreases the water fraction in the central and eastern part of the

delta (where the water fractions are highest) considerably. Furthermore, also shallow river arms and even coast-near areas of the Laptev Sea (Eicken et al., 2005) freeze to the bottom, so that we expect the true "open water" fraction relevant for microwave emission in winter to be significantly lower than the open water fractions obtained from summer imagery (see above) suggest. This is corroborated by the comparision to in-situ measurements for Samoylov Island (Fig. 3) situated in a relatively water-body-rich area where we find a satisfactory performance for GlobSnow. The largest impact on SWE retrievals is most likely during lake freezing and snow cover build-up in fall, when GlobSnow SWE retrievals must be considered highly uncertain. In the future, enhanced SWE retrieval algorithms taking the effect of water bodies explicity into account (e.g. Lemmetyinen et al., 2011) may become available.

- The spatial resolution of 25 km is insufficient to capture the considerable spatial variability of snow depths in the LRD, both on the modeling scale of 1 km and the considerably smaller scales where the snow distribution is strongly influenced by the microtopography (Boike et al., 2013). Studies with equilibrium models have demonstrated that the latter can to a certain degree be captured by statistical approaches that employ an (estimated) distribution of snow depths to obtain distributions of ground temperatures for each grid cell (Gisnås et al., 2014, 2015; Westermann et al., 2015). However, with the transient modeling scheme employed in these study, new issues arise that strongly complicate the application of a statistical representation of snow cover. First, spatial differences in snow depth will inevitably lead to a different timing of the snow melt which could influence in particular the modeled active layer thickness. Such small-scale differences of the snow start date cannot be captured by the 0.5 km scale MODIS SE product. Secondly, it is not clear how the distribution of snow depths can be translated to forcing time series of snow depths that are required for the CryoGrid 2 modeling. In some areas, snow depths may be relatively constant from year to year, while there may be strong interannual variations at other sites. Such temporal evolution is not contained in the distribution of snow depths, and computationally demanding deterministic snow redistribution models (e.g. Lehning et al., 2006) may be required to overcome such problems.

- In the coastal regions of the LRD, GlobSnow SWE does not provide a sufficient number of retrievals, so that the annual dynamics of the snow cover can be captured. In general, these regions must be excluded from the model domain. In this study, we chose to extrapolate the GlobSnow SWE retrievals to adjacent regions, so that more validation sites could be covered. The same issue applies to regions with pronounced topography which precludes the use of the modeling scheme for mountain permafrost area.

- The snow density is a crucial parameter, as it controls both the snow depth (since SWE is used as driving input data), the snow heat capacity and the snow thermal conductivity. In this study, the snow density was assumed to be constant in time and space, with the values determined by in-situ measurements (similar to Westermann et al., 2013; Langer et al., 2013). While this may be adequate for the relatively small model domain of the LRD, spatially distributed information on typical snow densities (e.g. Sturm et al., 1995) would be required for application on larger scales.

– The end and start of the snow cover have been determined at a comparatively high spatial resolution of 1 km using the MODIS SE product (Fig. 4), which corresponds to a downscaling of the coarsely resolved GlobSnow SWE product for these important periods. Furthermore, the performance of the GlobSnow SWE product is relatively poor for very shallow snow depths and for wet (melting) snow (Pulliainen, 2006) which is to a certain extent moderated by prescribing the snow start and end dates.

5.2 The CryoGrid 2 model

In this study, CryoGrid 2 is employed for a relatively short period of approx. 15 years, so that the model initialization deserves a critical discussion (Westermann et al., 2013). A model spin-up to periodic steady-state conditions was performed for the first five years of forcing data, i.e. from summer 2000 to summer 2005. Ground temperatures in deeper soil layers are strongly influenced by the choice of the initial condition, and the modeled temperatures should not be interpreted further. Therefore, we restrict the comparison to in-situ measurements to the uppermost three meters of soil and for the period following 2002 for active layer measurements (Figs. 6, 11) and after 2006 for ground temperatures in 2-3 m depth (Figs. 7, 8). In both cases, the model results are sufficiently independent of the initialization (Langer et al., 2013) which must therefore be considered a minor source of uncertainty.

The applied ground stratigraphy has a significant direct influence on the simulations results, both on ground temperatures and thaw depths (compare Westermann et al., 2016). For this study, three landscape units with associated “typical” stratigraphies were defined, which facilitate capturing the observed large-scale differences in particular for the thaw depth (Sect. 4.2.2). However, a significant small-scale variability of ground properties is superimposed on these large-scale differences which give rise to a significant variability of thaw depths and ground temperatures that are not captured at 1 km scale. An example is the in-situ record of thaw depths measurements at 150 points on Samoylov Island for which the model scheme can capture the interannual variations of the mean very well (Fig. 11). However, with an average standard deviation of 0.06 m the measurements feature a considerable spread (Boike et al., 2013), which is most likely explained by small-scale differences in ground properties, surface temperature and possibly snow cover. Another example is the borehole site on Samoylov Island, for which the “typical” ground stratigraphy for the first terrace is clearly not applicable (Fig. 8). In principle, such subgrid effects could be captured by running the model scheme not only for a single realization per grid cell, but for an ensemble of model realizations reflecting the statistical distribution of ground stratigraphies and properties within a grid cell. Such a scheme could also be extended to account for a subgrid distribution of snow depths by assigning different snow depths (according to a defined distribution, e.g. Gisnås et al., 2015) to the ensemble members. In addition to a considerable increase in computation time (e.g. a factor of 100 for 100 ensemble members), field data sets with statistical information on ground stratigraphies are generally lacking for the LRD. A simpler way could be aggregating high-resolution landcover data sets (e.g. Schneider et al., 2009) to the 1 km grid, so that fractional information on the landcover can be obtained. Assuming that each landcover class can be assigned a typical subsurface stratigraphy, the model scheme could be run for all landcover classes/stratigraphies present within one 1 km grid cell.

The model physics of CryoGrid 2 does not account for a range of processes that may influence the ground thermal regime in permafrost areas, such as infiltration of water in the snow pack and soil (Weismüller et al., 2011; Westermann et al., 2011a; Endrizzi et al., 2014), or thermokarst and ground subsidence due to excess ground ice melt. The latter can strongly modify the ground thermal regime, as demonstrated by Westermann et al. (2016), which makes a comparison of model results to in situ measurements at thermokarst-affected sites (Kurungnakh, Sardakh, Sect. 4.2.1) challenging. Furthermore, small water bodies and lakes can strongly modify the ground thermal regime both in the underlying ground and in the surrounding land areas (Boike et al., 2015; Langer et al., 2015), so that the model results are questionable in areas with a high fraction of open-water areas (Muster et al., 2012). While more sophisticated model schemes (Plug and West, 2009; Westermann et al., 2016) can simulate the ground thermal regime of such features, a spatially distributed application is challenging: in general, higher-complexity models require additional input data and model parameter sets (e.g. precipitation for a water balance model, Endrizzi et al., 2014) for which the spatial and temporal distributions are poorly known. Furthermore, the model sensitivity may vary in space depending on the interplay of different model parameters and input data (Gubler et al., 2011) which makes it harder to judge the uncertainty of model results.

5.3 The modeled ground thermal regime

The validation results suggest a model accuracy of $1^{\circ}\mathrm{C}$ to $1.5^{\circ}\mathrm{C}$ for multi-annual average ground temperatures (Fig. 10) and around 0.1-0.2 m for annual maximum thaw depths (Table 2). On the one hand, warm ground temperatures are modeled along the large river channels in the southern part of the LRD. These areas also feature high average surface temperatures (Fig. 5) which could at least partly be related to warm water advected by the Lena river. If this interpretation is correct, surface temperatures derived from remote sensors have a significant advantage over data sets derived from atmospheric modeling, which in general cannot reproduce such effects. On the other hand, the modeled ground temperatures are clearly influenced by ground stratigraphy. As evident in Fig. 12, the second terrace is systematically warmer than the adjacent first terrace, which is not visible in the temperature forcing (Fig. 5). This finding is corroborated by the sensitivity analysis (Table 3) which showcases the importance of a sound representation of ground thermal properties, in particular in and just below the active layer, for correct modeling of ground temperatures. These differences are at least partly related to stratigraphy-dependent thermal offsets between average ground surface and ground temperatures caused by seasonal changes of subsurface thermal conductivities due to freezing and thawing (Osterkamp and Romanovsky, 1999).

Thaw depths are to an even larger extent determined by the ground stratigraphy. On the third terrace, a comparatively dry organic-rich layer with low thermal conductivity limits the heat flux so that the underlying ice-rich layers experience only a limited amount of thawing. As a consequence, the thaw progression hardly extends below the uppermost layer, yielding thaw depths of around 0.3 m and less. On the first terrace, this effect is somewhat reduced (thinner and wetter organic top layer and lower water ice contents below), while the second terrace lacks the organic top layer and as a consequence experiences considerably deeper thawing than the two other stratigraphic units. In addition, the summer surface forcing strongly impacts thaw depths. Within the first terrace, the model results yield a pronounced north-south gradient of thaw depths (Fig. 13) which is related to the pattern of thawing degree days (Fig. 5).

5.4 Towards remote detection of ground temperature and thaw depth in permafrost areas?

The presented model approach can adequately reproduce both ground temperatures and thaw depths for an area of more than 10 000 km^2, largely based on remotely sensed data sets. Other than in satellite-based approaches with much simpler steady-state models (Hachem et al., 2009; Westermann et al., 2015), the time evolution of the ground thermal regime is explicitly accounted for in the transient approach using CryoGrid 2. Our results suggest that the annual temperature regime is adequately captured, while a longer time series is needed to evaluate and secure multi-annual trends, in particular since the first part of the model period is affected by the initialization. However, with the ever extending record of high-quality satellite data, remote detection of trends in permafrost temperatures may become feasible within the coming years.

Sufficient computational resources provided, the presented scheme could in principle be extended to the entire Northern Hemisphere, for which GlobSnow retrievals are available. However, at present such application is limited by a number of shortcomings and complications: first, the model scale of 1 km^2 may be sufficient to represent the ground thermal regime in lowland tundra landscapes like the LRD, but is significantly too coarse for heterogeneous terrain, e.g. in mountain areas (Fiddes et al., 2015). Since the grid cell size is determined by the spatial resolution of the remotely sensed land surface temperatures, it could only be improved with the deployment of higher-resolution remote sensors for surface temperature (which must also feature a high temporal resolution). Furthermore, remotely sensed data sets of snow water equivalent are lacking in many regions, in particular in coastal and mountain areas (compare Fig. 5), and the spatial resolution of 25 km is hardly sufficient to capture the spatial distribution of snow in the terrain in complex landscapes. **Furthermore, operational SWE retrievals are associated with considerable uncertainty in lake-rich tundra areas (Takala et al., 2011).** In many permafrost areas, this can be expected to results in a strongly reduced accuracy so that significantly simpler schemes (Westermann et al., 2015) might provide similar results. Another crucial issue is the lack of a standardized pan-arctic product on subsurface properties, which combines spatially resolved classes with information on subsurface stratigraphies and thermal properties. There exists a variety of such products on the regional and local scales, but they strongly differ in their quality and classes which are derived for different purposes. A pan-arctic homogenization effort similar to what has been accomplished for permafrost carbon stocks (Hugelius et al., 2013) is therefore needed in order to obtain meaningful results with a transient ground thermal model, such as CryoGrid 2.

Despite such challenges, the presented satellite-based model scheme offers great prospects for permafrost monitoring in remote areas that are not covered by in-situ measurements. The good performance regarding thaw depths and the timing of the seasonal thaw progression (Fig. 13) suggests that the results may even help estimating the release of greenhouse gases as a consequence of active layer deepening in a warming climate (Schuur et al., 2015).

6 Conclusions

We present a modeling approach that can estimate the evolution of the ground thermal regime in permafrost areas at 1 km spatial and weekly temporal resolution, based on a combination of satellite data and reanalysis products. The scheme is applied to an area of 16 000 km^2 the Lena River Delta in Northeastern Siberia where measurements of ground temperatures and thaw depths are available to evaluate the performance. The approach is based on the 1D ground thermal model CryoGrid 2 which calculates the time evolution of the subsurface temperature field based on forcing data sets of surface temperature and snow depth for each grid. As forcing data, we synthesize weekly average surface temperatures from MODIS Land Surface Temperature products and near-surface air temperatures from the ERA-interim reanalysis. For snow depth, low-resolution remotely sensed GlobSnow Snow Water Equivalent data are combined with higher-resolution satellite observations of snow extent which facilitates an adequate representation of the snow start and end dates in the model. For the subsurface domain, a classification based on geomorphological mapping has been compiled, which can resolve the large-scale differences in e.g. ground-ice and soil-water contents. The model was subsequently run for a period of 14 years (2000-2014) and the results compared to observations of the ground temperatures and thaw depths at in total nine sites.

- The forcing data sets in general agree very well with multi-year in-situ observations. Monthly average surface temperatures are reproduced within $1°\mathrm{C}$ or less, while the snow start and end dates in most years agree within one week. In a few years, larger deviations of up to three weeks occur.

- The comparison of model results to in-situ measurements suggests that the approach can reproduce the annual temperature amplitude very well. Multi-annual averages of ground temperatures at 2 to $3\,\mathrm{m}$ depth are reproduced with an accuracy of 1 to $1.5°\mathrm{C}$, **while comparison of monthly averages yielded an overall RMSE of 1.1°C and a cold-bias of 0.9°C for the model results. However, due to the small number of validation sites, this accuracy assessment must be considered preliminary.**

- Modeled thaw depths in general agree with in-situ observations within 0.1 to $0.2\,\mathrm{m}$. At one site, comparison with a multi-annual time series of thaw depth measurements suggests that the model scheme is capable of reproducing interannual differences in thaw depths with an accuracy of approx. $0.05\,\mathrm{m}$.

- A sensitivity analysis showcases the influence of the subsurface stratigraphy on both ground temperatures and thaw depths, with temperature differences up to $2°\mathrm{C}$ and thaw depth differences of a factor of three between classes for the same forcing data.

- The warmest average ground temperatures are modeled for grid cells close to the main river channels and areas featuring sandy sediments with low organic contents in the northwestern part of the Lena River Delta. The coldest modeled ground temperatures occur in the eastern part of the delta towards the coastline, and in areas with ice-rich Yedoma sediments.

- The lowest thaw depths are modeled for Yedoma in the southern parts of the delta, as well as in areas with both low snow depths and cold summer surface temperatures in the Northeastern part. The deepest thaw depths are found in areas where the stratigraphy assigns mineral ground with low ice and organic contents.

The results of this study indicate that satellite-based modeling of the ground thermal regime in permafrost areas could eventually become feasible even on continental scales. The largest obstacles are the lack of a standardized classification product on subsurface stratigraphies and thermal properties, as well as shortcomings and limitations of the currently available remote products on snow depth and snow water equivalent. If such limitations can be overcome, remote sensing-based methods could complement and support ground-based monitoring of the ground thermal regime.

Acknowledgements. This work was funded by SatPerm (project no. 239918; Research Council of Norway), COUP (project no. 244903/E10; JPI Climate; Research Council of Norway), Page21 (contract no. GA282700; European Union FP7-ENV), ESA GlobPermfrost (www.globpermafrost.info) and the Department of Geosciences, University of Oslo, Norway. We gratefully acknowledge the support of the Russian-German research station on Samoylov Island. We thank Günther Stoof, Sofia Antonova and Sascha Niemann for their contributions to fieldwork.

References

Abnizova, A., Siemens, J., Langer, M., and Boike, J.: Small ponds with major impact: The relevance of ponds and lakes in permafrost landscapes to carbon dioxide emissions, Global Biogeochemical Cycles, 26, GB2041, doi:10.1029/2011GB004237, 2012.

Antonova, S., Duguay, C. R., Kääb, A., Heim, B., Langer, M., Westermann, S., and Boike, J.: Monitoring Bedfast Ice and Ice Phenology in Lakes of the Lena River Delta Using TerraSAR-X Backscatter and Coherence Time Series, Remote Sensing, 8, 903, doi:10.3390/rs8110903, 2016.

Biskaborn, B. K., Lanckman, J.-P., Lantuit, H., Elger, K., Streletskiy, D. A., Cable, W. L., and Romanovsky, V. E.: The new database of the Global Terrestrial Network for Permafrost (GTN-P), Earth System Science Data, 7, 245–259, doi:10.5194/essd-7-245-2015, http://www.earth-syst-sci-data.net/7/245/2015/, 2015.

Boike, J., Kattenstroth, B., Abramova, K., Bornemann, N., Chetverova, A., Fedorova, I., Fröb, K., Grigoriev, M., Grüber, M., Kutzbach, L., Langer, M., Minke, M., Muster, S., Piel, K., Pfeiffer, E.-M., Stoof, G., Westermann, S., Wischnewski, K., Wille, C., and Hubberten, H.-W.: Baseline characteristics of climate, permafrost and land cover from a new permafrost observatory in the Lena River Delta, Siberia (1998–2011), Biogeosciences, 10, 2105–2128, doi:10.5194/bg-10-2105-2013, http://www.biogeosciences.net/10/2105/2013/, 2013.

Boike, J., Georgi, C., Kirilin, G., Muster, S., Abramova, K., Fedorova, I., Chetverova, A., Grigoriev, M., Bornemann, N., and Langer, M.: Thermal processes of thermokarst lakes in the continuous permafrost zone of northern Siberia – observations and modeling (Lena River Delta, Siberia), Biogeosciences, 12, 5941–5965, doi:10.5194/bg-12-5941-2015, http://www.biogeosciences.net/12/5941/2015/, 2015.

Borge, A. F., Westermann, S., Solheim, I., and Etzelmüller, B.: Strong degradation of palsas and peat plateaus in northern Norway during the last 60 years, The Cryosphere Discussions, 2016, 1–31, doi:10.5194/tc-2016-12, http://www.the-cryosphere-discuss.net/tc-2016-12/, 2016.

Brasnett, B.: A global analysis of snow depth for numerical weather prediction, Journal of Applied Meteorology, 38, 726–740, 1999.

Burgess, M., Smith, S., Brown, J., Romanovsky, V., and Hinkel, K.: The Global Terrestrial Network for Permafrost (GTNet-P): permafrost monitoring contributing to global climate observations, Current Research 2000-E, Geological Survey of Canada, 2000.

Calonne, N., Flin, F., Morin, S., Lesaffre, B., du Roscoat, S. R., and Geindreau, C.: Numerical and experimental investigations of the effective thermal conductivity of snow, Geophysical Research Letters, 38, doi:10.1029/2011GL049234, 2011.

Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES land-surface model, Geoscientific Model Development, 8, 1493–1508, doi:10.5194/gmd-8-1493-2015, http://www.geosci-model-dev.net/8/1493/2015/, 2015.

Dall'Amico, M., Endrizzi, S., Gruber, S., and Rigon, R.: A robust and energy-conserving model of freezing variably-saturated soil, The Cryosphere, 5, 469–484, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

Derksen, C., Silis, A., Sturm, M., Holmgren, J., Liston, G. E., Huntington, H., and Solie, D.: Northwest Territories and Nunavut snow characteristics from a subarctic traverse: Implications for passive microwave remote sensing, Journal of Hydrometeorology, 10, 448–463, 2009.

Derksen, C., Toose, P., Lemmetyinen, J., Pulliainen, J., Langlois, A., Rutter, N., and Fuller, M.: Evaluation of passive microwave brightness temperature simulations and snow water equivalent retrievals through a winter season, Remote Sensing of Environment, 117, 236–248, 2012.

Eicken, H., Dmitrenko, I., Tyshko, K., Darovskikh, A., Dierking, W., Blahak, U., Groves, J., and Kassens, H.: Zonation of the Laptev Sea landfast ice cover and its importance in a frozen estuary, Global and planetary change, 48, 55–83, 2005.

Ekici, A., Beer, C., Hagemann, S., and Hauck, C.: Simulating high-latitude permafrost regions by the JSBACH terrestrial ecosystem model, Geoscientific Model Development, 7, 631–647, 2014.

Endrizzi, S., Gruber, S., Dall'Amico, M., and Rigon, R.: GEOtop 2.0: simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects, Geoscientific Model Development, 7, 2831–2857, doi:10.5194/gmd-7-2831-2014, 2014.

Fedorova, I., Chetverova, A., Bolshiyanov, D., Makarov, A., Boike, J., Heim, B., Morgenstern, A., Overduin, P., Wegner, C., Kashina, V., et al.: Lena Delta hydrology and geochemistry: long-term hydrological data and recent field observations, Biogeosciences, 12, 345–363, 2015.

Fiddes, J., Endrizzi, S., and Gruber, S.: Large-area land surface simulations in heterogeneous terrain driven by global data sets: application to mountain permafrost, The Cryosphere, 9, 411–426, doi:10.5194/tc-9-411-2015, http://www.the-cryosphere.net/9/411/2015/, 2015.

Gallo, K., Hale, R., Tarpley, D., and Yu, Y.: Evaluation of the relationship between air and land surface temperature under clear-and cloudy-sky conditions, Journal of Applied Meteorology and Climatology, 50, 767–775, 2011.

Gisnås, K., Westermann, S., Schuler, T. V., Litherland, T., Isaksen, 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, http://www.the-cryosphere.net/8/2063/2014/, 2014.

Gisnås, K., Westermann, S., Schuler, T. V., Melvold, K., and Etzelmüller, B.: Small-scale variation of snow in a regional permafrost model, The Cryosphere Discussions, 9, 6661–6696, doi:10.5194/tcd-9-6661-2015, http://www.the-cryosphere-discuss.net/9/6661/2015/, 2015.

Grigoriev, M.: Cryomorphogenesis of the Lena River mouth area, Russian Academy of Science, Siberian Branch, Yakutsk, Russia (in Russian), 1993.

Grigoriev, M., Imaev, V., Imaeva, L., Kozmin, B., Kunitzkiy, V., Lationov, A., Mikulenko, K., Skrjabin, R., and Timirsin, K.: Geology, seismicity and cryogenic processes in the arctic areas of Western Yakutia, Yakut Scientific Centre, Russian Academy of Science, Siberian Branch, Yakutsk, Russia (in Russian), 1996.

Gubler, S., Fiddes, J., Keller, M., and Gruber, S.: Scale-dependent measurement and analysis of ground surface temperature variability in alpine terrain, The Cryosphere, 5, 431–443, doi:10.5194/tc-5-431-2011, http://www.the-cryosphere.net/5/431/2011/, 2011.

Gunn, G. E., Duguay, C. R., Derksen, C., Lemmetyinen, J., and Toose, P.: Evaluation of the HUT modified snow emission model over lake ice using airborne passive microwave measurements, Remote sensing of environment, 115, 233–244, 2011.

Hachem, S., Allard, M., and Duguay, C.: Using the MODIS land surface temperature product for mapping permafrost: an application to northern Québec and Labrador, Canada, Permafrost and Periglacial Processes, 20, 407–416, 2009.

Hudson, S. R. and Brandt, R. E.: A look at the surface-based temperature inversion on the Antarctic Plateau, Journal of Climate, 18, 1673–1696, 2005.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J., Kuhry, P., and Swanson, D.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth System Science Data, 5, 3–13, 2013.

Jafarov, E. E., Marchenko, S. S., and Romanovsky, V. E.: Numerical modeling of permafrost dynamics in Alaska using a high spatial resolution dataset, The Cryosphere, 6, 613–624, 2012.

Jones, B. M., Grosse, G., Arp, C., Jones, M., Walter Anthony, K., and Romanovsky, V.: Modern thermokarst lake dynamics in the continuous permafrost zone, northern Seward Peninsula, Alaska, Journal of Geophysical Research: Biogeosciences, 116, doi:10.1029/2011JG001666, 2011.

Jury, W. and Horton, R.: Soil physics, John Wiley & Sons, Hoboken, New Jersey, USA, 2004.

Kitaev, L., Kislov, A., Krenke, A., Razuvaev, V., Martuganov, R., and Konstantinov, I.: The snow cover characteristics of northern Eurasia and their relationship to climatic parameters, Boreal Environment Research, 7, 437–445, 2002.

Kryamyarya, V., Mokhov, V., Bol'shiyanov, D., and Rekant, P.: State Geological Map of the Russian Federation, scale 1:1 000 000, map of Quaternary formations, S-52 Lena River Delta, 2011.

Kutzbach, L., Wagner, D., and Pfeiffer, E.-M.: Effect of microrelief and vegetation on methane emission from wet polygonal tundra, Lena Delta, Northern Siberia, Biogeochemistry, 69, 341–362, 2004.

Kutzbach, L., Wille, C., and Pfeiffer, E.-M.: The exchange of carbon dioxide between wet arctic tundra and the atmosphere at the Lena River Delta, Northern Siberia, Biogeosciences, 4, 869–890, doi:10.5194/bg-4-869-2007, http://www.biogeosciences.net/4/869/2007/, 2007.

Langer, M., Westermann, S., and Boike, J.: Spatial and temporal variations of summer surface temperatures of wet polygonal tundra in Siberia-implications for MODIS LST based permafrost monitoring, Remote Sensing of Environment, 114, 2059–2069, 2010.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The surface energy balance of a polygonal tundra site in northern Siberia Part 1: Spring to fall, The Cryosphere, 5, 151–171, 2011a.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The surface energy balance of a polygonal tundra site in northern Siberia Part 2: Winter, Cryosphere, 5, 509–524, 2011b.

Langer, M., Westermann, S., Heikenfeld, M., Dorn, W., and Boike, J.: Satellite-based modeling of permafrost temperatures in a tundra lowland landscape, Remote Sensing of Environment, 135, 12–24, 2013.

Langer, M., Westermann, S., Walter Anthony, K., Wischnewski, K., and Boike, J.: Frozen ponds: production and storage of methane during the Arctic winter in a lowland tundra landscape in northern Siberia, Lena River delta, Biogeosciences, 12, 977–990, 2015.

Lawrence, D. M., Slater, A. G., and Swenson, S. C.: Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4, Journal of Climate, 25, 2207–2225, doi:10.1175/JCLI-D-11-00334.1, 2012.

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M., and Zappa, M.: ALPINE3D: a detailed model of mountain surface processes and its application to snow hydrology, Hydrological Processes, 20, 2111–2128, 2006.

Lemmetyinen, J., Kontu, A., Kärnä, J.-P., Vehviläinen, J., Takala, M., and Pulliainen, J.: Correcting for the influence of frozen lakes in satellite microwave radiometer observations through application of a microwave emission model, Remote Sensing of Environment, 115, 3695–3706, 2011.

Liljedahl, A. K., Boike, J., Daanen, R. P., Fedorov, A. N., Frost, G. V., Grosse, G., Hinzman, L. D., Iijma, Y., Jorgenson, J. C., Matveyeva, N., et al.: Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology, Nature Geoscience, 9, 312–318, 2016.

Liu, Y., Key, J., Frey, R., Ackerman, S., and Menzel, W.: Nighttime polar cloud detection with MODIS, Remote Sensing of Environment, 92, 181–194, 2004.

Luojus, K., Pulliainen, J., Takala, M., Derksen, C., Rott, H., Nagler, T., Solberg, R., Wiesmann, A., Metsamaki, S., Malnes, E., et al.: Investigating the feasibility of the GlobSnow snow water equivalent data for climate research purposes, in: Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, doi:10.1109/IGARSS.2010.5741987, 2010.

Morgenstern, A., Rohr, C., Grosse, G., and Grigoriev, M.: The Lena River Delta - inventory of lakes and geomorphological terraces, Alfred Wegener Institute for Polar and Marine Research - Research Unit Potsdam, doi:10.1594/PANGAEA.758728, 2011.

Morgenstern, A., Ulrich, M., Günther, F., Roessler, S., Fedorova, I. V., Rudaya, N. A., Wetterich, S., Boike, J., and Schirrmeister, L.: Evolution of thermokarst in East Siberian ice-rich permafrost: A case study, Geomorphology, 201, 363–379, 2013.

Muster, S., Langer, M., Heim, B., Westermann, S., and Boike, J.: Subpixel heterogeneity of ice-wedge polygonal tundra: a multi-scale analysis of land cover and evapotranspiration in the Lena River Delta, Siberia, Tellus B, 64, 17 301, doi:10.3402/tellusb.v64i0.17301, 2012.

Muster, S., Heim, B., Abnizova, A., and Boike, J.: Water body distributions across scales: a remote sensing based comparison of three arctic tundra wetlands, Remote Sensing, 5, 1498–1523, 2013.

Nitze, I. and Grosse, G.: Detection of landscape dynamics in the Arctic Lena Delta with temporally dense Landsat time-series stacks, Remote Sensing of Environment, 181, 27–41, 2016.

Østby, T. I., Schuler, T. V., and Westermann, S.: Severe cloud contamination of MODIS Land Surface Temperatures over an Arctic ice cap, Svalbard, Remote Sensing of Environment, 142, 95–102, 2014.

Osterkamp, T. and Romanovsky, V.: Evidence for warming and thawing of discontinuous permafrost in Alaska, Permafrost and Periglacial Processes, 10, 17–37, 1999.

Panda, S. K., Marchenko, S. S., and Romanovsky, V. E.: High-Resolution Permafrost Modeling in Denali National Park and Preserve, Natural Resource Technical Report, National Park Service, Fort Collins, Colorado, pp. 1–44, https://irma.nps.gov/App/Reference/Profile/2208990, 2014.

Plug, L. J. and West, J.: Thaw lake expansion in a two-dimensional coupled model of heat transfer, thaw subsidence, and mass movement, Journal of Geophysical Research: Earth Surface (2003–2012), 114, F01 002, doi:10.1029/2006JF000740, 2009.

Pulliainen, J.: Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations, Remote sensing of Environment, 101, 257–269, 2006.

Pulliainen, J. T., Grandell, J., and Hallikainen, M. T.: HUT snow emission model and its applicability to snow water equivalent retrieval, IEEE Transactions on Geoscience and Remote Sensing, 37, 1378–1390, 1999.

Rachold, V. and Grigoriev, M. N.: Russian-German Cooperation SYSTEM LAPTEV SEA 2000: The Lena Delta 1998 Expedition, Berichte zur Polarforschung (Reports on Polar Research), 315, 1–259, 1999.

Raleigh, M. S., Landry, C. C., Hayashi, M., Quinton, W. L., and Lundquist, J. D.: Approximating snow surface temperature from standard temperature and humidity data: New possibilities for snow model and remote sensing evaluation, Water Resources Research, 49, 8053–8069, 2013.

Rees, A., Derksen, C., English, M., Walker, A., and Duguay, C.: Uncertainty in snow mass retrievals from satellite passive microwave data in lake-rich high-latitude environments, Hydrological Processes, 20, 1019–1022, doi:10.1002/hyp.6076, 2006.

Romanovsky, V., Smith, S., and Christiansen, H.: Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007–2009: A synthesis, Permafrost and Periglacial processes, 21, 106–116, 2010.

Sachs, T., Giebels, M., Boike, J., and Kutzbach, L.: Environmental controls on CH4 emission from polygonal tundra on the microsite scale in the Lena river delta, Siberia, Global Change Biology, 16, 3096–3110, 2010.

Schirrmeister, L.: Expeditions in Siberia in 2005, Berichte zur Polar-und Meeresforschung (Reports on Polar and Marine Research), 550, 289, 2007.

Schirrmeister, L., Grosse, G., Schwamborn, G., Andreev, A., Meyer, H., Kunitsky, V. V., Kuznetsova, T. V., Dorozhkina, M., Pavlova, E., Bobrov, A., et al.: Late Quaternary history of the accumulation plain north of the Chekanovsky Ridge (Lena Delta, Russia): a multidisciplinary approach, Polar Geography, 27, 277–319, 2003.

Schirrmeister, L., Grosse, G., Schnelle, M., Fuchs, M., Krbetschek, M., Ulrich, M., Kunitsky, V., Grigoriev, M., Andreev, A., Kienast, F., et al.: Late Quaternary paleoenvironmental records from the western Lena Delta, arctic Siberia, Palaeogeography, Palaeoclimatology, Palaeoecology, 299, 175–196, 2011.

Schneider, J., Grosse, G., and Wagner, D.: Land cover classification of tundra environments in the Arctic Lena Delta based on Landsat 7 ETM+ data and its application for upscaling of methane emissions, Remote Sensing of Environment, 113, 380–391, 2009.

Schuur, E., McGuire, A., Schädel, C., Grosse, G., Harden, J., Hayes, D., Hugelius, G., Koven, C., Kuhry, P., Lawrence, D., Natali, S., Olefeldt, D., Romanovsky, V., Schaefer, K., Turetsky, M., Treat, C., and Vonk, J.: Climate change and the permafrost carbon feedback, Nature, 520, 171–179, 2015.

Schwamborn, G., Dix, J., Bull, J., and Rachold, V.: High-resolution seismic and ground penetrating radar-geophysical profiling of a thermokarst lake in the western Lena Delta, Northern Siberia, Permafrost and Periglacial Processes, 13, 259–269, doi:10.1002/ppp.430, 2002a.

Schwamborn, G., Rachold, V., and Grigoriev, M. N.: Late Quaternary sedimentation history of the Lena Delta, Quaternary International, 89, 119–134, 2002b.

Strauss, J., Schirrmeister, L., Grosse, G., Wetterich, S., Ulrich, M., Herzschuh, U., and Hubberten, H.-W.: The deep permafrost carbon pool of the Yedoma region in Siberia and Alaska, Geophysical Research Letters, 40, 6165–6170, 2013.

Sturm, M., Holmgren, J., and Liston, G. E.: A seasonal snow cover classification system for local to global applications, Journal of Climate, 8, 1261–1283, 1995.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., and Bojkov, B.: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sensing of Environment, 115, 3517–3529, 2011.

Ulrich, M., Grosse, G., Chabrillat, S., and Schirrmeister, L.: Spectral characterization of periglacial surfaces and geomorphological units in the Arctic Lena Delta using field spectrometry and remote sensing, Remote Sensing of Environment, 113, 1220–1235, 2009.

Weismüller, J., Wollschläger, U., Boike, J., Pan, X., Yu, Q., and Roth, K.: Modeling the thermal dynamics of the active layer at two contrasting permafrost sites on Svalbard and on the Tibetan Plateau, The Cryosphere, 5, 741–757, 2011.

Westermann, S., Boike, J., Langer, M., Schuler, T., and Etzelmüller, B.: Modeling the impact of wintertime rain events on the thermal regime of permafrost, The Cryosphere, 5, 1697–1736, 2011a.

Westermann, S., Langer, M., and Boike, J.: Spatial and temporal variations of summer surface temperatures of high-arctic tundra on Svalbard - Implications for MODIS LST based permafrost monitoring, Remote Sensing of Environment, 115, 908–922, 2011b.

Westermann, S., Langer, M., and Boike, J.: Systematic bias of average winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway, Remote Sensing of Environment, 118, 162–167, 2012.

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

Westermann, S., Østby, T. I., Gisnås, K., Schuler, T. V., and Etzelmüller, B.: A ground temperature map of the North Atlantic permafrost region based on remote sensing and reanalysis data, The Cryosphere, 9, 1303–1319, doi:10.5194/tc-9-1303-2015, http://www.the-cryosphere.net/9/1303/2015/, 2015.

Westermann, S., Langer, M., Boike, J., Heikenfeld, M., Peter, M., Etzelmüller, B., and Krinner, G.: Simulating the thermal regime and thaw processes of ice-rich permafrost ground with the land-surface model CryoGrid 3, Geoscientific Model Development, 9, 523–546, doi:10.5194/gmd-9-523-2016, http://www.geosci-model-dev.net/9/523/2016/, 2016.

Wille, C., Kutzbach, L., Sachs, T., Wagner, D., and Pfeiffer, E.: Methane emission from Siberian arctic polygonal tundra: eddy covariance measurements and modeling, Global Change Biology, 14, 1395–1408, 2008.

Yang, D., Kane, D. L., Hinzman, L. D., Zhang, X., Zhang, T., and Ye, H.: Siberian Lena River hydrologic regime and recent change, Journal of Geophysical Research: Atmospheres, 107, doi:10.1029/2002JD002542, 2002.

Yen, Y.-C.: Review of thermal properties of snow, ice and sea ice, Tech. rep., DTIC Document, 1981.

Yershov, E., Kondrat'yeva, K., Loginov, V., and Sychev, I.: Geocryological map of Russia and neighbouring republics, Faculty of Geology, Chair of Geocryology, Lomonosov Moscow State University, 1991.

Yi, S., Wischnewski, K., Langer, M., Muster, S., and Boike, J.: Freeze/thaw processes in complex permafrost landscapes of northern Siberia simulated using the TEM ecosystem model: impact of thermokarst ponds and lakes, Geoscientific Model Development, 7, 1671–1689, 2014.

Zhang, Y., Wang, X., Fraser, R., Olthof, I., Chen, W., Mclennan, D., Ponomarenko, S., and Wu, W.: Modelling and mapping climate change impacts on permafrost at high spatial resolution for an Arctic region with complex terrain, The Cryosphere, 7, 1121–1137, 2013.

Zubrzycki, S., Kutzbach, L., Pfeiffer, E., and Vakhrameeva, P.: Variability of Soil Organic Carbon Stocks of Different Permafrost-Affected Soils: Initial Results from a North-South Transect in Siberia, in: Proceedings of the Tenth International Conference on Permafrost, Salekhard, Russia, pp. 485–490, 2012.

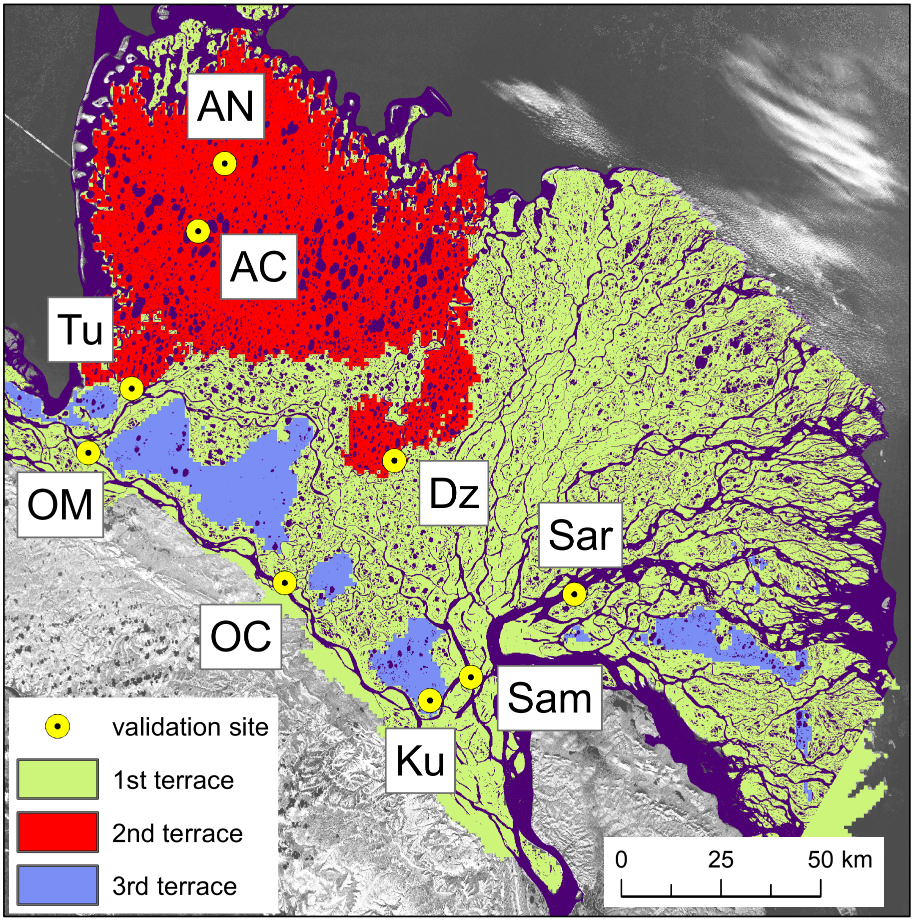

Figure 1. The Lena River Delta with the three stratigraphic classes distinguished in the ground thermal modeling (Sect. 3.2) and sites with in-situ observations (Sect. 2.2.2) employed for model validation. AN: Arga Island, north; AC: Arga Island, center; Dz: Dzhipperies Island; Ku: Kurungnakh Island; OC: Olenyokskaya Channel, center; OM: Olenyokskaya Channel, mouth; Sam: Samoylov Island; Sar: Sardakh Island; Tu: Turakh Island.

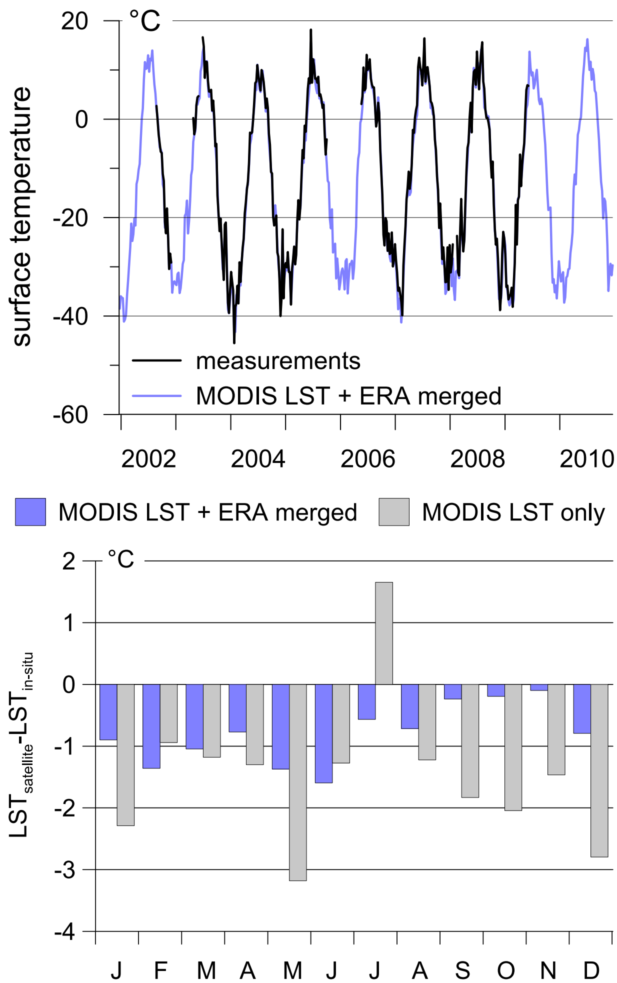

Figure 2. Top: daily average surface temperatures measured on Samoylov Island (Langer et al., 2013; Boike et al., 2013) vs. surface temperatures synthesized from MODIS LST and ERA reanalysis. Bottom: difference between satellite-derived LST and in-situ measurements for monthly averages of periods when in-situ measurements are available (see top figure). See text.

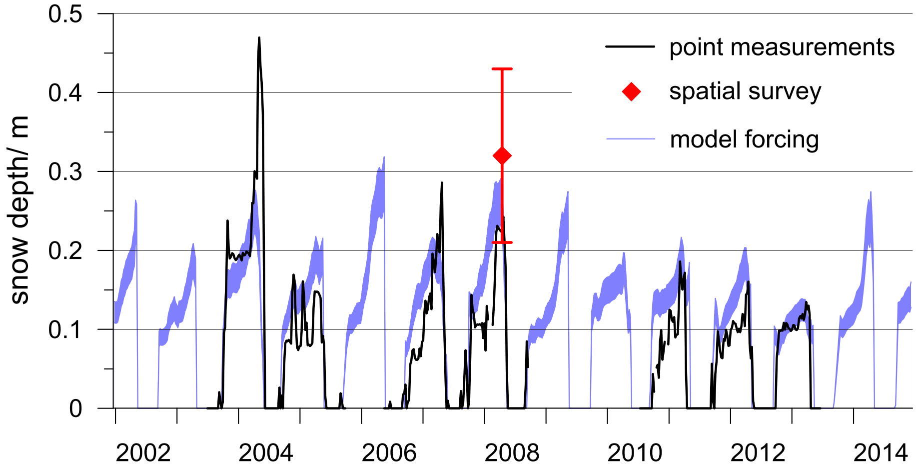

Figure 3. Modeled and measured snow depths on Samoylov Island (Boike et al., 2013). **The point measurements are conducted with an ultrasonic ranging sensor (data smoothed with running average filter with window size of one week, corresponding to the temporal resolution of the model forcing), the spatial survey is based on manual measurements at 216 points in polygonal tundra conducted between 25 April and 2 May 2008 (Fig. 6a, Boike et al., 2013).** The blue area depicts the spread between model runs with snow densities of 200 and 250 $\mathrm{kg\,m^{-3}}$.

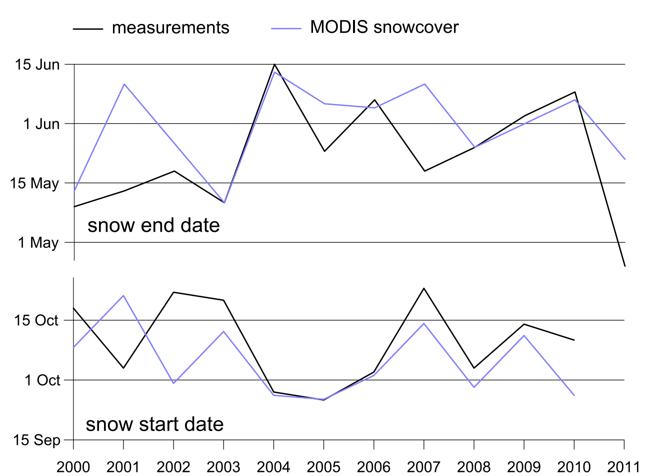

Figure 4. Modeled and measured snow start and end on Samoylov Island (Boike et al., 2013).

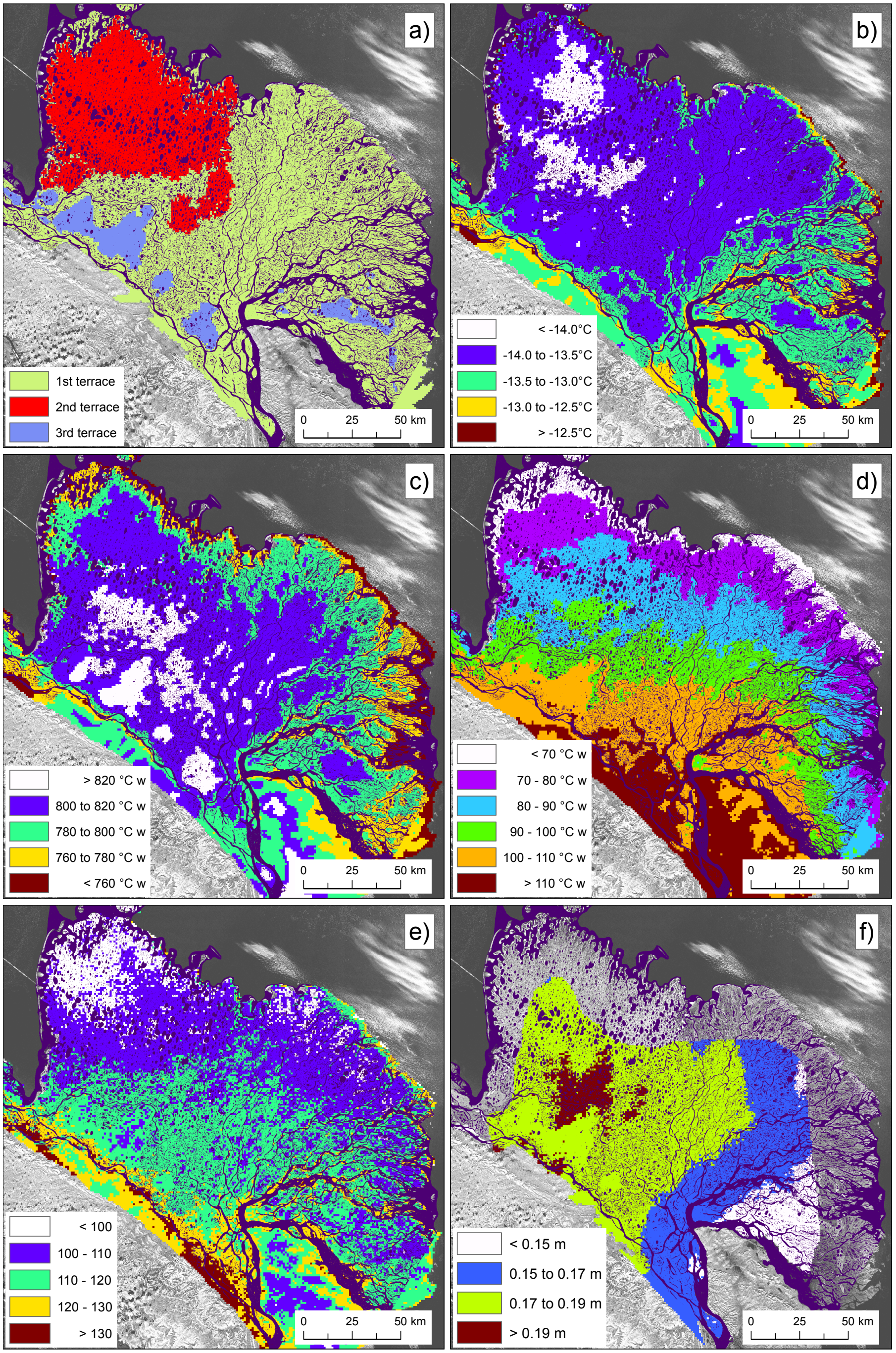

Figure 5. Spatial distribution of model input data sets in the LRD (Sects. 3.2, 3.3): a) subsurface classification (compare Table 1); b) average surface temperature 2004-2013; c) average freezing degree **weeks** 2004-2013; d) average thawing degree **weeks** 2004-2013; e) average number of snow-free days 2004-2013; f) average snow depth 2004-2013 for a snow density of $225 \mathrm{kg\,m^{-3}}$.

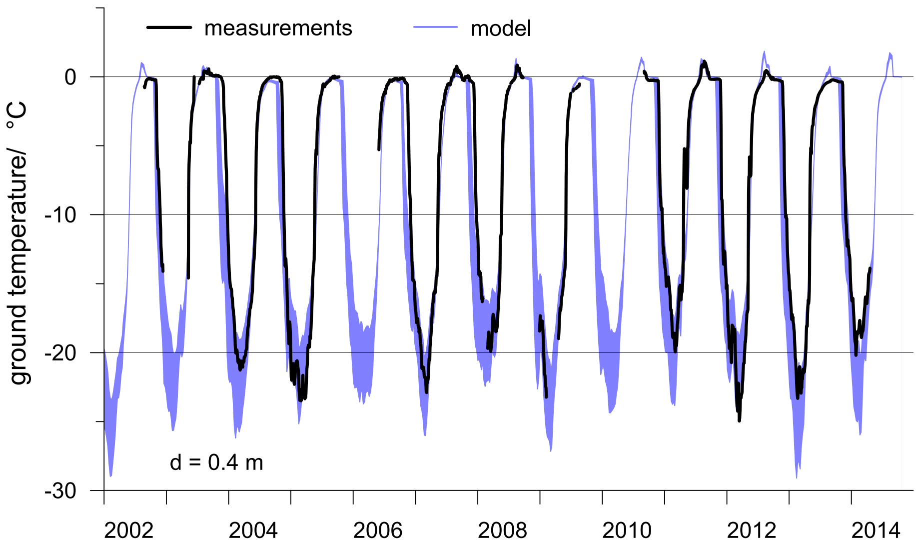

Figure 6. Modeled and measured ground temperatures at a depth of $0.4\,\mathrm{m}$ at a wet polygon center on Samoylov Island (Boike et al., 2013). The blue area depicts the spread between model runs with snow densities of 200 and $250\,\mathrm{kg\,m^{-3}}$. **The temperature sensor drifted by about -0.2°C (at 0°C) in the shown period.**

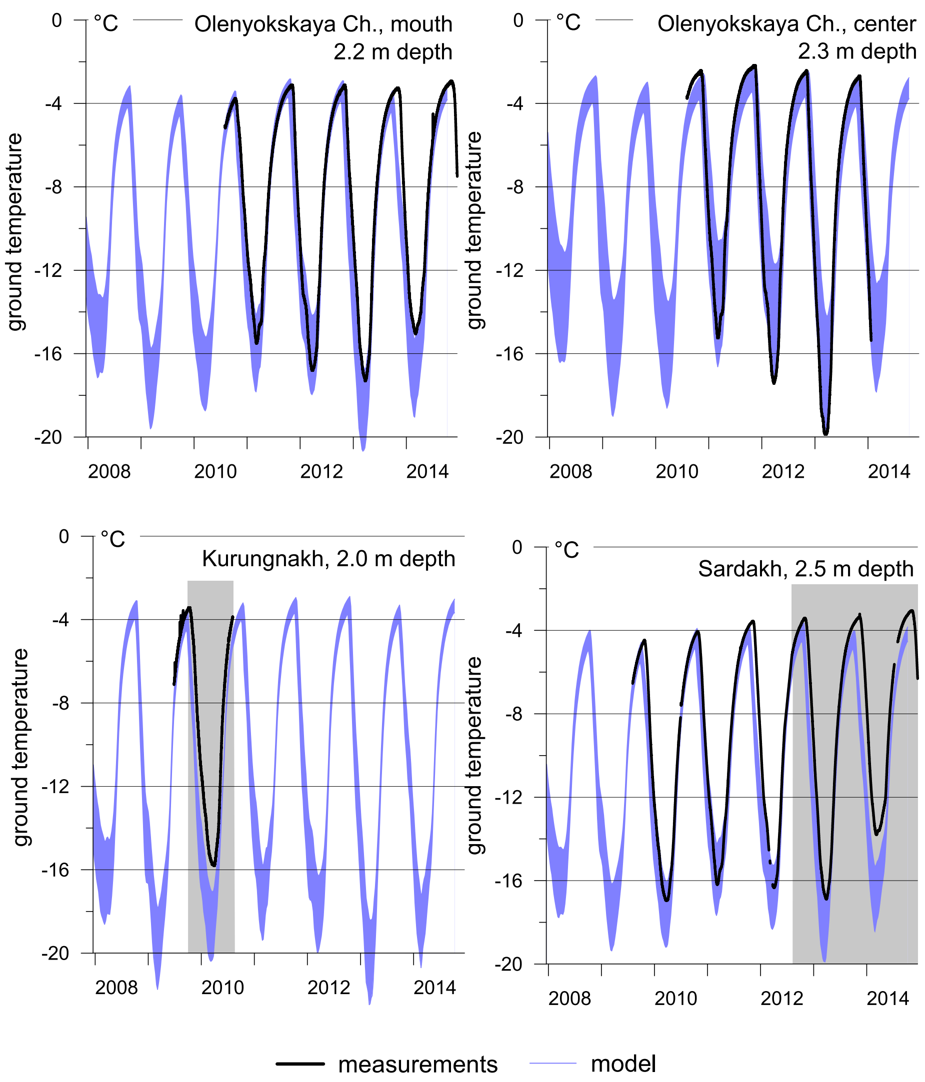

Figure 7. Modeled and measured ground temperatures at depths of 2.0-2.5 m at four locations in the LRD. The blue area depicts the spread between model runs with snow densities of 200 and 250 $\mathrm{kg\,m^{-3}}$. **Periods for which in-situ data are affected by thermokarst are marked in grey. These should not be used for comparison, see text.**

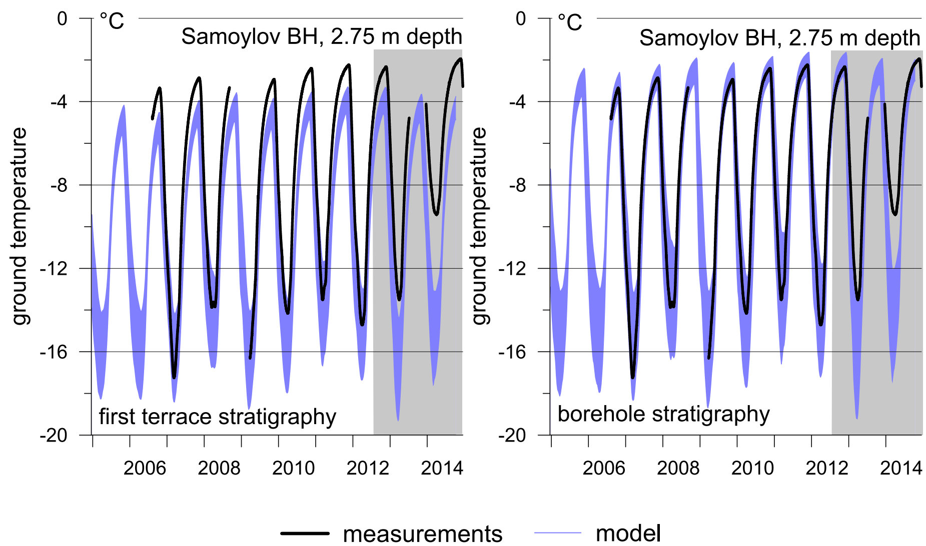

Figure 8. Modeled and measured ground temperatures for the borehole on Samoylov Island. Left: subsurface stratigraphy of the first terrace (Table 1). Right: stratigraphy adapted to the true ground conditions at the borehole (0-0.5 m: 30% water/ice, 10% air, 60% mineral, sand; 0.5-9 m: 40% water/ice, 60% mineral, sand; deeper layers as for first terrace, Sect. 4.2.1). The blue area depicts the spread between model runs with snow densities of 200 and 250 $\mathrm{kg\,m^{-3}}$. **Periods for which in-situ data are affected by new installations at the Samoylov station are marked in grey. These should not be used for comparison, see text.**

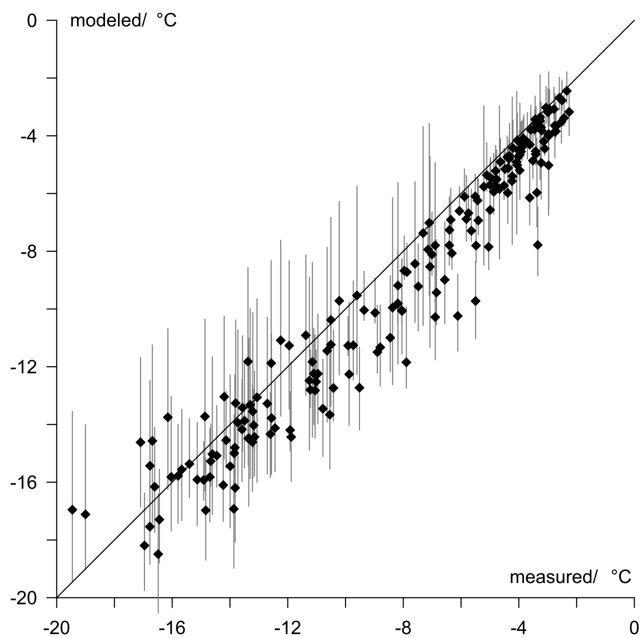

Figure 9. Modeled and measured monthly average ground temperatures for the LRD boreholes and 1:1 line (n=185, data as shown in Figs. 7 and 8 right). Olenyokskaya Channel mouth and center: full time series; Kurungnakh Island: time series until September 2009; Samoylov Island: time series until August 2012, model data with borehole stratigraphy (Fig. 8 right); Sardakh Island: time series until August 2012. Vertical bars: spread between model runs with snow densities of 200 and 250 $\mathrm{kg\,m^{-3}}$; diamonds: model run with snow density 225 $\mathrm{kg\,m^{-3}}$.

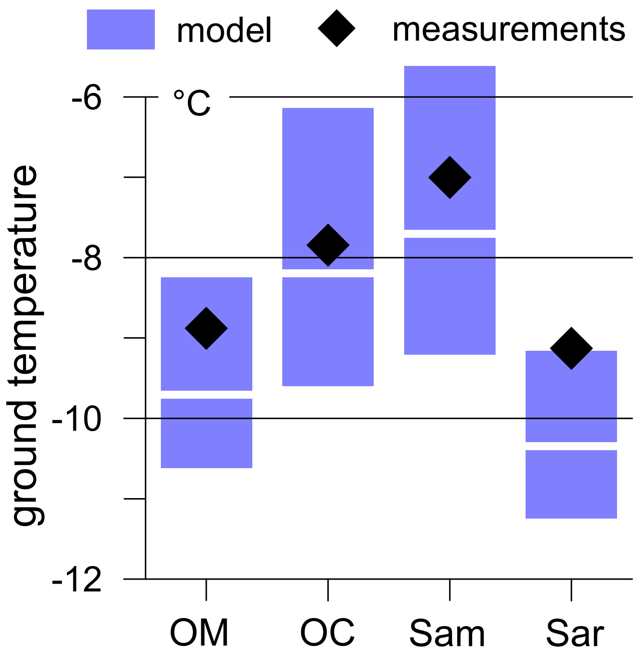

Figure 10. Modeled and measured annual average ground temperatures for the LRD boreholes for the two-year period September 2010 to August 2012 (OM: Olenyokskaya Channel mouth; OC: Olenyokskaya Channel center; Sam: Samoylov Island borehole; Sar: Sardakh Island). Blue bar: spread between model runs with snow densities of 200 and 250 $\mathrm{kg\,m^{-3}}$; white line: model run with snow density 225 $\mathrm{kg\,m^{-3}}$. The ground temperatures correspond to the depths given in Figs. 7 and 8, for Samoylov, the simulations for the borehole stratigraphy (Sect. 4.2.1, Fig. 8 right) are presented.

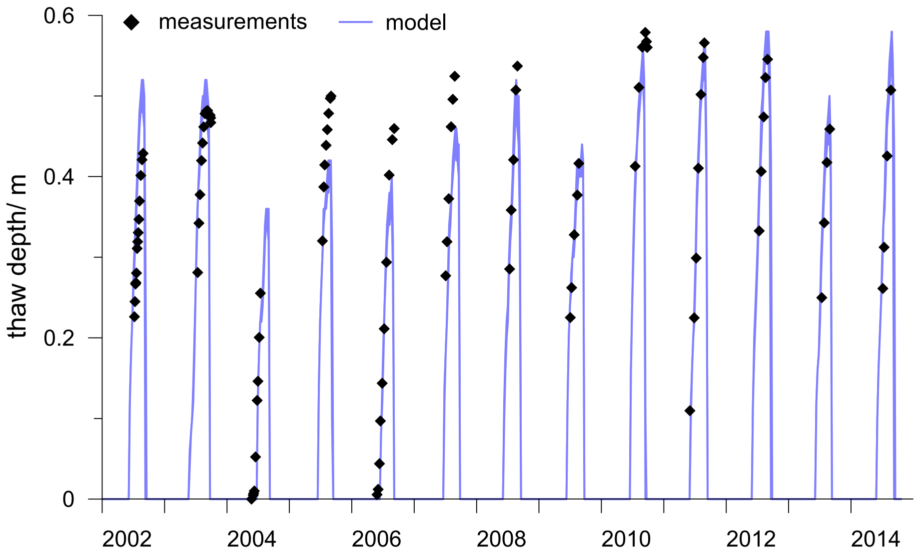

Figure 11. Modeled and measured thaw depths on Samoylov Island. The measurements correspond to the average of 150 locations on Samoylov Island (Boike et al., 2013). The average standard deviation of the measurements (i.e. the spatial variability of thaw depths) is 0.06 m. The blue area depicts the spread between model runs with snow densities of 200 and 250 $\mathrm{kg\,m^{-3}}$.

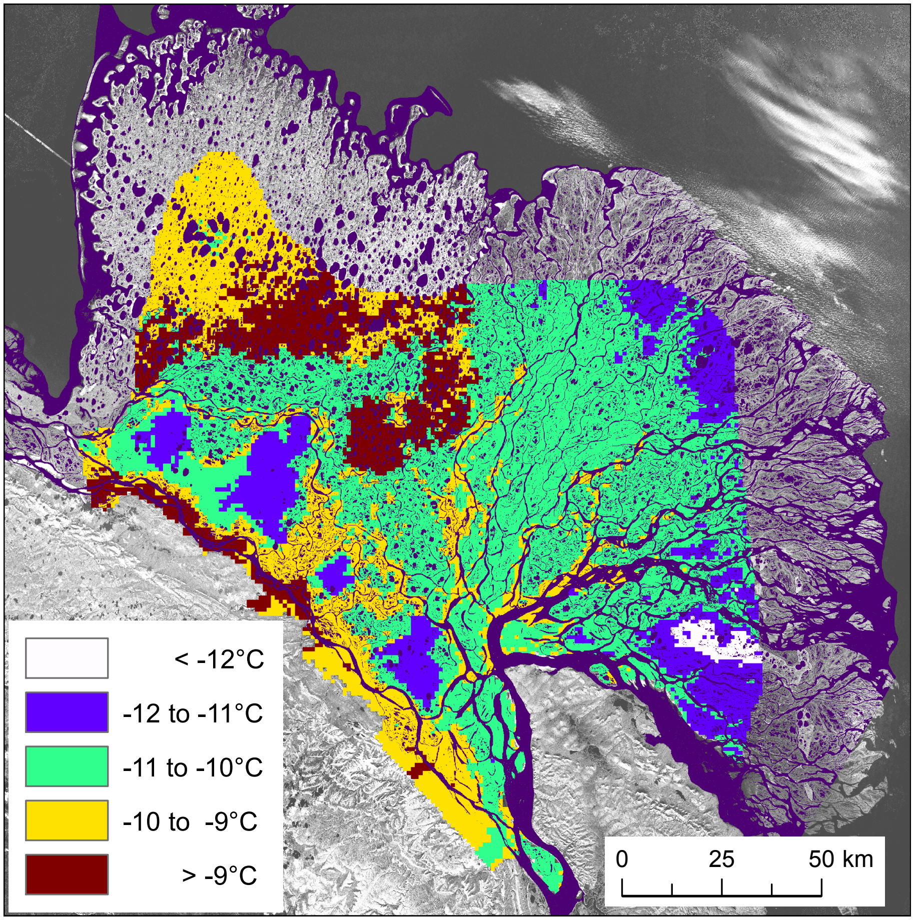

Figure 12. Modeled average ground temperatures at $1\,\mathrm{m}$ depth for the period 2004-2013, with a snow density of $225\,\mathrm{kg\,m^{-3}}$.

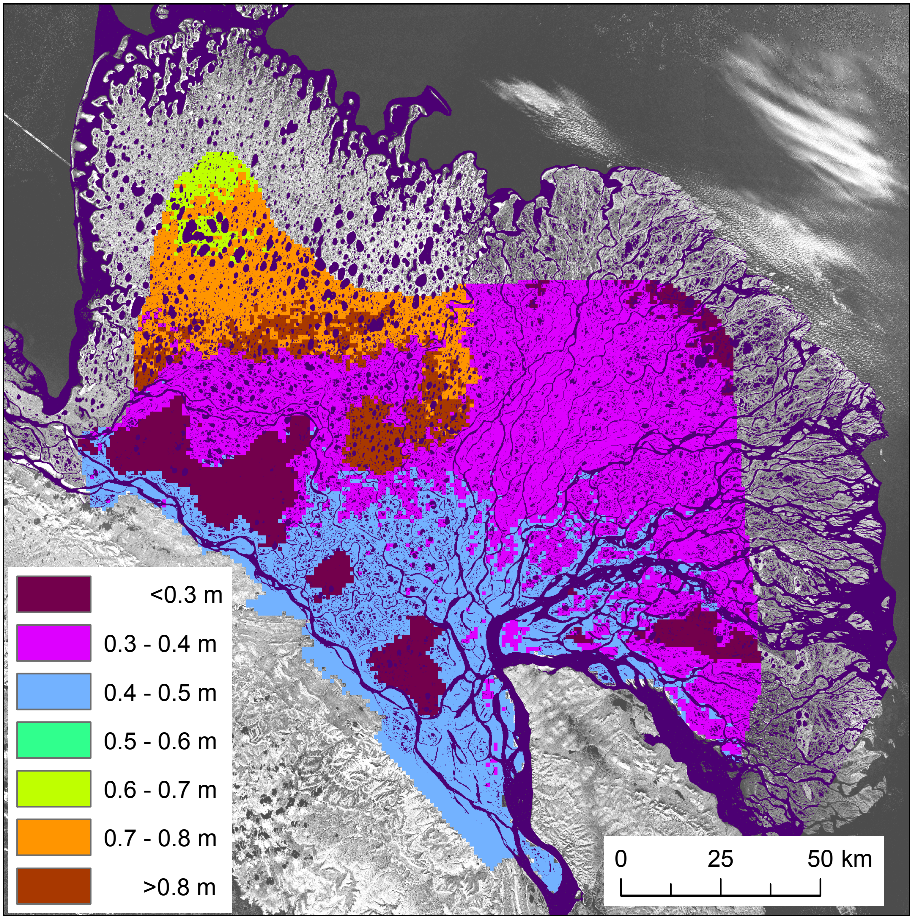

Figure 13. Modeled average maximum thaw depths for the period 2004-2013, with a snow density of $225\,\mathrm{kg\,m^{-3}}$.

Table 1. Subsurface stratigraphies for the three LRD terraces with volumetric fractions of the soil constituents and sediment type assigned to each layer.

depth [m]	water/ice	mineral	organic	air	type
First Terrace					
0–0.15	0.6	0.1	0.15	0.15	sand
0.15–9	0.65	0.3	0.05	0.0	silt
>9	0.3	0.7	0.0	0.0	sand
Second Terrace					
0-10	0.4	0.6	0.0	0.0	sand
>10	0.3	0.7	0.0	0.0	sand
Third Terrace - Yedoma					
0–0.15	0.3	0.1	0.1	0.5	sand
0.15–20	0.7	0.25	0.05	0.0	sand
>20	0.3	0.7	0.0	0.0	sand

Table 2. Modeled and measured thaw depths in the LRD for confining snow depths of $200 \mathrm{kg\,m^{-3}}$ and $250 \mathrm{kg\,m^{-3}}$.

Site	date	measured	modeled	
			$200 \mathrm{kg\,m^{-3}}$	$250 \mathrm{kg\,m^{-3}}$
Samoylov Island	2002-2014	see Fig. 11 for detailed comparison		
Olenyokskaya Ch., center	16 Aug 2010	0.6 m	0.55 m	0.51 m
Arga Island, North	11 Aug 2010	0.9-1.0 m	0.84 m	0.80 m
Arga Island, Center	3 Aug 1998	0.6 m	0.61 m	0.60 m
			average 3 Aug, 2001-2010	
Dzhipperies Island	23 Jul 1998	0.7 m	0.68 m	0.64 m
			average 23 Jul, 2001-2010	
Turakh Island	20-29 Aug 2005	1.0-1.1 m	0.74 m	0.70 m
Olenyokskaya Ch., mouth	14 Aug 2010	0.2 m	0.29 m	0.27 m
Kurungnakh Island	14/15 Jul 2013	0.12-0.18 m	0.19-0.20 m	0.19-0.20 m
(9 sites,	9/10 Aug 2013	0.16-0.22 m	0.26-0.28 m	0.20-0.21 m
6 grid cells)	26 Aug 2013	0.21-0.26 m	0.29-0.30 m	0.28-0.29 m

Table 3. Sensitivity of modeled average ground temperatures at 1 m depth and average maximum thaw depth over the period 2004-2013. All simulations with snow density 225 $\mathrm{kg\,m^{-3}}$.

Site	ground temperature/ $^{\circ}\mathrm{C}$			thaw depth/ m		
	1st	2nd	3rd	1st	2nd	3rd
	terrace stratigraphy			terrace stratigraphy		
Arga Island, north	-11.6	-10.3	-12.2	0.30	0.69	0.19
Arga Island, center	-11.3	-10.0	-12.1	0.30	0.71	0.19
Dzhipperies Island	-10.6	-9.0	-11.5	0.39	0.86	0.24
Kurungnakh Island	-10.6	-9.0	-11.5	0.46	0.96	0.28
Olenyokskaya Ch., mouth	-9.7	-8.0	-10.8	0.43	0.93	0.26
Olenyokskaya Ch., center	-9.5	-7.9	-10.6	0.45	0.96	0.28
Samoylov Island	-10.2	-8.6	-11.1	0.46	0.97	0.28
Sardakh Island	-10.5	-9.0	-11.3	0.41	0.90	0.25
Turakh Island	-10.7	-9.2	-11.6	0.38	0.94	0.22

5. Supplementary Material

Supplementary material

Transient modeling of the ground thermal conditions using satellite data in the Lena River Delta, Siberia

Sebastian Westermann, Maria Peter, Moritz Langer, Georg Schwamborn, Lutz Schirrmeister, Bernd Etzelmüller, and Julia Boike

1 Comparison of CryoGrid 2 snow cover forcing to independent model data sets

The model forcing of snow depths is synthesized from GlobSnow SWE (Luojus et al., 2010, Takala et al., 2011) and MODIS snow extent. Comparison to in-situ data from Samoylov Island in the SE part of the LRD yielded a satisfactory performance of the model forcing (Sect. 4.1.1). However, observations are not available from other parts of the LRD, so that significant uncertainty remains about the spatial pattern of SWE, in particular considering the abundance of water bodies which affects microwave emission and thus SWE retrievals (Sect. 5.1.2). Therefore, we compare the spatial distribution to independent data sets obtained from atmospheric model schemes, which do not make use of passive microwave retrievals and are thus fully unaffected by water bodies. First, we employ the Canadian Meteorological Centre Snow Depth Analysis (hereafter referred to as CMC), which provides SWE values at a spatial resolution of 24km (Brasnett, 1999; Brown & Brasnett, 2015), comparable to GlobSnow SWE. In CMC, a background field of SWE is calculated from snowfall in an atmospheric circulation model, which is subsequently updated by assimilating in-situ snow depth measurements from WMO (World Meteorological Organization) ground stations. Both CMC and GlobSnow therefore make use of snow data from Tiksi which is likely to affect the absolute values, but not the spatial pattern, as no other WMO station is located close-by to the W or N of the LRD. For the comparison to the CryoGrid 2 snow forcing, we have used the CMC monthly SWE data set for the period 2004 to 2013, corresponding to the period displayed in Fig. 5. The result of the spatial comparison is shown in Fig. S1. Both products show a similar spatial pattern and absolute values generally agree to within 10 mm, with CMC generally featuring lower values than the snow cover model forcing. When interpolated to the 1 km scale of the model forcing data, a significant correlation between the data sets is found (r^2=0.71). In CMC, the highest values occur in the coastal regions in the N and W, while the CryoGrid 2 forcing features the highest values in the more central areas of the LRD, although the difference to the coastal areas in the W and N is rather negligible. However, it is unclear which of the data sets is a better representation of true conditions, and only systematic in-situ measurements could clarify this issue.

The coarse-scale pattern in the LRD is further backed up by precipitation output from the ERA-interim reanalysis, which is fully independent of in-situ snow measurements (although measurements of atmospheric variables form Tiksi are assimilated). Sea ice concentrations and sea surface temperatures from satellite retrievals are prescribed as boundary conditions in ERA-interim (Dee et al., 2011), which is important for modeling moisture uptake of the atmosphere. Fig. S2 displays the annual average of precipitation falling in the months October to June at 2m-air temperatures of less then 0°C, which is a coarse proxy for the snowfall. Despite of the very coarse resolution of the land-sea mask, the pattern is similar to both the CMC and the snow cover forcing, with lowest values in the SE part and a values increasing towards the W. For areas cells flagged as sea, highly variable precipitation is modeled, possibly related to sea ice concentration in the Laptev Sea. While the absolute values are not directly comparable to both CMC and CryoGrid 2 forcing, we argue that the values can be reconciled: without considering sublimation, the total winter precipitation would correspond to maximum annual SWE. Assuming the annual snow build-up is roughly a triangular function, the average SWE would be about half of the maximum SWE (assuming the snow cover build-up starts at the first day and the snow cover lasts until the last day – the true average SWE would therefore be somewhat lower). However, a study of the surface energy balance on Samoylov Island suggests substantial sublimation (Langer et al., 2011), with latent heat fluxes from October to March in one winter season corresponding to an accumulated water equivalent of approx. 30 mm. Using these coarse estimates, ERA-interim-derived winter precipitation (Fig. S2) corresponds to a similar magnitude of SWE as CMC and the CryoGrid 2 forcing.

While these comparisons to independent model data sets do not constitute a validation of the CryoGrid 2 snow cover forcing in a strict sense, they indicate that the large-scale pattern derived from GlobSnow is not an artifact of the GlobSnow SWE retrieval, but related to regional gradients in winter precipitation.

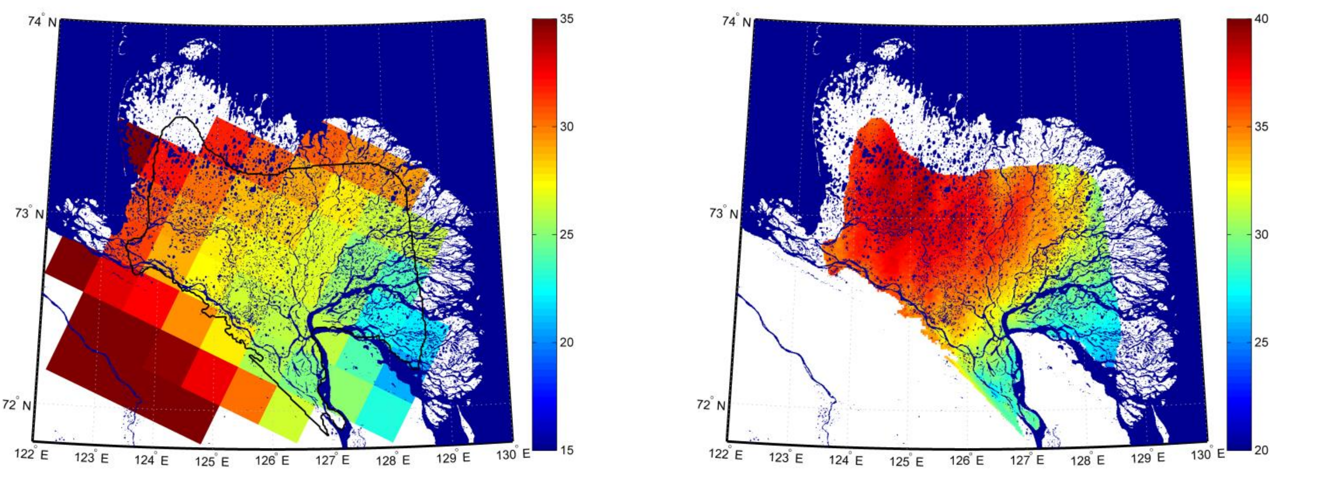

Fig. S1: Average SWE [mm] in the months October to June from 2004 to 2013. Left: Canadian Meteorological Centre (CMC) Snow Depth Analysis Data, 24 km resolution; the black line corresponds to the outline of the model domain. Right: CryoGrid 2 forcing data (1km resolution) based on GlobSnow SWE and MODIS snow cover (same data set as displayed in Fig. 5 f). Note the offset of 5 mm between the color scales.

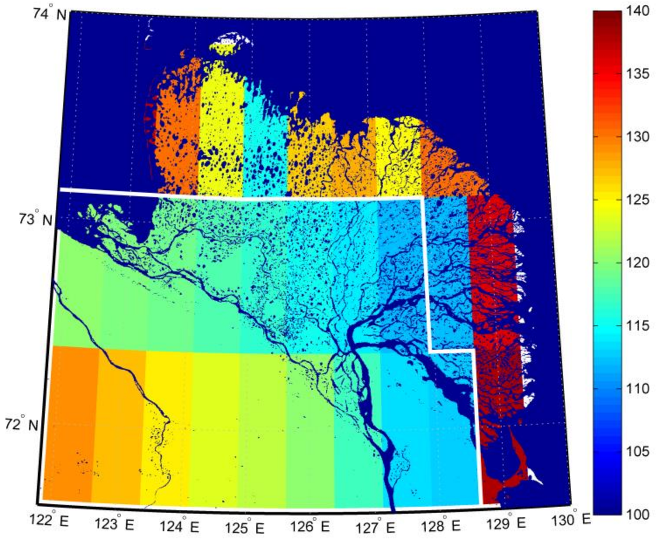

Fig. S2: Annual averages of the total precipitation [mm] falling at 2m-air temperatures of less than 0°C for the months October to June for 2004-2013, based on the ERA-interim reanalysis at 0.75° resolution. The land-sea mask is indicated by a white line (land areas in the bottom part).

2 Documentation of borehole sites

To facilitate a better impression of the borehole locations, their surroundings and thermokarst occurrence, we provide images of all boreholes. Fig. S3 shows the borehole locations “Olenyokskaya channel center” and “Olenyokskaya channel mouth”, both of which are located in a relatively homogenous flat landscape. The same is true for Kurungnakh Island (Fig. S4 left). Here, a thermokarst pond developed around the borehole already in the following summer. On Sardakh Island, a thermokarst around the borehole was for the first time recorded in summer 2012, and a larger pond had developed by 2014 (Fig. S5). On Samoylov Island, new structures and buildings were erected in the direct vicinity of the borehole in summer 2012 (Fig. S6).

Fig. S3: Left: Olenyokskaya ch. center, borehole site, August 2010. Right: Olenyokskaya ch. mouth, borehole site, August 2010. Photos: Jennifer Sobiech.

Fig. S4: Left: Kurungnakh Island, borehole site, July 2009, photo: Julia Boike. Right: Kurungnakh Island, borehole site with thermokarst pond, August 2010, photo: Jennifer Sobiech

 

Fig. S5: Left: Sardakh Island, borehole site after drilling, July 2009, photo: Julia Boike. Right: Sardakh Island, borehole site with thermokarst pond, August 2014, photo: Steffen Frey.

Fig. S6: Samoylov Island borehole, July 2016, photo: Niko Bornemann. All structures and buildings visible in the background were erected in summer 2012.

References

Brasnett, B.: A Global Analysis of Snow Depth for Numerical Weather Prediction, Journal of Applied Meteorology, 38, 726–740, 1999.

Brown, R., and Brasnett, B.: Canadian Meteorological Centre (CMC) Daily Snow Depth Analysis Data, Environment Canada, 2010, Boulder, Colorado USA: National Snow and Ice Data Center, 2015, updated annually.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The surface energy balance of a polygonal tundra site in northern Siberia, Part 2: Winter, Cryosphere, 5, 509–524, 2011.