Anonymous Referee #1

The manuscript "Modeling the spatio-temporal variability in subsurface thermal regimes across a low-relief polygonal tundra landscape" by Kumar et al. presents high-resolution simulations of the ground thermal regime for tundra polygons in N Alaska. The manuscript seems publishable in TC after major revisions, although some crucial information on the model setup (mainly the exact choice of the lower boundary condition, see details under minor points) is missing, so that the soundness of the approach cannot be finally determined.

We thank the reviewer for critical review and feedback on our manuscript. We have presented a methodology for modeling of thermal-hydrologic processes in polygonal tundra ecosystem dominated by micro-topographic relief. 3-D modeling approach presented is a first critical step toward study of fine scale processes in tundra ecosystem. We have also presented the application of the model at four sites representing different polygonal types present at our sites at Barrow, Alaska. Our tests and evaluations were designed to validate the model using observations and analyze the model agreements and disagreements to gain insight in the processes at the site, and identify model deficiencies and gaps in data availability which can guide model improvements and data collection to enable improved modeling of the study site.

Major Points:

1. The simulations are driven by temperature measurements at the surface, as it was common in early 2000s-publications on 1D-heat conduction schemes. Such simple approaches are generally no longer publishable in a journal like TC today. The authors add 3D-coupling of both heat and water transfer as a new state-of-the-art feature, which removes many of the limitations inherent in earlier approaches. However, the results are rather mediocre at best. While Figs. 8 ff seems to suggest a rather OK fit with measurements, it is actually not the case for active layer thickness (ALT). Although it is hard to see in the figures, the model predicts an ALT of ca. 100 cm at Site A, while it is in reality <50 cm. The same is true for Sites B, C, and D. This is much worse than many traditional 1D-schemes (which may or may not be tuned) and makes the model results virtually useless for further applications, where ALT is of interest. Even worse, the authors can present only a single year, so that it is impossible to determine whether the model can reproduce interannual variability of ALT and ground temperatures, or (e.g. decadal) trends in these variables, which is crucial for applications in the context of climate change. If the bad performance for ALT is indeed true, it must be clearly stated and the reasons investigated and discussed. The conclusion of the manuscript should then be that the model in the presented set-up is NOT suitable for studying the ground thermal regime of polygonal tundra near Barrow.

We have added simulated estimates of thaw depths at four sites (Figures 14, 15, 16, 17). The thaw depths show significant variability across the study region. Compared to an observed average active layer thickness of 50 cm (with thickness of up to 74 cm observed by Hubbard et al. 2013; Peterson 2016; Figure 22), the model is biased towards deeper thaw depths. We have discussed these biases and potential reasons on Page 18 Line 4-Page 21 Line 9.

We were limited to single year period in our study for which all forcing and validation data sets were available. Thus we were not able to investigate the model performance to reproduce interannual variability. We have identified this limitation on Page 27 Lines 1-2 and noted our plan to address this particular issue as more data become available.

While we agree that the presented case studies for our modeling approach show biases in comparison with the observations at the sites, one of our objectives was to identify model deficiencies and data gaps. Thus, one key conclusion from our study is the need for better characterization of site via co-located measurements Page 26 Lines 33-35. "While the models demonstrated the ability to simulate the soil temperature at shallow depths, the deviations from observations in deep soils highlights the need for better soil characterization using deep cores in these ecosystems. Our study also highlights the need for co-located observations for accurate modeling and understanding of the tundra landscape."

2. I doubt that the model scheme presented by the authors is scalable due to the computational requirements, so that it could be included in ESM frameworks or similar schemes. So what do we learn from the simulations of the four polygons then? I do not see that the study can provide any new insight in processes or process parameterizations. The main message seems to be that the authors managed to launch a model scheme of unprecedented complexity and computational requirements, but new insight in cryospheric processes and model parameterizations hereof are largely absent in my view. An example for a slightly similar study that does a significantly better job in this respect is Weismüller et al. (2011). They demonstrate a coupled 1D-scheme for heat conduction and water flow and use this model to show that a heat-conduction-only model scheme is more or less sufficient to reproduce the ground thermal regime at their study sites. So it would for instance be highly interesting, if a comparatively simple 1D-heat conduction model (e.g. GIPL2) with year-averaged ground properties/water contents could yield a similar performance for the center/rim sites. The authors could investigate if 3D-coupling for both heat and water fluxes is really needed, or if 3D-coupling only for water fluxes is needed, or 3D-coupling for heat fluxes only. Such information would be crucial to help designing a robust scalable scheme for representation of polygonal tundra in ESMs.

PFLOTRAN is a massively parallel software which has been extensively tested and optimized on a number of Department of Energy's Leadership Computing Facilities (Hammond et al. 2014, 2012, 2008; Mills et al. 2009). While computationally intensive it has capability to make efficient use of state-of-the-art supercomputing to address large computational problems.
While our interest is to use this process based complex model to inform ESM scale, we do not propose to do that at resolution and complexity detailed in this paper. Work presented here is a building block in the larger scaling philosophy under NGEE-Arctic project which we have discussed on Page 27 Line 10-20.
We disagree that the model doesn't provide any new insights in cryospheric processes. In contrast to traditional 1D heat conduction model, we have presented a 3-dimensional model that resolves the advective and diffusive flow of mass, advective and conductive heat flow in a tightly coupled fashion. The complexity of the model adds to the computational as well as data requirements but it also provides insights in processes which have been often ignored in traditional approaches. We have demonstrated the ability to represent the microtopography effect in a 3-D system. The spatial patterns of Centers, Rims, and Troughs emerge from the topography-aware model are not prescribed. While our case studies were limited to small set of polygons (to allow for comparison with the observations), our approach enables extension of the study region of regional scale.

PFLOTRAN solves a coupled system of PDEs for flow of mass and energy, thus 3-D coupling for heat flux only is not possible. However, the analysis of 3-D flow patterns of mass and energy will give us insights in dominance of lateral vs vertical flow patterns.

Glenn E. Hammond, Peter C. Lichtner, and Richard T. Mills. Evaluating the performance of parallel subsurface simulators: An illustrative example with pflotran. *WATER RESOURCES RESEARCH*, 50:208-228, JAN 2014 doi: 10.1002/2012WR013483
Peter C. Lichtner and Glenn E. Hammond. Using high performance computing to understand roles of labile and nonlabile uranium(vi) on hanford 300 area plume longevity. *VADOSE ZONE JOURNAL*, 11(2), MAY 2012 doi: 10.2136/vzj2011.0097

G. E. Hammond, P. C. Lichtner, C. Lu, and Mills R.T. Pflotran: Reactive flow and transport code for use on laptops to leadership-class supercomputers. In Fan Zhang, G.T. Yeh, and Jack C. Parker, editors, *Groundwater Reactive Transport Models*, pages 141-159. Bentham Science Publishers, Sharjah, UAE, 2012 doi: 10.2174/978160805306311201 01

Richard Tran Mills, Glenn E. Hammond, Peter C. Lichtner, Vamsi Sripathi, G. (Kumar) Mahinthakumar, and Barry F. Smith. Modeling subsurface reactive flows using leadership-class computing. In H Simon, editor, *SCIDAC 2009: SCIENTIFIC DISCOVERY THROUGH ADVANCED COMPUTING*, volume 180 of *Journal of Physics Conference Series*, 2009. 5th Annual Conference of Scientific Discovery through Advanced

Computing (SciDAC 2009), San Diego, CA, JUN 14-18, 2009 doi: 10.1088/1742-6596/180/1/012062

Glenn E. Hammond, Peter C. Lichtner, Richard Tran Milis, and Chuan Lu. Toward petascale computing in geosciences: application to the hanford 300 area - art. no. 012051. In RL Stevens, editor, *SCIDAC 2008: SCIENTIFIC DISCOVERY THROUGH ADVANCED COMPUTING*, volume 125 of *JOURNAL OF PHYSICS CONFERENCE SERIES*, page 12051. US DOE Off Sci; Cray; IBM; Intel; HP; SiCortex, 2008. 4th Annual Scientific Discovery through Advanced Computing Conference (SciDAC 2008), Seattle, WA, JUL 13-17, 2008 doi: 10.1088/1742-6596/125/1/012051

Ref: 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, doi:10.5194/tc-5-741-2011, 2011.

3. Excess ground ice is a main driver for the evolution of polygons and melting of excess ground ice will lead to changes of the microtopography, which in turn changes the hydrological regime. Are such processes represented in the model scheme? This should be explicitly stated and commented upon in the manuscript. If yes, is the surface stable during the 1-year test period? Are there sites where excess ground ice melt is observed and which could be used to test the model performance? If not, a key element determining the evolution of tundra polygons is missing, and it should be clearly stated that the scheme is not suitable for climate change studies in polygonal tundra.

Ground ice and other cryostructures are not represented in the model, which has important consequences for modeling of thermal regime. We have included a discussion on this issue on Page 14 Line 18 - Page 15 Line 26.

“For example, while we know that presence of ground ice (like ice wedges, segregated ice, ice lens etc.) is common in the subsurface of Arctic tundra, their representation in the model is completely missing. Lack of representation of these cryostructures are potentially one of the reasons for warmer soils in our simulations. While PFLOTRAN has the ability to capture and model such cryostructures (via heterogeneous subsurface structure and properties but not their formation and evolution), we lack any quantitative data to characterize them for representation in the model. Ongoing efforts under NGEE--Arctic project by Kneafsey and Ulrich 2016 and Dafflon 2016 using X-ray computed tomography (CT) scanner technology on ice cores from BEO can potentially provide detailed 3-dimensional soil structure and density information and help address this missing piece. ”

Evolution of tundra polygons through the process of ice wedge degradations is a complex process which has not been studied widely and approach for modeling them has been limited. Limitation of the presented model to represent dynamic geomorphological changes due to ice-wedge degradation has been noted on Page 26 Line 27-28.

4. The authors should state more clearly that the presented model is only a very first step towards a physical model of energy and water transfer within tundra polygons. Many of the key drivers of spatial variability in the system are implicitly prescribed by the forcing data (2cm temperature measurements) and not modeled. The authors present curves of snow depths at the various sites, but which factors lead to these differences? (How) could this be modeled? The same is true for vegetation, surface energy balance, evapotranspiration, etc. The authors state that coupling to CLM is planned, but many crucial processes (e.g. wind drift of snow) are not contained in CLM, since it is mainly designed for large-scale applications.

We agree with that the presented work is a first step towards a process based model for Arctic tundra ecosystem processes. We have added a statement on Page 27 Line 5-10 to highlight that.
"While we have not addressed all the deficiencies in model process representation and parameterization identified and reported here in this study, we believe we have developed and presented a process rich modeling framework as a first critical step that would enable such studies. The modeling approach developed in this study will allow accurate modeling of permafrost thermal hydrology and will help identify and guide the future observations required for improved modeling and understanding of the polygonal tundra ecosystem."

We have also added a description of NGEE-Arctic scaling philosophy (Page 27 Line 10-25) that the present study will serve as a critical building block for.

"While the knowledge gained by developing and evaluating fine-scale 3D simulations is valuable from the perspective of increased understanding of complex process interactions, the explicit long-term goal of the NGEE--Arctic project is to improve predictions of Arctic ecosystem processes at scales relevant to coupled climate and Earth system simulation. One element of our strategy to migrate knowledge across scales is to improve the grid and sub-grid representations in the land model component of our Earth system models to capture observed modes of variability in physical, biological, and biogeochemical processes. For example, our new top-level grid topology for global-scale land modeling follows watershed boundaries instead of the typical and arbitrary rectangular gridcell arrangement (Tesfa et al. 2014). Sub-grid schemes are being developed that represent topographic variation within basins, and our goal is to apply those methods in the micro-topographic setting of polygonal tundra to capture the variation in thermal, hydrologic, and biogeochemical regimes, and interactions with

vegetation communities. The current study is one step toward identifying the relevant modes of variation among diverse landforms in the polygonal tundra region. Another element of our scaling strategy is to use, to the full extent possible, a common set of modeling tools to construct simulations at various spatial scales. Even though many processes that can be represented explicitly at the finest scales (such as lateral flows of energy and water) must be parameterized for efficiency in a larger-scale simulation, having a common underlying set of equations helps to reduce unintentional loss of information across scales due, for example, to aggregation and disaggregation operators."

Minor Points:

Fig. 1: strange values given in the color bar, units ([m a.s.l.]?) should be provided.

Yes, elevations are in meter above mean sea level. The caption has been updated to clarify that.

Fig 2: zero-degree-line missing in b

The figure has been updated to include the zero degree line

P. 8, L25: This is a major design flaw of the study which questions the use of such a sophisticated model scheme. Are there plans to obtain such data sets in the future?

Unfortunately it is a common practice that field observations are often designed and conducted by disciplinary teams based on their objectives. Under the NGEE-Arctic project we are making efforts to coordinate a large multi-disciplinary team of modelers and field scientists. While in this study we employed observations by various teams (and often from literature), many of which were not coordinated or co-located. Under NGEE-Arctic Phase II we are engaging closely with various observational teams for coordinated and co-located observations needed for modeling studies (beyond the one presented here). Thus, we expect to be able to better parameterize our models at our study sites.

Sect. 3.1.3 How about the vertical discretization of the system?

Variable resolution was used to discretize the the system vertically, fine near the surface to better resolve the active layer and coarser in the deep subsurface. Table 2 describing the scheme has been added to the text and referenced on Page 10 Line 10.

Sect. 3.2: please provide a clear overview of the processes and parameterizations that are considered in the model (and the ones that are not), i.e. heat conduction, saturated flow, unsaturated flow, water vapor transport, how is the freeze curve determined, etc., etc.

Section 3.2 describes the governing equations for flow of mass and energy in variably saturated porous media modeled by PFLOTRAN. Formulation for three phase (ice, water, vapor) flow in PFLOTRAN are based on works by Painter (2011), Painter and Karra (2014), and Karra et al. (2014).

P.14, I. 10: Why -1 degree, that appears to be much too warm??
-1C degree temperature was used as the initial condition for the spin up phase of the simulations and had no effect on the final periodic steady state conditions after spin up. A statement to clarify the choice has been added on Page 13 Line 15-20.
"3-D subsurface models for each of the four sites were initialized by freezing the entire modeling domain at a temperature of -1.0 degC. The models were spun up to a thermal periodic steady state using a time series of mean daily temperatures applied to the top of the domain (ground surface). Spin up simulations were conducted for a period of by cycling annual time series of forcing. Spin up simulations were continued until a periodic steady state was achieved (i.e. close to zero interannual variability in annual thermal regime). Spin up duration of 10 years was used at all the sites and was determined to be sufficient. We conducted a series of initialization simulations by varying initial temperatures at start of spin up and found them to not have any significant impact on the final periodic steady state, besides simulation period required to reach that steady state."

P. 14, I. 10: At what depth is the deep bottom boundary? How does the choice of the lower boundary condition interfere with the selected spin-up-procedure? What are the resulting temperature gradients below the depth of zero annual amplitude? Why not use a heat flux as lower boundary condition, and perform a spin-up so that steady-state conditions are reached in the entire model domain?

The bottom boundary was at the depth of 50 m with variable resolution scheme for vertical discretization (Table 2). The choice of the bottom boundary temperature and the initial temperature has direct impact on the spin up simulation required to achieve thermal steady state condition.
Heat flux measurements in deep permafrost soil were not available at our sites, thus a fixed temperature boundary was used. We selected our model bottom boundary to be at significant depth of 50 m to best avoid the boundary condition effect on the simulated thermal states. We have identified heat flux measurements from deep cores as one of the data set necessary for modeling thermal regimes at the site (Page 25 Line 31-33).

P. 14, I. 11: West Dock is several 100 km E of Barrow – how realistic is this assumption and in how far would errors in this temperature compromise the results. Does this temperature roughly correspond to a steady-state condition given the applied surface forcing, or does it introduce a heat sink/source at the bottom? See also comment Above.

Romanovsky et al. (2010) analyzed permafrost thermal state in the polar northern hemisphere and found the permafrost temperature of -8 to -10 C around 70N latitude. We used temperature from the West Dock which was closest site at comparable latitude studied by Romanovsky et al. 2010. In our simulation, we did not observed any source/sink effect introduced by the bottom boundary at deep 50m.

P. 14, I. 23: What is “thermal hydrology”, and why is soil moisture not discussed? This doesn’t make sense to me since it is one of the assets of the new model, that the 3D-interplay between moisture and heat fluxes is explicitly considered. The authors should investigate their model results further to show how for instance water fluxes change the thermal properties of the system, which in turn affects heat conduction.

Term “thermal hydrology” refers to coupled processes of water and energy flow. Focus of the presented work was primarily on thermal regimes (Page 14 Line 5-6). We agree that the interplay of soil moisture and heat fluxes is a key process represented in the model. However, on Page 13 Line 29-31 we have acknowledged the caveat of our choice of flow boundary condition due to lack of data on drainage patterns. In Figure 21 we have presented the results of simulated maximum water table across our site. While we have discussed the patterns of soil moisture qualitatively in Sec 4.2 Page 22-23, we have also highlighted the implications of the boundary condition on thermal regimes on Page 17 Line 3.

P. 14. L. 5: How is it determined that periodic steady-state conditions are reached?

The spin up simulations were continued until close to zero interannual variability in the thermal regime was achieved (i.e. annual pattern of soil temperature were same between one year to next). In our analysis, we found 10 years to be sufficient period to achieve that state across all sites.

P. 14, I.9ff: The authors should quantify the magnitude of the advective heat flow, and set them in relation to the conductive heat fluxes. Could a similar model accuracy (considering Figs. 8 ff) be achieved when such fluxes are neglected, as it is done in most model approaches? See major comments.

PFLOTRAN does not compute the advective/conductive heat fluxes separately. Thus we do not have a quantitative way to set and comment on the differences in magnitude. While we anticipate that conductive fluxes dominate the convective, our aim was to improve beyond existing simplified conductive only approaches and provide a process rich model. While simple models can be parameterized and tuned for accuracy against observations, our approach, while complex, would enable understanding of fine scale processes.

P. 14, l. 20: Any idea on the accuracy of the precipitation measurements? And why do the authors use daily precipitation, not better resolved in time? Are only daily values available?

Sensors at the site were checked for accuracy every two weeks to a month interval and precipitation measurements compared with the NOAA CRN facility. However, quantitative accuracy assessments for the precipitation measurements were not available to us. The observations were available at hourly interval, however we aggregated them to daily for input to the models. While model time steps are of seconds to 30 minutes size, daily time series was used to smooth the forcings to the model for faster numerical convergence. Use of finer in time time series for the model is however possible.

Fig. 13: Why is the thermal conductivity for saturated soils higher with some liquid water compared to fully frozen conditions? Is that a real physical process, or an artifact of the employed parameterization? If it's the latter, what is the effect on the simulation results?

Behavior of thermal conductivity noted by reviewer is an effect of parameterization. The issue is that we have estimates of conductivity in the pure states, that is, pure ice or pure liquid water. However we have soils at a wide range of intermediate states. We have used a standard method for blending the pure state information from published literature by Painter et al. 2012, Karra et al. 2014, Painter and Karra et. al. 2014.

P. 20., l. 1ff: I very much agree with this statement! Therefore, many of the results could be strongly influenced by the particular parameterization of the thermal conductivity chosen by the authors. It is a standard parameterization used in many models, but it is not based on first principles and could thus be prone to biases at the particular study site. This is in particular crucial since the authors attempt to reproduce fine-scale ground temperatures, rather than provide a coarse assessment of the ground thermal regime with a simple thermal model.

We agree the parameterization of thermal conductivity can be prone to bias. Formulation implemented in our model is based on the Painter 2011.

P. 23, Sect. 5.1: I like that the authors present non-optimal fits between measurements and model, rather than tuning the model to fit the available ground data perfectly. In addition, the authors could/should present a sensitivity analysis at least for some of the crucial model parameters, to show which parameters need to be better estimated in order to improve results. But this is probably difficult due to the model complexity and computational requirements?

We thank the reviewer for appreciating the value of non-optimal results. We believe that understanding the reasons (missing processes or data) behind the mismatch and

addressing them should come first before we calibrate the model, which is essential for understanding the processes at the site.
Focus of our presented work was to develop a modeling framework and our case studies without any calibration essentially demonstrates the developed capability using the best off the shelf data sets available to us. Mechanistic representation of processes in our approach does add complexity, increases data requirement for parameterization and computational requirements. However, PFLOTRAN is highly scalable in high performance computing environment and well placed to address such problem by through efficient use of increasingly available computational resources.

While we have not presented a systematic model sensitivity and calibration study, such efforts has been undertaken and recently published by our colleagues Atchley et al, 2015, Harp et. al. 2016 and are complementary to our work. We have added references to their relevant work on Page 25 Line 15-16.

P. 24, l. 3: Not sure this is explainable by missing soil properties, etc. The bias is systematic, and the authors should also investigate and comment on the short time period (1y) of their runs and the way they handle spin-up and the lower boundary condition (see comment above).

Page 26 Line 9-11 we have acknowledged these limitations and commented on the need for heat flux observations to better handle the lower boundary conditions.

P. 26, l. 3: The authors do not provide any quantitative evidence that the hydrology is really reproduced. In a qualitative way, it is (high rims are dry, depressed centers wet, etc.), but quantitative validation information is not presented. Therefore, this statement should be formulated more carefully.

We agree that we have only presented qualitative and not quantitative evidence for reproducing hydrology. One of our goals in the study was to identify the gaps in observations and motivate future data collections. In absence of data to correctly inform our flow boundary condition, instead of making assumptions to match the observations we have chosen to use a simple no flow boundary condition and analyze the implications of such choice, and thus motivating the need for necessary observations.

P. 26, l. 7: The authors must present evidence (e.g. sensitivity analysis) that the bias in temperatures is really explainable by deep soil properties (see comments above). If so, they should elaborate on which soil parameters have the largest influence on simulation results.
Recent complementary work and papers by our NGEE-Arctic colleagues have investigated this issues in detail. Atchley et al. 2015 conducted a comprehensive 1-D model based calibration study at Site C and identified limitations due to lack of lateral flows that can be addressed by a 3-D model. Harp et. al 2016 extended the work of

Atchley et al. 2015 to conduct a Null-Space Monte Carlo method based systematic uncertainty analysis to quantify the effect of soil property uncertainties on permafrost thaw under CESM projected RCP 8.5 scenario from year 2006 to 2100.

P. 26, I. 7: The statement on C fluxes is misplaced in this discussion.

Zona et al. 2016 studied cold season emissions in Arctic tundra, highlighting the need for not just understanding active layer thickness, but soil temperature even during winter months. We have revised the text to be more clear (Page 25 Line 25-30).

"Modeling soil temperatures, beyond the high level estimation of thaw depth (or active layer thickeness) is important to understand the thermal regime of permafrost soil and its behavior under warming conditions. For example, during winter seasons even when the soils are completely frozen, variability in in soil temperatures (Figure7, 8, 9, 10) exist and may impact carbon fluxes from the system even during the winter season when soils are frozen (Zona et al. 2016)."