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: 10.2174/97816080530631120101

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, l. 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, l. 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, l. 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, l. 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, l.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, l. 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)."

Anonymous Referee #2

The authors present a new model (PFLOTRAN), in the framework of the New Generation Ecosystem Experiments initiative, which can assess the thermal hydrology for permafrost regions. The authors apply the model at four study sites in Barrow, where they compare the model performance with in situ collected data. The model presented is interesting and physically sound, and the the authors carefully run tests to evaluate model against measurements, On the other hand, some information on the model is missing, and the structure of the paper seems sometimes unclear to me. Therefore I recommend the manuscript for publication after major revisions.

Major points:

1. There is a problem in the data-model comparison that the authors fail to mention In Figures 8-11, the authors state that the model performs well against data except from the deeper soil layer. This is not actually true, since the model seems to largely overestimate the soil temperature in the summer months (May - October) also in the upper layers, as measurements and model only agree well at 5 cm depth. This seems to me to be a major issue, since this influences the estimation of key variables as the active layer depth. This problem, of course, would greatly limit the applicability of the model and its potential coupling to any biogeochemical model for estimation, e.g., of methane emissions. The authors should at least discuss this point in the paper. Does PFLOTRAN provide estimations of the active layer depth? If yes, how do they compare against measurements? What can be done to improve this key issue? And why does the model overestimate the temperature in the first place? It is clear that biogeochemistry is outside of the purposes of the paper, but nevertheless there seems to be a relevant issue, since this overestimation occurs in all four sites.

PFLOTRAN solves for soil temperature as primary variables and provides fraction of water in ice, water and vapor states in every computational cell. It does not directly provide the estimates of active layer depth. However, the depth of thawed layer (instead of active layer since we have not addressed BGC) can be estimated using soil temperatures and/or fraction of frozen vs unfrozen water content. We used 0°C soil temperature to define the threshold temperature for thaw and calculated the thaw depth time series at each site. Figure 14, 15, 16 and 17 show the temporal pattern of thaw depths during the simulation period and the spatial distribution of maximum thaw depth. Spatial distribution of the maximum thaw depths are four sites show strong correlation with the micro-topography. Model does show a warm bias in soil temperatures which translates to a bias towards deeper than observed thaw depths at all the study sites. Tables 3-6, 8-11 presents the validation statistics and bias when compared against observations. We have discussed the potential reasons for these bias in Section 4.1. We believe the primary reason for this bias is the parameterization of soil thermal and hydraulic properties, which in absence of data from our sites, were derived from Hinzman

et al. 1998 collected at a different site. With the difference in model representation of subsurface structure from the real world, the periodic steady state achieved at the end of model spin up phase had a warm bias. Poorly constrained flow boundary conditions may also be contributing towards this warm bias. Based on our analysis we were able to identify a deficiency in the model (for example missing representation of ground ice), data to characterize the subsurface and to constrain the boundary condition. We plan to use the modeling results to guide the collection of additional data at our sites and improve the simulated estimates of thermal and hydrologic states at our sites.
Warm bias in thermal regime would certainly lead to errors in biogeochemical processes, however, we believe these biases can addressed by our planned additional data collection and through model calibration process as demonstrated by Atchley et al. 2015 (Page).

2. Another issue is the interannual variability. The authors state that 2013-2014 was the only year in which the needed information was available at the desired time step. It would be though helpful to show simulations also of other years, if possible, also against partially complete datasets. If only one year is shown, it is difficult to assess the validity of the model in dealing with potential differences in, for example, precipitation regimes. The authors state that the model is not calibrated for a specific year and a specific time, but nevertheless, this issue should be addressed.
We were limited to a single year for this study due the availability of data sets. While we plan to investigate the important issue of interannual variability as data becomes available, we have added a discussion at Page 27 Line 4.
"Present study was limited to single year when we had all the necessary data for model forcings and validation available, thus was not able to investigate and address the role of interannual variability. We plan to address this important problem as more data from our sites become available."

3. On this note: how would such a topography-based model react to topography changes due to, say, ice-wedge degradation? This seems to be an important issue, as highlighted by Liljedahl et al., 2016. This is a big issue in the sense that if the model cannot cope to topography transformations, it would only work under present-day conditions, and would not be very useful for future simulations.
Ice wedge degradation is an important process in polygonal tundra, however, while presented model can model the effect of topography it currently does not have the capability to model dynamically evolving topography. We have added statement at Page 26 Lines 28-29 to discuss this limitation.
"Model developed here does not have the ability to simulate the dynamic changes in microtopography expected due to ice-wedge degradation (Liljedahl et al. 2016}).."

4. There is also a problem of scalability of the results. The model seems to be very complex and computationally demanding, since it is a 3D representation of thermal regimes working at a very small spacial scale. The authors mention a potential coupling with CLM. How this coupling could work is not clear to me. Also, upscaling the detailed information of the 3D model at "just" the ecosystem-scale would be already a significant first step. How do the authors imagine such an upscaling? A stochastic approach would may be help, but how to link the very detailed and small-scale information needed to initialize the model to the larger scale ecosystem dynamics?

We agree that upscale transfer of knowledge gained through fine-scale simulations to inform and improve simulations at larger spatial scales is a critical issue. While we can not comprehensively address that research problem in the current study, we have added text to indicate how the fine-scale simulation capability explored in this study contributes to a broader scaling strategy for the NGEE Arctic project. The following text has been added at Page 27 Lines 11-25:
"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. 2015). 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."

- Tesfa TK, H Li, LYR Leung, M Huang, Y Ke, Y Sun, and Y Liu. 2014. "A Subbasin-based framework to represent land surface processes in an Earth System Model." Geoscientific Model Development 7(3):947-963. doi:10.5194/gmd-7-947-2014

Minor points:

1. Page 1, Line 19: blank between words and hyphen.

Blank space added between words and hyphen at page 1 Line 19. Rest of the document updated for similar consistency as well.

2. Page 2, Line 20: the link does not help readability. I suggest to insert the link in the Appendix.
We have moved the URL to be part of the citation in the bibliography.

3. Page 2, Line 25: Please check the citation, it should be Cresto Aleina et al., 2013.
We have update the citation correctly.

4. Figure 2: The Paper has a large amount of Figures, and I am not sure if all of them are needed. This Figure, for example, only shows data that are used in the model, but might as well be shown in the Appendix.

We have moved several figures to appendix and referenced them in the text as needed. We have also reduced the number of subfigures in Figures 8-11 to show plots at 5cm ,10cm, 50cm, and 150 cm depths only. Tables showing statistics at all 16 sensor depths has been included in the appendix (Tables 8-11).

5. Page 7, Line 11: "the features": which features? Please, be specific.
Updated to "polygonal features".

6. Page 7, Line 23: the information about the contribution of Dr. Craig Tweedie can be inserted in the Acknowledgments.
Citation to the public archive of the LiDAR data has been included in the text
Page 5 Line 27.
"High-resolution LiDAR data (25 cm resolution) were collected on October 4, 2005 by Tweedie (2010)."

7. Page 8, Line 13: You do not enforce any information on polygonal shape. But this information seems to me to be needed for simulation of other properties, such as, for example, water table dynamics. What do you mean then to scale the results to the whole region? Please elaborate this sentence, since it is not clear how this scaling would work.

In our approach characterization of microtopographic features (Center, Ridge, Trough) is only used to parameterize the soil properties and determine the surface boundary conditions to applied in the 3-D PFLOTRAN model. Polygonal shape information is not used by PFLOTRAN, rather it simulates the water and energy dynamics from first principles in 3-D where topography drives the simulated patterns. While our test cases were focused on only two polygons to allow for fair comparison with observed data, our approach would allow application of the model (thus scaling up of the problem) to larger regions of interest where high resolution elevation data is available.

8. Figure 3 and Figure 4 could be incorporated.
We have consolidated Figures 3 and 4 in one.

9. Page 11, Line 21 and further: please highlight in the text which ones are the prognostic variables you are evaluating in the equations and which ones the parameters. What is the model time-step? I might have missed it, but it should be clearly stated here.

Liquid pressure and bulk temperature are the prognostic variables in the model.
Page 12 Line 8 "the liquid pressure P and the bulk temperature T are the unknown variables" has been changed to "the liquid pressure P and the bulk temperature T are the primary unknown variables."

The time stepping scheme is Backward-Euler and the time step size is dynamically varied to balance error and the solvability of the nonlinear system. In practice this amounts to a time step size on the order of seconds when a phase transition occurs, and 30 minutes otherwise.
Page 10 Line 22-23 "The PDEs are spatially discretized using a finite volume technique, and backward Euler scheme is used for implicit time discretization."

10. Page 14: The first paragraph of Chapter 4 is about Methodology (initial and boundary conditions) and should should be moved there.

Subsection: Initial and boundary conditions has been moved to Section: Methodology

11. Page 14, Line 17: How long is the spin up?
Spin up simulations were conducted for a period of 10 years. A statement clarifying that has been added to Page 13, Line 17-21.
"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."

12. Figures 8 to 11: Please check the style of the caption. There isa a racket at the beginning of the sentence that does not make sense.
Caption has been updated.

13. Figure 12: Why do you only show Site A? Is the behavior representative for the other Sites?
Similar spatio-temporal variability was observed in simulations at the sites. Results for Site A has been included in the text for illustrations, however, we have added the results for other sites too in the Appendix D1 Figure 18.

14. Figure 13: Is this figure really needed? In any case, if yes, it should go in the methodology. The color scale should also be changed, to improve readability for colorblind readers.
Figure 13 illustrates the approach for calculation of effective thermal conductivity in the model. It has been moved to methodology as Figure 7.

15. Tables 2 to 5. I do not think that all this information is needed. I suggest to give the information only at the depths showed in Figures 8 to 11. In this way, the authors could summarize the 4 tables in only one, improving readability.

We have reduced the detail to show results only at 5, 10, 50, 150 cm depths. Detailed tables has been added to the Appendix D1.

16. Page 21, Line 6: Are such surface processes then implicitly included in the model? If yes how?
Surface processes are not represented in the model. We have updated the statement on Page 21 Line 6 to "not modeled in our study" to clarify that.

17. Page 25, Line 3: Which parameters could be tuned? It would be interesting to understand how tuning which parameter would improve model-data comparison.

A statement pointing to the parameter table have been added in model description. Page 14 Line 4 "Key parameters for the model relevant for current study are described in Table 7." We have also added a reference to relevant parameter sensitivity and calibration analysis study by Atchley et al. 2015.

18. Figure 15 does not seem to be needed in the text and be moved to the Appendix.
Figure has been moved to the Appendix.

19. Page 27, Lines 4 and further: This discussion can be moved in the Conclusions, since it outlines ongoing and future work, which is not part of the paper.

We have moved the discussion to the conclusion section.

20. Page 27, Line32: please change “models” in “model”.
“Models” has been changed to “model”.

References:

Liljedahl et al., Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology, Nature Geosciences, 2016. doi:10.1038/ngeo2674

Anonymous Referee #3

J. Kumar et al. presented a pilot study using PFLOTRAN model to investigate the role of micro-topography in soil thermal dynamics of different types of ice-wedged tundra, which is important for further studying the responses of large-amount of frozen soil carbon to warming. Field measurements were provided for parameterization and validation. Therefore, I think the topic is important and the method is appropriate .

The 3-D modeling is computing-intensive, it is hard, if not impossible, to be coupled in large-scale climate or terrestrial ecosystem models to investigate the effects of fine scale heterogeneity. Meanwhile, it is well-known that micro-topography of ice-wedged tundra ecosystem plays an important role in redistribution of surface water and vegetation growth. Therefore, the manuscript should focus on quantitatively assessing the role of micro-topography in soil thermal dynamics by comparing sensitivity tests with and without 3-D heat transfer.

We agree with reviewer comments. While the role of micro-topography in tundra ecosystem, effort to develop high resolution process based models for these processes have been limited in the literature. In the presented study we have developed a framework for numerical rigorous high resolution modeling of hydrologic processes. We have demonstrated the ability of the models to capture the effect of micro-topography. We have presented qualitative and quantitative results for a set of simple case studies. Our study provides the base framework to make possible the further detailed analysis suggested by the reviewer. We are actively working on these issues and will report the results in future publications. And we also have made our best effort to archive and publicly release all source codes, workflows, and input/output data sets from the study to allow others interested in conducting such studies.

Unfortunately, the manuscript reached two main conclusions: 1) the 3-D modeling can properly simulate the soil thermal dynamics under the complex micro-topography, which is good; and 2) microtopography is important, which is already known without the 3-D modeling.
The role of microtopography in tundra ecosystems has been studied primarily through field-based investigation. Modeling studies using traditional 1D approaches have been focused around particular sites and thus may or may not be extensible and/or applicable to new or larger regions of interest. The 3-D modeling approach developed in our study enables the modeling of these processes from first principles using a microtopography-resolving model. Patterns due to polygonal microtopography emerge in the model organically, without explicit spatio-temporal flow patterns defined in the model. This enables such studies in any region of interest where high resolution DEM are available to capture the topography in the model. We believe this is key and a new contribution of the presented work beyond published literature.

I believe the authors can do better work with this 3-D model and provide readers more informative results than the current one. I do not recommend publication of the manuscript in the current form.

I would suggest the authors to split the manuscript into two since there were already too much content in the current manuscript. The first deals with model description, model validation and detailed sensitivity tests on only one of the four sites. The second manuscript then deals with differences among different types of ice-wedged tundra ecosystems and upscaling to larger regions.

We thank the reviewer for suggesting a set of analysis where the presented 3-D modeling framework can applied to gain important insights. Under the NGEE-Arctic project we have been working on addressing these questions. However, it's a multi-faceted and complex problem that requires a series of designed steps.
The presented study is the first key step towards that goal and is focused on developing an end-to-end modeling framework to simulate the thermal flow processes in this micro-topographic environments. We applied the methodology at four different study sites of interest and developed methods to synthesize and use the best available data for a field scale study. We have identified a number of key processes and analysis that are needed but were not included but are subject of ongoing work and will be reported in the future.

In close coordination with observation team we currently are working on addressing the data gaps identified in our study which would allow us to conduct a watershed to regional scale application of our framework at Barrow, Alaska.

While we agree with and are working on majority of analysis suggested by the reviewer, we believe we have designed our steps slightly different. The presented approach is our first step that provides us with the numerical framework to conduct the other investigations.

More specifically, I would suggest to do the following model runs on one site in the first manuscript: 1) shut down the lateral heat flow in the 3-D model and compare the results with those using fully 3-D heat transfer. This work is to demonstrate the importance of lateral heat exchange; 2) in one simulation, use the same soil texture for all microtopography positions, e.g. rim and center. Compare results with different soil textures; 3) prepare future climate data using GCM outputs under different scenarios. You might also need to convert atmospheric driving to near surface soil temperatures for different micro-topographic components. The long-term simulating might reveal some modeling issues, e.g. lateral boundary conditions; 4) implement different amount of

excess ice in soil column to test whether excess ice causes disagreement between simulated or measured soil temperatures.

We would also like to highlight a number of recent studies conducted by our NGEE-Arctic colleagues that complement ours and address some of the suggestions identified by the reviewer. 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. Findings from their study would help guide the parameterization in our model.

A. Atchley
aatchley@lanl.gov

Comment by Adam L. Atchley, Dylan R. Harp, Ethan T. Coon, Cathy J. Wilson

The authors present an impressive modeling effort investigating 3D water and energy simulations of polygonal tundra. Their research is of interest to the community as it undoubtedly yields insight into the thermal hydrology of polygonal tundra. Furthermore, a modeling effort informed by extensive field measurements provides a unique opportunity to validate and properly shape the process rich models currently being developed for terrestrial Arctic applications. It is for this particular reason, that we are first interested in this manuscript and second concerned with the message in section 5.1. In particular, we refer to line 2-3 on page 25, "At our study sites, while calibration may compensate for lack of data, it does not improve our understanding of the system."
In section 5.1 the authors provide reasons for not calibrating the model to the observed data available at the study site, specifically that process rich models have high degrees of freedom and therefore are plagued with non-uniqueness (equifinality). In other words, there are multiple combinations of parameters, or more generally model structures (Beven, 2006) that can produce optimized results which fit observed data equally well. While the authors do not quantitatively demonstrate the existence of equifinality here, non-unique parameter combinations certainty exist in this situation, as has been systematically identified for thermal hydrological models at the same site by Atchley et al. (2015) using multi-try calibration and rigorously quantified by Harp et al., (2016) using Null-Space Monte Carlo. However, it is our understanding that the literature addressing equifinalty does not argue for giving up calibration as a lost cause, but rather strongly suggests that additional efforts are required to account for a set or distribution of parameter combinations consistent with observations (e.g. Vrugt et al., 2009, Vrugt and Ter Braak, 2011; Bárdossy, 2007; Tonkin, 2009) and model structural error (Beven, 2005; Clark et al., 2008; Fenicia et al., 2011; Larson et al., 2014).

The research behind calibration and model optimization has long since evolved from simple parameter fits to more strategic calibration methods (Hill, 1998). Therefore, we believe that equifinality does not provide a justification to avoid calibration, especially if the objective of the modeling exercise is to improve understanding of system and model behavior. On the contrary, it has been our experience that, while difficult, time-consuming, and computationally expensive, extensive, systematic multi-try calibration can yield important system understanding and identify model capabilities and limitations. The work presented in Atchley et al., (2015) at the same site at the Barrow Environmental Observatory, shows that systematic multi-try calibration can be used as a tool to reduce model structural error and achieve system understanding. For example, calibration efforts led to the recognition of the importance of the representation of snow distribution and depth hoar formation in our models. These insights are not simply better

model parameters, but are physical representations of system components; this effort led to better system understanding. Furthermore, quantifying the equifinality of the combined model and represented system then allows for quantification of model uncertainty, where for example the projected ALT uncertainty attributed to parameter uncertainty can be measured and compared to meteorological and/or climate model uncertainty (Harp et al., 2016). Moreover, the parameter sensitivity quantified by such exercises has, in our opinion, provided valuable information for reducing model uncertainty. Porosity and material thermal conductivity measurements are shown to have the greatest potential to reduce projected ALT uncertainty (Harp et al., 2016), thereby directing which additional field data and process understanding are necessary to reduce uncertainty.

In the context of the model presented in this manuscript, we realize that exhaustive model calibration may be computationally infeasible, and we also do not over look the valuable contribution presented here as the 3D representation of energy and water fluxes in freeze-thaw polygonal tundra indeed pushes the boundaries of process-rich mechanistic modeling. Therefore, it is not our wish to force model calibration and parameter sensitivity analysis on the current manuscript. However, we strongly encourage the authors to reconsider the stated view of model calibration and to discuss how calibration and parameter sensitivity may provide insight into model performance as well as system understanding in polygonal tundra.

We thank Atchley et al. for their interest in our work and insightful comments. We completely agree with the comments above. Model calibration has since long been used a powerful and insightful tool in hydrology and it was certainly not our intention to overlook its power, usefulness and validity. Objective of our presented study was to apply the modeling framework to the study site and synthesize and use available observational data sets for the presented case studies. Work by Atchley et al. 2015 and Harp et al. 2016 provides methodology for calibration, insights in parameter sensitivity and calibrated parameters for three phase thermal hydrology model similar to ours and would certainly complement and guide our future modeling studies. We have updated our manuscript to add a reference to Atchley et al. 2015, Harp et al. 2016 and relevance of their work for our study (Page 24 Lines 16-18).

References

Atchley, A., S. Painter, D. Harp, E. Coon, C. Wilson, A. Liljedahl, and V. Romanovsky (2015a), Using field observations to inform thermal hydrology models of permafrost dynamics with ATS (v0.83), Geoscientific Model Development, 8, 2701-2722. Doi: 10.5194/gmd-8-2701-2015.

Bárdossy, A. "Calibration of hydrological model parameters for ungauged catchments." Hydrology and Earth System Sciences Discussions 11.2 (2007): 703-710.

Beven K. 2005. On the concept of model structural error. Water Science and Technology. 52(6): 167-175

Beven K. 2006. A manifesto for the equifinality thesis. Journal of Hydrology. 320: 18-36. doi:10.1016/j.jhydrol.2005.07.007.
Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., ... & Hay, L. E. 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44(12) W00B02, doi:10.1029/2007WR006735,
Fenicia, F., Kavetski, D., & Savenije, H. H. 2011. Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. Water Resources Research, 47(11). W11510, doi:10.1029/2010WR010174
Harp, D., A. L. Atchley, S. L. Painter, E. Coon, C. Wilson, V. Romanovsky, and J. Rowland (2015), Effect of soil property uncertainties on permafrost thaw projections: a calibration-constrained analysis, The Cryosphere 10(3), 1-18. doi: 10.5194/tc-10-1-2016.
Hill, Mary C. Methods and guidelines for effective model calibration. Denver, CO, USA: US Geological Survey, 1998.
Kavetski, D., & Fenicia, F. 2011. Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. Water Resources Research, 47(11). W11511, doi:10.1029/2011WR010748,
Keating, Elizabeth H., et al. "Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality." Water Resources Research 46.10 (2010).
Printer-friendly version
Larsen L., Thomas C., Eppinga M. 2014. Exploratory modeling: extracting causality from complexity. EOS, 95(12) 285-292.
Vrugt, Jasper A., Cajo JF Ter Braak, Hoshin V. Gupta, and Bruce A. Robinson. "Equi-finality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?." Stochastic environmental research and risk assessment 23, no. 7 (2009): 1011-1026. TCD
Tonkin, Matthew, and John Doherty. "CalibrationâĂŘconstrained Monte Carlo analysis of highly parameterized models using subspace techniques." Water Resources Research 45, no. 12 (2009). Interactive
comment
Vrugt, Jasper A., and C. J. Ter Braak. "DREAM (D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems." Hydrology and Earth System Sciences 15, no. 12 (2011).

Modeling the spatio-temporal variability in subsurface thermal regimes across a low-relief polygonal tundra landscape

Jitendra Kumar[1], Nathan Collier[2], Gautam Bisht[3], Richard T. Mills[4], Peter E. Thornton[1], Colleen M. Iversen[1], and Vladimir Romanovsky[5]

[1]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[3]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[4]Intel Corporation, Hillsboro, OR, USA
[5]Geophysical Institute, University of Alaska Fairbanks, AK, USA

Correspondence to: Jitendra Kumar (jkumar@climatemodeling.org)

Abstract. Vast carbon stocks stored in permafrost soils of Arctic tundra are under risk of release to atmosphere under warming climate. ~~Ice–wedge~~ Ice – wedge polygons in the ~~low–gradient~~ low – gradient polygonal tundra create a complex mosaic of microtopographic features. The microtopography plays a critical role in regulating the fine scale variability in thermal and hydrological regimes in the polygonal tundra landscape underlain by continuous permafrost. Modeling of thermal regimes of this sensitive ecosystem is essential for understanding the landscape behavior under current as well as changing climate. We present here an end-to-end effort for high resolution numerical modeling of thermal hydrology at real–world field sites, utilizing the best available data to characterize and parameterize the models. We develop approaches to model the thermal hydrology of polygonal tundra and apply them at four study sites at Barrow, Alaska spanning across low to transitional to high-centered polygon, representing a broad polygonal tundra landscape. A ~~multi–phase~~ multi – phase subsurface thermal hydrology model (PFLOTRAN) was developed and applied to study the thermal regimes at four sites. Using high resolution LiDAR DEM, microtopographic features of the landscape were characterized and represented in the high resolution model mesh. Best available soil data from field observations and literature was utilized to represent the complex heterogeneous subsurface in the numerical model. Simulation results demonstrate the ability of the developed modeling approach to model the complex thermal regimes across the sites. Our study provides insights into the critical role of polygonal tundra microtopography in regulating the thermal dynamics of the carbon rich permafrost soils. Study also highlights the importance of field–based observations of soil thermal and hydraulic properties for ~~modeling–based~~ modeling – based studies of permafrost thermal dynamics and provides motivation for future observations by identifying gaps in our current understanding of the system.

1 Introduction

Coastal Arctic ~~landscapes–dominated~~ landscapes – dominated by wetlands and patterned ground–cover approximately ~~5–10~~5 – 10% of Earth's land surface and play an important role in the hydrology, geomorphology, biogeochemistry and vegetation dynamics of the vast Arctic region. The low-gradient topography of the polygonal tundra characteristic of these landscapes

is a complex mosaic of ~~microtopographic~~ micro-topographic features created by ice wedge polygons. This ~~microtopography~~ micro-topography leads to strong ~~fine–scale~~ fine – scale variability in thermal and hydrological regimes of landscapes underlain by continuous permafrost. Permafrost landforms like drained-lakes, low-centered polygons and high-centered polygons retard surface runoff after snowmelt, leading to increased surface water storage (in form of lakes, ponds and wetlands) (Kane et al. (2003)). Complex surface drainage patterns lead to heterogeneous soil moisture and substrate conditions supporting a wide range of vegetation composition across the landscape. Arctic tundra soil pools are estimated to contain 190 Pg of carbon (Post et al. (1982)), much of which is under risk of rapid release to the atmosphere in a warming climate. Hobbie et al. (2000) studied the controls over carbon storage and turn-over in Arctic soils and found temperature, micro-topography and vegetation composition to be the primary controls at regional scale.

Changes in the surface geomorphology which lead to the creation of ice-wedge polygons are induced by thermal disequilibrium and permafrost degradation. Lowland polygonal relief is dominated by low-centered polygons and high-centered polygons. Low-centered polygons are the most common polygonal landscape feature and are characteristic of poorly drained tundra. They consist of a raised rim with a wet central depression. Raised rims are the result of growing ice wedges that push material away from the center of the ice wedges to the sides (French (2007)). The standing water in the ice-wedge troughs leads to thermal erosion (i.e., accelerated thawing) along the rim. This preferential thaw may cause the ridge to collapse and form trenches surrounding the polygon center, inverting the relief to form high-centered polygon. High-centered polygon are well-drained with often dry centers, leading to low peat accumulation and deeper active layers. The ~~microtopographic~~ micro-topographic relief and associated heterogeneity in soil moisture support a diverse distribution of vegetation in the Arctic, with wet centers and troughs of low-centered polygon covered by mosses and sedges, while drier rims, and centers of high-centered polygon and rims of low-centered polygon are dominated by mosses, lichen and dwarf shrubs. These diverse land-cover types can also alter the surface energy balance and thermal properties through changes in albedo, surface roughness, and evapotranspiration (Langer et al. (2011)).

Large scale climate and terrestrial ecosystem processes are represented in global to regional scale climate and ecosystem models. However, most of these models lack the representation of fine scale heterogeneity in surface and subsurface processes at subgrid scale that exercise significant control on the landscape scale behavior. Representation of the fine scale heterogeneity is important to model the non-linear processes involved (Cresto Aleina et al. (2013)).

Accurate characterization and modeling of subsurface thermal regimes in polygonal tundra is critical for our understanding of this sensitive system and our ability to predict its fate under climate change. In this study we developed approaches to 1) characterize the surface ~~microtopography~~ micro-topography and subsurface structure of the polygonal tundra, 2) represent heterogeneous subsurface stratigraphy and hydraulic and thermal properties, 3) numerically model permafrost hydrology, and 4) combine the above to simulate the permafrost thermal regime at field sites in a polygonal tundra region near Barrow, Alaska.

2 Study Area

The study area is located within the Barrow Environmental Observatory (BEO), (Figure 1) which lies 6 km East of Barrow, Alaska (71°18'N, 156°35'W), and is a field site of the U.S. Department of Energy's Next Generation Ecosystem Experiments (NGEE) – Arctic project. The BEO spans 32.21 km^2 of natural tundra, lakes, and wetlands, and is reserved for scientific research. The landscape has low topographic relief, with elevations ranging from 0 to 7 m above sea level and low hydraulic gradient present across the region. Barrow has a polar maritime climate with mean annual air temperature of -12.0°C and 3.3°C during summer (June – August) (Liljedahl et al. (2011)). The winter snowpack averages 20 to 40 cm, but snow accumulation is spatially variable due to variations in terrain roughness and drifting from strong easterly winds (Bockheim et al. (2001)). Annual adjusted precipitation is 173 mm, with the majority of precipitation falling during summer months (Liljedahl et al. (2011)). The polygonal tundra landscape is punctuated by thermokarst lakes and drained lake basins, with grass, moss, and sedge as dominant vegetation types. Basins at Barrow are underlain by permafrost within 1 m of the surface and are classified as Gelisols, with an organic-rich surface layer underlain by a horizon of silt and clay to silt-loam textured mineral material and a frozen organic-rich mineral layer (Bockheim et al. (2001)). The seasonal active layer thickness ranges between 30 to 70 cm at the BEO.

The various stages of geomorphologial and ecological change from low to transitional to high-centered polygon, lakes and drained lakes are all represented at the BEO. Following a "space for time" philosophy, NGEE–Arctic intensive field sites at BEO were chosen across the landscape to observe and study polygonal landscapes at all stages of transition. Table 1 shows the characteristics of four sites (A, B, C, D) where our current study is focused.

Table 1. Areas A, B, C, D polygonal features and environmental characteristics.

Area	Characteristics	Relative Elevation (Min / Max / Median m)
A	Low center polygons (with ridges and troughs)	Low (4.5 / 4.9 / 4.6)
B	High center polygons	High (4.5 / 5.1 / 4.8)
C	Transitional low center polygons (with ridges and troughs)	Moderate (4.3 / 4.9 / 4.6)
D	Low center polygons (no troughs)	Low (4.1 / 4.6 / 4.3)

Relative elevation is qualitative summary of topography in the region, while Min/Max/Median are minimum, maximum and median elevations.

A suite of observations are being collected at each of these intensive sites. Since 2012, meteorological data (including air temperature, summer precipitation, snow depth, relative humidity, wind speed and radiation) are being collected at all four sites (Hinzman et al. (2014b)). Surface temperature data is being collected along a transect from the center of the polygon to ridge to trough. Each location on the transect consists of nine soil temperature sensors ranging in depth from 2 cm to 150 cm (Romanovsky and Cable (2012)). Figure 2 shows the observed time series of hourly air temperature and liquid precipitation at our four sites for the period of October 1, 2013 – September 30, 2014. Data for the one year period (October 1, 2013 – September 30, 2014) were chosen for this study since it was only complete year for which all the necessary observations

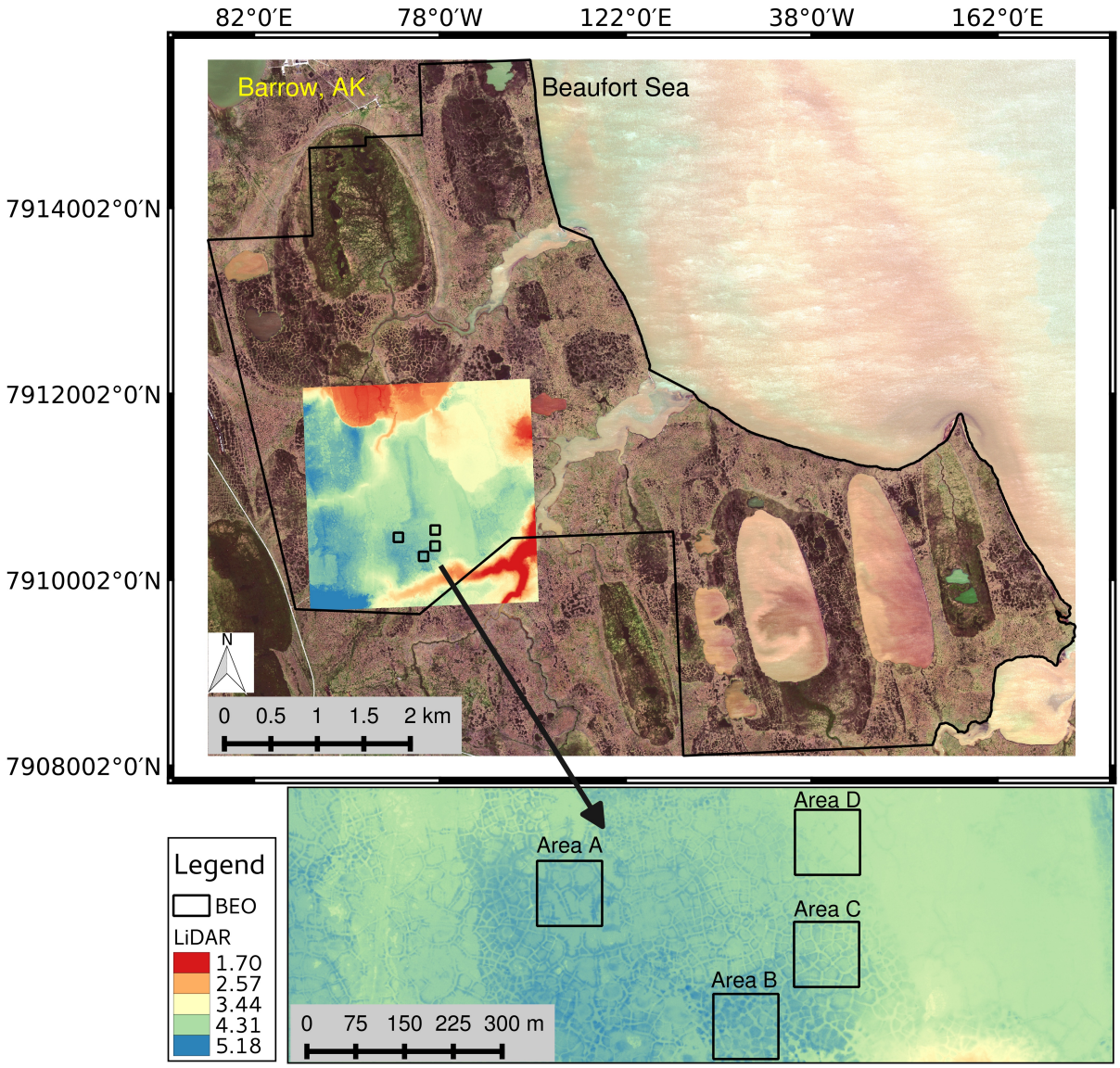


Figure 1. NGEE-Arctic Field Sites at BEO (LiDAR elevations are in unit of meter above mean sea level)

were available. Figure ~~??~~ 3a, d, g, j shows high resolution imagery of the sites where the observations have been collected. Figure ~~??~~ 3a, d, g, j shows the boundary of the region around the intensive sites where we conducted the detailed modeling study presented here.

~~Aerial imagery of the study areas at Sites A, B, C, D.~~

3 Methodology

To model the thermal regimes of the heterogeneous polygonal tundra ecosystem we developed approaches to 1) characterize the surface ~~microtopography~~ micro-topography and subsurface structure of the polygonal tundra 2) represent heterogeneous subsurface stratigraphy and hydraulic and thermal properties, and 3) numerically model permafrost hydrology.

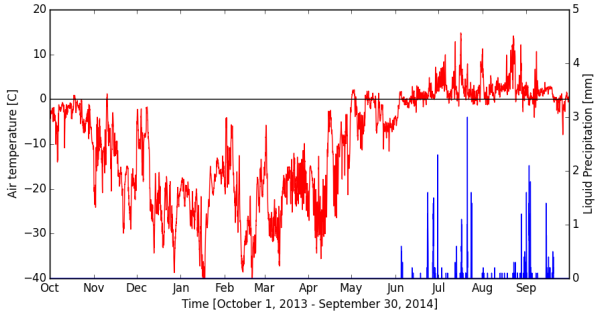
3.1 Representation of landscape heterogeneity

Accurate representation of polygonal tundra in the model requires 1) identification of ~~microtopographic~~ micro-topographic features on the landscape and 2) characterization of soil stratigraphy and properties across the landscape.

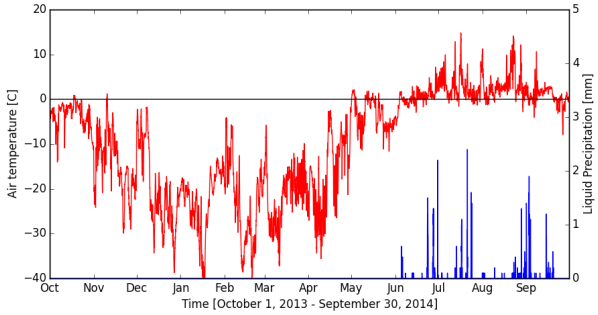
3.1.1 Identification of polygonal features

The human eye can discern polygonal patterns and features in satellite high resolution satellite imagery with relative ease. However, automated recognition and delineation of ~~such~~ polygonal features are challenging due to the variability in their spectral appearance, irregularity of polygon shape, dimension and orientation and lack of unique spectral signatures associated with the features (Skurikhin et al. (2013)). Muster et al. (2012) investigated the subpixel heterogeneity in Landsat satellite imagery over ice-wedge polygonal tundra using a range of multi-scale data (field measurements and remote sensing) and concluded that resolutions of 4 m or less are necessary to map the fine-scale landscape elements of polygonal tundra. Skurikhin et al. (2013) used a combination of segmentation and shape-based classification approaches using high-resolution WorldView-2 satellite imagery (60 cm resolution) to identify the landscape elements within the BEO. While they reported an overall accuracy of 95%, their study region was limited to a 1000 $\times$ 1100 pixels subimage. The scalability of such a specialized algorithm based on high-resolution satellite imagery (of limited availability) is untested and difficult for application for landscape–scale studies like ours.

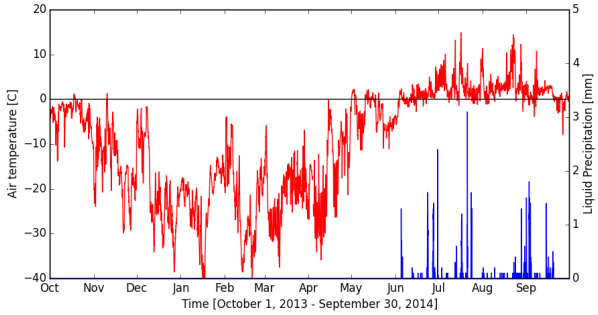
Thus, with landscape–scale application in mind, we employed a relatively simple and generic approach using a high resolution Digital Elevation Model (DEM) that exploits the relative difference in surface elevations that distinguish the polygonal features (center, ridges and troughs). High-resolution LiDAR data (25 cm resolution) were collected on October 4, 2005 by ~~Dr. Craig Tweedie of the University of Texas, El Paso.~~ Tweedie (2010) . The LiDAR data horizontal and vertical accuracy were approximately 30 and 15 cm respectively. Covering an approximately 2.5 $km \times 2.5$ km area, the data set encompasses all of the NGEE-Arctic intensive sampling sites (Figure 1) where our study was focused. Using a high resolution DEM created from this data set, elevation contours (10 cm interval) were developed to segment and classify the landscape in centers, ridges and troughs.



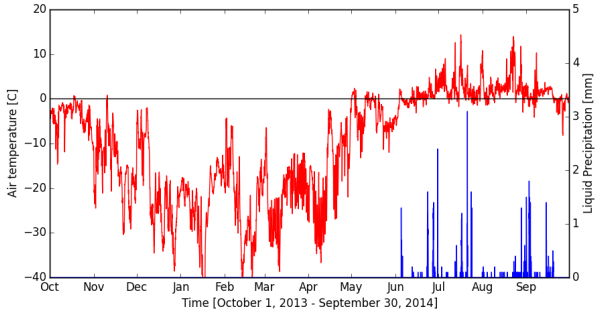
(a) Site A



(b) Site B



(c) Site C



(d) Site D

Figure 2. Meteorological conditions at study sites. Red lines: hourly air temperature, Blue histograms: hourly summer time liquid precipitation

- *Site A*: Elevation at Site A ranged from a minimum of 4.5 m to a maximum of 4.9 m (Figure 3(a)). Low elevation depressions (4.5 – 4.6 m) were classified as Center, surrounded by elevated rim (4.7 – 4.9 m), and deep troughs adjacent to them (4.6 – 4.7 m). Center, Rim and Troughs occupied approximately 35%, 24% and 41% area respectively (Figure 3(b)).

- *Site B*: At Site B with high-centered polygons, elevation ranges from 4.5 – 5.1 m (Figure 3(c)). High elevation areas (4.8 – 5.1 m) were classified as Center, low elevation (4.7 – 4.8 m) rim and deep troughs (4.5 – 4.7 m), occupying 39%, 21%, and 40% of total area respectively (Figure 3(d)).

- *Site C*: A wider range (4.3 – 4.9 m) of elevations are present at Site C (Figure 3(e)) which are often considered *low* or flat-centered polygon. Low elevation areas (4.5 – 4.6 m) were classified as Center, with raised rims (4.6 – 4.9 m) and deep (4.3 – 4.5 m) troughs. Center, Rim and Trough occupied 35%, 36%, and 29% of the total area (Figure 3(f)).

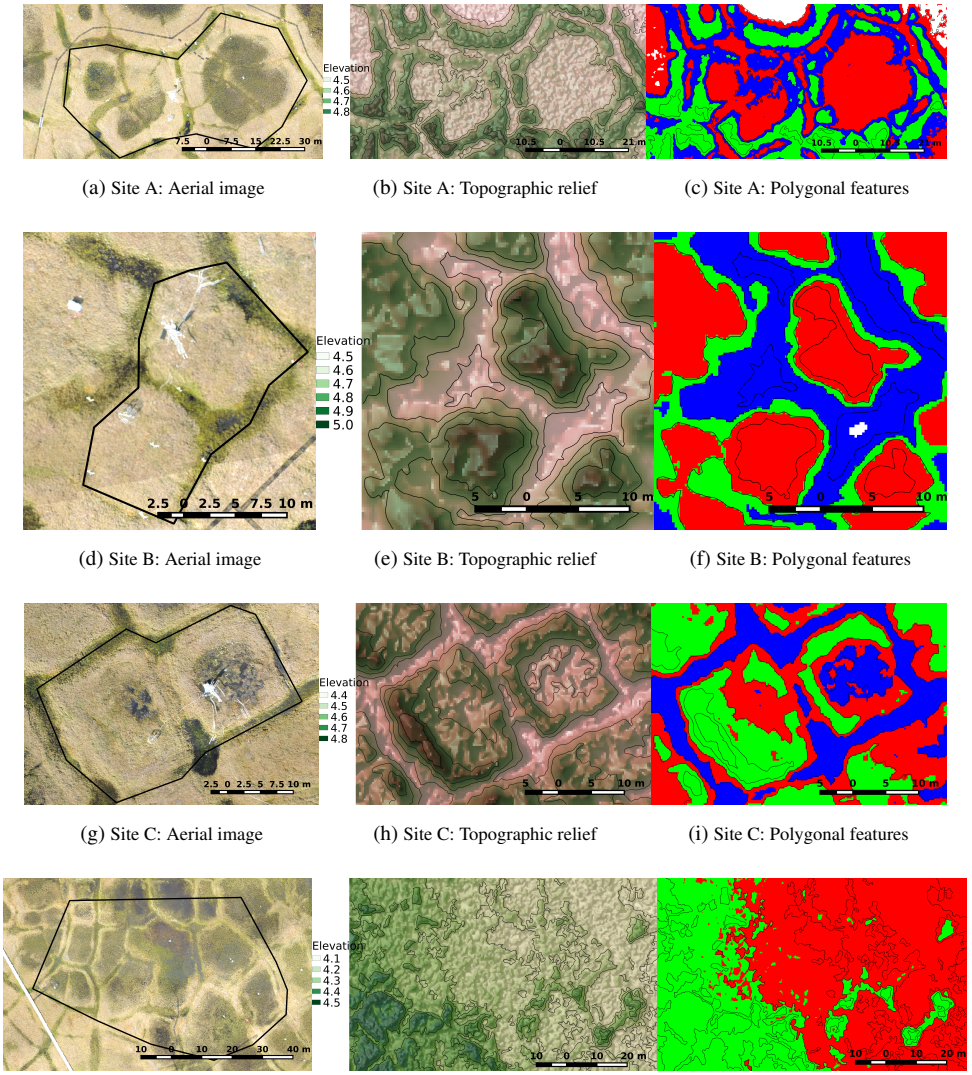
- *Site D*: Site D is relatively flat and is thus identified as flat-centered polygons, with the entire area within a narrow elevation range of 4.1 – 4.6 m. While polygonal features were evident in 0.25 m resolution aerial optical image (Figure **??**(d)), they were difficult to identify in the LiDAR DEM (Figure 3(g)) due to the limitations of the vertical accuracy of LiDAR. Trough features in flat-centered polygons are not well pronounced. Thus the area was classified only as Center (4.1 – 4.3 m) and Rim (4.3 – 4.6 m) features. About 72% of the area was classified as Center while 28% as Rim (Figure 3(h)).

We did not apply any specialized rules to enforce any shape, dimension and/or patterns of the polygon features (Center, Ridge, Trough), allowing us to scale our approach to the entire region where high resolution DEMs were available.

3.1.2 Subsurface characterization

The structure and properties of subsurface soils are important factors controlling the pattern and variability of permafrost thermal processes in the tundra environment, and accurate characterization and representation of the heterogeneous subsurface properties is critical to understanding and modeling the subsurface thermal dynamics. However, the limited availability of soil properties in tundra environments and at our sites at the BEO presents a significant challenge.

During the period July 31 – August 3, 2012 a field campaign was conducted by NGEE-Arctic researchers to collect soil cores at one replicate polygon at each of the sites A, B, C, and D at three micro-topographic positions (Center, Ridge, Trough) per polygon. Cores were collected using a hammer and a 5.08 cm diameter corer to collect one soil core per location to a depth of 30 cm. The soil horizons (moss, organic layer, mineral layer) were measured for each core to the nearest centimeter. A deep organic layer was found at several of the locations. However the total depth of the deep organic layer was not determined if it extended beyond the 30 cm core depth. Cores were collected for the purpose of biogeochemical analysis and thus no soil hydraulic or thermal properties were measured by the team. Figure 4 illustrates the subsurface soil horizons based on observations that we used in all our modeling studies presented here. In absence of co-located observations for soil hydraulic and thermal properties we derived the data for use in our studies from the published literature in tundra regions (Hinzman et al.



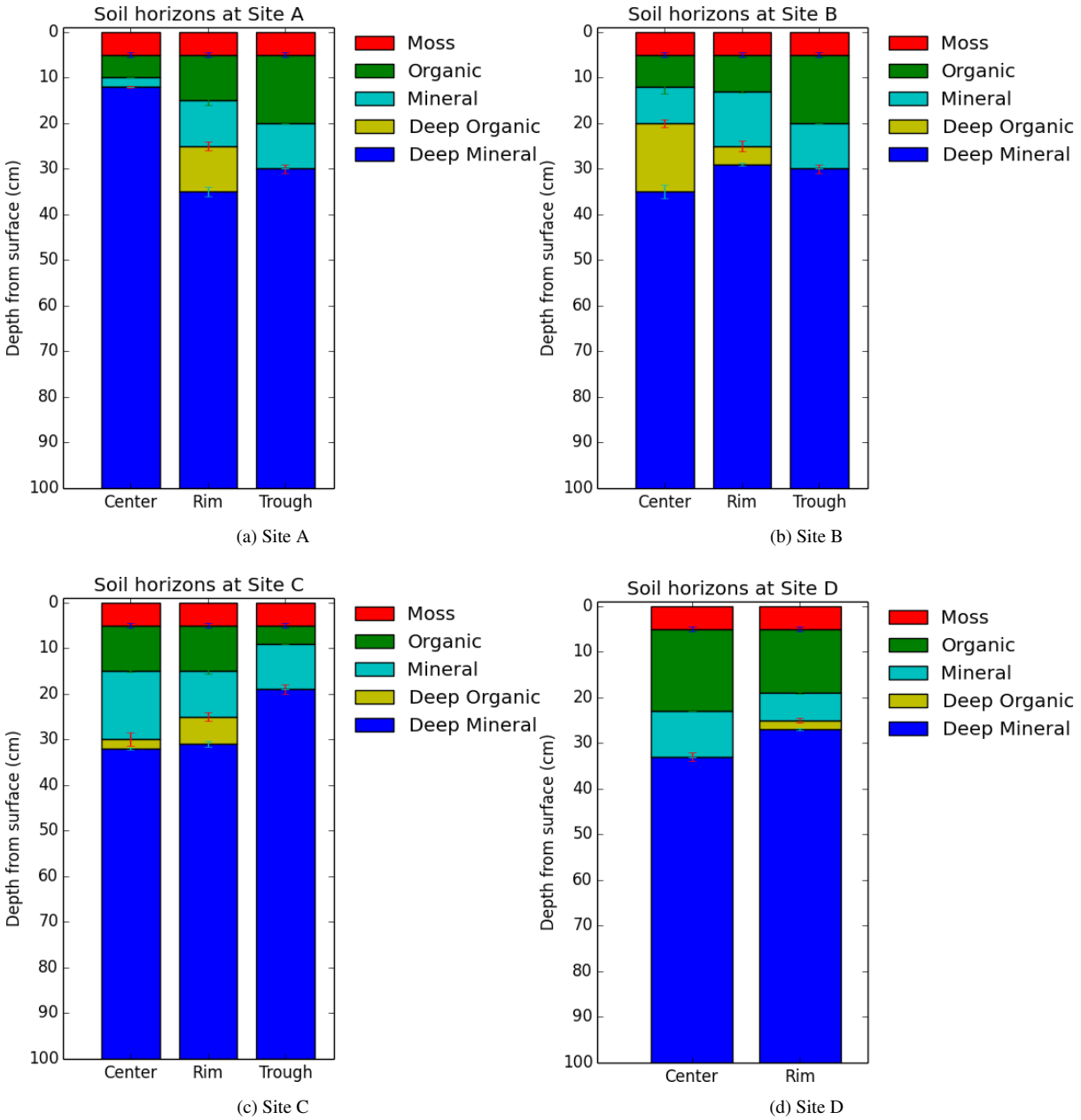
(a) Site A: Aerial image (b) Site A: Topographic relief (c) Site A: Polygonal features

(d) Site B: Aerial image (e) Site B: Topographic relief (f) Site B: Polygonal features

(g) Site C: Aerial image (h) Site C: Topographic relief (i) Site C: Polygonal features

(j) Site D: Aerial image (k) Site D: Topographic relief (l) Site D: Polygonal features

Figure 3. Elevation contour based classification of the study areas at Sites A, B, C, D. In subfigures (~~b~~c, ~~d~~,f, i, and ~~h~~l) the colors reflect polygon type (red: center; green: ridge; blue: trough).



(a) Site A

(b) Site B

(c) Site C

(d) Site D

Figure 4. Subsurface soil horizons across ~~microtopographic~~ micro-topographic positions at Sites A, B, C, D.

(1991), Hinzman et al. (1998)) and a recent parameter calibration study conducted at one of our sites (site C) (Atchley et al. (2015)). Table 7 shows the soil hydraulic and thermal properties used in our modeling study.

Table 2. Mesh resolution in vertical column

Vertical depth	Resolution
0 – 0.5 m	5 cm
0.5 m - 1.0 m	10 cm
1.0 m – 5.0 m	25 cm
5.0 m – 10.0 m	50 cm
10.0 m – 50.0 m	1.0 m

3.1.3 Development of ~~microtopography~~ micro-topography resolving computational mesh

Collier and Kumar (2016) developed MeshMaker, a Python-based meshing framework to create high resolution computational meshes for use in numerical simulation of permafrost thermal hydrologic processes at our polygonal tundra study sites. The meshing framework uses a high-resolution DEM and landscape classification (Figure 3) to develop Triangulated Irregular Networks (TINs). Non-uniform locally refined TINs adapt to the topographic complexity to create fine-resolution elements in areas with sharp changes in topography while creating coarser elements elsewhere, thus creating a high quality micro-topography resolving mesh (Figure 6). Variable resolution in vertical column was employed as described in Table 2.

Data from Sections 3.1.1 and 3.1.2 were embedded within the generated meshes (Figure 5) to represent the heterogeneity in the thermal hydrology models. By overlaying the TIN mesh with classified maps from Section 3.1.1, micro-topographic position (Center/Ridge/Trough of a polygon) of each element in the mesh was identified. Polygon type and micro-topographic specific soil horizons data (Section 3.1.2) were used to determine the soil horizons in the model mesh. While our data set was limited to a single replicate for each polygon type and location, significant spatial heterogeneity exists in reality. We assumed a variability of 10% in soil horizon (moss, organic, mineral and deep organic soil) depths and stochastically generated the soil horizon depths at each spatial location in the modeling domain.

3.2 Three phase model for permafrost hydrology

In this study, we will use the open–source code PFLOTRAN to model the flow of mass and energy in the subsurface. PFLOTRAN (Hammond et al. (2016, 2014)) is a state-of-the-art, massively parallel subsurface flow and reactive transport code. PFLOTRAN solves a system of generally nonlinear partial differential equations (PDEs) describing multiphase, multicomponent and multiscale reactive flow and transport in porous materials. The PDEs are spatially discretized using a finite volume technique, and backward Euler scheme is used for implicit time discretization. One system of PDEs which PFLOTRAN implements is a three-phase, thermal-hydrology model (the TH process model in PFLOTRAN parlance) which describes a balance of mass

$$\frac{\partial}{\partial t}\left[\varphi\left(s_\ell\eta_\ell + s_i\eta_i + s_g\eta_g\right)\right] + \nabla\cdot\left(\eta_\ell\mathbf{q}_\ell\right) = Q_M \tag{1}$$

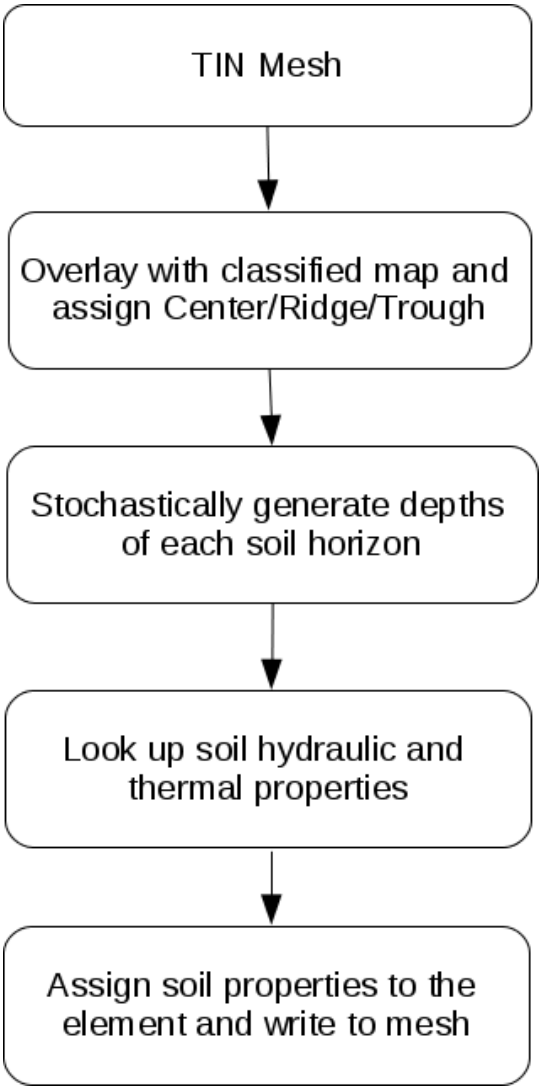
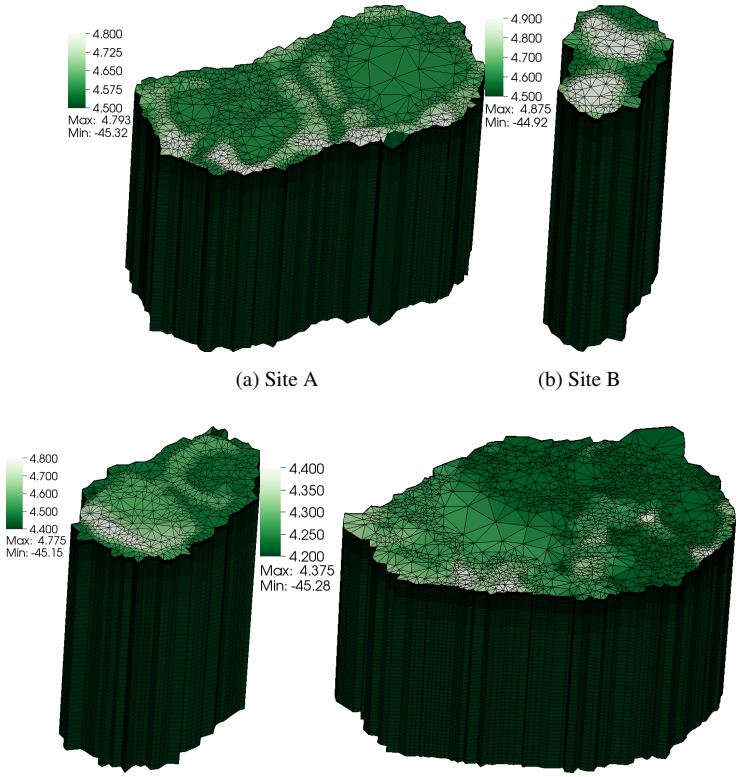


Figure 5. Workflow for heterogeneous soil parameter assignment in the computational mesh.



(a) Site A (b) Site B

(c) Site C (d) Site D

Figure 6. ~~Microtopography~~ Micro-topography resolving unstructured meshes. The coloring reflects the ground surface elevation.

and energy,

$$\frac{\partial}{\partial t}\left[\varphi\left(s_\ell \eta_\ell U_\ell + s_i \eta_i U_i + s_g \eta_g U_g\right) + (1-\varphi)\rho_s c_s T\right]$$
$$+\nabla\cdot\left(\eta_\ell \mathbf{q}_\ell H_\ell - \kappa_{eff}\nabla T\right) = Q_E \tag{2}$$

in which the liquid pressure P_ℓ and the bulk temperature T are the primary unknown variables. In equations (1) and (2), φ refers to porosity, s to percent saturation, η to molar density, U to internal energy, ρ to mass density, c to specific heat, and H to enthalpy. The subscripts $\{\ell, g, i\}$ refer to the liquid, gas, and ice phases of water, respectively, and the subscript s to the soil matrix. The Darcy velocity is given by,

$$\mathbf{q}_\ell = -\frac{k k_r}{\mu_\ell}\nabla\left(P_\ell - \rho_\ell g z\right) \tag{3}$$

where k denotes intrinsic permeability, k_r relative permeability, μ viscosity, g unsigned gravity, and z the vertical component of the position vector $\mathbf{x}$. The effective thermal conductivity is expressed as

$$\kappa_{eff} = K_i\kappa_i + K_\ell\kappa_\ell + (1 - K_i - K_\ell)\kappa_g \tag{4}$$

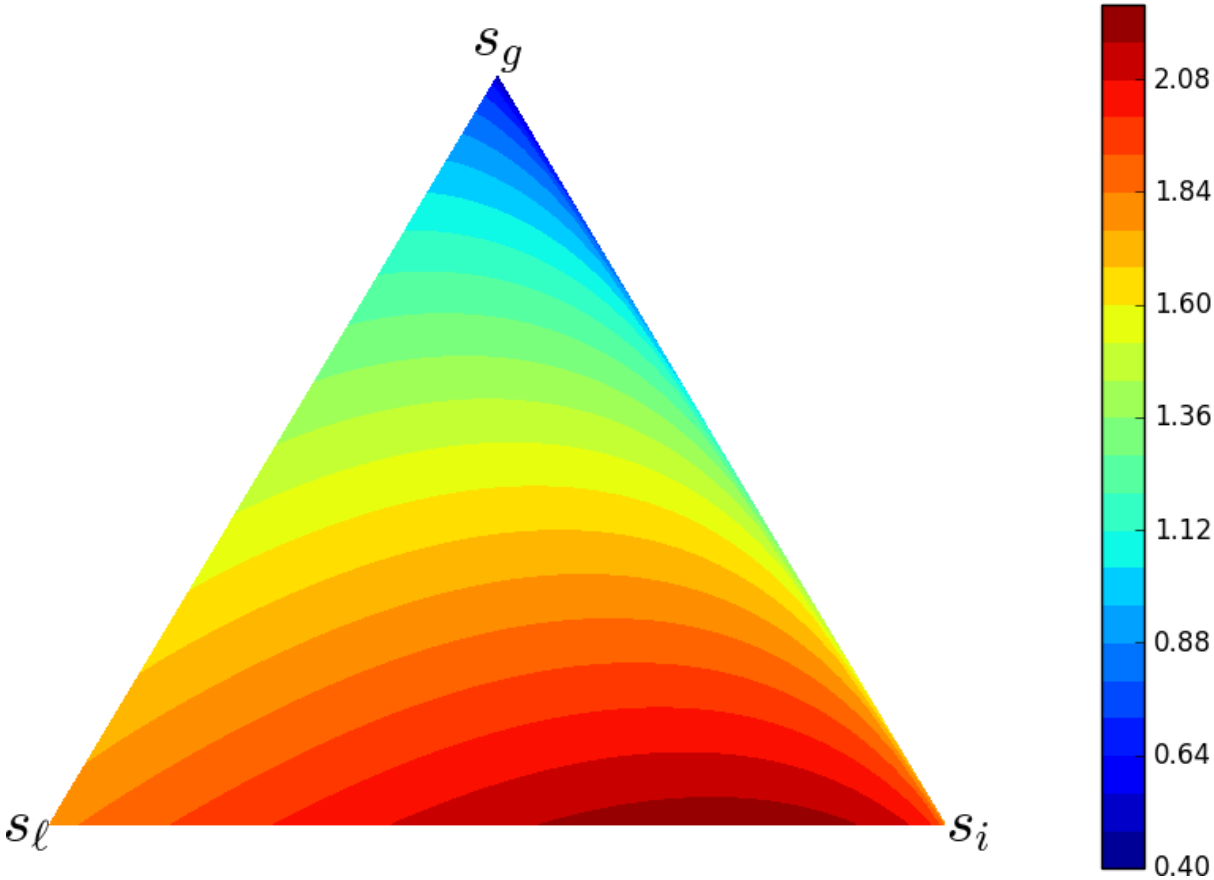


Figure 7. Model dependence of effective thermal conductivity on liquid (s_ℓ), ice (s_i) and gas phase (s_g) fraction of water

where κ are the thermal conductivities of each pure phase and K represents the Kersten number of the frozen and unfrozen phase,

$$K_i = (s_i + \epsilon)^{\alpha_i} \tag{5}$$

$$K_\ell = (s_\ell + \epsilon)^{\alpha_\ell} \tag{6}$$

where $\epsilon = 1 \times 10^{-6}$ and α_i and α_ℓ are parameters of the assumed power law. Figure 7 shows the modeled dependence of effective thermal conductivity (κ_{eff}) on fraction of water present in liquid (s_ℓ)), ice (s_i) and gas (s_g) phases.

The variables Q_M and Q_E represent generic mass and energy sources and sinks. We emphasize that the saturations, densities, and internal energies are all nonlinear functions of the liquid pressure and temperature and include the latent heat of fusion associated with change of phase. We also note that PFLOTRAN implements several choices of constitutive models for relating the saturations to the liquid pressure and bulk temperature. ~~For more details on the formulation, see~~ While only a brief overview of the numerical formulation has been provided here, we would refer to Painter (2011), Painter and Karra (2014), and Karra

et al. (2014) for detailed discussion of the formulation. Key parameters for the model relevant for current study are described in Table 7.

3.3 Initial and boundary conditions

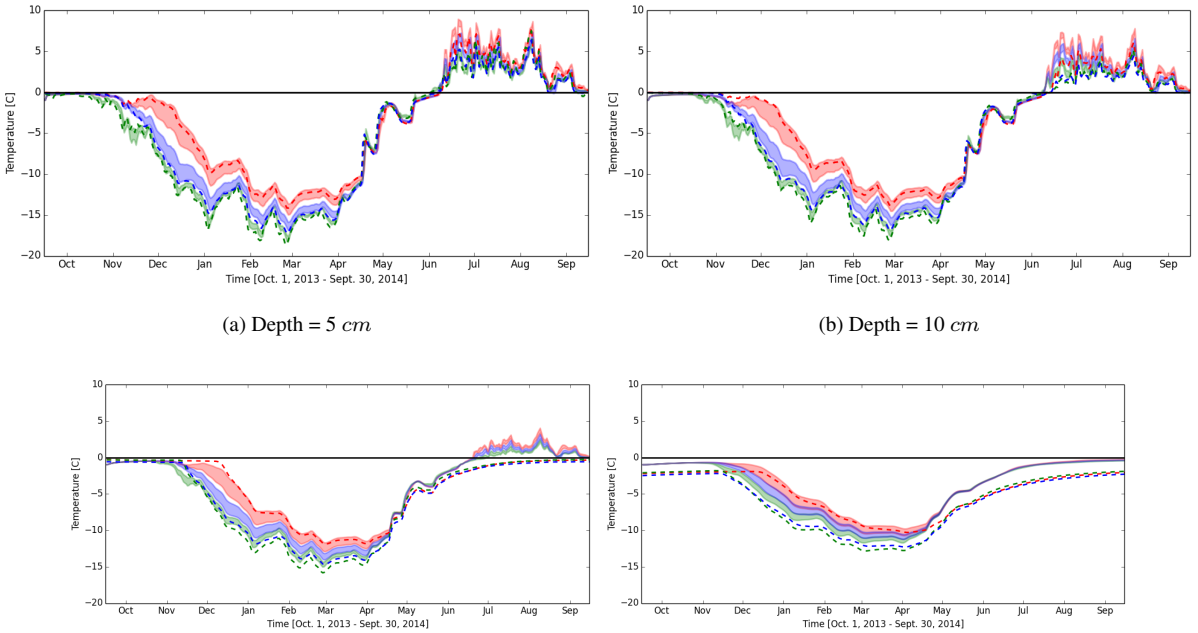
3-D subsurface models for each of the four sites were initialized by freezing the entire modeling domain at a temperature of -1.0 °C. The models were spun up to a thermal periodic steady state using a time series of mean daily temperatures applied to the top of the domain (ground surface). Spin up simulations 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 inter annual 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.

Mean daily near-surface temperature time series for period October 1, 2013 – September 30, 2014 were derived from hourly in situ temperatures from sensors located at 2 cm depths. At all four sites, using sensors installed at center, ridge and troughs, three different time series were prepared. Using the classification (center, ridge, trough) embedded in the model (Section 3.1.1), these micro-topography specific temperature time series were applied in a spatially heterogeneous, micro-topography aware ~~manner~~ fashion to simulate the complex thermal hydrologic regimes in permafrost soils at the BEO. A no–flow boundary condition was applied to the sides of the domain, while the deep bottom boundary was held at constant -10 °C, based on the temperature from West Dock site (Figure 3 of Romanovsky et al. (2010)) which is located at a comparable latitude (70.4°N) to the BEO (71.29°N). It would be important to note that due to lack of observations of drainage pattern at the site we opted to use a no-flow boundary condition, however, surface runoff in and out of the region occurs in reality. Thus the soil moisture states were not very well constrained in the model and also has consequences for the simulated thermal regime. Surface processes (such as vegetation cover and snow) play an important role in regulating the thermal regimes of permafrost soils. While the surface processes are not represented in our model, use of near ground surface (2 cm depth) temperature as the boundary condition for the simulation allows us to isolate (though not completely) the effect of surface processes.

4 Modeling permafrost thermal hydrology

4.1 Simulation of permafrost thermal regimes

After the models were spun up to periodic steady state condition, the simulation was continued for another year and outputs were used for validation and analysis. Soil temperature observations from the thermal sensors at 2 cm depths at the sites for the period October 1, 2013 – September 30, 2014 were used to drive the time dependent (Dirichlet) boundary condition at the top (ground surface) of the model. In addition, mean daily time series (October 1, 2013 – September 30, 2014) of liquid (summer time) precipitation was also applied as moisture input to the model. Groundwater infiltration was considered to be zero if ground surface temperature was below freezing or if the domain was fully saturated. While soil moisture plays an important



(a) Depth = 5 cm

(b) Depth = 10 cm

(c) Depth = 50 cm

(d) Depth = 150 cm

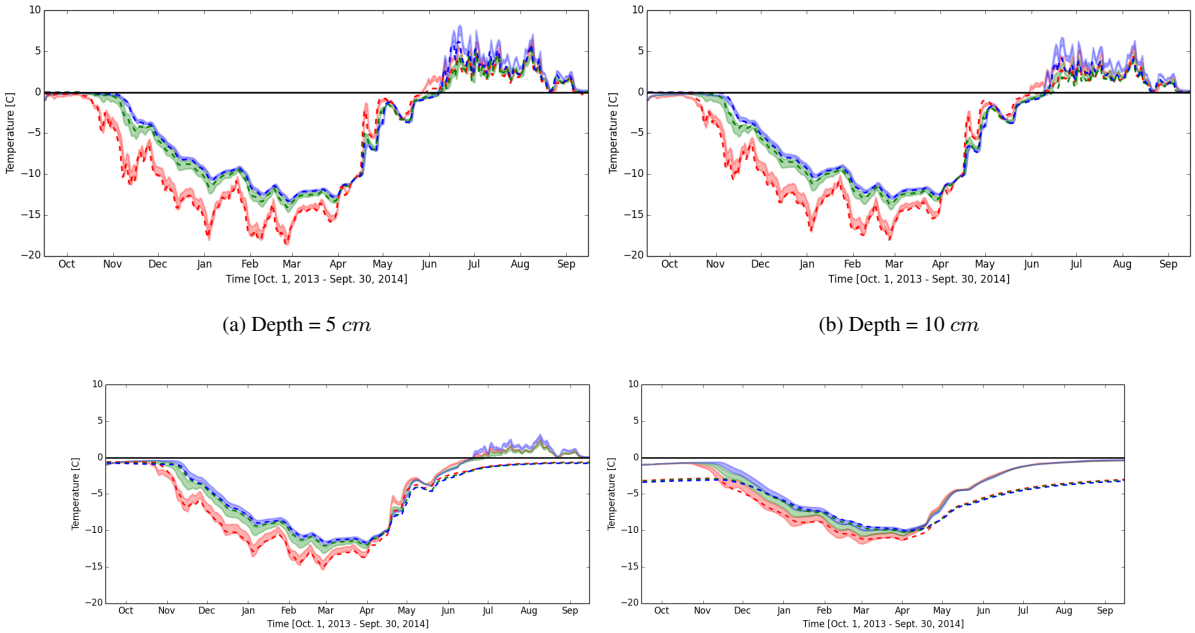
Figure 8. Simulated vs observed soil temperatures at Site A

dotted lines represent observed data at center (in red), rim (green), and trough (blue) locations, while shaded curves show mean +/- standard deviation of simulated daily soil temperatures across the domain.

role, the focus of this study was on thermal ~~hydrology~~ regime and thus all results and discussions presented are focused on soil temperature. Section D3 show the spatial pattern of maximum water table depth across study region and exhibits strong correlation with micro-topography. Soil temperature data from all the sensors (~~nine~~ sixteen sensors at depths from 5 cm to 150 cm) were used to evaluate the accuracy of the models. While we have selected sensors at four depths 5 cm, 10 cm, 50 cm and 150 cm for discussion the results at all depths can be found in Section D1.

At all the sites, simulated soil temperatures in the top most soil layer (5 cm thick) compared well with the observed near surface temperature at 5 cm depth (Figures 8(a), 9(a), 10(a), 11(a)). Simulated soil temperatures in deeper soils were also in good agreement with the observed temperatures at the sensors (coefficient of determination R^2 0.93 – 0.99) (Tables ~~8,9,10,11~~3, 4, 5, 6). Model matched the observations with a Root Mean Square Error ($RMSE$) of 0.60 – 0.99 $^\circ C$ near surface, with an increasing errors at deeper soils. While modeled temperature bias was in range of -0.30 – 0.10 $^\circ C$ near surface, a warm bias of up to 1.0 – 1.8 $^\circ C$ was found at deep soils.

Figures 8,~~9,10,~~ 9, 10, 11 shows the comparison between the simulated and observed soil temperatures at several select depths (5 cm, 10 cm, ~~20 cm,~~ 50 cm, ~~100 cm and~~ 150 cm from surface) at sites A, B, C, and D respectively. Simulated temperatures across all the sites matched ~~very~~ well with the observed temperatures at shallow depths, but showed a deviation towards warmer than observed temperatures in deep soils. A number of factor may be contributing to this bias in simulations, including



(a) Depth = 5 cm

(b) Depth = 10 cm

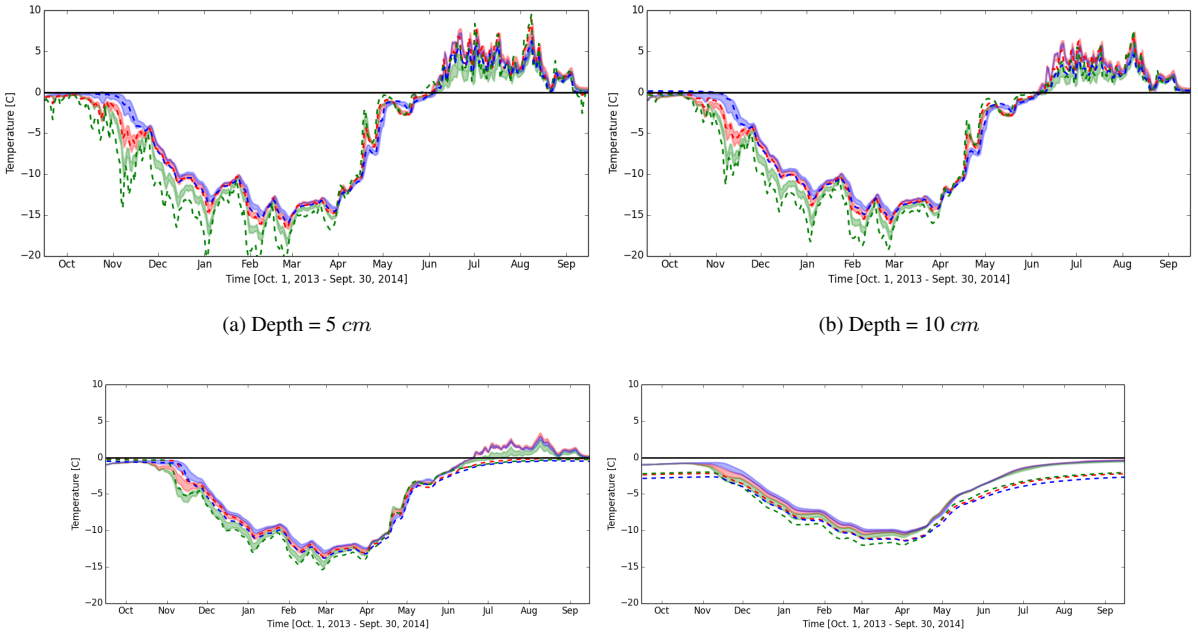
(c) Depth = 50 cm

(d) Depth = 150 cm

Figure 9. Simulated vs observed soil temperatures at Site B

dotted lines represent observed data at center (in red), rim (green), and trough (blue) locations, while shaded curves show mean +/- standard deviation of simulated daily soil temperatures across the domain.

applied boundary conditions (Section 3.3). We believe that insufficient characterization and parameterization of heterogeneous properties due to limited data availability is one of the key reasons for this bias. Model was evolved to a periodic thermal steady state through a spin up process and soil temperatures in vertical profile are strongly dependent on the soil properties. Forced with surface temperature boundary condition, while the soil temperature in the model after spin up stage was close to the observed, a warm bias was observed in the deeper soils. That warm bias was carried over to the final stage of the simulation resulting in bias in the simulated soil temperatures and thaw depths reported here. With soil cores collected at the sites limited to top 30 cm of the soil, our understanding of structure and physical and thermal properties of deeper soils is limited. For example, while we know that presence of ground ice (like ice wedges, segregated ice, ice lens etc.) is common in 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 et al. (2016) using X-ray computed tomography (CT) scanner technology on ice cores from BEO can potentially provide detail 3-dimensional soil structure and density information and help address this missing piece.



(a) Depth = 5 cm

(b) Depth = 10 cm

(c) Depth = 50 cm

(d) Depth = 150 cm

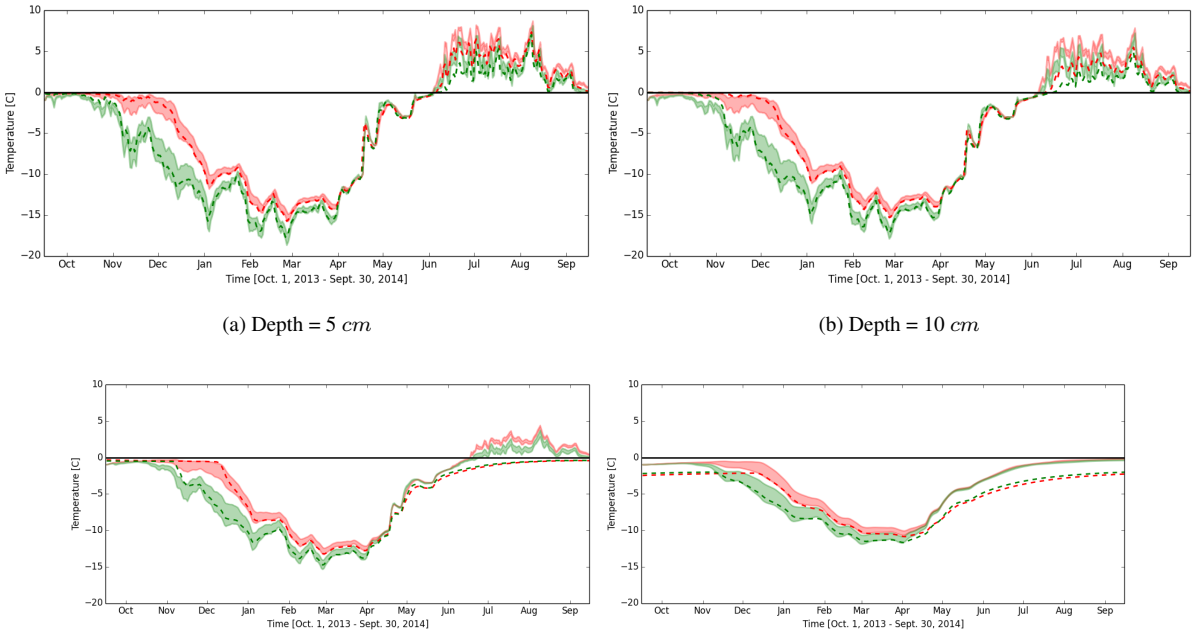
Figure 10. Simulated vs observed soil temperatures at Site C

dotted lines represent observed data at center (in red), rim (green), and trough (blue) locations, while shaded curves show mean +/- standard deviation of simulated daily soil temperatures across the domain.

Spatial variability in soil temperatures was observed in the simulations (Figures 8,~~9,10,~~ 9, 10, 11) arising in part due to three dimensional heat flow and heterogeneous subsurface structure and soil properties represented in the model. Simulated soil temperatures also show a seasonal pattern of spatial variability with high variability during the cold winter season and lower spatial variability during summer. Figure 12 shows time series of spatial variability (standard deviation) in soil temperature at Site A during the simulation period, showing strong seasonality, the magnitude of which is reduced at deeper soils. Similar patterns of variability was observed at Site B, C, and D (Figures 18, 19, 20).

Heat flow in the permafrost soils which are frozen for a significant part of the year occurs primarily due to conduction. Thermal conductivity of the soil is sensitive to the temperatures and thus the fraction of water present in liquid vs ice phase (Equation (4), Figure 7). Effective thermal conductivity (κ_{eff}) of the soil is higher during the winter months when almost the entire soil domain is in a frozen state (thus high ice saturation (s_i) and low liquid saturation (s_ℓ)), compared to summer months when active layer is in thawed state. ~~Figure 7 shows the modeled dependence of effective thermal conductivity (κ_{eff}) on fraction of water present in liquid (s_ℓ)), ice (s_i) and gas (s_g) phases.~~ Higher conductivity of the soil and thus higher conductive heat flows during the winter season in a heterogeneous soil domain leads to higher temperature variability in model simulations of the permafrost thermal regimes at our sites.

~~Model dependence of effective thermal conductivity on liquid (s_ℓ), ice (s_i) and gas phase (s_g) fraction of water~~



(a) Depth = 5 cm

(b) Depth = 10 cm

(c) Depth = 50 cm

(d) Depth = 150 cm

Figure 11. Simulated vs observed soil temperatures at Site D

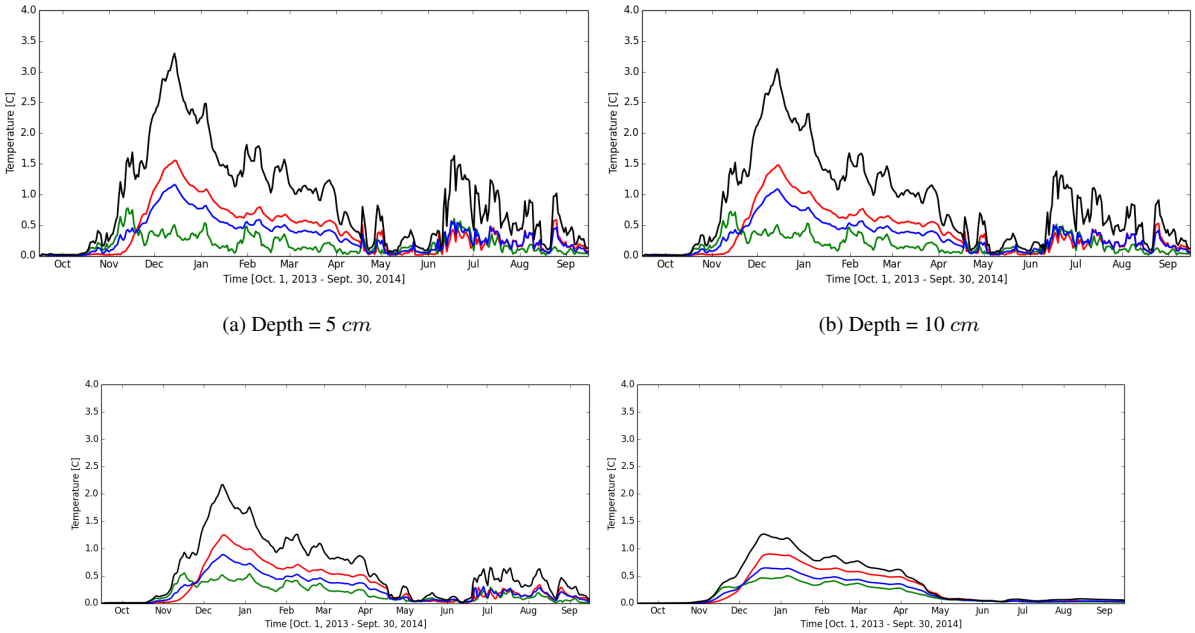
dotted lines represent observed data at center (in red), rim (green), and trough (blue) locations, while shaded curves show mean +/- standard deviation of simulated daily soil temperatures across the domain.

Table 3. Model performance statistics at Site A compared to observed soil temperatures

Depth from surface [m]	Center			Rim			Trough		
	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$
0.05	0.73	0.99	-0.06	0.80	0.99	0.13	0.91	0.99	0.48
0.10	0.85	0.98	0.03	0.72	0.99	0.31	1.13	0.98	0.67
0.50	1.30	0.94	0.39	1.01	0.99	0.72	1.35	0.98	1.16
1.50	1.42	0.94	1.01	1.55	0.99	1.52	1.95	0.99	1.92

$RMSE$ = Root Mean Squared Error, R^2 = Coefficient of determination, $Bias$ = $negative$ bias indicates cold bias in the model while $positive$ indicates a warm bias

During the summer season, advective heat flow processes occur within the thawed soil layers. The study sites also receive liquid precipitation and thus infiltration during the summer season (Figure 2) which leads to vertical as well as horizontal flows in the thawed soil. For example, at Site A horizontal velocities are close to zero during the early summer (Figure 13(a)) when soil temperatures are close to freezing. After the ground has thawed, liquid precipitation events during summer (Figure 13(b, c)) lead to significant lateral flows. High elevation rim regions drain to the center and trough of the low-centered polygons.



(a) Depth = 5 cm

(b) Depth = 10 cm

(c) Depth = 50 cm

(d) Depth = 150 cm

Figure 12. Spatio temporal variability (standard deviation) in simulated soil temperature at Site A.
Red, green and blue lines represent center, rim and trough respectively, and black line represent the standard deviation across the Site A.

Table 4. Model performance statistics at Site B compared to observed soil temperatures

Depth from surface $[m]$	Center			Rim			Trough		
	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$
0.05	0.83	0.99	0.34	0.50	0.99	-0.08	0.68	0.99	0.23
0.10	0.80	0.99	0.38	0.46	0.99	0.04	0.72	0.99	0.31
0.50	1.02	0.98	0.71	0.98	0.98	0.39	1.23	0.98	0.81
1.50	1.89	0.96	1.49	1.95	0.92	1.34	2.09	0.94	1.63

$RMSE$ = Root Mean Squared Error, R^2 = Coefficient of determination, $Bias$ = $negative$ bias indicates cold bias in the model while $positive$ indicates a warm bias

As ground thaws with the rise of summer temperature, significant lateral flow occurs even in the absence of rainfall events (Figure 13(d)). No flow boundary condition applied in our model prevents flow out of the modeling domain, leading to increased soil moisture (Figure 21) in low elevation areas where the higher warm bias in the model is observed (Figure 14, 15, 16, 17).

In permafrost environment, the active layer is the top layer of soil that thaws during the summer and freezes again during the autumn. While PFLOTRAN solves for soil temperature as primary variable, we derived active layer depth (or thaw depth) as sum of thickness of soil layers above freezing temperature (0°C). Figure 14, 15, 16, 17 show the temporal dynamics of

Table 5. Model performance statistics at Site C compared to observed soil temperatures

Depth from surface $[m]$	Center			Rim			Trough		
	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$
0.05	0.80	0.99	0.18	1.84	0.96	0.50	0.69	0.99	0.29
0.10	0.63	0.99	0.12	1.23	0.98	0.35	0.78	0.99	0.34
0.50	0.96	0.97	0.42	0.75	0.99	0.42	1.13	0.98	0.91
1.50	1.35	0.99	1.27	1.23	0.99	1.19	1.81	1.00	1.76

$RMSE$ = Root Mean Squared Error, R^2 = Coefficient of determination, $Bias$ = $negative$ bias indicates cold bias in the model while $positive$ indicates a warm bias

Table 6. Model performance statistics at Site D compared to observed soil temperatures

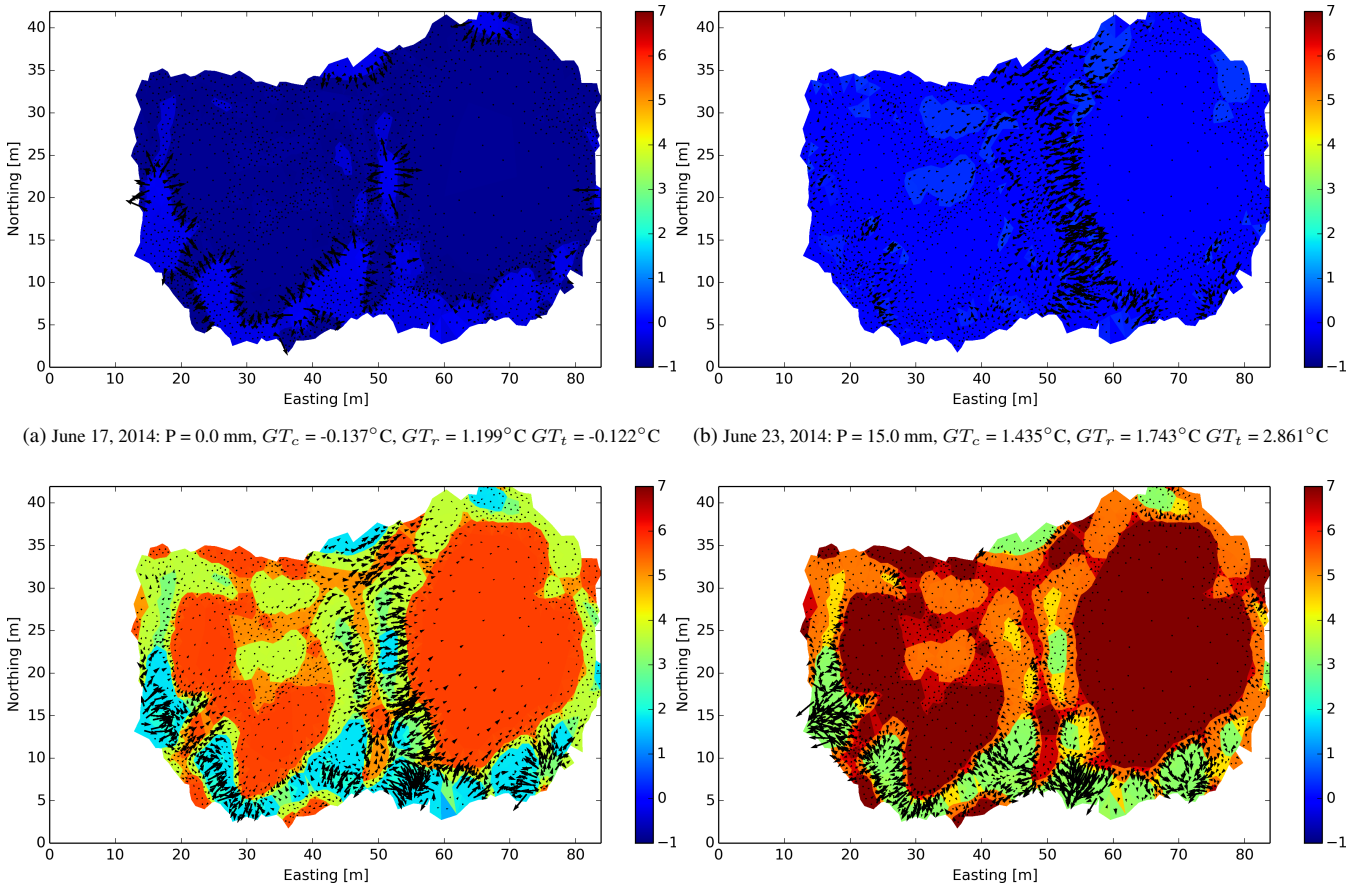
Depth from surface $[m]$	Center			Rim		
	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$
0.05	0.80	0.99	0.34	0.86	0.99	0.26
0.10	1.03	0.98	0.45	0.96	0.98	0.36
0.50	1.49	0.95	0.86	1.02	0.98	0.57
1.50	1.65	0.97	1.47	1.29	0.98	1.14

$RMSE$ = Root Mean Squared Error, R^2 = Coefficient of determination, $Bias$ = $negative$ bias indicates cold bias in the model while $positive$ indicates a warm bias

thaw depth during the period of simulation and spatial pattern of maximum thaw depth across the region. A wide spatial variability in the thaw depth was observed in the simulations which were strongly correlated to the micro-topography. The variability is primarily derived by the micro-topography and the subsurface heterogeneity. The warm bias in soil temperature in the model translates to a bias towards deeper thaw depths as compared to the observations at the site (Section E, 22). While the average active layer thickness at site was 50 cm, depths up to 74 cm were observed by Hubbard et al. (2013) and Peterson (2016) indicating the variability due the heterogeneity at the site. Active layer thickness of 25 to ~100 cm were reported at Kuparuk watershed where thermal properties derived from Hinzman et al. (1998) and used in this study were collected.

4.2 Understanding the thermal regimes of polygonal tundra

Micro-topography of the polygonal tundra exerts critical controls on the flow of water and energy at local to regional scales which further influences the ecological and biogeochemical processes on the landscape. Surface processes (not ~~explicitly~~ modeled in our study) like vegetation and snow cover also play a critical role in regulating the subsurface thermal regimes through thermal insulation effects. In our modeling approach we represented the micro-topographic features Center, Rim and



(a) June 17, 2014: P = 0.0 mm, GT_c = -0.137°C, GT_r = 1.199°C GT_t = -0.122°C

(b) June 23, 2014: P = 15.0 mm, GT_c = 1.435°C, GT_r = 1.743°C GT_t = 2.861°C

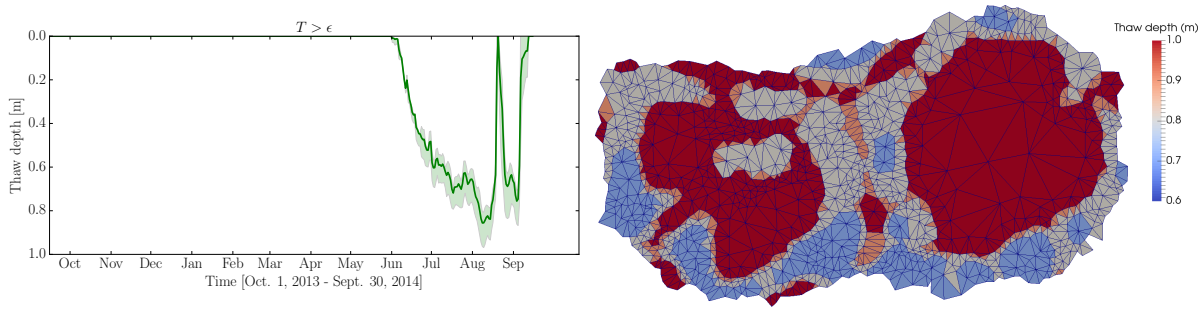
(c) June 30, 2014: P = 7.4 mm, GT_c = 9.780°C, GT_r = 4.134°C GT_t = 6.927°C

(d) July 06, 2014: P = 0.1 mm, GT_c = 10.513°C, GT_r = 4.435°C GT_t = 8.289°C

Figure 13. Lateral water flow velocity fields at Site A during early summer. In the horizontal cross section plot for Site A, background color shows the soil temperature distribution while the vector arrows show the magnitude and direction of lateral flow fields. (P = total daily precipitation, GT_c, GT_r, GT_t = mean ground surface temperature at center, rim and trough respectively.

Trough across four low to transitional to high-centered polygons. PFLOTRAN successfully simulated the pattern of thermal regimes in Center, Rim and Trough across four sites A, B, C and D (Table 1).

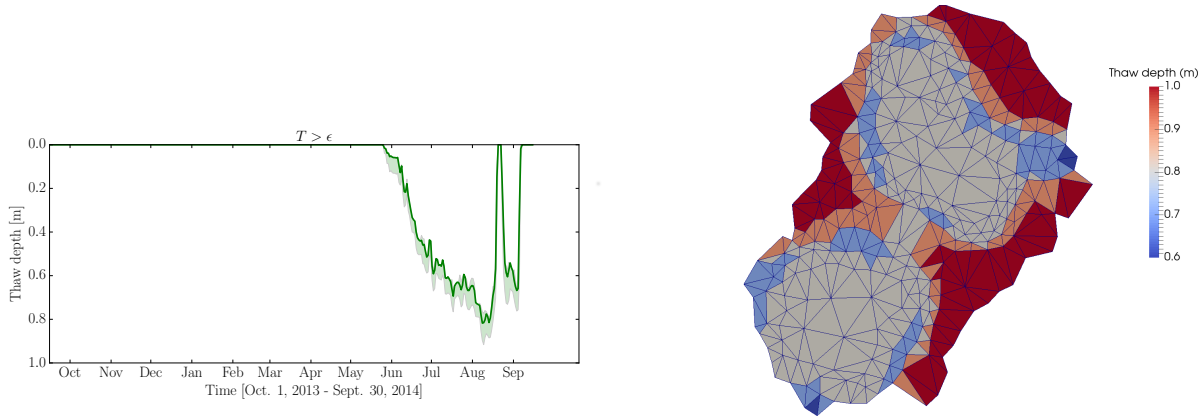
- *Site A*: Site A is located in a poorly drained region dominated by low-centered polygons with low elevation centers, raised rims and troughs. Center areas are warmer than rim and trough areas while rims are coldest (Figure 8). Centers in low-centered polygon are often ~~innundated~~ inundated and relatively wet (Figure 21) most of the year and support vegetation (mosses and sedges). Low elevation centers also receive higher snow cover. Vegetation and snow cover provide thermal insulation to the ground keeping the center region warmer compared to rim and trough. Dry rims (Figure 21) with low vegetation cover and low snow accumulation (Figure 23) are most exposed to the winter temperatures and are thus the coldest.



(a) Temporal pattern of thaw depth

(b) Spatial pattern of maximum thaw depth

Figure 14. Temporal (left) patterns of simulated thaw depths (defined as thickness of soil layers above 0°C) and spatial pattern of maximum thaw depth (right) at Site A. Bold line represents the mean thaw depth across the site, while shaded curve represents standard deviation



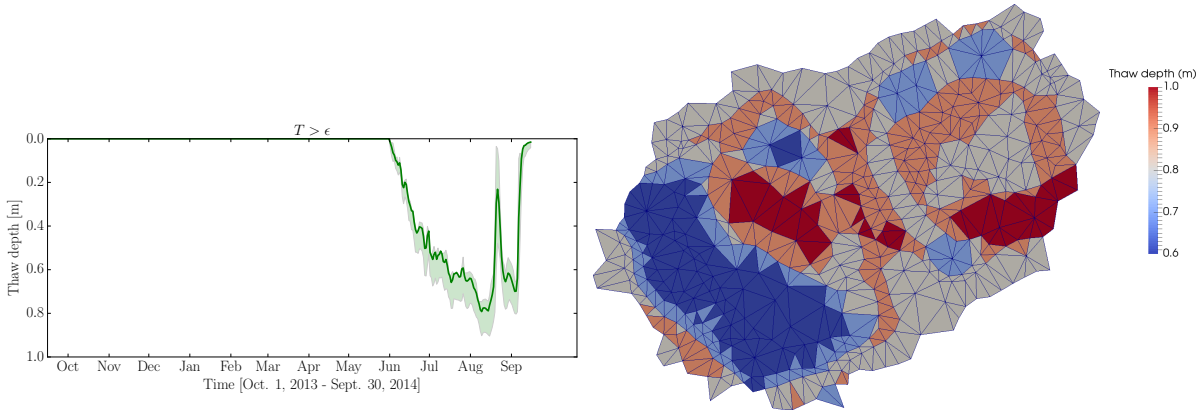
(a) Temporal pattern of thaw depth

(b) Spatial pattern of maximum thaw depth

Figure 15. Temporal (left) patterns of simulated thaw depths (defined as thickness of soil layers above 0°C) and spatial pattern of maximum thaw depth (right) at Site B. Bold line represents the mean thaw depth across the site, while shaded curve represents standard deviation

- *Site B*: Site B is dominated by well drained high-centered polygons with relatively dry elevated centers and deep troughs (Figure 21) . Vegetation in high-centered polygons are dominated by lichens, moss and dwarf shrubs. In contrast to the low-centered polygons, centers and dry-tundra graminoids have low vegetation and snow cover (Figure 23), are most exposed to the changes in air temperatures, and are thus colder than rim and trough which show relatively warmer soil temperature regimes during the winter. (Figure 9)

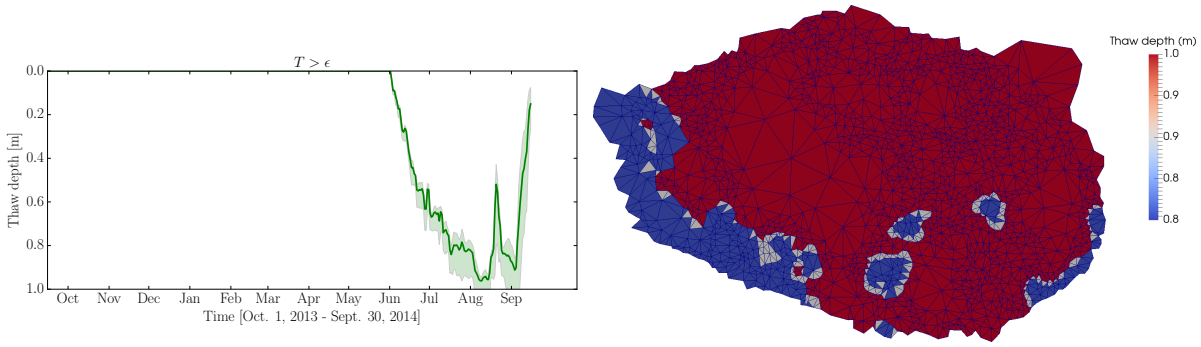
- *Site C*: Site C is located in an area of geomorphological transition from low to high-centered polygons, characterized as flat-centered polygons. They consist of shallow flat centers, deep troughs and raised rim regions. Soils in deep troughs are thermally insulated by higher snow cover (Figure 23) and thus show warmer soil temperature regimes compared to



(a) Temporal pattern of thaw depth

(b) Spatial pattern of maximum thaw depth

Figure 16. Temporal (left) patterns of simulated thaw depths (defined as thickness of soil layers above 0°C) and spatial pattern of maximum thaw depth (right) at Site C. Bold line represents the mean thaw depth across the site, while shaded curve represents standard deviation



(a) Temporal pattern of thaw depth

(b) Spatial pattern of maximum thaw depth

Figure 17. Temporal (left) patterns of simulated thaw depths (defined as thickness of soil layers above 0°C) and spatial pattern of maximum thaw depth (right) at Site D. Bold line represents the mean thaw depth across the site, while shaded curve represents standard deviation

centers and rims (Figure 10). Center and rim regions show similar thermal regimes with centers being slightly warmer due to higher vegetation and snow cover.

– *Site D*: Site D is characterized as low centered polygons with no pronounced rims. Site D is wettest among the four study areas, with low elevation center areas that remain ~~innundated~~ inundated for most of the summer season (Figure 21). While snow accumulation was fairly uniform (Figure 23) across the flat region, vegetation cover plays an important role. Wet centers supports rich vegetation, leading to a warmer soil temperatures as compared to the trough regions (Figure 11).

~~Observed snow depths at different topographic positions center (in red), rim (in green), trough (in blue) at our study sites (Hinzman et al. (2014a) .~~

5 Model uncertainties and limitations

5.1 Why not to calibrate

Accurate simulation of permafrost thermal regimes requires the mechanistic representation of thermal hydrologic processes in the model. However, equally important is the accurate representation of subsurface structure and soil properties, model parameters and initial and boundary conditions. Given the lack of co-located observations for the soil properties, in this study we used soil thermal and hydraulic properties data from a different tundra site based on Hinzman et al. (1998) . While PFLOTRAN was able to simulate the thermal hydrologic processes and match fairly well the soil temperature observations at the sites across a range of polygonal landscape and micro-topography features, simulated soil temperatures show deviations from the observed temperatures at times. Simulated temperatures show warm bias in deep soils where data for soil characterization and properties are almost completely missing. Parameter calibration is a popular technique that has been widely used in hydrologic modeling to determine model parameters and properties to optimize the model fit to target observations. The high resolution 3-D PFLOTRAN thermal hydrology model used in this study includes many degrees of freedom and parameters, which combined with complex non-linearity of hydrologic processes poses a complex high dimensional optimization problem. While a wide range of calibration approaches (Heuvelmans et al. (2006), Madsen (2000), Shafii and De Smedt (2009), Singh and Minsker (2008)) are available to determine optimal model parameters to fit the observed data (soil temperatures in this study), we face the problem of non-uniqueness (equifinality). A diverse set of possible parameter values can lead to similar model performance (Beven and Freer (2001)). The issue of non-uniqueness is especially pronounced in tundra ecosystem due to poor availability of data and thus poor bounds on parameters which leads to a high degree of uncertainty in the models with or without calibration.

While systematic calibration can help identify effective parameters for the model, transfer of parameters across models and modeling domains is difficult ~~(Bárdossy (2007)). At our study sites , while~~ (Bárdossy, 2007) . At data limited study sites like ours, while a ill-constrained calibration may compensate for lack of data, it does not improve our understanding of the system. In this study we choose not to calibrate the model parameters to achieve better fit with the observations, instead ~~using~~ use the

uncalibrated results to diagnose the potential model deficiencies and identify the characterization and data needed to better represent the real world in our simulations.

When properly constrained using observed data, the model calibration is a powerful tool that can provide improved physical representation of processes and parameters and new insights. Atchley et al. (2015) and Harp et al. (2016) has successfully demonstrated the use of such techniques for three phase hydrology models at one of our sites (site C) and would inform our future studies.

5.2 Identifying model and data gaps

While the agreement between modeled and observed simulated soil temperature demonstrates the ability of the model to simulated the thermal hydrologic processes in the polygonal tundra, disagreements help us identify the existing gaps in data and model.

Modeling results highlight the need for co-located measurements of soil thermal and hydraulic properties for accurate modeling of hydrologic processes. While most soil core observations, including ~~these~~ those used in our study, are focused on the shallow active layer, characterization of deeper permafrost soils is essential for understanding the thermal regimes and potential changes expected under warming climate. ~~Even during the winter season when soils are frozen, thermal dynamics~~ Modeling soil temperatures, beyond the high level estimation of thaw depth (or active layer thickness) 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 (Figure 8, 9, 10, 11) exist and may impact carbon fluxes from the system (Zona et al. (2016)). Warmer than observed soil temperature in deep soils in our thermal periodic steady state solutions are due to inaccurate soil characterization and poorly bounded boundary conditions at the bottom of the modeling domain. Heat flux observations in deep permafrost, while hard to measure, would help provide accurate bounds for the thermal hydrology model. In addition to rainfall events (which was captured in our models), surface drainage processes provides inputs to the ground water system (not captured in our models). Surface drainage observations in the local catchments (not available to us) are needed to appropriately model and constrain this process~~.~~

~~While primary focus of the current study was to model and understand the subsurface thermal hydrology of polygonal tundra ecosystem, surface energy balance and snow processes also plays an important role. While beyond the current scope, efforts are underway within the NGEE–Arctic project to couple PFLOTRAN with a terrestrial biosphere model, the Community Land Model (CLM). The coupled CLM–PFLOTRAN model will provide an integrated treatment of surface and subsurface thermal hydrologic processes, and will be reported in future studies~~.

6 Summary and conclusions

Low-relief polygonal tundra ecosystems consist of micro-topographic features that controls the local scale hydrology. The water and energy flow patterns on the landscape in turn regulate biogeochemical processes and vegetation dynamics. The objective of this study was to develop an end-to-end modeling approach for landscape scale modeling of permafrost thermal hy-

drology of real world polygonal tundra sites, improving our ability to model and to understand the patterns of thermal regimes. Using the best available data for our study sites we developed techniques to characterize the polygonal ~~microtopographic~~ micro-topographic features and represent the heterogeneous soil hydraulic and thermal properties at the sites. These data sets were embedded within topography–following high–resolution meshes to simulate the thermal hydrologic processes in PFLOTRAN thermal hydrology model. We employed detailed surface meteorology and subsurface soil temperature observations from the site to simulate and analyze the thermal regimes at four representative sites in polygonal tundra ecosystem at the Barrow Environmental Observatory.

Our modeling–based study reveals the role of ~~microtopographic~~ micro-topographic features in regulating the permafrost thermal dynamics across heterogeneous polygonal tundra landscape. Simulation results at four sites across the polygonal tundra landscape demonstrate the effectiveness of the developed approach to model the thermal regimes. Thermal regimes of center, rim and trough features of polygonal tundra exhibit distinct patterns in low to transitional to high-centered polygon landscape, which are governed by the ~~microtopography~~micro-topography, surface and subsurface hydrology and surface processes (like air temperature, snow cover, and vegetation). Our PFLOTRAN–based modeling approach was able to successfully simulate these patterns at four study sites. In a warming world, wet low-centered polygon landscapes are expected to go through geomorphological change to drier high-centered polygonn landscapes. Modeling approach developed in this study would enable future investigations in impact of thermal hydrologic changes in these landscapes under projected climate scenarios. Model developed here does not have the ability to simulate the dynamic changes in micro-topography expected due to ice-wedge degradation (Liljedahl et al. (2016)). While beyond the scope of the current study, ongoing developments in biogeochemical modeling within PFLOTRAN (Tang et al. (2015)) in combination with our thermal hydrology model developments will also allow modeling of the terrestrial carbon cycle in this sensitive landscape under future warming scenarios.

We compared and validated the simulated soil temperatures against observations. While the ~~models~~ model 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. Model disagreements with the observations in this study may partially be due to use of soil properties from literature in absence of site–based measurements. Under the NGEE–Arctic project we are working with field scientists for improved co-located measurements. Present study was limited to single year when we had all the necessary data for model forcings and validation available, thus was not able to investigate and address the role of interannual variability. We plan to address this important problem as more data from our sites become available. 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.

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.

Acknowledgments

The Next Generation Ecosystem Experiments (NGEE) – Arctic project (http://ngee-arctic.ornl.gov/) is supported by the Office of Biological and Environmental Research in Department of Energy Office of Science. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Author contributions

JK designed the study and carried out the simulations. NC, GB, JK and RM developed the thermal hydrology model in PFLOTRAN. NC and JK developed the meshing, characterization and parameterization techniques. CI collected and analyzed the soil cores and assisted with vegetation characterization. VR designed and maintained meteorological station and thermal sensors at the sites. JK led the analysis and writing of manuscript with contributions from NC, GB, RM, PT, CI and VR.

Appendix: ~~Appendix I: Model Reproducibility~~

Appendix A: Model Reproducibility

Reproducibility, rigour, transparency and independent verification are cornerstones of the scientific method (Nature (2014)). We document here model/software and computational platform used in the work reported in this study.

Software:

Subsurface thermal hydrology modeling was conduced using PFLOTRAN, which is an open source, state-of-the-art massively parallel subsurface flow and reactive transport code. PFLOTRAN (http://www.pflotran.org/) code is developed under GNU LGPL license and publicly available at https://bitbucket.org/pflotran.

```
PFLOTRAN version used:
Repository URL: https://bitbucket.org/pflotran/pflotran-dev
Changeset: 18ec488fc6ac
```

PFLOTRAN employs parallelization through domain decomposition using the MPI-based PETSc framework with *pflotran-dev* tracking the *git maint* branch of PETSc available through https://bitbucket.org/petsc/petsc.

```
PETSc version used:
Repository URL: https://bitbucket.org/petsc
Changeset: c41c7662de68b036bda6be236f939e8b55959cb0
Version: v3.5.2-137-gc41c766
```

Computational platform: All simulations were conducted on the *Titan Cray XK7* at Oak Ridge Leadership Computing Facility (https://www.olcf.ornl.gov/computing-resources/titan-cray-xk7/). GNU compilers were used to compile PFLOTRAN and PETSc.

Appendix: ~~Appendix II: Data archiving and distribution~~

Appendix B: Data archiving and distribution

Model input files for all simulations reported along with the forcing data files, and computational mesh are publicly available through NGEE–Arctic long term data archive http://dx.doi.org/10.5440/1184018 (Kumar et al. (2016)). Summary outputs and statistics presented in various figures in this article are also available as part of the data collection. While the long term archiving of complete PFLOTRAN simulation outputs reported here was not possible due to large data volume, they can be obtained by contacting the lead author.

Table 7. Soil hydraulic and thermal properties used in the models

Parameter [Unit]	Moss	Organic	Mineral	Deep organic	Deep mineral	Data source
Porosity [–]	0.90	0.86	0.60	0.86	0.54	Hinzman et al. (1991)
Hydraulic conductivity [m/s]	1.94	1.04	0.376	1.08	0.14	Hinzman et al. (1991)
Bulk density [g/cm^3]	0.15	0.18	1.39	0.18	1.33	Hinzman et al. (1991)
VG Alpha (α) [1/Pa]	1.5 $\times 10^{-4}$	1.5 $\times 10^{-4}$	1.5 $\times 10^{-4}$	1.5 $\times 10^{-4}$	1.5 $\times 10^{-4}$	–
VG Lambda (λ) [–]	0.23	0.95	0.33	0.95	0.33	–
Residual saturation [–]	0.05	0.34	0.20	0.34	0.20	–
κ_l [–]	0.45	0.43	0.8	0.43	0.8	Hinzman et al. (1998)
κ_i [–]	1.81	1.73	3.2	1.73	3.2	Hinzman et al. (1998)
κ_g [–]	1.81	1.73	3.2	1.73	3.2	Hinzman et al. (1998)
α_ℓ [–]	0.45	0.45	0.45	0.45	0.45	–
α_i [–]	0.97	0.07	0.97	0.97	0.97	–
Specific heat [$J/kg^\circ C$]	1.04 $\times 10^4$	8.65 $\times 10^4$	2.36 $\times 10^3$	3.19 $\times 10^2$	2.46 $\times 10^3$	–

Appendix: ~~Appendix III: Soil hydraulic and thermal properties~~

Appendix C: Soil hydraulic and thermal properties

Table 7 shows the soil hydraulic and thermal properties used in our modeling study.

Appendix D: Model validation

D1 Simulated soil temperature

Tables 8, 9, 10 and 11 presents the simulated soil temperature validation statistics at sixteen sensors at sites A, B, C and D respectively.

D2 simulated variability in soil temperature

D3 Water table depth

Figure 21 show the spatial distribution of maximum water table elevation during the simulation period at sites A, B, C and D. Ground surface is at 50 m elevation at all the sites.

Appendix E: Active layer thickness observation at BEO

Table 8. Model performance statistics at Site A compared to observed soil temperatures

Depth from surface $[m]$	Center			Rim			Trough		
	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$
0.02	0.97	0.98	-0.32	0.70	0.99	-0.09	0.81	0.99	0.10
0.05	0.73	0.99	-0.06	0.80	0.99	0.13	0.91	0.99	0.48
0.10	0.85	0.98	0.03	0.72	0.99	0.31	1.13	0.98	0.67
0.15	1.35	0.96	0.23	0.87	0.99	0.39	1.32	0.97	0.76
0.20	1.52	0.94	0.28	0.90	0.99	0.43	1.47	0.97	0.89
0.25	1.69	0.93	0.37	0.98	0.98	0.50	1.55	0.96	0.98
0.30	1.39	0.94	0.32	1.00	0.98	0.58	1.36	0.97	1.01
0.35	1.34	0.94	0.31	1.02	0.98	0.64	1.35	0.98	1.07
0.40	1.41	0.93	0.34	1.03	0.98	0.65	1.38	0.98	1.11
0.50	1.30	0.94	0.39	1.01	0.99	0.72	1.35	0.98	1.16
0.60	1.27	0.94	0.44	1.01	0.99	0.78	1.38	0.98	1.23
0.70	1.18	0.95	0.47	1.03	0.99	0.86	1.38	0.99	1.27
0.80	0.93	0.97	0.47	1.14	0.99	0.98	1.48	0.99	1.38
1.00	1.08	0.96	0.59	1.17	0.99	1.08	1.57	0.99	1.52
1.25	1.28	0.94	0.79	1.36	0.99	1.30	1.76	0.99	1.72
1.50	1.42	0.94	1.01	1.55	0.99	1.52	1.95	0.99	1.92

$RMSE$ = Root Mean Squared Error, R^2 = Coefficient of determination, $Bias$ = $negative$ bias indicates cold bias in the model while $positive$ indicates a warm bias

Hubbard et al. (2013) and Peterson (2016) collected repeat observations of active layer thickness at a 500 m transect close to (and representative of) our sites A, B, C, D. Peterson (2016) provides the repeat observations collected on July 02, 2014, August 16, 2014, and September 23, 2014. Figure 22 show the maximum active layer thickness observed at each of the point along the transect. While the average active layer thickness was around 50 cm, a significant amount of variability was observed with active layer depths up to 74 cm being observed.

Appendix F: Observed distribution of snow

Table 9. Model performance statistics at Site B compared to observed soil temperatures

Depth from surface $[m]$	Center			Rim			Trough		
	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$
0.02	0.99	0.98	0.06	0.54	0.99	-0.05	0.83	0.98	-0.07
0.05	0.83	0.99	0.34	0.50	0.99	-0.08	0.68	0.99	0.23
0.10	0.80	0.99	0.38	0.46	0.99	0.04	0.72	0.99	0.31
0.15	0.90	0.99	0.43	0.77	0.99	0.14	1.11	0.98	0.49
0.20	0.96	0.99	0.49	0.90	0.98	0.20	1.27	0.97	0.60
0.25	1.05	0.98	0.52	1.01	0.98	0.29	1.43	0.96	0.72
0.30	0.95	0.98	0.59	0.84	0.98	0.27	1.20	0.97	0.70
0.35	0.95	0.98	0.63	0.83	0.98	0.30	1.17	0.97	0.72
0.40	1.03	0.98	0.64	0.96	0.98	0.34	1.26	0.97	0.77
0.50	1.02	0.98	0.71	0.98	0.98	0.39	1.23	0.98	0.81
0.60	1.06	0.98	0.75	1.07	0.97	0.47	1.28	0.98	0.88
0.70	1.11	0.98	0.84	1.12	0.97	0.55	1.31	0.98	0.94
0.80	1.10	0.99	0.97	1.06	0.98	0.66	1.22	0.99	0.99
1.00	1.35	0.98	1.09	1.41	0.96	0.85	1.51	0.98	1.16
1.25	1.67	0.97	1.28	1.71	0.94	1.08	1.81	0.96	1.36
1.50	1.89	0.96	1.49	1.95	0.92	1.34	2.09	0.94	1.63

$RMSE$ = Root Mean Squared Error, R^2 = Coefficient of determination, $Bias$ = $negative$ bias indicates cold bias in the model while $positive$ indicates a warm bias

Table 10. Model performance statistics at Site C compared to observed soil temperatures

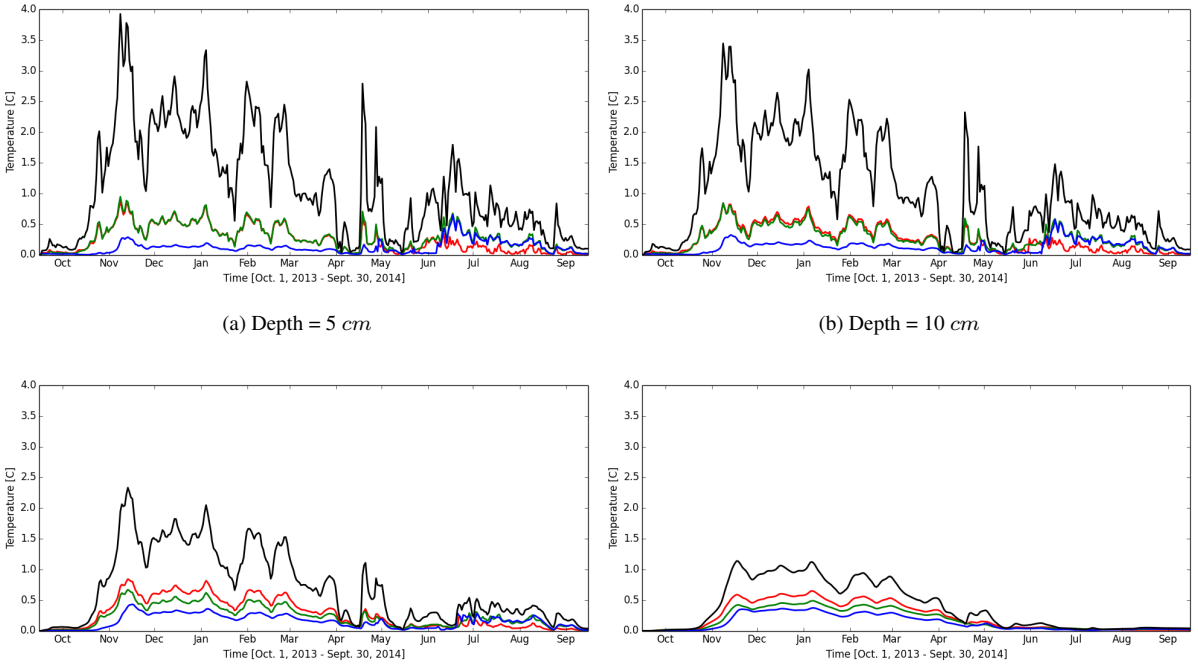
Depth from surface $[m]$	Center			Rim			Trough		
	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$
0.02	0.90	0.99	-0.08	0.95	0.98	0.15	0.67	0.99	-0.04
0.05	0.80	0.99	0.18	1.84	0.96	0.50	0.69	0.99	0.29
0.10	0.63	0.99	0.12	1.23	0.98	0.35	0.78	0.99	0.34
0.15	0.85	0.99	0.21	0.78	0.99	0.31	1.09	0.98	0.57
0.20	0.93	0.98	0.26	0.80	0.99	0.37	1.11	0.98	0.62
0.25	0.98	0.98	0.29	0.80	0.99	0.36	1.16	0.98	0.68
0.30	0.81	0.98	0.29	0.86	0.99	0.41	1.06	0.98	0.70
0.35	0.87	0.98	0.32	0.83	0.99	0.41	1.12	0.98	0.80
0.40	1.03	0.97	0.38	0.80	0.98	0.40	1.16	0.98	0.84
0.50	0.96	0.97	0.42	0.75	0.99	0.42	1.13	0.98	0.91
0.60	0.91	0.98	0.44	0.73	0.99	0.46	1.15	0.99	0.98
0.70	0.87	0.98	0.53	0.73	0.99	0.52	1.17	0.99	1.06
0.80	0.78	0.99	0.64	0.77	0.99	0.56	1.20	0.99	1.14
1.00	0.93	0.99	0.80	0.86	0.99	0.76	1.34	0.99	1.29
1.25	1.13	0.99	1.02	1.02	0.99	0.95	1.58	1.00	1.52
1.50	1.35	0.99	1.27	1.23	0.99	1.19	1.81	1.00	1.76

$RMSE$ = Root Mean Squared Error, R^2 = Coefficient of determination, $Bias$ = $negative$ bias indicates cold bias in the model while $positive$ indicates a warm bias

Table 11. Model performance statistics at Site D compared to observed soil temperatures

Depth from surface $[m]$	Center			Rim		
	$RMSE$	R^2	$Bias$	$RMSE$	R^2	$Bias$
0.02	0.87	0.99	-0.15	1.19	0.98	0.29
0.05	0.80	0.99	0.34	0.86	0.99	0.26
0.10	1.03	0.98	0.45	0.96	0.98	0.36
0.15	1.53	0.96	0.65	1.25	0.97	0.43
0.20	1.65	0.94	0.72	1.29	0.97	0.50
0.25	1.78	0.93	0.80	1.32	0.96	0.53
0.30	1.56	0.94	0.77	1.12	0.97	0.56
0.35	1.68	0.93	0.85	1.14	0.97	0.54
0.40	1.61	0.94	0.84	1.12	0.97	0.56
0.50	1.49	0.95	0.86	1.02	0.98	0.57
0.60	1.46	0.95	0.92	1.02	0.98	0.61
0.70	1.39	0.96	0.95	0.96	0.98	0.63
0.80	1.21	0.97	0.99	0.87	0.99	0.72
1.00	1.20	0.98	1.02	0.97	0.99	0.80
1.25	1.47	0.97	1.27	1.13	0.98	0.94
1.50	1.65	0.97	1.47	1.29	0.98	1.14

$RMSE$ = Root Mean Squared Error, R^2 = Coefficient of determination, $Bias$ = $negative$ bias indicates cold bias in the model while $positive$ indicates a warm bias



(a) Depth = 5 cm

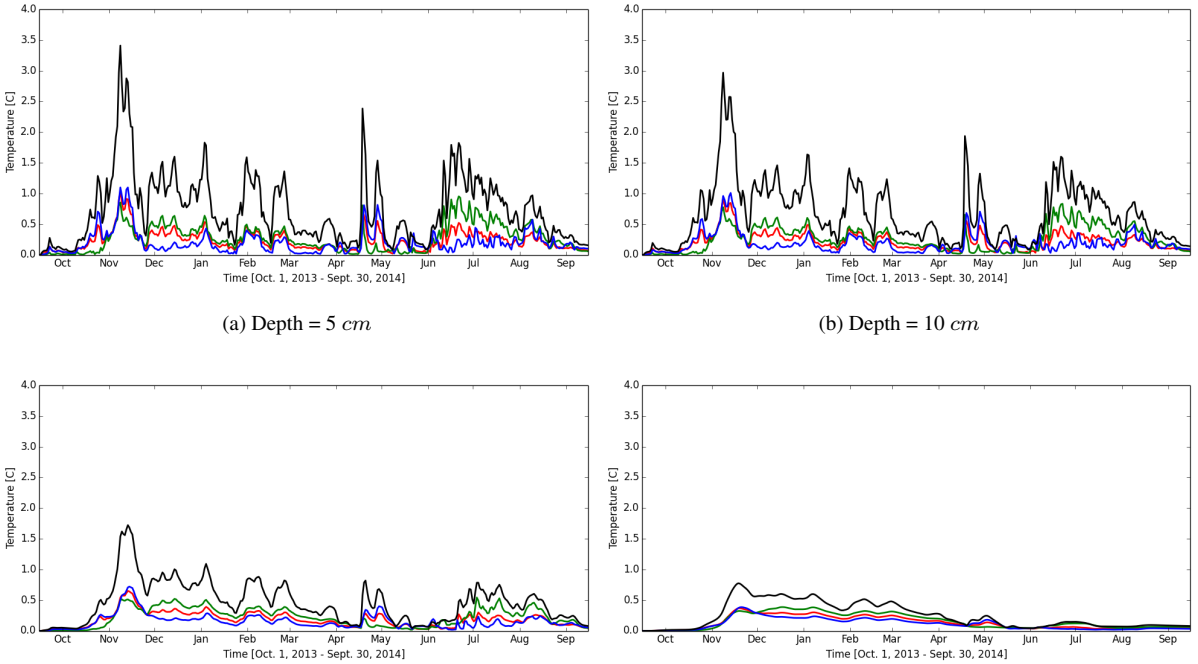
(b) Depth = 10 cm

(c) Depth = 50 cm

(d) Depth = 150 cm

Figure 18. Spatio temporal variability (standard deviation) in simulated soil temperature at Site B.

Red, green and blue lines represent center, rim and trough respectively, and black line represent the standard deviation across the Site B.



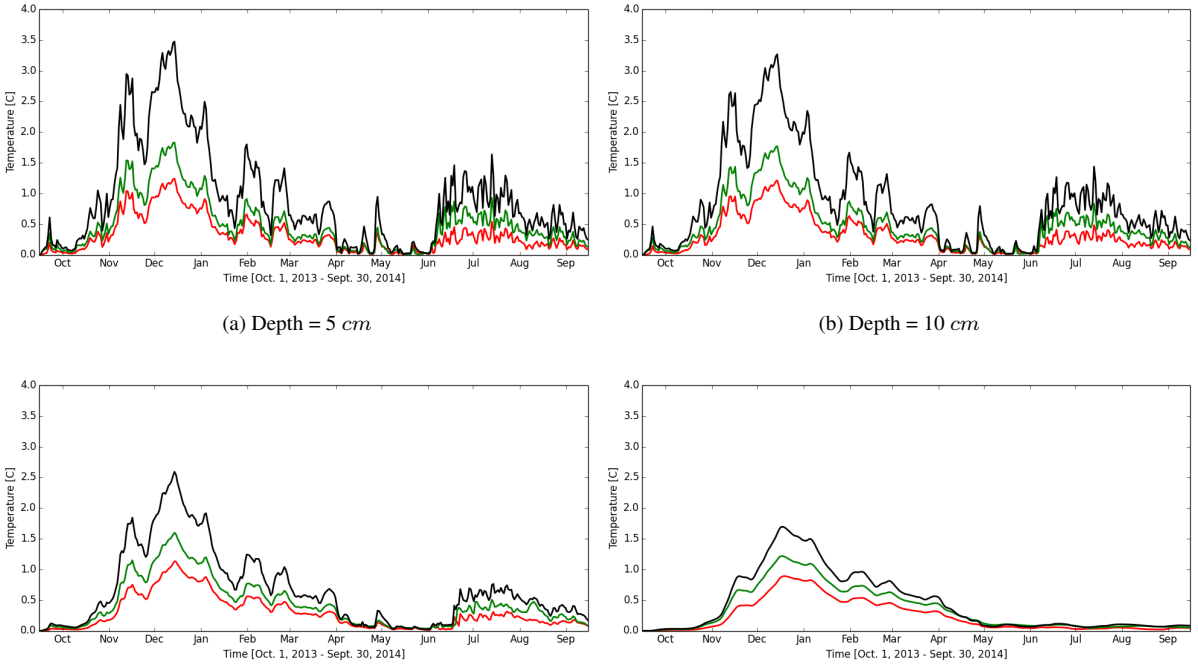
(a) Depth = 5 cm

(b) Depth = 10 cm

(c) Depth = 50 cm

(d) Depth = 150 cm

Figure 19. Spatio temporal variability (standard deviation) in simulated soil temperature at Site C.
Red, green and blue lines represent center, rim and trough respectively, and black line represent the standard deviation across the Site C.



(a) Depth = 5 cm

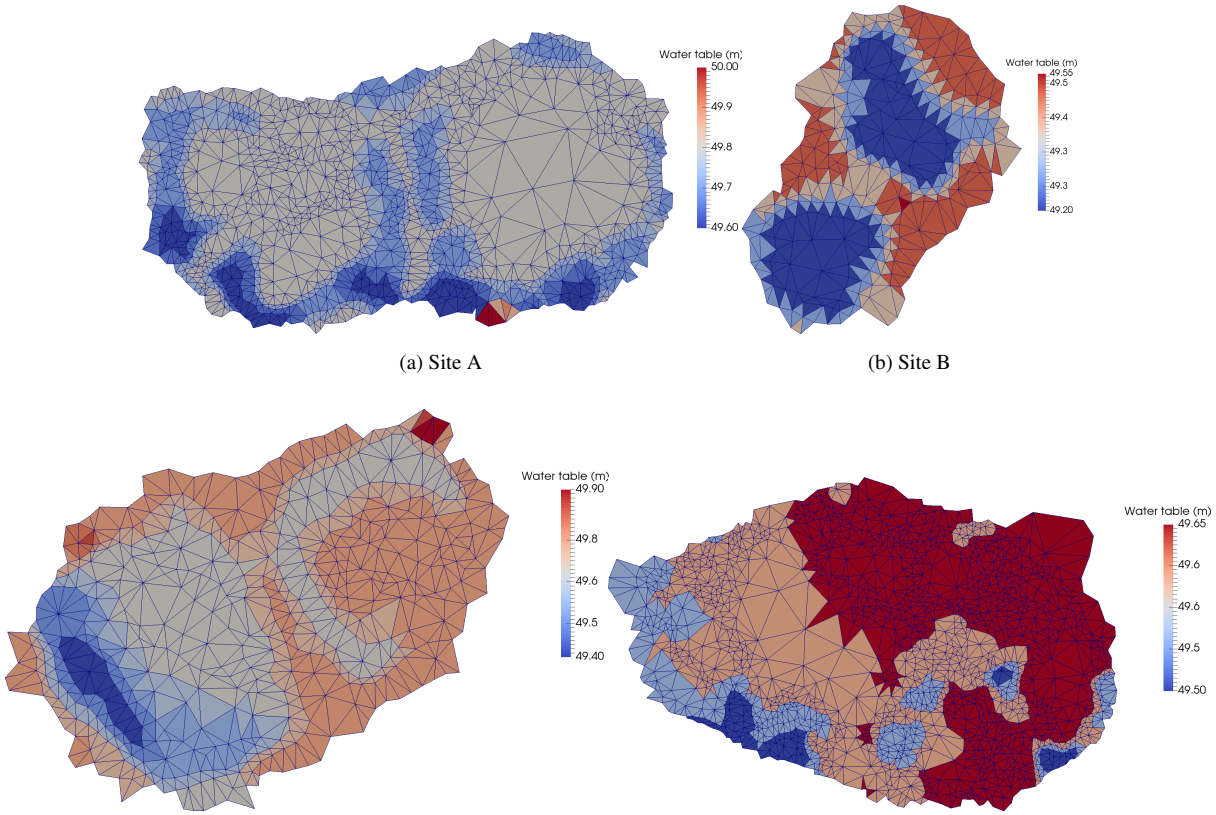
(b) Depth = 10 cm

(c) Depth = 50 cm

(d) Depth = 150 cm

Figure 20. Spatio temporal variability (standard deviation) in simulated soil temperature at Site D.

Red, green and blue lines represent center, rim and trough respectively, and black line represent the standard deviation across the Site D.



(a) Site A

(b) Site B

(c) Site C

(d) Site D

Figure 21. Elevation of maximum water table across the study region. Ground surface is at 50 m elevation. Bold line represents the mean thaw depth across the site, while shaded curve represents standard deviation

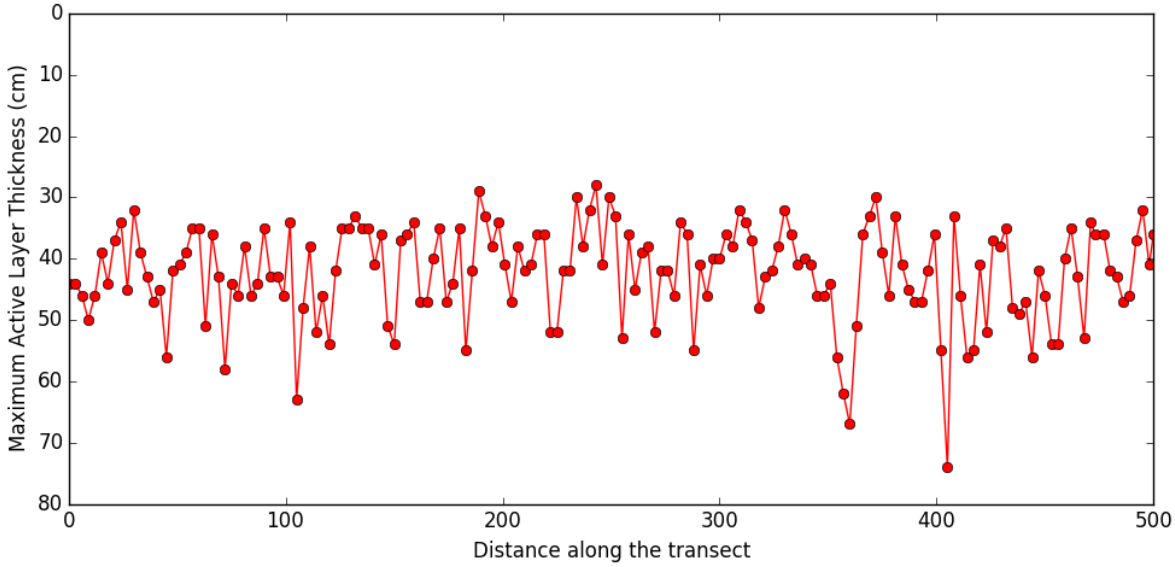
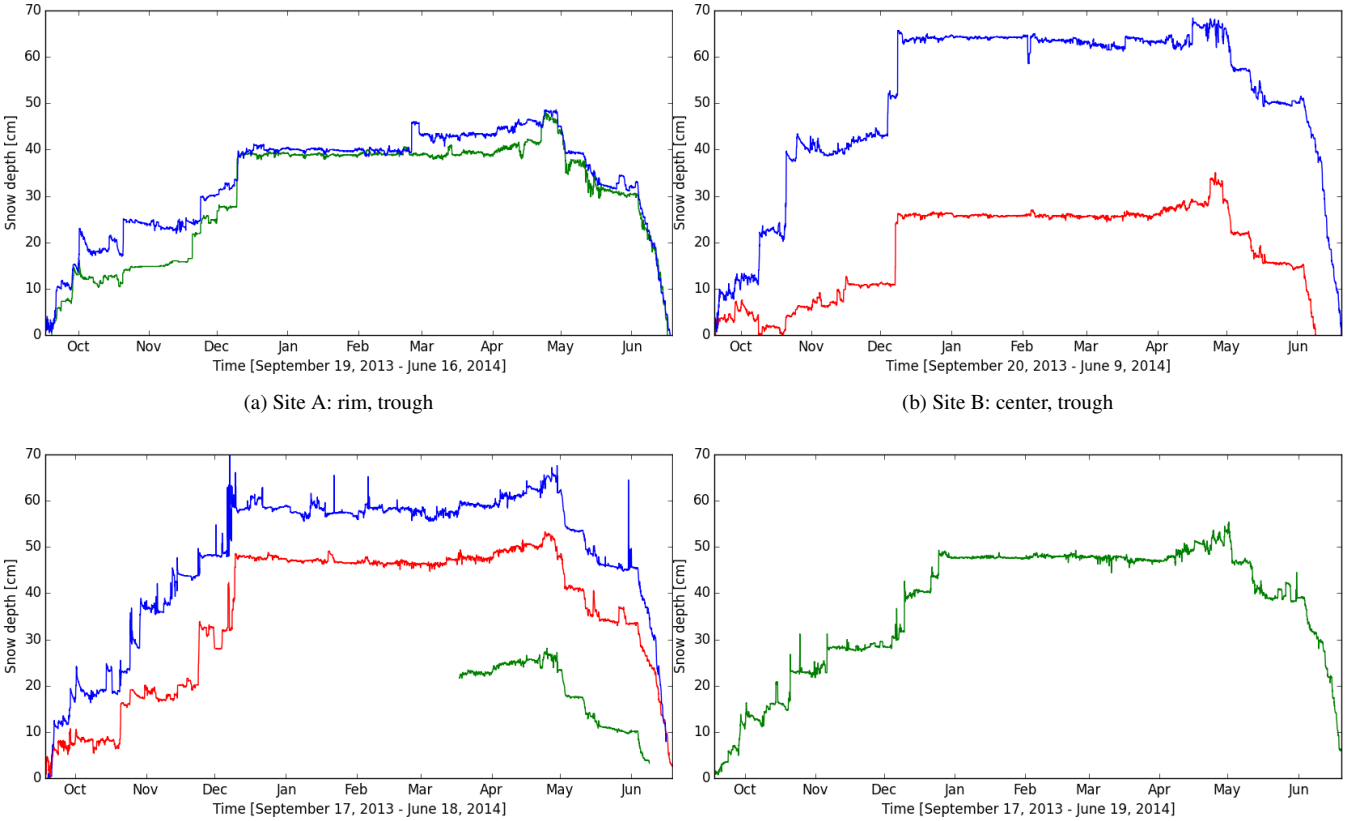


Figure 22. Observed maximum active layer thickness at NGEE-Arctic Site 0 (data collected on July 02, 2014, August 16, 2014, September 23, 2014)



(a) Site A: rim, trough

(b) Site B: center, trough

(c) Site C: center, rim, trough

(d) Site D: rim

Figure 23. Observed snow depths at different topographic positions center (in red), rim (in green), trough (in blue) at our study sites (Hinzman et al. (2014a) .

References

Atchley, A. L., Painter, S. L., Harp, D. R., Coon, E. T., Wilson, C. J., Liljedahl, A. K., and Romanovsky, V. E.: Using field observations to inform thermal hydrology models of permafrost dynamics with ATS (v0.83), Geoscientific Model Development, 8, 2701–2722, doi:10.5194/gmd-8-2701-2015, 2015.

Bárdossy, A.: Calibration of hydrological model parameters for ungauged catchments, Hydrology and Earth System Sciences, 11, 703–710, doi:10.5194/hess-11-703-2007, 2007.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11 – 29, doi:10.1016/S0022-1694(01)00421-8, 2001.

Bockheim, J. G., Hinkel, K. M., and Nelson, F. E.: Soils of the Barrow region, Alaska, Polar Geography, 25, 163–181, doi:10.1080/10889370109377711, 2001.

Collier, N. and Kumar, J.: MeshMaker: Configurable Meshing Framework for Eco-Hydrology Models, Tech. Rep. ORNL/TM-2016/46, Oak Ridge National Laboratory, doi:10.5440/1237353, 2016.

Cresto Aleina, F., Brovkin, V., Muster, S., Boike, J., Kutzbach, L., Sachs, T., and Zuyev, S.: A stochastic model for the polygonal tundra based on Poisson–Voronoi diagrams, Earth System Dynamics, 4, 187–198, doi:10.5194/esd-4-187-2013, 2013.

Dafflon, B., Hubbard, S., Ulrich, C., Peterson, J., Wu, Y., Wainwright, H., and Kneafsey, T. J.: Geophysical estimation of shallow permafrost distribution and properties in an ice-wedge polygon-dominated Arctic tundra region, Geophysics, 81, WA247–WA263, doi:10.1190/geo2015-0175.1, http://dx.doi.org/10.1190/geo2015-0175.1, 2016.

French, H. M.: Thermokarst, pp. 186–215, John Wiley & Sons Ltd,., doi:10.1002/9781118684931.ch8, 2007.

Hammond, G., Andre, B., Bisht, G., Collier, N., Karra, S., Kumar, J., Lichtner, P., and Mills, R.: PFLOTRAN: A Massively Parallel Reactive Flow and Transport Model for describing Surface and Subsurface Processes, online, http://www.pflotran.org/, 2016.

Hammond, G. E., Lichtner, P. C., and Mills, R. T.: Evaluating the performance of parallel subsurface simulators: An illustrative example with PFLOTRAN, Water Resources Research, 50, 208–228, doi:10.1002/2012WR013483, 2014.

Harp, D. R., Atchley, A. L., Painter, S. L., Coon, E. T., Wilson, C. J., Romanovsky, V. E., and Rowland, J. C.: Effect of soil property uncertainties on permafrost thaw projections: a calibration-constrained analysis, The Cryosphere, 10, 341–358, doi:10.5194/tc-10-341-2016, http://www.the-cryosphere.net/10/341/2016/, 2016.

Heuvelmans, G., Muys, B., and Feyen, J.: Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets, Journal of Hydrology, 319, 245 – 265, doi:10.1016/j.jhydrol.2005.07.030, 2006.

Hinzman, L., Kane, D., Gieck, R., and Everett, K.: Hydrologic and thermal properties of the active layer in the Alaskan Arctic, Cold Regions Science and Technology, 19, 95–110, doi:10.1016/0165-232X(91)90001-W, 1991.

Hinzman, L., Romanovsky, V., Cable, W., and Busey, B.: Continuous Snow Depth, Intensive Site 1, Barrow, Alaska [Data], Next Generation Ecosystem Experiments Arctic Data Collection, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.5440/1163347, 2014a.

Hinzman, L., Romanovsky, V., Cable, W., and Busey, B.: Surface Meteorology, Barrow, Alaska, Area A, B, C and D [Data], Next Generation Ecosystem Experiments Arctic Data Collection, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.5440/1164893, 2014b.

Hinzman, L. D., Goering, D. J., and Kane, D. L.: A distributed thermal model for calculating soil temperature profiles and depth of thaw in permafrost regions, Journal of Geophysical Research, 103, 28 975–28 991, doi:10.1029/98JD01731, 1998.

Hobbie, S. E., Schimel, J. P., Trumbore, S. E., and Randerson, J. R.: Controls over carbon storage and turnover in high-latitude soils, Global Change Biology, 6, 196–210, doi:10.1046/j.1365-2486.2000.06021.x, 2000.

Hubbard, S., Gangodagamage, C., Dafflon, B., Wainwright, H., Peterson, J., Gusmeroli, A., Ulrich, C., Wu, Y., Wilson, C., Rowland, J., Tweedie, C., and Wullschleger, S.: Quantifying and relating land-surface and subsurface variability in permafrost environments using LiDAR and surface geophysical datasets, Hydrogeology Journal, 21, 149–169, doi:10.1007/s10040-012-0939-y, 2013.

Kane, D., Gieck, R., and L.C.Bowling: Impacts of surficial permafrost landforms on surface hydrology, in: Proceedings of the Eighth International Conference on Permafrost, edited by Philips, Springman, and Anderson, 2003.

Karra, S., Painter, S. L., and Lichtner, P. C.: Three-phase numerical model for subsurface hydrology in permafrost-affected regions (PFLOTRAN-ICE v1.0), The Cryosphere, 8, 1935–1950, 2014.

Kneafsey, T. and Ulrich, C.: CT Scans of Cores Metadata, Barrow, Alaska 2015 [Data], Next Generation Ecosystem Experiments Arctic Data Collection, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.5440/1170517, 2016.

Kumar, J., Collier, N., Bisht, G., Mills, R. T., Thornton, P. E., Iversen, C. M., and Romanovsky, V.: Modeling the spatio-temporal variability in subsurface thermal regimes across a low-relief polygonal tundra landscape: Modeling Archive [Data], Next Generation Ecosystem Experiments Arctic Data Collection, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.5440/1184018, 2016.

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, doi:10.5194/tc-5-151-2011, 2011.

Liljedahl, A. K., Hinzman, L. D., Harazono, Y., Zona, D., Tweedie, C. E., Hollister, R. D., Engstrom, R., and Oechel, W. C.: Nonlinear controls on evapotranspiration in arctic coastal wetlands, Biogeosciences, 8, 3375–3389, doi:10.5194/bg-8-3375-2011, 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., Necsoiu, M., Raynolds, M. K., Romanovsky, V. E., Schulla, J., Tape, K. D., Walker, D. A., Wilson, C. J., Yabuki, H., and Zona, D.: Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology, Nature Geoscience, 9, 312–318, http://dx.doi.org/10.1038/ngeo2674, 2016.

Madsen, H.: Automatic calibration of a conceptual rainfall–runoff model using multiple objectives, Journal of Hydrology, 235, 276 – 288, doi:10.1016/S0022-1694(00)00279-1, 2000.

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, doi:10.3402/tellusb.v64i0.17301, 2012.

Nature: Journals unite for reproducibility, Nature, 515, doi:10.1038/515007a, 2014.

Painter, S.: Three-phase numerical model of water migration in partially frozen geological media: model formulatoin, validation, and applications, Computational Geosciences, 15, 69–85, 2011.

Painter, S. and Karra, S.: Constitutive Model for Unfrozen Water Content in Subfreezing Unsaturated Soils, Vadose Zone Journal, 13, 2014.

Peterson, J.: Ground Penetrating Radar, Barrow, Alaska [Data], Next Generation Ecosystem Experiments Arctic Data Collection, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.5440/1171723, 2016.

Post, W. M., Emanuel, W. R., Zinke, P. J., and Stangenberger, A. G.: Soil carbon pools and world life zones, Nature, 298, 156–159, 1982.

Romanovsky, V. and Cable, W.: Subsurface Temperature, Moisture, Thermal Conductivity and Heat Flux, Barrow, Area A, B, C, D [Data], Next Generation Ecosystem Experiments Arctic Data Collection, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.5440/1126515, http://lapland.gi.alaskaedu/vdv/index.html, 2012.

Romanovsky, V. E., Smith, S. L., and Christiansen, H. 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, doi:10.1002/ppp.689, 2010.

Shafii, M. and De Smedt, F.: Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm, Hydrology and Earth System Sciences, 13, 2137–2149, doi:10.5194/hess-13-2137-2009, 2009.

Singh, A. and Minsker, B. S.: Uncertainty-based multiobjective optimization of groundwater remediation design, Water Resources Research, 44, doi:10.1029/2005WR004436, w02404, 2008.

Skurikhin, A. N., Gangodagamage, C., Rowland, J. C., and Wilson, C. J.: Arctic tundra ice-wedge landscape characterization by active contours without edges and structural analysis using high-resolution satellite imagery, Remote Sensing Letters, 4, 1077–1086, doi:10.1080/2150704X.2013.840404, 2013.

Tang, G., Yuan, F., Bisht, G., Hammond, G. E., Lichtner, P. C., Kumar, J., Mills, R. T., Xu, X., Andre, B., Hoffman, F. M., Painter, S. L., and Thornton, P. E.: Using reactive transport codes to provide mechanistic biogeochemistry representations in global land surface models: CLM–PFLOTRAN 1.0, Geoscientific Model Development Discussions, 8, 10 627–10 676, doi:10.5194/gmdd-8-10627-2015, 2015.

Tesfa, T. K., Li, H.-Y., Leung, L. R., Huang, M., Ke, Y., Sun, Y., and Liu, Y.: A subbasin-based framework to represent land surface processes in an Earth system model, Geoscientific Model Development, 7, 947–963, doi:10.5194/gmd-7-947-2014, http://www.geosci-model-dev.net/7/947/2014/, 2014.

Tweedie, C. E.: Barrow Area Remote Sensing - Digital Elevation Model [Data], Earth Observing Laboratory, National Center for Atmospheric Research, doi:10.5065/D6KS6PQ3, 2010.

Zona, D., Gioli, B., Commane, R., Lindaas, J., Wofsy, S. C., Miller, C. E., Dinardo, S. J., Dengel, S., Sweeney, C., Karion, A., Chang, R. Y.-W., Henderson, J. M., Murphy, P. C., Goodrich, J. P., Moreaux, V., Liljedahl, A., Watts, J. D., Kimball, J. S., Lipson, D. A., and Oechel, W. C.: Cold season emissions dominate the Arctic tundra methane budget, Proceedings of the National Academy of Sciences, 113, 40–45, doi:10.1073/pnas.1516017113, 2016.