We are thankful for the comments and suggestions which have significantly improved the manuscript. In the revised version of the manuscript, we have included an additional validation run for the subsurface stratigraphy of a well-developed thermokarst lake on Samoylov Island, for which measurements of lake bottom temperatures are available (described in Boike et al., 2015). The comparison of the model run to the in-situ measurements yielded a cold exceeding 2°C which was mainly caused by too cold summer temperatures. As a result of the not optimal performance, we have slightly modified the water body scheme of CryoGrid 3 Xice. The modifications concern exclusively the summer heat transfer, i.e. during the times when the water surface is unfrozen. In the old version, the increased heat transfer during summer due to mixing of the water column was phenomenologically simulated by artificially increasing the thermal conductivity of the water grid cells. We find that the performance of the scheme is considerably improved, if we apply "true" mixing to the water column instead of an increased thermal conductivity. In the revised model version, which has been employed in the simulations shown in the revised manuscript, we assign a single temperature value to all water grid cells which is computed as the weighted average of all water grid cells after each time step, i.e. after surface energy balance and conductive heat transfer are applied. The modifications to the model are thoroughly described in Sect. 2.7 of the revised manuscript, and the comparison to in-situ measurements of lake bottom temperatures is described in Sect. 5.1 and the new Fig. 6. The modifications to the water body scheme of CryoGrid 3 Xice have also made a recalculation of the long-term susceptibility runs necessary. As a result, Figs. 7-9 (6-8 in the old version) are slightly changed in the revised version. However, all conclusions and discussion points remain fully valid.

In the following, we give point-by-point replies to all comments raised by the Reviewer (in bold font):

This manuscript is comprised of three fairly different components: model description, model validation, and future experiments. These components are each very interesting on their own, but I don't feel that they fit together as well as they could in the current version, and I might suggest that the authors try to make a more unified manuscript by better linking the various pieces. Overall, two main areas of mechanistic development in the model description are explored in the manuscript: surface energy fluxes, and excess ice/subsidence. The more novel of these is the description of the excess ice parameterization (section 2.7). The observational testing of the model (section 5.1) is limited to the thermal dynamics and some surface energy budget calculations; are there other observations that could be used to test certain aspects of the subsidence model?

We have inserted a new Sect. 6.2 in the revised version of the manuscript in which we provide further evidence for the performance of the model scheme by comparing to findings from the Günther et al. (2015)-study:

"In the long-term thaw susceptibility runs for Samoylov Island, ground subsidence does not occur at or before present-day which is in agreement with in-situ observations (Boike et al., 2013). A validation of the modeled future subsidence rates is therefore not possible in a strict sense. However, the study of Günther et al. (2015) conducted for Muostakh Island off the coast of the Lena River Delta (150 km SE of Samoylov Island, dominated by Pleistocene sediments similar to the "Ice Complex" stratigraphy), where strong ground subsidence has occurred in the last 50 years, facilitates comparing measured and modeled properties of ground subsidence in a qualitatively way. The site has experienced significant warming most likely related to changes in the sea ice conditions in the past 50 years, and Günther et al. (2015) document ground subsidence rates of up to 5m/50 y which is comparable in magnitude to the simulations for the Ice Complex stratigraphy for the next 50 years. Furthermore, the study can clearly document an inverse relationship between active layer thickness and subsidence rates (Fig. 10, Günther et al., 2015) which they relate to the differences in the ground ice content below the active layer. They document a ground subsidence of on average 2m/50 y at a thaw depth of 0.4m, about half this value for a thaw depth of 0.65m, and close to no subsidence for active layer depths exceeding 0.8m. In our simulations for the three stratigraphies, we can qualitatively reproduce this relationship. We model a current active layer thickness of about 0.4m with future ground subsidence rates 6-8m/50 y (Ice Complex stratigraphy) and 0.65m with 4m/50 y subsidence for the Samoylov stratigraphy. For the Arga stratigraphy lacking excess ground ice, the modeled active layer thickness exceeds 1m, while no subsidence occurs. The exact subsidence rates necessarily depend on the applied forcing (which is certainly different for the past 50 years on Muostakh Island than for the next 50 years on Samoylov Island), but the qualitative agreement with the long-term observations suggests that CryoGrid 3 Xice can capture crucial dependencies on subsurface parameters that have been observed in nature.”

For example, how does the simulated soil thermal regime when submerged below a shallow water layer compare with observations of soil temperatures below shallow lakes?

We are thankful for this important comment that has triggered more work and improvements to the CryoGrid3 Xice model scheme (see also introductory statements above). In the revised version of the manuscript, we have conducted a validation run with a subsurface stratigraphy of a 6m deep thermokarst lake on Samoylov Island. The lake bottom is instrumented with a temperature sensor, which we compared to CryoGrid 3 Xice model results. As mentioned above, the comparison of the old scheme with increased summer thermal conductivity yielded a significant cold-bias of 2°C in the annual average, as well as a somewhat biased representation of the annual dynamics. This is considerably improved in the new version of the model scheme. As a result of the newly included comparison with measured lake bottom temperatures, the manuscript has slightly increased in length (one Figure and one Table added), but we feel that the additional information have led to significant improvements.

How do surface energy budgets of subaerial or subaqueous soils compare to each other, both in the model and observations (if available)?

Systematic observations on these parameters are not available for the study area. We emphasize, however, that the model results are now validated for the two end-points of the thermokarst simulations, a land and a deep thermokarst lake situation.

How do surface energy budgets compare between the more vegetated versus less vegetated terrains in the Lena Delta? Can the comparison between model and data in figure 3 be broken out more (for example as a function of time) to give more insight into the dynamics?

The only field observations on the surface energy balance in the Lena River Delta are obtained on Samoylov Island. Therefore, a comparison for sites outside Samoylov (e.g. the Arga site with less vegetation) is not possible. Moreover, the future simulations for these sites leave the surface parameters affecting the surface energy balance unchanged, since they are exclusively meant as a sensitivity study for the different ground stratigraphies. Modeling the surface energy balance of different areas on Samoylov Island is beyond the scope of this study.

It is indeed possible to break up Fig. 3 in time, at least for some of the components of the surface energy balance, for which relatively gap-free time series exist (see details in Langer et al. 2011a,b). Below, we show a comparison of daily average values of measured (with eddy covariance) and modeled (validation run) latent and sensible heat fluxes from Samoylov Island for a two-year period. In general, the model can reproduce the measurements very well, particularly keeping the large uncertainties associated with the measurements in mind (Langer et al. 2011a, b). We have chosen not to show these time-resolved comparisons in the actual manuscript, since they would add significantly in length and thus distract from the overall focus of the study. While it is highly important that the model can reproduce all the different model variables (surface energy balance, surface temperature, etc.), the aim of the study is clearly on the ground thermal regime. We feel that displaying the averages of the surface energy balance (as in Fig. 3) is an adequate compromise, as it clearly demonstrates the performance of CryoGrid 3 with respect to the average surface energy balance.

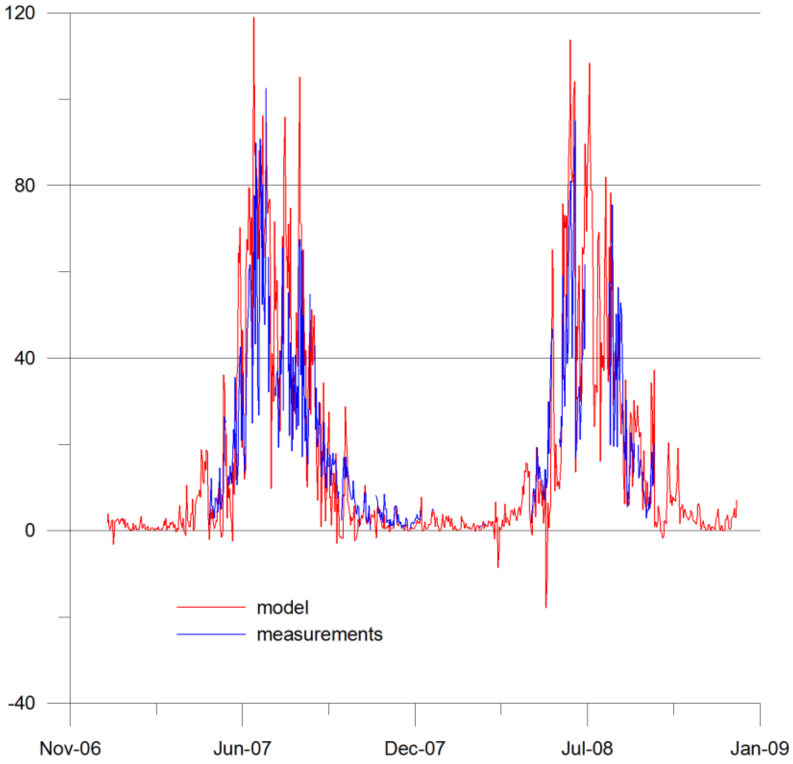

Fig 1: Comparison of modeled and measured latent heat fluxes for Samoylov Island. X-Axis: time; y-axis: fluxes in W/m^2. See text.

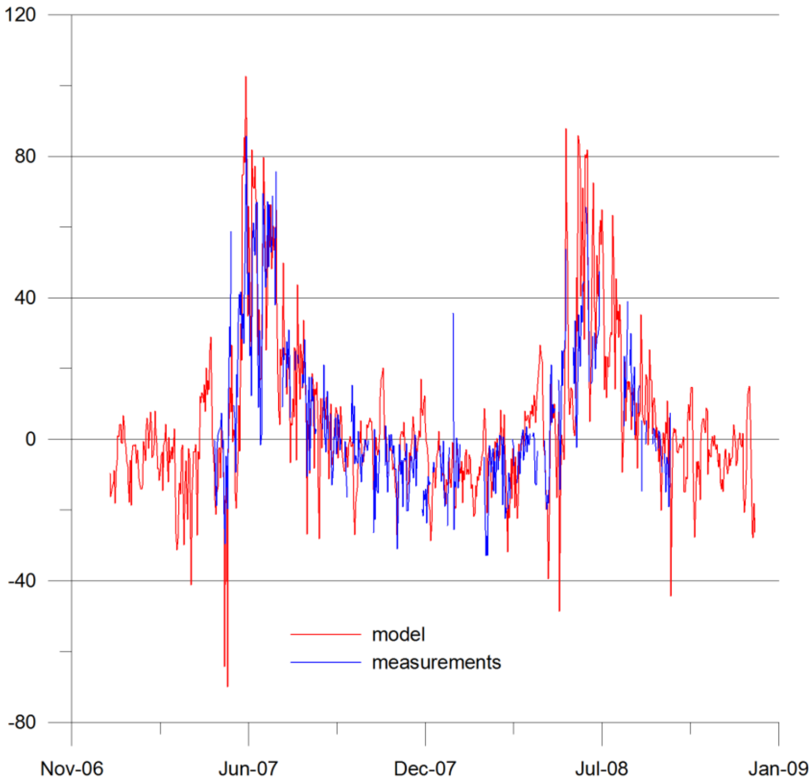

Fig 2: Comparison of modeled and measured sensible heat fluxes for Samoylov Island. X-Axis: time; y-axis: fluxes in W/m^2. See text.

The simulations to explore future sensitivity (section 5.2) are very interesting but not clearly related to the observations shown in section 5.1, other than being at the same location. What observations would best allow the testing of the excess ice parameterizations?

We have supplemented two further pieces of information: 1. the qualitative comparison to measured long-term subsidence rates on Muostakh Island off the Lena River Delta described in Günther et al. (2015). The CryoGrid 3 Xice scheme can reproduce the measured correlation between subsidence rates and active layer thickness, indicating that model is based on the correct physical dependencies (see new Sect. 6.2). 2.we provide a comparison of modeled temperatures at the bottom of a well-developed thermokarst lake to model results, indicating that CryoGrid 3 Xice can to a large extent reproduce the ground thermal regime on the ground of such water bodies which is a crucial prerequisite for correctly modeling excess ice thaw and talik evolution.

What observations would be required to run the excess ice model at larger scales?

For meaningful application on larger spatial scales, the spatial variability of ground ice contents and the lateral fluxes of water, snow and heat resulting from excess ground ice thaw must be represented in the model scheme. In essence, the steps described in Sect. 6.3 must be tackled, before large-scale runs are meaningful.

The discussion of metastable states (section 6.3) is also very interesting, but I wonder if the discussion is missing something by being informed by a model that only considers the ice loss processes and not the ice formation processes. What role may ice formation processes play along with the feedbacks described here in setting slow oscillatory behavior of ice wedge growth and decay under stable climate regimes?

We fully agree with the statement – processes that lead to the formation of ground ice are very likely to play a strong role in the stabilization of permafrost states, in particular in case of stable climate regimes. In the current model version of CryoGrid 3 Xice, such processes are not represented and we have explicitly mentioned this limitation in the manuscript. We think that it is ultimately a question of the timescales of the processes involved. The physical processes that lead to ground ice formation are inherently different from the processes that lead to the melting, and as a result the environmental drivers and the associated timescales are different. Compared to the potential rates of excess ice loss modeled with CryoGrid 3 in the future climate (several meters/100years, which was also observed by Günther et al. 2015), formation of excess ground ice can be expected a slow process. The ice complex in the Lena River Delta (~40m thickness) formed during the last glacial cycle (~100ky) in a period of intensive ground ice formation, which corresponds to a rate of about 4cm/100years. Even if one assumes that more and less intensive periods of ground ice formation alternated, it is highly likely that ground ice melting will strongly override ground ice formation in the 21st century, so that it may be possible to neglect the latter at least in first-order modeling applications, as in this study.

Also, are there ways of using models such as this to identify where the thresholds of rapid ice loss lie?

We think that this may be possible in the future, but an integrated approach including advances in modeling and field investigations is required. As the model results appear to be rather sensitive to various parameters, long-term observations of e.g. ground subsidence must be established. Another possibility could be multi-decadal data sets from areas, where ground subsidence has already led to significant landscape modifications in the past, as the mentioned data set from Muostakh Island (Günther et al., 2015). However, high-quality forcing data sets for land-surface schemes in general do not exist for such areas, but advanced downscaling methods for reanalysis data, such as WRF modeling, may be a way to overcome such difficulties.

How are the positions of these thresholds sensitive to landscape features such as the depth of organic horizon, depth to massive ice, water content of excess ice, etc?

In the presented simulations, the threshold of excess ground ice thaw is reached when the thaw depth deepens and for the first time extends in into layers with excess ground ice, so the depth to excess ground ice layers in relation to the thaw depth is an important parameter. The thaw depth itself is determined by many factors, among others the meteorological forcing in particular during summer and the thermal properties of the active layer (which in turn are determined e.g. by the depth of the organic horizon). In this study, we have therefore chosen three different subsurface stratigraphies that lead to different thicknesses of the active layer which in turn influences the susceptibility to excess ground ice thaw.

It would be really interesting to use this model to explore these sorts of phase spaces and identify particularly vulnerable or resistant landscapes to warming.

We fully agree, and this is exactly one of the main scientific topics we want to explore with the new modeling tool CryoGrid 3. In order to do so, formulations for small-scale variability of ground properties (see above) and lateral exchange of water, heat ad snow must be developed and incorporated in the model, as mentioned in Sects. 6.3 and 6.4. Such studies are already ongoing, but presenting results is beyond the scope of the present manuscript.

Minor comments:

p. 6935 l. 21: If hydrology is not specified, are soils held at a fully-saturated state?

The soils are held in the state specified in Table 1. We have clarified this by inserting a sentence in Sect. 2.4: “The subsurface thermal properties are specified as a stratigraphy of the volumetric contents of soil mineral and organic material, air and the sum of ice and water (as in Table 1).”

p. 6945, line 9: How are the saturated/unsaturated dynamics considered if hydrology is not prognostic? Where does the air come from?

The air content is prescribed constant (see response above). In general, the soil water balance cannot be realistically described in 1D for our study area (see also Boike et al., 2013), so that lateral fluxes need to be specified in order to achieve realistic results. For our 1D model, it is therefore preferable to specify approximate volumetric contents for the different soil constituents. In many studies applying permafrost models, this procedure has been adopted (e.g. Jafarov et al., 2012, Westermann et al., 2013).

The model as described here uses only conductive heat transfer, so I am not sure I understand the criticism of conductive heat transfer in line 11 of page 6964 (and also the first sentence of the abstract). The real issue isn’t conductive versus nonconductive heat transfer but the complex hydrology and volume changes associated with excess ice.

Our criticism aims at model simulations that are EXCLUSIVELY based on conductive heat transfer (l. 11, p. 6964). Conductive heat transfer is naturally the main process of energy transfer in the soil, and so it is necessarily the basis for modeling schemes. However, we show that processes following melting of excess ground ice (e.g. the complex hydrology and volume changes) can completely modify future thaw trajectories. This cannot be accounted for in “conductive heat transfer”-ONLY schemes. To be more clear, we have inserted “and processes following excess ground ice thaw” in the sentence before line 11 in the revised version of the manuscript.

On behalf of all authors,

Sebastian Westermann