ModEx Reviewer #1

This study presents application of the Arctic Terrestrial Simulator (ATS) to simulate ice wedge dynamics near Barrow, Alaska. The subject matter is timely as the ability to model the complex interactions between water and heat in arctic grounds is currently lacking. As such, the study presents a nice step forward in advancing the science and our ability to model permafrost dynamics. Further, the study does well to combine observational data with modeling simulation. The study is well presented and well written. With that, I have only some minor comments for the authors to consider.

We are grateful for the reviewer's recognition in the quality and timeliness of this work and thank the reviewer for the insightful comments and recommendations.

In general, I appreciate the use of the ModEx cycle approach. An a priori assumption of a modeling structure is ubiquitous and often clouds the potential for process insight across the current generation of hydrological (let alone permafrost) models. It would be good to see a bit more reference in discussion to other approaches (e.g., FUSE modeling from Clark or FLEX from Fenicia) that allow for model structure flexibility. This will make for a richer consideration of the current field of modeling and increase connection to existing research beyond arctic regions.

We have now included a discussion of the FUSE and FLEX modeling approach at Lines 176-179, in the new manuscript. We furthermore thank the reviewer from bringing this literature to our attention as it provides a good tie to literature dealing with t calibration and module structure reduction.

It is interesting to settle on a root mean square error response function. Were any other functions considered? There is marked bias in the RMSE toward high-end errors in estimates that cold impact the calibration procedure. It warrants consideration of various response functions or optimization approaches here. For example, limits of likelihood or Pareto front approaches could be interesting in a multi-objective sense. That said, such full optimization procedure consideration is outside the scope of this study. However, the potential impacts or limitations of selecting RMSE could be presented and discussed.

The ability of the RMSE approach to target high end errors was, we believe, beneficial to the overall calibration process, specifically for the errors that occurred during the summer months when ALT is evolving. As shown in Figure 7, using the RMSE for a gradient based calibration resulted in a substantial decrease in error and eliminated much of the summer time temperature differences. A Pareto front would be interesting for a multi-objective approach, however, during the subsurface calibration when a calibration response surface was used, only subsurface temperatures were considered for calibration targets. Calibration parameters such as porosity or thermal conductivity could have been used to limit objective functions by creating Pareto optimality, which would have prevented over calibration. However, allowing the parameters a large possible range enabled the calibration procedure to identify structural error in the model. For example, consistently

calibrating to unrealistic parameters for porosity and thermal conductivity of the coupled calibration is section 3.3 diagnosed the need to include unsaturated conditions for the centers. This is why, as is discussed in lines 13-16 on page 3241, calibration parameters were allowed a large possible range.

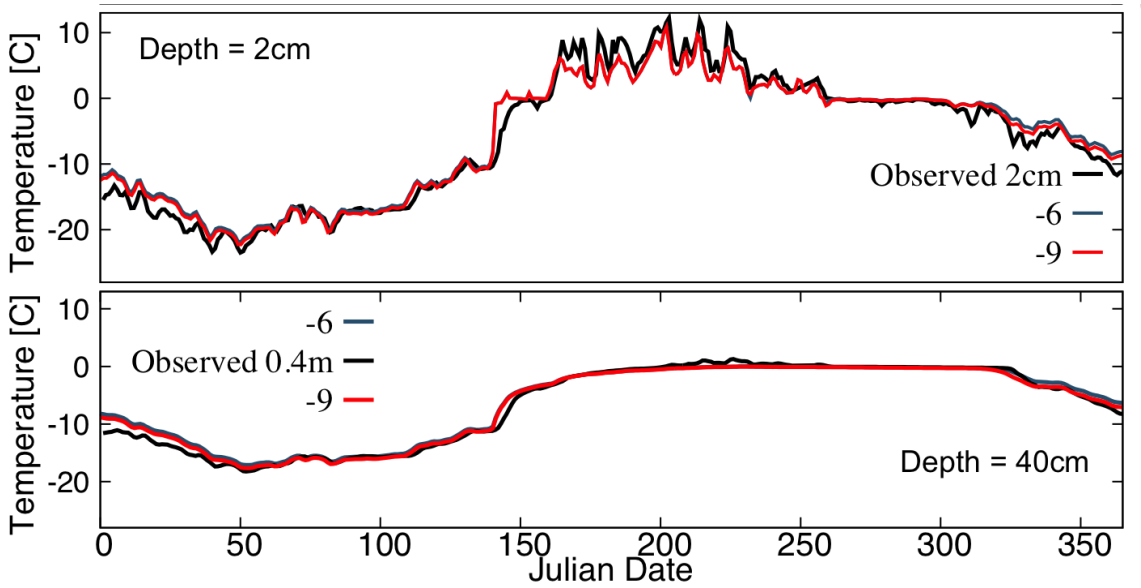
It is somewhat interesting that there is no consideration of the impact of uncertainty in the parameter definitions on the modeling performance. Clearly, this is a complex model with various interactions (hence the ModEx approach adopted). With that, it would be interesting to understand better the role of uncertainty in defining a given parameter on the subsequent model performance. Specifically, this is the case with regards to taking field observations into the modeling environment. A simple sensitivity analysis would be helpful in this regard. As it is currently presented, the modeling comes across as extremely site specific. Of course, there is some consideration of a mixed-scale approach to couple this detailed modeling into a larger scale system. However, without understanding the uncertainty impacts associated with defining the parameterization in ATS (let alone how it can shift across scale) there may be difficulty in generalization of the findings. Since the manuscript is rather dense and should not be overly extended, I recommend the authors take up some more discussion on these aspects (in particular surrounding parameter identifiability and observational un-certainty).

We agree that uncertainty is important and that it should be thoroughly addressed. So much so in fact that our original aim was identify how to best identify parameter uncertainty and specifically what parameters contribute to model uncertainty. However, we soon discovered that properly calibrating and creating a process rich model of thermal hydrology systems which includes site-specific field data was a difficult but rewarding task that deserved its own place in literature. We therefore decided to write a manuscript devoted to the model creation and calibration process. The subsequent parameter uncertainty and sensitivity analysis has recently been submitted to 'Cryosphere.' Never-the-less, we've decided to add a small discussion about the importance of a parameter uncertainty analysis and now point to how future uncertainty analysis will provide a greater breadth of information by adding, ***"****Further modeling efforts that focus on uncertainty analysis and environmental parameters sensitivity to provide information which parameters govern model outcome will inform observational efforts*." to lines 718-720 in the new manuscript.

Specific Comments

Page 3243: It is not completely clear to me why a constant temperature of -6◦C is set for the bottom boundary at 50m depth. Is this based on some observation, was it somehow calibrated, and how could this affect the results?

The -6◦C bottom boundary condition was chosen because it represents a far field constant low temperature gradient. However, simulations with colder bottom boundary conditions were performed and had little to no affect of ALT formation or shallow soil temperatures.



This figure shows soil temperature time series for the observed soil temperature at 2cm and 40cm depth, and simulations with a -6 and -9 bottom boundary condition. Only small temperature differences are found at either depth.

We now clarify in the new manuscript at Lines 273-275 that, "A far field bottom boundary condition was held constant at -6°C to represent the average deep permafrost temperature in the North Slope of Alaska (Romanovsky, et al., 2010)."

Pages 3245-3246: The two models for thermal conductivity were calibrated for fully saturated conditions and the BPC model resulted in unrealistic parameter values and was discarded. However, the next section tells that unsaturated conditions are likely for two of three boreholes and that this would affect the resulting simulated temperatures. It is not clear from the text why it is enough to evaluate the two thermal conductivity models against each other for only fully saturated conditions, if unsaturated/surface energy balance processes do indeed affect these results.

This is a very intuitive observation from the reviewer and one that the authors considered as well, and as such deserves some additional discussion here and in the Manuscript. Because the goal is to arrive at a realistic and calibrated model, rather then to exhaustively explore all modeling options it would be better to move forward and not posthumously retesting prior model structural decisions. We also felt that because the MC thermal model was more physical as described in section 2.3, where each component; soil material, ice, liquid, and gas contributes to the thermal conductivity of the subsurface, the affect of unsaturated conditions especially transient saturation would provide a better system representation and therefore calibration parameters. However, it is also important to admit that not all decisions were straightforward and completely quantitative as stated at Line 17-19, page 3239. For this reason and for better clarity in the final manuscript we have added, "Here we only tested unsaturated

conditions using the MC thermal model rather then to posthumously retesting prior model structural decisions, as the MC model was thought to be more physically accurate." to the text at Line 416-417, to inform the reader why we made our decision, as well as admit, that the BCP approach may be adequate.

Page 3248, line 17: "…a single layered of snowpack…", should read "…a single layer of snowpack…"?

Sentence now reads, "…Appendix B are applied on a single layer snowpack."

Page 3251, line 7: "…consistently lower then…" should read "…consistently lower than…"?

Made change in new manuscript.

Reviewer #2:

This paper describes an excessively detailed assessment of how to model a set of temperature measurements done at different depths in an Arctic landscape.

The topic area of this study and model development is important. However, this paper seems to overshoot the goal of providing a straightforward and useable modeling approach for these systems.

We thank the reviewer for their time and effort on this review and specifically agree that this is a timely topic that deserves much attention.

Our primary goal for this manuscript, as stated in the title, is to demonstrate how field observations can be incorporated into the development of process-rich models, both in terms of building confidence in the process representations and in systematically inferring model parameters that are not directly measurable. As noted by Kurylyk and Watanabe (2013) a need for better representations of permafrost environments in a warming climate has motivated the development of fine-scale thermal models. These emerging models are complements to the reduced complexity models used at regional and pan-Arctic scales and consider the wide range of coupled processes that are needed to model the permafrost environment at the level of detailed required for comparing to observations at their native spatial and temporal scales. Our manuscript fills a gap between these new process models that have been evaluated against laboratory data and regional models with coarse spatial resolution, which are poorly constrained by direct observation. By iterative calibration in what is termed the 'ModEx' cycle we are able to evaluate competing representations for processes governing ALT, calibrate the most successful representations, and then incorporate those process representations in the model development. This paper is an important step in the longer-term goal of refining and building confidence in land-surface models of permafrost affected regions and indeed far surpasses the

objective, 'to model a set of temperature measurements'. Moreover, the iterative calibration and model refinement process documented in our manuscript has broader applicability to the development of environmental systems models in that a detailed guide for developing process rich models with available field data is presented and is of interest to readers of Geoscientific Model Development.

Given that the reviewer had missed the primary goals of this work, we have revised the manuscript to be more specific in the abstract, introduction and conclusions so that readers can clearly identify the purpose of the paper. First in the abstract we now state the goals of the this work and have added the reworded to now say, "A recently developed surface/subsurface model for permafrost thermal hydrology, the Advanced Terrestrial Simulator (ATS), is used in combination with field measurements to achieve the goals of constructing a process rich model based on plausible parameters and to identify fine scale controls of ALT in ice wedge polygon tundra in Barrow, Alaska." Then in the introduction we now say, "We use repeated calibration of model parameters against site-specific field measurements and iterative model adjustments of the model structure to reduce mismatch between model predictions and measurements in order to attain a viable model of thermal hydrological conditions." And we conclude the introduction section with a summary of our approach and have reworded part of it to say, "In this paper we summarize our ModEx experience involving the detailed use of subsurface temperature and snow cover field data to develop and test process-rich simulations of ALT dynamics, such that observational data and necessary physical dynamics are incorporated into the model." Finally in the conclusion section we now restate how our ModEx approach achieved this works objectives with, "The particular variant of the ModEx approach combined calibration with iterative refinement of the model structure; parameter feasibility and model-observation mismatch were used as metrics to achieve the objective of model development and identification of viable representations of key thermal hydrological process." We also end the paper with a discussion regarding our approach to merging observations and model development, and how similar approaches may be useful in other applications.

It is unclear how this kind of model simulations could be used to inform models at a regional scale.

There is significant interest in using fine-scale models to challenge and improve the coarse parameterizations used in regional and global land surface models. In particular, fine-scale models can represent processes and heterogeneities in greater detail and at the native spatial scales of field observations, and can thus bridge spatial scales and generally build confidence in coarse-scale models. This work is not focused on regional scale models, but by combining fine-scale models and observations to identify appropriate representations of key processes and appropriate parameterizations of those processes, the work indirectly informs regional scale models. Based on our results, we did suggest on page 3256 Lines 11-

16 in the conclusion section that multiscale models that use overland flow to establish ponded depth in conjunction with subsurface thermal process models are a good approximation for simulating ALT at scale. Furthermore, our work describes in detail what processes our model found to be important for representing ALT. By employing the iterative ModEx calibration process we found that 1) representing thermal conductivity as dependent on material properties and saturation states are necessary to propagate thermal changes in the subsurface (Section 3.2). 2) The dominant and transient saturation states are also necessary, especially considering how thermal conduction depends on both the phase and saturation state of the subsurface (Sections 3.4 and 4.3) 3) The representation of snow distribution, snow deformation, i.e ageing and depth hoar formation (Section 4.4). These are physical representations that may be important for large scale or multiscale models to consider, and in the case of subsurface thermal conductivity, may require extensive calibration that can be achieved using fine scale models. For greater clarity regarding how our work can be used to inform larger scale models, we also now describe in the introduction a general manor of how fine scale models such as the one presented here can be used to inform larger scale models, "Improved fine-scale simulation capabilities can inform the representation of soil thermal processes in regional to global scale models by identifying appropriate representations of key processes governing ALT, and by providing calibrated model parameterization."

Why spend so much effort on the detailed parameterisation of thermal properties if lateral heat flow might be important, which is then not included.

We acknowledge that lateral heat and water flow might have an influence on the system (Page 3251 Line 25) and future 3-D modeling and fieldwork is necessary to quantify what the influence of lateral heat flow might be (Page 3256 Lines 17-26).

However, 1D calibration and parameterization is beneficial in that the computational time to simulate a 1D problem allows for the many simulations necessary to sufficiently explore parameters space in order to identify what thermal properties are necessary to simulate ALT. Our work also shows that without a representation of lateral heat flow, we are able to match subsurface temperatures consistently for rims and centers, and with the exception of early snow melt and fall freeze, the simulated trough temperatures match the observed temperature with plausible subsurface properties.

It seems awkward to fix the lower temperature boundary, it is unclear what this is based upon.

The lower boundary temperature is a far field boundary condition that is within the average permafrost temperature for the North Slope Alaska (Romanovsky, et al., 2010). A fixed boundary condition is a reasonable approximation as seasonal temperature variations generally to not penetrate deeper the 10 to 16 meters and deep permafrost will see only negligible temperature increases over the course of a calibration that spans only a few years.

The paper is written very densely, but still does not contain enough information to fully appreciate what it is that has been carried out, and how. On the other hand this contains to much information without detailed description resulting in a difficult to read.

> *We believe that the additions to the manuscript, which now clearly state the objective of the work provide a bases for the level of detail in the manuscript. (See response to the reviewer number 2's first* comment).

One is left with a feeling that the authors invested a lot of effort to develop a unusually
detailed model but then fail to carry out a sensitivity or uncertainty study to evaluate the need for the complex model construction presented here. Could the same fit be obtained with a much simpler model too? In other words what is the sensitivity of the model fit to model complexity?

> *We acknowledge that a lot of effort was indeed invested to develop this model as is necessary for such complex and dynamic process representation. We further appreciate the question of needed complexity as distinguishing the governing processes from those that can be neglected is a central component to scaling the representation of thermal hydrology up to a larger scale model. Our manuscript documents how added process representation, i.e transient saturation, phase change, and the tightly coupled surface thermal conduction is needed to capture the subsurface thermal regime. In other words, our work started with a simple model, and as a result of incorporating observational data into the iterative calibration procedure, we were able improve model performance and identified the level of processes representation needed. For example, the addition of transient saturation in section 3.4 and 4.3, and refinement of the snow representation is section 4.4, demonstrate that without the additional process representation model fit and the plausibility the parameters used in the model would not be possible. Nevertheless, appropriate model complexity needs to be addressed, especially when attempting to find the appropriate level of process representation in larger scale models. We therefore added a comment about complexity in the conclusion section, "Further evaluations of the 1-D representations against 3-D model representations are needed to identify addition process representation and the appropriate level of model complexity to capture scale dependencies of thermal dynamics."*

Other comments are provided in the attached annotated PDF.

Location in Original Text Page 3237, Line 16-17: "These local-to-intermediate scale processes are under-resolved or completely missing in ESMs. Therefore, improved fine-scale simulation capabilities can inform the representation of soil thermal processes in regional to global scale models."

Reviewer #2 Comment: This makes it sound like the upscaling of what you find at the small scale to the ESM scale is obvious. I am sure it isn't. The term 'inform' is somewhat ambiguous. How are the ESM's informed exactly?

*We agree that this sentence is somewhat ambiguous and have therefore re-worded that sentence to be more specific and describe a general way in which fine-scale modeling efforts can inform larger-scale simulations. The sentence now states: "*Improved fine-scale simulation capabilities can inform the representation of soil thermal processes in regional to global scale models by identifying the appropriate representations of key processes governing ALT, and by providing calibrated model parameterization.*"*

Location in Original Text Page 3239, Line 1-2: "Additionally, correct model structure representation is typically not known a priori."

Reviewer #2 Comment: how is this defined? what is correct?

*We now define what is meant by correct model structural representation, and have changed the sentence to, "*Additionally, correct model structure representation, capable of representing the system based on known physical relationships while using plausible model parameters, is typically not known *a priori*.*"*

Location in Original Text Page 3239, Line 5-7: Therefore, when dealing with a coupled system of complex processes, it is imperative that the conceptual model is refined during the calibration process to increase model structure adequacy (Gupta et al., 2012)."

Reviewer #2 Comment: Does this not all depend on the objective of the modeling study? Is the objective here to correctly estimate, forecast ALT?

The objective of this study is two fold, first to incorporate observational data into a process rich model representation of ALT dynamics, second is to identify the appropriate representations of governing processes that control the thermal hydrological dynamics that form ALT. Meeting these goals then creates a model that is capable of estimating and forecasting ALT.
Many models may be capable of simulating ALT, however, without rigorous testing and comparison to observed variables (in this case, the plausibility of calibrated parameters) models may simulate the correct ALT for the wrong reasons. A model that is calibrated using the wrong structure, i.e., conceptual model, can result in erroneous forecasts, especially if conditions change such as the case for climate change. The problem of over-fitting and relying on models that are 'calibrated', but do not use plausible parameters is discussed in the previous sentence. If however, the model is tested and refined to produce both an accurate ALT and plausible calibrated parameters, some of the structural error can be reduced and confidence in ALT projections is increased.

Location in Original Text Page 3240, Line 14-16: "Here the ModEx procedure moves beyond the standard calibration by assuming the model itself is unknown, but can be refined through successive comparison to observation (outer loop in Fig. 2)."

Reviewer #2 Comment: how can something unknown be refined?

We now clarify the intent of this sentence by, "Here the ModEx procedure moves beyond the standard calibration by assuming the model itself is

uncertain, but can be further constrained through successive comparison to observation (outer loop in Figure 2)."

Location in Original Text Page 3240, Line 18: "These improved model runs then inform the observation process by specifying the data needs, either through informal numerical experimentation or through more formal data worth exercises."
Reviewer #2 Comment: What are these?

Changed sentence to be more specific, "These improved model runs then inform the observation process by specifying the data needs, either through further calibration or through informal numerical experimentation."

Location in Original Text Page 3240, Line 20-22: "We implement ModEx model refinement by focusing on the plausibility of calibrated parameters in addition to the mismatch between field measurements and simulated responses."
Reviewer #2 Comment: Is this a manual process?

Calibration is automatic as described in the manuscript. The model refinement was done manually. We evaluate the set of calibrated parameters against the range of appropriate parameter values compiled from literature (See Appendix C) or field measurements. In doing so, insight about model behavior is gained which can them be used to improve the model and reshape the calibration response surface. For better clarity we have replaced the work 'focusing' with 'evaluating'.

Location in Original Text Page 3241, Line 12: "However, in the case of a complex model with high dimensionality, multiple local minima may exist, which results in gradient-based calibrations finding many solutions to the problem (Beven, 2006)."
Reviewer #2 Comment: non-uniqueness this is often called.

Yes. For clarification we now say, "However, in the case of a complex model with high dimensionality, multiple local minima may exist, which causes gradient-based calibrations to find non-unique solutions."

Page 3241, Line 16-17: "It is important to extend calibration boundaries beyond the acceptable parameter range in order to both diagnose model inadequacy and avoid boundary effects caused by automated calibration algorithms."
Reviewer #2 Comment: It is unclear what this exactly means, at least to me.

We now re-worded the sentence for clarity, to say, "Model structure error can also cause the response surface to slope to a parameter boundary indicating that over-fitting is necessary to calibrate to observed data (Beven, 2005). Therefore, it is important to extend calibration boundaries beyond the acceptable parameter range to allow the optimization algorithm to travel into the infeasible range when the response surface dictates an implausible combination of parameter values, indicating an inadequate model."

Location in Original Text Page 3242, Line 23: "The focus of the model development chronicled here is NGEE-Arctic site "Area C" (Fig. 1), which is characterized by ~ 50 cm deep troughs, rims and shallow low centers."
Reviewer #2 Comment: I have a sense that Fig 1 is first mentioned after Fig 2, but I

did not check in depth.

Figure 1 is introduced first on page 3239 line 15, while Figure 2 is introduced on line 17, after figure 1.

Location in Original Text Page 3243, Line 9-10: "The bottom boundary condition was held constant at a temper1ature of -6 ℃."
Reviewer #2 Comment: Why was a constant temperature boundary chosen? Surely the temperature at this depth cannot be considered constant over a 20 year period during which long term GST changes occur.

The bottom boundary condition is set as a far field boundary condition at depth of 50 meters. The boundary condition temperature is within the range of permafrost temperatures of the North Slope, which has seen between 0 to 2 degrees warming between the years of 1975 and 2010 (Romanovsky et al., 2010) for the entire permafrost zone. Furthermore, seasonal temperature changes do not penetrate to deep permafrost (See figure 4 in Romanovsky et al., 2010). Therefore, a calibration exercise over the course of a few years, such as this one will see negligible deep (greater than 15 meters) permafrost warming.

We now clarify that this is based on an average far field boundary condition and have changed the manuscript to now say, "The underlying mineral soil was a silty loam to a total depth of 50 m. A far field bottom boundary condition was held constant at -6℃ to represent the average deep permafrost temperature in the North Slope of Alaska (Romanovsky, et al., 2010)."

Location in Original Text Page 3243, Line 23: "In this model liquid water can coexist with ice below 0 C, as observed (Watanabe and Wake, 2009), which is an important process to represent in soils with rapid freeze thaw cycles in order to prevent unrealistic rapid cooling of the subsurface (Romanovsky and Osterkamp, 2000; Nicolsky et al., 2007)."
Reviewer #2 Comment: what process? the coexisting of liquid water and ice below 0 C is not a process, that is a phenomenon as a result of a process. But the process remains unnamed here. Liquid and ice partitioning is the process? I am not sure that is a process either. What is causing this?

We agree that 'process' is a poor word choice here. None-the-less, some water in pore space remains as liquid below 0℃ due to surface forces and pore geometry (Dash et al., 1995). In a thermal model it is also important to accurately represent the phases of water, which have different thermal conductivities. We therefore have re-worded the sentence to be more specific and now state, "In this model liquid water can coexist with ice below 0^{o}C, as is well known (e.g. MILLER, 1980; Williams and Smith 1991), which occurs due to soil surface forces and pore geometry."

Location in Original Text Page 3251, Line 25-28: "A possible reason for the underestimated soil moisture is that the 1-D surrogate model neglected lateral surface- and subsurface flow that could be flowing on to the column, especially for troughs that are connected to an extensive trough-network."
Reviewer #2 Comment: Indeed, so what is the point of all this detailed calibration

for thermal properties?

The point of the detailed calibration of thermal properties is to identify dominant controls of ALT and to best represent those processes in models. If adjustment to the conceptual model is warranted in order to attain both good fit to calibration targets and plausible parameters we changed the model accordingly to build a process rich model and noted why and how we think the model improvement is necessary. If we thought that additional process consideration may only slightly improve model performance, we noted what that process might be and how it could improve the model. Documenting the model development process in this way is important to 1) demonstrate the thought necessary for process rich model development and 2) add to literature the reasons why some processes are included and others are not necessary.

Here our model representation of the system was found to be good with plausible parameters for most times and depths with the exception of spring and fall periods in the trough. Lateral flow could contribute to the mismatch between observations and simulation in the troughs. However, if representing lateral flow were to improve the simulation, it would only improve the trough representation and for a small percentages of time (Figure 9).

Using Field Observations to Inform Thermal Hydrology Models of Permafrost Dynamics with ATS (v0.83).

Adam L. Atchley[1], Scott L. Painter[2], Dylan R. Harp[1], Ethan T. Coon[1], Cathy J. Wilson[1], Anna K. Liljedahl[3,4], V. E. Romanovsky[5]

[1] {Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA}
[2] {Climate Change Science Institute, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA}
[3] {Water and Environmental Research Center, Univ. of Alaska Fairbanks, USA}
[4] {International Arctic Research Center, Univ. of Alaska Fairbanks, USA}
[5] {Geophysical Institute, University of Alaska Fairbanks, USA}
Correspondence to: Adam. L. Atchley (aatchley@lanl.gov)

Abstract

Climate change is profoundly transforming the carbon-rich Arctic tundra landscape, potentially moving it from a carbon sink to a carbon source by increasing the thickness of soil that thaws on a seasonal basis. However, the modeling capability and precise parameterizations of the physical characteristics needed to estimate projected active layer thickness (ALT) are limited in Earth System Models (ESMs). In particular, discrepancies in spatial scale between field measurements

and Earth System Models challenge validation and parameterization of hydrothermal models. A recently developed surface/subsurface model for permafrost thermal hydrology, the Advanced Terrestrial Simulator (ATS), is used in combination with field measurements to achieve the goals ~~calibrate~~of constructing a process rich model based on plausible parameters and to identify fine scale controls of ALT in ice wedge polygon tundra in Barrow, Alaska. An iterative model refinement procedure that cycles between borehole temperature and snow cover measurements and simulations functions to evaluate and parameterize different model processes necessary to simulate freeze/thaw processes and ALT formation. After model refinement and calibration, reasonable matches between simulated and measured soil temperatures are obtained, with the largest errors occurring during early summer above ice wedges (e.g. troughs). The results suggest that properly constructed and calibrated one-dimensional thermal hydrology models have the potential to provide reasonable representation of the subsurface thermal response and can be used to infer model input parameters and process representations. The models for soil thermal conductivity and snow distribution were found to be the most sensitive process representations. However, information on lateral flow and snowpack evolution might be needed to constrain model representations of surface hydrology and snow depth.

I. Introduction

In Arctic tundra, the thickness of the soil layer that reaches above 0°C temperatures, defined as the active layer thickness (ALT), largely determines the volume of carbon stores available for decomposition. Predicting ALT is therefore critical when characterizing potential climate feedbacks due to greenhouse gas release into the atmosphere from decomposition of organic soil carbon (McGuire et al, 2009; Koven et al., 2011; Schneider von Deimling et al., 2012). Current long-term predictions of ALT generally use large-scale Earth System Models (ESMs) with simplified representations of the hydrothermal processes, and are thus producing results with significant uncertainty (Schaefer et al., 2009; Slater & Lawrence, 2013; Koven et al., 2014). The freeze-thaw dynamics that determine the ALT function on a vertical scale of centimeters and vary horizontally on a scale of meters across the characteristic microtopography of polygonal tundra (Painter et al., 2013). Freeze-thaw dynamics are also strongly controlled by local inundation state (Muster et al., 2012), which can vary over a horizontal extent of meters to hundreds of meters. These local-to-intermediate scale processes are under-resolved or completely missing in ESMs. ~~Therefore, improved fine-scale simulation capabilities can inform the representation of soil thermal processes in regional to global scale models.~~ Improved fine-scale simulation capabilities can inform the representation of soil thermal processes in regional to global scale models by identifying appropriate representations of key processes governing ALT, and by providing calibrated model parameterization.

Previous efforts have been made to characterize ALT using field, lab, and numerical experiments (e.g. Osterkamp and Romanovsky, 1996; Romanovsky and Osterkamp, 1997). Site-specific properties of Arctic soils, such as porosity, bulk thermal conductivity, and water retention characteristics have been measured in lab settings from samples taken in the field (Hinzman et al., 1991; Letts et al., ~~2010~~2000). Those field and lab measured properties were then used in ESMs in order to predict future ALT and permafrost conditions (Beringer et al., 2001; Lawrence and Slater, 2008; Subin et al., 2013). However, such regional and global scale projections are difficult to constrain by measurements of soil properties made at vastly smaller scales of observation. This scale-gap between the governing fine-scale physical processes and large-scale simulations impedes direct model validation against measurements, which has motivated development of fine to intermediate- scale hydrothermal models (e.g. Hinzman et al., 1998; Hansson et al., 2004; Daanen et al., 2007; Mckenzie et al., 2007; Painter 2011; Karra et al. 2014; Endrizzi et al., 2014; Yi et al., 2014) for a review see Kurylyk and Watanabe (2013). Numerical experiments using high-resolution coupled hydrothermal models, which are calibrated against fine-scale measurements, can play a fundamental role in understanding the governing physical processes of ALT formation.

Simulating thermal hydrology in polygonal tundra systems is a challenging endeavor that requires simultaneous representation of multiple physical processes including phase change and highly nonlinear constitutive relationships (e.g. Painter, 2011). Soil thermal conductivity alone depends on volumetric water content,

mineral composition, porosity, density, and temperature (Farouki, 1981). In soils experiencing freeze-thaw cycles, the phase of water also affects bulk thermal conduction (e.g. Johansen, 1977; Peters-Lidard et al., 1997). Latent heat of fusion and evaporation impart further control on the propagation of the freezing front and therefore thermal conduction. Thermally driven vapor transport can slowly change ice content and thus thermal conduction in partially and fully frozen soils (Grimm and Painter, 2009; Karra et al., 2014). Characterizing subsurface properties for modeling is further complicated due to variability in microtopography and cryoturbated soil that create a heterogeneous surface and subsurface in polygonal tundra systems. In addition, coupling of the soil to the atmosphere involves a balance among multiple energy transfer processes, which occur across interfaces of snow, water, ice and exposed ground. All of the above attributes describing soil structure, surface energy balances, and processes of phase change result in a tightly coupled hydrothermal system. Therefore, numerical experiments using high-fidelity representations of fine-scale processes require calibrated parameters that are able to effectively link dependent processes.

Despite the model gains of calibrating thermal properties (Romanovsky and Osterkamp, 1997; Nicolsky et al., 2009), relatively few hydrothermal modeling studies of Arctic systems have documented calibration procedures, with the noted exception of Tang and Zhuang, (2011) and Jiang et al., (2012). Additionally, correct model structure representation, capable of representing the system based on known physical relationships while using plausible model parameters, is typically not

known *a priori*. Calibration of a model with an inadequate model structure may result in over-fitting and unreliable forward simulations that incorrectly predict system behavior based on faulty processes representation (e.g. Beven, 2005; Gupta et al., 2012). Therefore, when dealing with a coupled system of complex processes, it is imperative that the conceptual model is refined during the calibration process to increase model structure adequacy (Gupta et al., 2012).

Iterative modeling approaches that use repeated model runs with different combinations of parameters, governing mechanisms, or process representation can help fundamental system understanding (Clark et al., 2008; Kavetski and Fenicia, 2011; Fenicia et al., 2011; Larsen et al., 2014). Here we use an iterative procedure that integrates finely resolved models with field observations and measurements to develop a process-rich model with physical mechanisms and parameters consistent with measurements from the DOE Office of Science Next Generation Ecosystem Experiment (NGEE-Arctic) site Barrow Environmental Observatory (BEO), Barrow, Alaska (Figure 1). The iterative process of using field observations to inform model development and subsequent simulations to inform new data needs is referred to here as the model-observation/experiment or ModEx cycle (Figure 2). Clearly, there is no unique way to approach iterative modeling procedures (Larsen et al. 2014), which is intrinsically subjective and highly dependent on expert knowledge. Well-documented examples of successful applications of model refinement are thus invaluable for building the required experience base. We use repeated calibration of model parameters against site-specific field measurements and iterative model

adjustments of the model structure to reduce mismatch between model predictions and measurements in order to attain a viable model of thermal hydrological conditions.

In this paper we summarize our ModEx experience involving the detailed use of subsurface temperature and snow cover field data to ~~calibrate~~ develop and test process-rich simulations of ALT dynamics, such that observational data and necessary physical dynamics are incorporated into the model. In order to calibrate and refine model structure in a tractable fashion, the model development first focuses on a series of subsurface-only calibrations in section 3 before moving onto a series of coupled surface energy balance and subsurface calibrations in section 4. The end result is a set of calibrated thermal and hydrological parameters for moss, peat, and mineral soil layers, along with a consistent model structure, employed for various microtopographic positions characteristic of polygonal tundra. We demonstrate how the detailed calibration and model development effort informs understanding of the key processes that define the ALT in polygonal ground. We further complete the ModEx cycle by discussing how future data needs can reduce system uncertainty and refine our understanding of process behavior.

II. Methods

2.1 ModEx Process Applied to Thermal Hydrology Processes in Permafrost

Our variant of the ModEx approach is shown schematically in Figure 2. Starting with site identification and characterization, field observations and measurements

begin to form the modeling activity by providing model parameter inputs and targets for the model calibration process. Standard model calibration – denoted by the inner loop– aims to match simulations to field measurements by varying parameters while keeping the model structure fixed. Here the ModEx procedure moves beyond the standard calibration by assuming the model itself is uncertain, but can be further constrained through successive comparison to observation (outer loop in Figure 2~~unknown, but can be refined through successive comparison to observation (outer loop in Figure 2)~~. These improved model runs then inform the observation process by specifying the data needs, either through further calibration or through informal numerical experimentation~~ or through more formal data worth exercises~~. Such model refinement is not a unique process, and can be achieved through multiple avenues. For example, flexible modeling approaches have been used in understand structural errors by combining functional aspects of several models (Clark et al., 2008; Kavetski and Fenicia, 2011; Fenicia et al., 2011). We implement ModEx model refinement by ~~focusing~~ evaluating~~on~~ the plausibility of calibrated parameters in addition to the mismatch between field measurements and simulated responses.

The calibration process uses a multi-dimensional response surface to evaluate the plausibility of parameters and the degree of mismatch between simulated results and observed data. Sets of parameters values are mapped to the response surface with the respective mismatch between simulated results and field

observations/measurements, quantified by the root-mean-squared error (RMSE), which determines the shape of the responses surface. RMSE is given by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{T}_i(q) - T_i\right)^2} \quad (1)$$

where θ is a vector comprised of a combination of parameter values, $\hat{T}_i(\theta)$ is the ith simulated temperature given θ, and T_i is the ith calibration measured temperature target, and N is the number of calibration targets. Simulations with a poor fit to data have high RMSE and a corresponding high value on the response surface. Conversely simulations with a good fit to data have a low RMSE and therefore a low value on the response surface and may constitute a minimum in the response surface. A minimum in the response surface indicates that a possible calibration has been achieved. However, in the case of a complex model with high dimensionality, multiple local minima may exist, which ~~results~~ causes in gradient-based calibrations to find~~ing many solutions to the problem~~non-unique solutions (Beven, 2006). Model structure error can also cause the response surface to slope to a parameter boundary indicating that over-fitting is necessary to calibrate to observed data (Beven, 2005). Therefore, it is important to extend calibration boundaries beyond the acceptable parameter range to allow the optimization algorithm to travel into the infeasible range when the response surface dictates an implausible combination of parameter values, indicating an inadequate model. ~~It is important to extend calibration boundaries beyond the acceptable parameter range in order to both diagnose model inadequacy and avoid boundary effects caused by automated calibration algorithms.~~ By altering the model itself, and not just model parameters

Formatted: Normal

the ModEx process can work to reduce model structure error and reshape the response surface such that the simulated system matches the observed data and calibrated parameters are realistic.

Formatted: Font color: Blue

The ModEx process is facilitated by two software components. First, for calibrating a given model to determine an optimal match to the measurements we use PEST (Doherty, 2004), which implements the Levenberg-Marquardt algorithm (Marquardt, 1963). This method uses gradient descent to determine (from a high-dimensional space of calibration parameters) a set of parameters that (in a local sense) minimize the forward model's error in predicting observed data. Second, the ModEx process requires iterative exchange, comparison, and addition of process models, which is greatly facilitated by a dynamically configured model with many process options. Therefore a framework that manages complexity and allows for rapid development of new physical representations is critical. To this end, we have implemented the Advanced Terrestrial Simulator (ATS), version 0.83, as a collection of physics modules managed by the Arcos multiphysics framework (Coon et al. 2015b). At run-time, Arcos dynamically forms a dependency graph where each variable identifies its data requirements, allowing the automation of model evaluation. Process kernels (i.e. a single PDE, such as mass balance) are coupled to form complex systems of equations in which each term or component can easily be replaced. The ease of swapping and adding processes makes model verification and evaluation more tractable, and facilitates the ModEx process by allowing the model structure to be easily changed and extended.

2.2 Site Description and Initial Conceptual Model Set-up

The lowland, cold continuous permafrost tundra at BEO was established as the end-member of the NGEE-Arctic sites, which follow a bioclimatic gradient that extends to the warm discontinuous permafrost, shrub tundra environment of the Seward Peninsula. The site supports the NGEE-Arctic goal to improve climate model predictions through advanced understanding of coupled processes in Arctic terrestrial ecosystems. NGEE-Arctic scientists are collecting multiscale in-situ field measurements and remote sensing observations of polygonal tundra. A range of polygon types including low center polygons, which are surrounded by rims and, in some areas shallow troughs, and high center polygons with deep troughs as a result of ice wedge degradation. The focus of the model development chronicled here is NGEE-Arctic site "Area C" (Figure 1), which is characterized by~50 cm deep troughs, rims and shallow low centers. The site was chosen because it serves as a representative state that polygonal tundra may develop into as permafrost degrades. Three one-dimensional (1D) model domains represent the main ice-wedge polygon sub-features: center, rim, and trough. Each domain includes a unique model structure and parameterization (Figure 1 & 3). Nine soil temperature sensors (0.1 to 1.5m depth) from three soil profiles representing center, rim, and trough, respectively, were used to compare simulated to measured soil temperatures (http://lapland.gi.alaska.edu/vdv/vdv_historical.php?station_id=20&page_id=-1&direct=1). The shallowest soil temperature sensor (2cm depth), located just

under a layer of green moss, provided the subsurface model with an upper boundary condition. Each column had unique near-surface soil temperature forcing, measurements for calibration and assigned peat layer thicknesses typical of the micro-topographical features. The center-, rim- and trough- columns had an organic peat layer of 10, 6 and 14 cm respectively. The underlying mineral soil was a silty loam to a total depth of 50 m. A far field bottom boundary condition was held constant at -6°C to represent the average deep permafrost temperature in the North Slope of Alaska (Romanovsky, et al., 2010)~~The bottom boundary condition was held constant at a temperature of -6 C°~~. All columns were initialized by first freezing the entire column from the bottom with a no flux upper boundary condition and then spun-up to a cyclical steady state using a "decadal average" year of daily values looped for 20 simulation years. The decadal average year was made by averaging the daily mean temperature from 10/1/1998 to 9/30/2009 at Barrow, AK for each day of the year to produce forcing data that represented seasonal trends. Each calibration parameter combination was then simulated for an additional year using the same decadal average year before the in-situ soil temperature forcing data at 2cm depth was applied.

2.3 Model Description

The ATS solves water and energy flow in variably saturated soils at temperatures above and below freezing using the conservation equations described by Karra et al. (2014) (see also Painter, 2011; Coon et al., 2015a). Liquid and ice partitioning is represented by the model of Painter and Karra (2014). In this model liquid water

can coexist with ice below 0°C, as is well known~~observed~~ (e.g. Miller, 1980; Williams and Smith, 1991)~~{Williams, 1991 #121}{Williams, 1991 #121}(Watanabe and Wake, 2009)~~, which occurs due to soil surface forces and pore geometry. ~~is an important process to represent in soils with rapid freeze thaw cycles in order to prevent unrealistic rapid cooling of the subsurface (Romanovsky and Osterkamp, 2000; Nicolsky et al., 2007).~~ Ice/water partitioning is related to the soil water characteristic curve under unfrozen conditions. Thus, soil moisture characteristic curve parameters directly contribute to thermal conduction regimes when the soil is saturated and frozen. Two variations of a three-phase thermal conductivity model (Painter 2011), both an extension of Johansen (1977), were used to relate bulk thermal conductivity to ice and liquid contents. The three-phase thermal conductivity model is described in detail in Appendix A. The first thermal conductivity model variant is a simplification of the Johansen method and is referred to as the Bulk Phase Component model (BPC). The BPC model has porosity and the bulk-phase unfrozen saturated thermal conductivity ($K_{sat,uf}$) and bulk-phase dry thermal conductivity (K_{dry}) as input parameters to be calibrated (equation A-3 in Appendix A). The third bulk-phase component, saturated frozen thermal conductivity ($K_{sat,f}$) (equation A-3) is then calculated based off an empirical relationship with $K_{sat,uf}$ shown by equation A-8 in Appendix A. The second option for thermal conductivity is denoted the Material Component (MC) model. The MC model has porosity and the solid material thermal conductivity K_{soil} as input parameters; $K_{\mathrm{sat,uf}}$ and K_{dry} are then calculated using functional relationships shown in equation A-6 and A-11, respectively. Material components ice, water, and gas are

Formatted: Font: Not Bold, Not Italic

Formatted: Font: Not Bold, Not Italic

fixed material thermal conductivities in the MC model. Switching from the BPC model to the MC model reduces the dimensionality of parameter space by one. Perhaps more importantly, the MC model calculates all bulk-phase components as a function of soil porosity; thus, porosity is more correlated to thermal conductivity in the MC model as compared to the BPC model.

2.4 Parameter starting values and ranges from literature

Parameter value ranges for moss, peat, and mineral soils of Arctic tundra systems were drawn from literature and field observations at the NGEE-Arctic site (NGEE-Arctic data portal, http://ngee-arctic.ornl.gov, see references in appendix C). Estimates of reasonable calibration ranges are listed in Table 1. Depending on the thermal model being calibrated, seven to eight parameters for both peat and mineral soil were calibrated creating a 14-16 dimensional parameter space. Based on the literature and assigning greater weight to study sites with characteristics and proximity to Barrow, AK, a probable parameter guess was selected as one starting point of the calibration process, along with seven additional starting calibration parameter sets located near the boundary of parameter space. Together the eight starting calibration parameter sets determined the dependence of calibration results on starting location (i.e. the degree of non-uniqueness in the calibration results).

III. Subsurface ModEx Results

3.1 ModEx Applied to the Subsurface System

Our experience with the ModEx cycle applied to the coupled subsurface hydrothermal system at the BEO is shown in process flow form in Figure 4. In this cycle the ATS model only included subsurface processes, and the shallowest measurement of temperature (2cm depth) was used as a time-dependent upper boundary condition to force the model. Measurements at deeper locations (from 0.1 to 1.5m) (Figure 3) represented the calibration targets. In the initial iteration, calibration was performed using the BPC model for thermal conductivity and assumed full saturation of the soil column. That calibration resulted in parameters being out of range. In the second iteration, the thermal conductivity model was changed to an alternative model (the MC model), which resulted in improved parameter values but inferior match to measured soil temperatures. In the final iteration, surface pressure was calibrated at the borehole locations, which determines liquid saturation that affects near surface thermal conductivity. The iteration to calibrate surface pressure resulted in a calibration that was judged to be adequate for continuation of a coupled surface energy balance-subsurface calibration and model development (see section 4). Details of the subsurface calibration and model development are discussed in the remainder of this section.

3.2 Subsurface BPC vs. MC Thermal model

The first subsurface calibration attempt used the BPC model (Figure 4) and resulted in unrealistic parameters sets. The response surface of the center and rim columns resulted in calibrated peat porosities to move to the lower parameter boundary (Figure 5). With a few exceptions, the thermal conductivities for peat in the center,

rim, and trough calibrated outside the acceptable parameter range to the lower boundary for peat. The first calibration iteration produced unrealistic parameter values and indicated that the BPC model is not an adequate calibration tool for subsurface hydrothermal modeling.

In the second iteration of our model/data integration cycle, subsurface thermal conductivity was simulated using the MC model instead of the BPC model, which reshaped the calibration response surface such that calibrated porosities spread out across parameter space and away from the parameter boundary. Calibrating with the MC model generally kept the porosity parameters within the acceptable range and improved the thermal conductivity parameters, however, RMSE increased for all columns (Table 2). Yet, the MC model was selected for the remainder of the paper because calibrated parameters were reasonable.

Formatted: Font: Not Italic, No underline

Formatted: No underline

3.3 Simultaneous Calibration of Center, Rim, and Trough.

Up scaled parameters for larger scale models were calibrated by coupling all three columns to find a single set of peat and mineral soil hydrothermal parameters. The calibration was coupled by combining objective function results from each microtopographical feature in the PEST Levenberg-Marquardt algorithm to inform the next parameter update that is then applied to all 1D columns. The initial application of the coupled calibration resulted in unrealistic parameter values and motivated a reformulation of the conceptual model to include near-surface

365 unsaturated conditions necessary for center and trough simulations. The saturated
366 condition response surface decreased the K_e for the peat layer, and maintained or
367 increased heat conduction for mineral soil. Peat porosity and peat $K_{sat,uf}$ calibrated
368 to the lower calibration boundary of 0.59 and 0.33 [W/m K] respectively and
369 mineral porosity calibrated to a higher value (0.65) than the peat porosity, while the
370 mineral $K_{sat,uf}$ calibrated to 1.04 [W/m K]. An unsaturated near surface could
371 conversely result in a reduced thermal conductivity for the peat layer while
372 maintaining thermal conduction for the mineral soil layer.
373
374 **3.4 Variably Saturated versus Unsaturated Soils.**
375 The fourth iteration of the ModEx cycle allowed the surface pressure to be a
376 calibration parameter for the center and trough columns, which were previously
377 assumed fully saturated for the duration of the year. A surface pressure less than
378 atmospheric results in an unsaturated condition at the top of the soil column, and
379 introduces air with low thermal conduction, creating a gradient of increasing K_e
380 with depth. The surface pressure in the rim, which did not manifest the issues
381 described above, was still fixed at 25% gas saturation. It is important to note that
382 calibrating a top pressure for this set of subsurface calibrations does not allow the
383 near surface saturation to vary throughout the year and therefore, the saturation
384 state is only a function of pressure and ice content. Figure 6 illustrates how K_e of
385 peat decreases with lower surface pressure. Decreasing surface pressure results in
386 decreased K_e, but the effect is especially large during the winter. Ice has a large

thermal conductivity compared to either water or gas; any variation in the amount of ice in the domain will cause a large change in K_e.

The eight calibration starting locations for the uncoupled column calibration were then re-tested for the center and trough by calibrating surface pressures (Figure 4). Here we only tested unsaturated conditions using the MC thermal model rather then to posthumously retesting prior model structural decisions, as the MC model was thought to be more physically accurate. The new conceptual model with unsaturated conditions at the soil surface became the second model refinement, which resulted in a reshaped parameter response surface. More calibrated center porosity values were within the acceptable parameter range when surface pressures were calibrated, but more trough peat porosities calibrated to the upper peat boundary. Both the center and trough had more calibrated $K_{dry,\ material}$ within the realistic range. The increase in calibrations resulting in porosities outside their acceptable range for the trough may be indicative of the trough being more saturated than the center, or being fully saturated. However, unsaturated conditions reduced the RMSE for both the center and trough indicating a better model fit (Table 2). The increased model fit with more realistic parameters suggests that it is necessary to capture characteristic saturation states of the dominant topographical features (center, rim, and trough) to constrain model calibration. Furthermore, the single coupled center-rim-trough calibration, where surface pressures were calibrated, also resulted in realistic parameters with surface pressures at 95440.9 and 97638.2 Pa for the center and trough respectively (Table

5). Moreover, the revised coupled calibration found a low RMSE of 0.554 °C and the temperature time-series results fit measured data near the point of the active layer depth (Figure 7).

IV. Coupled Surface/Subsurface Model

4.1 Surface Methods

After the calibration of subsurface thermal properties, a 2 cm moss layer was added to each of the three columns and a surface energy balance model was used to calibrate both the thermal properties of the moss layer and parameter values for the surface energy balance in a second set of ModEx iterations (Figure 8). Parameters from the subsurface calibration were used in the coupled snow-surface energy balance-subsurface simulation. The ranges of hydrothermal parameters for moss are listed in Table 1. The surface energy balance, described in detail in appendix B, is implicitly coupled with subsurface thermal hydrology and is based on the work of Hinzman et al., (1998) and Ling and Zhang (2004). Simulated snow deformation and snow density changes described by equation B-6 and B-7 in Appendix B are applied on a single layer~~ed of~~ snowpack. The center, rim, trough columns had unique maximum head boundary conditions of 8, 0.7, and 15cm respectively, were water spills off each column at or above the specified head heights. The maximum head boundary conditions were selected according to relative elevation differences observed in polygonal tundra.

For the surface energy balance calibration each column was spun-up over a 10-year loop using decadally averaged air temperature along with shortwave radiation, relative humidity, and windspeed data from 10/1/1998 to 9/30/2009 at Barrow, AK, where meteorological data from each day in the ten years was averaged together. After spin-up, daily meteorological data from 2010-2013 were used to drive the model. This forcing data was compiled from several sources; the incoming solar radiation is from the Atmospheric Radiation Measurement (ARM) Climate Research Facility (ARM, 1993; 1996); rainfall and snowfall is from Barrow Airport (Station GHND:USW00027502 National Weather Service, National Atmospheric and Oceanic Administration); air temperature, relative humidity and wind speed are from individual research projects at the BEO (Liljedahl et al. 2011, Zona et al. 2014); and landscape-averaged end-of-winter snow depth from the Circumpolar Active Layer Monitoring (CLAM) Program (Shiklomanov et al., 2012). Daily rain and snowfall were adjusted for undercatch according to Yang et al. (1998). A second adjustment was applied to the snowfall where the average ratio between the 1997-2006 CALM observations and the undercatch-adjusted NWS snow accumulation was applied to respective daily precipitation events. The simulation results from 2013 were then compared with measured subsurface temperature data, at a 2cm depth below the moss layer. The runtime increased when including the surface energy balance component model such that automated calibration algorithms could no longer be employed. Manual calibration was used with 2 cm soil temperature borehole measurements and observed ALT, as calibration targets.

4.2 ModEx Applied to the Coupled Surface Energy Balance System

The second set of ModEx cycle iterations is presented in Figure 8 in process flow form. The focus of the second set of ModEx cycles is process identification and calibration of the moss layer and surface energy balance parameters. The first iteration of the cycle coupled the surface energy balance model and 2cm moss layer to the previously calibrated and refined subsurface model. The initial iteration matched surface temperatures well in all three columns, however soil temperatures were generally under simulated for center and trough columns, especially during winter (Figure 9). The second iteration added a microtopography-informed snow depth from measurements between utm coordinates: Northing 7910330-7910350, Easting 585900-585930, which encompasses the borehole temperature locations. Center and trough near-surface winter temperatures substantially improved, which also resulted in late summer ALT to be in or near the observed ALT range. However, near-surface winter rim temperatures were colder than measured because microtopography-informed snow distribution produces less snow on rims and results in less snow cover insulation. The third iteration of the ModEx cycle added a depth hoar representation in the snowpack, which resulted in a better representation of winter rim soil temperatures and caused the rim ALT to be within the range of observed ALT. In the final ModEx iteration hydrothermal properties of moss and surface energy balance parameters were hand calibrated within the plausible range of parameters space, which resulted in only slight improvements of near surface temperature simulations. Details of how each iteration of the ModEx

cycle (for the coupled surface energy balance – subsurface model) informed both model development and future data needs are presented below.

4.3 Importance of Surface energy balance governing saturation time series

Forcing the subsurface thermal propagation through a surface energy balance in the second set of ModEx cycles attempts to capture variable surface thermal conductivities due to changing surface saturation states as pulses of precipitation enter the subsurface and subsequently dry from evaporation. Modeling studies that do not explicitly model surface energy balance processes may not adequately capture near-surface saturation states and have reported the greatest error during the summer when highly variable soil moisture states occur (Romanovsky and Osterkamp, 1997; Jiang et al., 2012). It is known that soil moisture influences soil temperature in addition to meteorological controls, by governing the amount of latent heat of fusion necessary to freeze/thaw and evaporate water from soils (Johansen, 1977; Farouki, 1981; Peters-Lidard et al., 1998; Subin et al., 2013). Consequently, the timing of the precipitation pulses and subsequent drying may have a significant impact on ALT because the highly variable saturation states coincide with summer soil warming. Therefore, the second set of ModEx cycles starts with a more detailed representation of transient soil moisture conditions, which is the third major model refinement. Simulation results showed that it is important to capture the freeze-up timing with the highly variable fall saturation state in order to set up near surface ice content and thermal conductivity during winter (Figure 10, plot A). Properly representing the freeze-up with transient soil

moisture is especially important giving that winter has the largest range of possible thermal conductivity values (Figure 6) and therefore is highly variable from year to year.

Simulating the surface energy balance for each column resulted in varied model fits to the measured 2cm soil temperature time series. For example, the simulated center and trough 2cm soil temperature during the summer is consistently lower ~~then~~ than the measured 2cm temperature (Figure 9, center and trough plots), especially for the early summer, which in turn lowers the simulated soil temperature at depth. However, simulated 2cm deep soil temperatures for the rim matched measured soil temperatures. The ability for the model to match measured summer surface temperatures for the rim versus the center and trough is most likely attributed to either the spatial differences and local microtopography of the three columns and/or the surface saturation state. The rim is higher and therefore drier than the center and trough columns (Figure 3). To mimic microtopographical differences in the three columns, unique maximum ponded water depths were assigned to each column, the rim had a negligible max ponded depth with effectively no standing water from snow melt compared to the center and trough columns. Unfortunately, limitations to our surrogate 1-D model exist and inherently contribute to model structural error. For example, the largest deviation of surface temperature for the trough occurred during the fall as the temperature dropped below freezing. The measured surface temperature at 2cm depth had a longer duration of the zero curtain, where soil temperatures are at 0°C as water freezes,

compared to the simulated surface temperature (Figure 9). One possible explanation for this difference is that there is greater soil moisture in the trough than was simulated, as added soil moisture will extend the time to freeze a block of soil. A possible reason for the underestimated soil moisture is that the 1-D surrogate model neglected lateral surface- and subsurface flow that could be flowing on to the column, especially for troughs that are connected to an extensive trough-network. Monitoring of lateral flow in polygonal tundra systems could help to constrain the conceptual model needed to understand soil moisture dynamics.

4.4 Snow Model Refinement

The largest gains from calibrating the surface energy balance portion of the model came from the fourth model refinement, which resulted from two additional ModEx iterations 1) updating the conceptual and numerical model to add snow depth variation informed by microtopography and 2) include a depth hoar representation in the snowpack model. The snowpack at Barrow, AK is scoured relatively flat due to strong winds (Benson and Sturm, 1993; Zhang et al., 1996) resulting in deeper snow in depressions such as troughs and low-centers. To match measured snow depths of the three topographical features (Table 3), snowfall was increased for the center and trough columns by 30% (3.6cm) and 82.5% (9.9 cm), respectively, and reduced for the rim to 87% (10.4cm) of the total adjusted snowfall (12cm) for the snow year of 2012-2013. Although manually distributing snow does not fully capture snowpack dynamics, especially year-to-year snowpack variation, simulated near surface (2 cm) winter temperature more accurately matched the measured

546 temperatures (Figure 9, center and trough plots). Summer ALT increased for both
547 the center and trough, which improved the model prediction to be within the
548 observed ALT range for the trough and closer to the observed ALT range for the
549 center column (Table 4). Conversely, the decreased snow depth over the rim cooled
550 the winter surface soil temperature below the measured soil temperatures.
551 Including a depth hoar layer in the model counteracted the reduced insulation of a
552 shallower snowpack on the rim. The combination of reduced snow depth and depth
553 hoar representation on the rim translated to a slightly shallower ALT, resulting in
554 the rim ALT to be within the observed ALT range.
555
556 Without snow re-distribution or depth hoar representation the snowpack evolved
557 to a density of 410 to 440 kg/m^3 by mid May and early June as determined from
558 equation B-26. At first, this seemed reasonable because the surface of tundra snow
559 forms a wind slab layer due to the wind scouring affect with densities between 400
560 – 500 kg/m^3 (Benson and Sturm, 1993; Dominé et al., 2002). Having a snowpack
561 surface with high densities is required to accurately capture snow surface albedo.
562 However, underneath the wind slab layer, a hoar layer forms during the winter with
563 a density between 100-250 kg/m^3, (Benson and Sturm, 1993; Zhang et al., 1996;
564 Zhang, 2005), which reduces the thermal conductivity of the snowpack. The single
565 layer snow model did not include the formation of a depth hoar layer and would
566 overestimate the thermal conduction of the snowpack and therefore, increase
567 winter cooling of the ground surface. The iterative ModEx process however,
568 encouraged us to formulate a way of both representing snowpack top densities in

569 order to properly simulate surface albedo, and capture a depth hoar layer to account
570 for lower snowpack thermal conduction. The new formulation, similar to the snow
571 classes used by Schaefer et al., (2009) and Sturm et al., (1995), employed in the
572 model runs plotted in Figure 9, calculates a new thermal conduction by assuming a
573 depth hoar layer forms for 15% of the snowpack with a calibrated density. Then a
574 harmonic mean snow density is taken between the depth hoar layer and rest of the
575 snowpack in order to calculate an adjusted thermal conductivity of the snowpack.
576 Because this process applies only to calculating the snowpack thermal conduction,
577 the simulation of snow albedo is unaffected. Center and Rim depth hoar densities
578 calibrated to 110 kg/m^3 and the trough depth hoar density calibrated to 190 kg/m^3.
579 The addition of the depth hoar also reduced end of winter (May 2nd) snowpack
580 densities from above 400 kg/m^3 to between 320 to 370 kg/m^3 (Table 3), which is
581 closer to the measured end-of-winter average snowpack density of 326 kg/m^3.
582
583 Adjusting the snow accumulation due to topographically informed snow
584 distribution and including a depth hoar representation increased the insulative
585 effect of the snowpack and had a clear impact on winter near surface temperatures
586 (Figure 9). In addition snow distribution and depth hoar representation improved
587 summertime ALT predictions (Table 4). Summertime changes in ALT due to winter
588 conditions highlights a memory trait of the system and the necessity to capture
589 dominant winter processes in order to simulate transient thermal conditions in
590 physically based models. Research by Hinkel and Hurd (2006) showed that large
591 snow drifts cause long term deepening of the ALT, due in part from the additional

592 insulation for the snow and the loss of cold thermal propagation into the subsurface.
593 Timing of snowpack accumulation and thickness has also been shown to govern
594 permafrost formation (Zhang, 2005). However at the scale of microtopographical
595 relief, where trough to rim vertical relief changes by 40cm within a horizontal
596 distance of a meter, questions regarding how snow thickness and associated melt
597 water inputs affect ALT formation remain. Results for this work show that
598 topographically informed snow distribution will change the spring and early
599 summer surface saturation state (Figure 10, plot D) due to distributed snow water
600 equivalence amounts (Table 3). The change in early summer surface saturation
601 state then affects the thermal conduction for early summer as well as adding greater
602 water mass that then requires a greater amount of energy to heat up (Hinkel and
603 Hurd, 2006). Moreover, studies have found that the depth hoar layer can be as thick
604 as 50% of the snowpack height in artic conditions (Sturm et al., 1995; Schaefer et al.,
605 2009). However, due to continuous wind slab and depth hoar formation significant
606 snowpack heterogeneities develop within and across topographical features (Sturm
607 and Benson, 2004; Sturm et al., 2004). Therefore, spatially distributed snow depth
608 measurements and snowpack density profiles that characterize local snowpack
609 variability and over microtopographical features can help constrain both modeled
610 snowpack thermal conduction representation, and surface water inputs.
611
612 **4.5 Surface Energy Balance Calibration**
613 In the final ModEx iteration and model refinement, attempts to increase the
614 simulated summer surface (2 cm) temperature were made (Figure 8). Special

615 attention was paid to the early summer wet conditions found in the center and
616 trough for the Julian dates between 150 and 200 (Figure 10, plots B and D), where
617 the biggest error in surface temperatures is found (Figure 9 center and rim plots). It
618 was thought that by calibrating parameters which control the amount of energy
619 entering the subsurface under wet conditions, such as the albedo of standing water
620 (see Appendix B for details), the surface temperature of the center and trough,
621 which are wet, will increase without affecting the relatively dry rim surface
622 temperature. However, variables specific to the surface energy balance and moss
623 properties had little effect of simulated soil temperature during the snow free
624 summer. The range of accepted albedo values for tundra varied from 0.12 to 0.17
625 based on wet or dry conditions (Grenfell and Perovich, 2004), and the albedo range
626 for standing water values ranged from 0.11-0.20 for the months of May through
627 September for latitude of 70^{o} near Barrow, AK (Cogley, 1979). Only slight gains in
628 simulated surface temperature were observed by decreasing albedo of standing
629 water from 0.14 to 0.11 and tundra from 0.15 to 0.12. This iteration of the ModEx
630 cycle shows that adjusted standing water albedo and roughness length within the
631 perceived parameter range did not substantially improve model fit, which suggest
632 that the model is lacking either a necessary process representation or the
633 calibration parameter range is not correct. One possible improvement would be a
634 distributed surface albedo representation that provides a unique albedo for centers,
635 rims, and troughs. Local-scale tundra albedo measurements can inform models of
636 spatially distributed albedo conditions. Another possible explanation is how
637 atmospheric mixing coefficients such as roughness length (noted as z_0 in equation B-

12 in appendix B) could change over microtopographical features. Specific exchange coefficients for each microtopographical feature would then produce unique sensible and latent heat fluxes. For example, rim surface temperatures were well matched under current roughness lengths. But topographically protected troughs and centers could have a different roughness length, which may result in changes to latent and sensible heat exchanges and higher surface temperatures. Observations of how microtopography affect near surface wind and associated atmospheric mixing could support an improved conceptualization of sensible and latent heat exchanges.

V. Summary & Conclusions

1-D thermal hydrology models of transient saturation and frozen states combined with a surface energy balance model were used to represent active layer dynamics in polygonal tundra at the Barrow Environmental Observatory. In the coupled model, surface water was allowed to pond to a specified maximum height but any additional water was removed (spill over condition). The surface model also includes a surface energy balance model for bare, snow-, ice- or water-covered ground. The model was used in combination with borehole temperature and snowpack field measurements in an iterative model-data integration (ModEx) framework to produce calibrated model parameters and refine constitutive models and process representations. The particular variant of the ModEx approach combined calibration with iterative refinement of the model structure; parameter feasibility and model-observation mismatch were used as metrics to achieve the

objective of model development and identification of viable representations of key thermal hydrological process.~~in the model refinement process.~~

Formatted: Font: Not Bold, Not Italic

The results demonstrate the effectiveness of using borehole temperature measurements to effectively develop and refine the model structure for hydrothermal models of permafrost-affected landscapes. Results also suggest that properly constructed and calibrated 1-D models coupled to a surface energy balance may be adequate for representing thermal response at a given location provided the maximum ponded depth (spill point) is known for that location. This suggests a multiscale modeling strategy that uses overland flow models to establish the spill point (maximum ponded depth) at each surface location in conjunction with a set of thermal hydrology simulations. Further evaluations of the 1-D representations against 3-D model representations are needed~~, however~~ to identify addition process representation and the appropriate level of model complexity to capture scale dependencies of thermal dynamics.~~.~~ In addition, it is important to note that the largest discrepancy between model and field measurements occurred during early summer in the troughs and that mismatch is likely indicating model structural error with inflow of water from upstream locations and/or unique surface energy balance conditions. Observations of water fluxes such as evapotranspiration, lateral flow, and snowmelt at the sub-polygon scale would help model representation, and in particular, role of advective lateral heat transport. However, the temperature mismatch was brief and confined to the trough location, and is thus not expected to have large consequences for integrated results such as thaw depth.

The model refinement process identified the representation of thermal conductivity – specifically the dependence of bulk thermal conductivity on porosity, water content, and ice content – as a constitutive model that affects model performance. Thus, field and laboratory work to better constrain hydrothermal representation and the governing model parameters would help reduce uncertainty in model projections. Further modeling efforts that focus on uncertainty analysis and environmental parameters sensitivity to provide information which parameters govern model outcome will inform observational efforts. Similarly, snowpack properties and snow distribution were found to be important. Investigations similar to Benson and Sturm (1993), Zhang et al., (1996) and Tape et al., (2010) that better define the relationship between depth hoar, microtopography and wind slab formation would help reduce uncertainty in projections. For example, snowpack dynamics and density profile observations at the NGEE-Arctic site will inform models of how the snowpack develops and how snow will distribute across microtopography.

More generally, these results demonstrated the utility of one particular approach to merging observations and models in environmental applications. In this particular iterative approach, formal parameter estimation methods are used iteratively. Each calibration run – the inner loop in Figure 2 – minimizes mismatch between data and model with fixed model structure. The "reasonableness" or feasibility of the calibrated parameters and the RMSE are performance metrics for the calibrated model. Model structural adjustment, the outer loop in Figure 2, is initiated when

calibrated parameters fall outside reasonable bounds. Although structural model adjustments were done in an ad-hoc manner guided by experience and knowledge of the system being modeled, the resulting refinements have produced robust representation of system response. Such an approach combining structural model adjustments drawing from literature, field observations and formal calibration exercises is likely to be useful in other environmental applications.

VI. Code Availability

The Advance Terrestrial Simulator (version 0.83) is a suite of physics modules managed within the Arcos metaphysics framework that couples multiple model components at run-time. ATS, Arcos, and the host software AMANZI is developed by Los Alamos National Labs and the source code is available upon request (ecoon@lanl.gov), interested parties should see http://software.lanl.gov/ats for more information. The input data and calibration results presented here can be obtained by contacting the lead author via e-mail, or accessed at the NGEE-Arctic data portal: http://dx.doi.org/10.5440/1167674

Appendix A. Thermal conductivity model

Farouki [1981] reviewed methods for calculating the thermal conductivity of soils and concluded that a modification to a method by Johansen [1977] was superior to other models in most conditions. Peters-Lidard et al. [1998] provide a clear summary of the modified Johansen approach. Following Painter [2011], we further

730 modify the approach to a form convenient for a three-phase model and to more
731 accurately represent thermal conductivity of peat and organic-rich soils.
732
733 Thermal conductivity in unfrozen soils is often written as (Johansen [1977]; Farouki
734 [1981]; Peters-Lidard [1998])
735 $$k_e = k_{dry} + \left(k_{sat,l} - k_{dry}\right) Ke_u \qquad \text{(A-1)}$$
736 where $Ke_u\left(s_l\right)$ is the Kersten number (Kersten, 1949) for unfrozen conditions, s_l is
737 the liquid saturation index, $k_{sat,l}$ is the liquid-saturated thermal conductivity and
738 $k_{dry}$is the dry conductivity.
739
740 For soils that are frozen and with no liquid water content, the corresponding
741 equation is
742 $$k_e = k_{dry} + \left(k_{sat,i} - k_{dry}\right) Ke_f \qquad \text{(A-2)}$$
743 where $Ke_f\left(s_i\right)$ is the Kersten number for frozen conditions, s_i is the ice saturation,
744 $k_{sat,i}$ is the thermal conductivity under ice-saturated conditions.
745
746 For a general-purpose three-phase code, thermal conductivity is needed as a
747 function of both s_l and s_i. To this end, bilinear interpolation in the Kersten numbers
748 may be used [Painter, 2011]
749 $$k_e = Ke_f k_{sat,f} - Ke_u k_{sat,u} + \left(1 - Ke_f - Ke_u\right) k_{dry} \qquad \text{(A-3)}$$
750 The Kersten numbers in Eqs. A-1 and A-2 are simply ratios of partially saturated
751 thermal conductivity to fully saturated thermal conductivity. Both range from 0 for

752 dry conditions to 1 for saturated conditions and are, in general, nonlinear functions
753 of the respective saturation indices.
754
755 A variety of empirical fits have been used to relate the Kersten numbers to
756 saturation indices for ice and liquid (see, e.g. Farouki [1981] for a summary). A
757 simple power-law function is assumed here as a convenient model [Painter, 2011]

758 $$Ke_u = \left(s_l + \epsilon\right)^{\alpha_u} \quad \text{(A-4)}$$

759 $$Ke_f = \left(s_l + \epsilon\right)^{\alpha_f} \quad \text{(A-5)}$$

760 where α_u and α_f are empirical exponents and $\varepsilon << 1$ is a regularization parameter
761 that prevents, for numerical reasons, the derivative with respect to s_l or s_i from
762 becoming unbounded at 0 when α_u and α_f are less than 1.
763
764 For saturated conductivity, geometric means are often used [Johansen, 1977]

765 $$k_{sat,u} = k_s^{1-f} k_w^f \quad \text{(A-6)}$$

766 and

767 $$k_{sat,f} = k_s^{1-f} k_i^f \quad \text{(A-7)}$$

768 where κ_i, κ_w, κ_s are thermal conductivities for water ice, liquid water, and soil solids,
769 respectively. We take $\kappa_{sat,u}$ as a property of the medium which can be measured or
770 calibrated, then assume

771 $$k_{sat,f} = k_{sat.uf} \left(\frac{k_i}{k_w} \right)^f \quad \text{(A-8)}$$

772 consistent with eqs A-6 and A-7.

We denote the model specified by equations A-3, A-4, A-5 and A-8 with input parameters, $\kappa_{sat,uf}$, κ_{dry}, α_u, and α_f as the BPC model.

An alternative model, which we denote the MC model, is obtained by relating κ_{dry} and $\kappa_{sat,uf}$ to the thermal conductivities of the material components (ice, liquid, gas, and soil solids). For κ_{dry} the following empirical fit has been suggested [Johansen, 1977]

$$k_{dry} = \frac{0.135 r_b + 64.7}{r_s - 0.947 r_b} \qquad \text{(A-9)}$$

where ρ_b and ρ_s are the dry bulk and solid densities, respectively, in kg m^{-3} and κ_{dry} is in W m^{-1} K^{-1}. Using $\rho_b = \rho_s(1-\varphi)$, this equation can be placed in the form

$$k_{dry} = \frac{0.135 r_s (1 - f) + 64.7}{r_s - (1 - d) r_s (1 - f)} = \frac{0.135(1 - f) + 64.7 / r_s}{f + d(1 - f)} \qquad \text{(A-10)}$$

where d is 0.053 (unitless). Equation 9 is problematic as a general model for two reasons. First, the thermal conductivity of air should be recovered as porosity approaches unity, which is not the case in Eq. 9. Second, the thermal conductivity of the soil solids should be recovered when the porosity is zero, which is also not the case for Eq. 9. Setting porosity to 0 results in a thermal conductivity of ~3 W/m-K for soil minerals with grain density of 2700 kg/m3, which is consistent with a "typical" value [van Wijk, 1963] of 2.9 W/m-K at ρ_s = 2700 kg/m3. However, setting ρ_s to the value of a typical organic material (1.3 kg/m^3) results in ~3.5 W/m-K,

793 which is more than an order of magnitude greater than a typical value for peat (0.25
794 W/m-K).
795
796 To better represent κ_{dry} for organic-rich soils, we thus modify equation 9 to be
797 $$k_{dry} = \frac{d(1- f)k_s + k_a f}{d(1- f) + f} \qquad \text{(A-11)}$$
798
799 where κ_a is the thermal conductivity of air and κ_s is the thermal conductivity of soil
800 solids. When porosity is 0, $\kappa_{dry} = \kappa_s$ is recovered from equation A-11. When porosity
801 is 1, $\kappa_{dry} = \kappa_a$. A comparison between equation A-11 and the Johansen equivalent (eq
802 A-9) for a mineral soil (r_s = 2700 kg/m^3 in Eq. A-9 and k_s = 2.9 W/m-K in Eq. A-
803 11). The Johansen fit and our modification, Eq. A-11, have only very minor
804 differences in this case. However, for peat material (ρ_s = 1300 kg/m3 in Eq. A-9 and
805 κ_s = 0.25 W/m-K in Eq. A-10), the two models diverge. The alternative
806 parameterization of using κ_s instead of ρ_s in Eq. A-11 provides enough flexibility to
807 produce reasonable values for dry thermal conductivity for both mineral soil and
808 peat.
809
810 In summary, two thermal conductivity models are available. The BPC model uses the
811 following parameters: thermal conductivity of dry soil, saturated thermal
812 conductivity in unfrozen conditions, the exponents α_u and α_{uf}, and porosity. The MC
813 model uses the following parameters: thermal conductivity of soil solid, the
814 exponents α_u and α_{uf} , and porosity. Although each of these may be determined by

laboratory measurements on core samples, the use of such small-scale measurements at the field scale is often confounded by multiscale heterogeneity. We thus use field-scale temperature measurements to estimate the parameters.

Appendix B Snow-surface-energy-balance model

The surface energy balance model is a coupled mass and energy balance simulator used to deliver energy fluxes and any water associated with snowmelt or precipitation to the ground surface simulated by the Advanced Terrestrial Simulator (ATS). The surface energy simulator is split into two parts depending on if a snowpack is present or absent. If a snowpack is present, the surface energy balance solves for the snow surface temperature (T_s) following the methods by Hinzman et al., (1998) and Ling and Zhang (2004). Energy fluxes are then delivered through a mass conservative evolving snowpack deformation model to the surface of the ground. In addition to energy, water mass is also delivered to ground surface. The surface energy balance equation for snow is:

$$0 = (1 - a) Q_{sw,met}^{In} + Q_{lw}^{In} + Q_{lw}^{Out}(T_s) + Q_h(T_s) + Q_e(T_s) + Q_c(T_s) \quad \text{(B-1)}$$

Q_{lw}^{In} and $Q_{sw,met}^{In}$ are incoming long and shortwave radiation respectively, Q_{lw}^{Out} is out going long-wave radiation. Q_h is sensible heat, Q_e is latent heat, and Q_c is the conduction of heat from the snow surface through the snowpack to the ground surface. All energy balance components are in [W/m^2]. This method assumes the snowpack is in equilibrium with all energy fluxes going into and out of the snowpack. If no snow is present, the energy balance is calculated on the top of the surface water, bare tundra, or a gradation between the two, and the water and

energy fluxes are delivered to the subsurface portion of ATS. The ground surface energy balance equation without snow is:

$$Q_{gf} = (1 - a) Q_{sw}^{In} + Q_{lw}^{In} + Q_{lw}^{Out}(T_{gs}) + Q_h(T_{gs}) + Q_e(T_{gs}) \quad \text{(B-2)}$$

T_{gs} is the ground surface temperature and Q_{gf} is the flux of energy into the subsurface and because no snow is present, Q_c is no longer computed.

Components of the energy balance model that do not depend on the surface temperature are computed initially, Q_{lw}^{In} and $Q_{sw,met}^{In}$. Q_{lw}^{In} can be either read in from a data file or modeled based on an empirical equation for calculating the emissivity of air from Satterlund, (1979); and Fleagle & Businger, (1980):

$$Q_{lw}^{In} = e_a s T_a^4 \quad \text{(B-3)}$$

Where σ is the Stephan-Boltzmann Constant, 5.670676 x 10^{-8}[W/m^2 K^4], and T_a is the air temperature [K]. The emissivity of air (ε_a) is calculated by:

$$e_a = 1.08\left(1 - \exp^{-(0.01 e_a)^{\frac{T_a}{2016}}}\right) \quad \text{(B-4)}$$

Where e_a is the vapor pressure of air.

Q_{sw}^{In} in the surface energy balance model is the shortwave radiation absorbed by the surface, after a percentage of the total shortwave radiation from the meteorological data ($Q_{sw,met}^{In}$) has been reflected by the albedo (α) of the surface.

$$Q_{sw}^{In} = (1 - a) Q_{sw,met}^{In} \quad \text{(B-5)}$$

The albedo α in Barrow, Alaska can change spatially due to heterogeneous surface conditions and temporally due to the changing physical conditions of the surface (Grenfell and Perovich, 2004). The changing surface conditions between snow, ice, and water strongly influence incoming shortwave radiation by altering α; therefore its representation in the model plays a critical role in accurately simulating the arctic energy budget (Curry et al., 1995; Hansen and Nazarenko, 2004). Currently, there are four possible surfaces with unique α values 1) snow, 2) ice, 3) ponded water, and 4) tundra vegetation.

The α of snow is based on snow density (ρ_s) following the methods of Anderson, 1976; Ling and Zhang, 2004; and Peter ReVelle's thesis (2012) and reflects the aging process of snow deformation.

if $\rho_s <= 450$ kg/m^3

$$a = 1 - \left(0.247 + 0.16 + 110\left(\frac{r_s}{1000}\right)^4\right)^{0.5} \quad \text{(B-6)}$$

if $\rho_s > 450$ kg/m^3

$$a = 0.6 - \frac{r_s}{4600} \quad \text{(B-7)}$$

The snow deformation model is outlined in Martinec (1977).

The albedo of the four possible surfaces are listed in Table B-1.

Surface	Albedo	Range
ice †	0.44	0.27 - 0.49
water †	0.141	0.112 - 0.202
tundra †	0.135	0.12 - 0.17

† From Grenfell & Perovich 2004

æ From Cogley 1979

876 Table B-1

877 The α of ponded water is the average α of standing water at a latitude of 70º from
878 May through September. During freezing and thawing of the ground surface any
879 ponded water is subdivided into an unfrozen water fraction and a frozen water
880 fraction in ATS. The α values for this surface is then an average of water and ice α
881 values and are found to transition linearly between the two states (Grenfell and
882 Perovich, 2004) based on unfrozen water fraction. Transitional α values between
883 each type of surface can occur and are triggered when the snowpack height is less
884 then 2 cm, or the standing water height is less then 10 cm. The transition height for
885 ponded water is based on the penetration depth of shortwave radiation in ice
886 (10cm). Transitional α weighting values are calculated by:

$$Tran_{snow} = \left(\frac{Z_s}{Pen_s}\right)^2$$

887 $$Tran_{water} = \frac{Z_w}{Pen_w}\left[1 - Tran_{snow}\right]$$

$$Tran_{tundra} = \left[1 - Tran_{snow}\right] - Tran_{water} \qquad \text{(B-8)}$$

888 Where Z is the height of water or snow and Pen is the penetration depth of
889 shortwave radiation. The transitional α value is then calculated by:

890 $$a_{trans} = a_{snow}Tran_{snow} + a_{water}Tran_{water} + a_{tundra}Tran_{tundra} \qquad \text{(B-9)}$$

891 In this model, if snow is present it is always the top surface, and ponded water or
892 surface ice will always be below snow and above the tundra surface. Therefore, the

893 α value is set first by snow, if present, then by standing water and/or ice if present,
894 and finally by the tundra surface.
895
896 Once the incoming radiation components of the energy balance model are
897 computed, evaporative resistance (E_r) is then calculated by:
898 $$E_r = \frac{1}{R_{air} + R_{soil}} \quad \text{(B-10)}$$
899 where the air resistance term (R_{air}) is the inverse of the turbulent exchange of latent
900 and sensible heat (D_{eh}) and the stability function(ζ):
901 $$R_{air} = \frac{1}{D_{eh}\zeta} \quad \text{(B-11)}$$
902 $$D_{eh} = \frac{k^2 U_s}{\left(\ln\left(z_r / z_0\right)\right)^2} \quad \text{(B-12)}$$
903 κ is the von Karman Constant 0.41 [-], U_s is the wind speed at the reference height
904 (z_r) of the meteorological measurement location. z_0 is the roughness length. Due to
905 the changing conditions of the landscape at barrow, z_0 changes from 0.005 [m] for
906 wind swept snow (Wieringa and Rudel, 2002), to 0.04 [m] for polygonal tundra
907 (Weller and Holmgren, 1974; Hansen, 1993).
908
909 The stability function (ζ) accounts for both stable (ζ_{stable}) and unstable ($\zeta_{unstable}$)
910 atmospheric conditions (Price and Dunne, 1976)
911 $$\zeta_{stable} = \frac{1}{1+10R_i} \quad \text{or} \quad \zeta_{unstable} = 1 - 10R_i. \quad \text{(B-13)}$$

912 $\zeta_{unstable}$ conditions occur when the ground surface (T_s) is warmer than the air
913 temperature (T_a) causing more air to mix vertically. R_i defines atmospheric stability;
914 where R_i is positive in the stable condition and R_i is negative in an unstable condition.

915 $$R_i = \frac{g z_r \left(T_a - T_s\right)}{T_a U_s^2} \quad \text{(B-14)}$$

916 g is the acceleration due to gravity. R_{soil} [m/s] is calculated following the methods
917 used by Sakaguchi and Zeng (2009) and is only implemented during ground surface
918 evaporation when the saturation state of the upper most subsurface cell adjacent to
919 the domain surface is less than 1.

920 $$R_{soil} = \frac{L}{D} \quad \text{(B-15)}$$

921 Where D is vapor diffusion [m^2/s] calculated empirically (Moldrup et al., 2004;
922 Sakaguchi and Zeng, 2009) from the residual saturation (θ_r), saturation (θ_{sat}), and
923 the molecular diffusion coefficient of water vapor in the air (D_o), assumed to be
924 constant 2.2 X 10^{-5} [m^2/s] (Moldrup et al., 1999; Sakaguchi and Zeng, 2009).

925 $$D = D_o q_{sat}^2 \left(1 - \frac{q_r}{q_{sat}}\right)^{2+3b} \quad \text{(B-16)}$$

926 The exponent b in equation *B-16* is a Clapp and Hornberger, (1978) fitting parameter
927 for the soil water characteristic curve, assumed to be 1 for moss (Beringer et al.,
928 2001), which covers the tundra surface and is simulated as the top subsurface layer
929 for the tundra.
930 L is dry layer thickness or the length vapor must travel from the point of evaporation.

931 $$L = d_1 \frac{\exp\left[1 - \left(q_l / q_{sat}\right)^w\right] - 1}{e - 1} \quad \text{(B-17)}$$

932 Once all necessary components of the energy balance are calculated, either the snow
933 energy balance or surface energy balance is computed. The snow energy balance,
934 **eq. B-1**, is calculated if snow height (Z_s) is more than 2cm. The ground surface
935 energy balance, **eq. B-2**, is used if no snow is present. Between Z_s of 0 and 2cm, a
936 transition between the snow energy balance and the ground surface energy balance
937 is used where both surface conditions are solved. When calculating the energy
938 balance for the transitional regime, the snow energy balance assumes a Z_s of 2cm for
939 all components that depend on Z_s and an area-weighted average is used between the
940 ground surface and snow energy balance based on the actual Z_s that is equal to or
941 less than 2cm. Assuming a 2cm Z_s within the snow energy balance calculation
942 prevents unreasonable heat conduction through the snowpack (Q_c), calculated by:

943 $$Q_c = -\frac{k_s\left(T_s - T_g\right)}{Z_s} \quad \text{(B-18)}$$

944 where k_s is the effective thermal conductivity of snow [W/m K] and is calculated
945 from an empirical function of ρ_s used by Ling and Zhang, (2004), described by
946 Goodrich (1982)

947 $$k_s = 2.9 \times 10^{-6} r_s^2. \quad \text{(B-19)}$$

948 The snow and surface energy balance use the same formulation for Q_h and Q_{lw}^{Out}. Q_h
949 is:

950 $$Q_h = \rho_a C_p D_{eh} \zeta (T_a - T_s) \quad \text{(B-20)}$$

where ρ_a is the density of air 1.275 [kg/m^3], and C_p is the specific heat of air (1004 J/K kg). Q_{lw}^{Out} is:

$$Q_{lw}^{Out} = -\varepsilon_s \sigma T_s^4 \quad \text{(B-21)}$$

ε_s is the emissivity of the surface. The ε_s for snow and ice 0.98 [-], is taken from Liston and Hall, (1995), and the ε_s for tundra is 0.92 (Ling and Zhang, 2004) and for standing water is 0.979 (Robinson and Davis, 1972).

Q_e is slightly different between the snow and ground surface energy balance where the porosity (ϕ_s) of the top cell in the ground surface is included for the surface energy balance calculation.

$$Q_{e,snow} = \rho_a L_s E_r \left(0.622 \frac{e_a - e_s}{A_{pa}}\right)$$
$$Q_{e,ground\_surface} = f_s \rho_a L_e E_r \left(0.622 \frac{e_a - e_s}{A_{pa}}\right) \quad \text{(B-22)}$$

where E_r, the evaporation resistance as defined by eq. B-8 and R_{soil} is 0 in the case of snow, or condensation on the surface. L_s is the latent heat of sublimation for snow (2834000 J/kg) and L_e is the latent heat of evaporation for the ground surface (2497848 J/kg). e_s is the vapor pressure of the snow or surface, and A_{pa} is the atmospheric pressure (101.325 kPa).

Once the energy balance is calculated, then the water fluxes to the ground surface are calculated. In the case of snow, if the snow surface temperature (T_s) is greater than freezing, T_s is set to freezing and the snow surface energy balance is

recalculated with all excess energy assigned to the melting energy (Q_m), and a melting rate (Mr) [m/s] is calculated from:

$$M_r = \frac{Q_m}{r_w * H_f}, \quad \text{(B-23)}$$

where ρ_w is the density of water and H_f is the heat of fusion for melting snow 333500 [J/kg]). Condensation or sublimation of the snow surface is also calculated from Q_e, where the sublimation/condensation rate (S_r) is added to the total water flux. If T_a and $Z_s > 0$ and S_r is positive, then

$$Q_{water} = S_r + P_r$$
$$S_r = \frac{Q_e}{r_w L_s} \quad \text{(B-24)}$$

Sublimation is removed from the snowpack when S_r is positive. If only the ground surface energy balance is used then water is delivered to the ground surface as precipitation and condensation when S_r is negative. Water is evaporated from the surface/sub-surface when S_r is positive.

Snow water equivalence (SWE), Z_s, and ρ_s are tracked through the simulation of snowpack evolution and related by:

$$SWE = \frac{Z_s}{r_s} \quad \text{(B-25)}$$

Both Z_s and ρ_s are important in the snow energy balance equation for calculated Q_c and snow α, and both variables evolve as the snowpack ages through snowpack deformation simulated by (Martinec, 1977):

$$r_{settled} = r_{freshsnow}\left(SP_{age}\right)^{0.3} \quad \text{(B-26)}$$

where $\rho_{freshsnow}$ is assigned a density of 100 kg/m^3, SP_{age} is the age of the snowpack. The total snowpack density and Z_s are then calculated by a weighted average of 3 components: old settled snow, new snow accumulation, and any ice from condensation. The density of condensation is assigned 200 kg/m^3.

Appendix C. Parameter Literature Sources

Values for hydrothermal properties of moss were gathered from Hinzman et al., (1991); Letts et al., (2000); Quinton et al., (2000); Price et al., (2008); O'Donnell et al., (2009); and Zhang et al., (2010). Large-scale simulations including a moss layer were also considered and informed valid parameters ranges (Beringer et al., 2001). Peat properties were found in Hinzman et al., (1991); Hinzman et al., (1998); Letts et al., (2000); Quinton et al., (2000); Quintion et al., (2008); Nicolsky et al., 2009); Zhang et al., (2010) and the accompanying larger scale simulations (Beringer et al., 2001; Lawrence and Slater, 2008). Mineral soil properties were gathered from Hinzman et al., (1991); Hinzman et al., (1998); Beringer et al., (2001); Overduin et al., (2006); Lawrence and Slater, (2008); Nicolsky et al., (2009). van Genuchten parameters were fitted to the published soil water characteristics curves (Hinzman et al., 1991).

Acknowledgments. This work was supported by the Los Alamos National Laboratory, Laboratory Direction Research and Development project LDRD201200068DR and by the Next Generation Ecosystem Experiment (NGEE-Arctic) project. NGEE-Arctic is supported by the Office of Biological and

Environmental Research in the DOE Office of Science. We are also dearly indebted to all field personal, in particular Andy Chamberlain, William Cable, and Robert Busey, who braved freezing temperatures, polar bears, and mosquito swarms to provide the necessary field measurements to develop our models.

References

Anderson, E.A. 1976. A point energy and mass balance model of a snow cover. NOAA Tech. Rep. NWS-19.

Atmospheric Radiation Measurement (ARM) Climate Research Facility. 1996, updated hourly. Sky Radiometers on Stand for Downwelling Radiation (SKYRAD60S). 2010-01-01 to 2013-12-31, 71.323 N 156.609 W: North Slope Alaska (NSA) Central Facility, Barrow AK (C1). Compiled by V. Morris, M. Sengupta, A. Habte, I. Reda, M. Anderberg, M. Dooraghi, P. Gotseff, V. Morris, A. Andreas and M. Kutchenreiter. Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive: Oak Ridge, Tennessee, USA. Data set accessed 2014-05-19 at http://dx.doi.org/10.5439/1025281

Atmospheric Radiation Measurement (ARM) Climate Research Facility. 1993, updated hourly. Surface Meteorological Instrumentation (MET). 2010-01-01 to 2013-12-31, 71.323 N 156.609 W: North Slope Alaska (NSA) Central Facility, Barrow AK (C1). Compiled by J. Kyrouac and D. Holdridge. Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive: Oak Ridge, Tennessee, USA. Data set accessed 2014-05-19 at http://dx.doi.org/10.5439/1025220

Benson, C.S., Sturm, M. 1993. Structure and wind transport of seasonal snow on the Arctic slope of Alaska. Annals of Glaciology 18: 261-267.

Beringer, J., Lynch, A.H., Chapin III, F.S., Mack, M., Bonan, G.B. 2001. The representation of Arctic soils in the Land Surface Model: The importance of Mosses. Journal of Climate. 14: 3324-3335.

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.
~~2006.~~

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.

;

Clapp, R. B., Hornberger, G.M. 1978. Empirical equations for some soil hydraulic properties. Water Resources Research, 14(4): 601-604, doi:10.1029/WR014i004p00601.

Cogley, J.G. 1979. The albedo of water as a function of latitude. American Meteorological Society. 107(6), 775-781.

Coon, E.T., J.D. Moulton, M. Berndt, G. Manzini, R. Garimella, K. Lipnikov,and S.L. Painter. (*In Review, 2015a*), Coupled surface and subsurface hydrologic flow using mimetic finite differences. Advances in Water Resources.

Coon, E.T., J.D. Moulton, and S.L. Painter (*In Review, 2015b*) Managing Complexity in Simulations of Land Surface and Near-surface Processes. *Environmental Modelling and Software*

Curry, J.A., Schramm, J.L., Ebert, E.E. 1995. Sea ice-albedo climate feedback mechanism. Journal of Climate 8: 240-247.

Daanen, R. P., Misra, D., & Epstein, H. 2007. Active-layer hydrology in nonsorted circle ecosystems of the arctic tundra. Vadose Zone Journal, 6(4), 694-704.

Doherty, John. "PEST Model-Independent Parameter Estimation User Manual, Watermark Numerical Computing." Brisbane, Australia (2004).

Dominé, F., Cabanes, A., & Legagneux, L. 2002. Structure, microphysics, and surface area of the Arctic snowpack near Alert during the ALERT 2000 campaign. Atmospheric Environment, *36*(15), 2753-2765.

Endrizzi, S., Gruber, S., Dall'Amico, M., & Rigon, R. 2014. GEOtop 2.0: simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. Geoscientific Model Development, 7(6), 2831-2857.

Farouki, O. T. 1981. The thermal properties of soils in cold regions. Cold Regions Science and Technology, 5(1), 67-75.

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,

Formatted: Font: +Body (Cambria), 12 pt

Formatted: Indent: Left: 0 cm, Hanging: 0.79 cm

Formatted: Font: +Body (Cambria), 12 pt

Formatted: Font: +Body (Cambria), 12 pt

Formatted: Indent: Left: 0 cm, Hanging: 0.79 cm

Grenfell, T.C., Perovich, D.K. 2004. Seasonal and spatial evolution of albedo in a snow-ice-land-ocean environment. Journal of Geophysical Research 109(C1), C01001,doi: 10.1029/2003JC001866.

Grimm, R. E., & Painter, S. L. 2009. On the secular evolution of groundwater on Mars. Geophysical Research Letters, *36*(24).

Goodrich, LE. 1982. The influence of snow cover on the ground thermal regime. Canadian Geotechnical Journal. 19, 421-432.

Gupta J.V., Clark M.P., Vrugt J.A., Abramowitz G., Ye M. 2012. Towards a comprehensive assessment of model structural adequacy. Water Resources Research. 48: W08301, doi:1029/2011WR011044.

Hansen, S. V. 1993. Surface roughness lengths. ARL Technical Report U. S. Army, White Sands Missile Range, NM 88002-5501.

Hansen, J., Nazarenko, L. Soot climate forcing via snow and ice albedos. Proceedings of the National Academy of Sciences of the United States of America 101, 423–428.

Hansson, K., Šimůnek, J., Mizoguchi, M., Lundin, L. C., & Van Genuchten, M. T. 2004. Water flow and heat transport in frozen soil. Vadose Zone Journal, 3(2), 693-704.

Hinkel K.M., Hurd J.K.Jr. 2006. Permafrost destabilization and thermokarst following snow fence installation, Barrow, Alaska, U.S.A. Arctic, Antarctic, and Alpine Research. 38(4): 530-539.

Hinzman L.D., Kane D.L., Gieck R.E., Everett K.R. 1991. Hydrological and thermal properties of the active layer in the Alaskan Arctic. Cold Regions Science and Technology. 19: 95-110.

Hinzman L.D., Goering D.J., Kane D.L. 1998. A distributed thermal model for calculating soil temperature profiles and depth of thaw in permafrost regions. Journal of Geophysical Research 103(22): 28,975-28,991.

Jiang Y., Zhuang Q., O'Donnell J.A. 2012. Modeling thermal dynamics of active layer soils and near-surface permafrost using a fully coupled water and heat transport model. Journal of Geophysical Research. 117, D1110, doi:10.1029/2012JD017512.

Johansen, O. 1977. Thermal conductivity of soils (No. CRREL-TL-637). COLD REGIONS RESEARCH AND ENGINEERING LAB HANOVER NH.

Karra, S., Painter, S. L., & Lichtner, P. C. 2014. Three-phase numerical model for subsurface hydrology in permafrost-affected regions. The Cryosphere Discussions, 8(1), 149-185.

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,

Kersten, M. S. (1949). Thermal properties of soils. University of Minnesota, Institute of Technology, Engineering Experiment Station. Bulletin, (28).

Koven C.D., Ringeval B., Friedlingstein P., Ciais P., Cadule P., Khvorostyanov D., Krinner G., Tarnocai C. 2011. Permafrost carbon-climate feedbacks accelerate global warming. Proceeding o the National Academy of Sciences of the United States of America. 108(36): 14769-14774, doi:10.1073/pnas.1103910108.

Kurylyk, B. L., Watanabe, K. 2013. The mathematical representation of freezing and thawing processes in variably-saturated, non-deformable soils. Advances in Water Resources, *60*, 160-177, doi:10.1016/j.advwatres.2013.07.016.

Larsen L., Thomas C., Eppinga M. 2014. Exploratory modeling: extracting causality from complexity. EOS, 95(12) 285-292.

Lawrence D.M., Slater A.G. 2008. Incorporating organic soil into a global climate model. Climate Dynamics 30:145-160, doi:10.1007/s00382-007-0278-1.

Letts M.G., Roulet N.T., Comer N.T., Skarupa M.R., Verseghy D.L. 2000. Parameterization of peatland hydraulic properties for the Canadian land surface scheme. Atmosphere-Ocean. 38: 141-160.

Liljedahl, A.K., L.D. Hinzman, Y. Harazono, D. Zona, C. Tweedie, R.D. Hollister, R. Engstrom, and W.C. Oechel. 2011. Nonlinear controls on evapotranspiration in arctic coastal wetlands. Biogeosci., 8, 3375–3389, doi:10.5194/bg-8-3375-2011.

Ling ,F., Zhang, T. 2004. A numerical model for surface energy balance and thermal regime of the active layer and permafrost containing unfrozen water. Cold Regions Science and Technology 38, 1-15. doi:10.1016/S0165-232X(03)00057-0.

Liston, G.E., Hall, D.K. 1995. An energy balance model of lake ice evolution. Journal if Glaciology 41(138): 373-382.

Martinec, J. 1977. Expected snow loads on structures from incomplete hydrological data. Journal of Glaciology 19, 185-195.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial & Applied Mathematics, 11(2), 431-441.

McGuire, A. D., Anderson, L. G., Christensen, T. R., Dallimore, S., Guo, L., Hayes, D. J., ... & Roulet, N. 2009. Sensitivity of the carbon cycle in the Arctic to climate change. Ecological Monographs, 79(4), 523-555.

McKenzie, J. M., Voss, C. I., & Siegel, D. I. 2007. Groundwater flow with energy transport and water–ice phase change: numerical simulations, benchmarks, and application to freezing in peat bogs. Advances in water resources, 30(4), 966-983.

Miller, R.D. 1980. Freezing phenomena in soil. In: D. Hillel, editor, Application of soil physics. Academic Press, New York. p. 255–299.

Moldrup, P., Olesen, T., Yamaguchi, T., Schjønning, P., & Rolston, D. E. 1999. Modeling diffusion and reaction in soils: IX. The Buckingham-Burdine-Campbell equation for gas diffusivity in undisturbed soil. Soil Science, 164(8), 542-551.

Moldrup, P., Oleson, T., Yoshikawa, S., Komatsu, T., Rolston, D.E. 2004. Three-porosity model for predicting the gas diffusion coefficient in undisturbed soil. Soil Sci. Soc. Am. J., 68: 750-759.

Muster, S., Langer, M., Heim, B., Westermann, S., & Boike, J. (2012). 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*.

~~Nicolsky D.J., Romanovsky V.E., Alexeev V.A., Lawrence D.M. 2007. Improved modeling of permafrost dynamics in a GCM land-surface scheme. Geophysical Research Letters. 34, L08501, doi:10.1029/2007GL029525.~~

Nicolsky, D. J., Romanovsky, V. E., & Panteleev, G. G. 2009. Estimation of soil thermal properties using in-situ temperature measurements in the active layer and permafrost. Cold Regions Science and Technology, 55(1), 120-129.

Overduin P.P., Kane D.L., van Loon W.K.P. 2006. Measuring thermal conductivity in freezing and thawing soil using the soil temperature response to heating. Cold Regions Science and Technology. 45:8-22. doi:10.1016/j.coldregions.2005.12.003.

Osterkamp, T. E., & Romanovsky, V. E. 1996. Characteristics of changing permafrost temperatures in the Alaskan Arctic, USA. Arctic and Alpine research, 267-273.

Painter S.L. 2011. Three-phase numerical model of water migration in partially frozen geological media: model formulation, validation, and applications. Computational Geosciences. 15: 69-85, doi:10.1007/s10596-010-9197-z.

Painter, S. L., Moulton, J. D., & Wilson, C. J. 2013. Modeling challenges for predicting hydrologic response to degrading permafrost. Hydrogeology Journal, 1-4.

Painter, Scott L., and Satish Karra. 2014. Constitutive model for unfrozen water content in subfreezing unsaturated soils. Vadose Zone Journal 13.4 doi:10.2136/vzj2013.04.0071.

Peters-Lidard, C.D., Blackburn, E., Liang, X., Wood, E.F.: The effect of thermal conductivity parameterization on surface energy fluxes and temperatures. J. Atmos. 55, 1209–1224 (1998)

Price, A.D., Dunne T. 1976. Energy balance computations of snow melt in a sub-arctic area. Water Resources Research. 12, 686-689.

Price J.S., Elrick D.E., Strack M., Brunet N., Faux E. 2008. A method to determine unsaturated hydraulic conductivity in living and undecomposed Sphagnum moss. Soil Science Society of America Journal 72:487-491. doi:10.2136/sssaj2007.0111N.

Quinton W.L., Gray D.M., Marsh P. 2000. Subsurface drainage from hummock-covered hillslopes in the Arctic tundra. Journal of Hydrology 237: 113-125.

Quinton W.L., Hayashi M., Carey S.K., Myers T. 2008. Peat hydraulic conductivity in cold regions and its relation to pore size and geometry. Hydrological processes. 22(15): 2829-2837.

ReVelle, P. 2012. A snow model used to examine the affect of seasonal snow on an arctic environment. New Mexico Tech, Department of Earth and Environmental Science.

Robinson P.J., Davies J.A. 1972. Laboratory Determination of water surface emissivity. Journal of applied meteorology. 11: 1391-1393.

Romanovsky, V.E., and T.E. Osterkamp 1997. Thawing of the active layer on the coastal plain of the Alaskan Arctic, *Permafrost and Periglacial Processes*, 8(1), 1-22.

~~Romanovsky V., Osterkamp T. 2000. Effects of unfrozen water on heat and mass transport processes in the active layer and permafrost. Permafrost Periglacial Processes. 11: 219-239.~~

Romanovsky, V. E., Smith, S. L., & Christiansen, H. H. 2010. Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007–

2009: a synthesis. *Permafrost and Periglacial Processes, 21*(2), 106-116. DOI: 10.1002/ppp.689

Satterlund, D. R. 1979. An improved equation for estimating long-wave radiation from the atmosphere. Water Resources Research, *15*(6), 1649-1650.

Sakaguchi K., Zeng X. 2009. Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5). Journal of Geophysical Research. 114, D01107, doi:10.1029/2008JD010834.

Schaefer, K., Zhang, T., Slater, A. G., Lu, L., Etringer, A., & Baker, I. 2009. Improving simulated soil temperatures and soil freeze/thaw at high-latitude regions in the Simple Biosphere/Carnegie-Ames-Stanford Approach model. Journal of Geophysical Research: Earth Surface (2003–2012), 114(F2).

Schneider von Deimling T., Meinshausen M., Levermann A., Huber V., Frieler K., Lawrence D.M., Brovkin V. 2012. Estimating the near-surface permafrost-carbon feedback on global warming. Biogeosciences. 9: 649-665, doi:10.5194/bg-9-649-2012.

Shiklomanov, N I., F. E. Nelson, and D.A. Streletskiy (2012). The Circumpolar Active Layer Monitoring (CALM) Program: Data Collection, Management, and Dissemination Strategies. In Tenth International Conference on Permafrost Vol. 1: International Contributions, Hinkel K. M. (Ed.), The Northern Publisher, Salekhard, Russia, pp 377-382.

Sturm, M., & Benson, C. 2004. Scales of spatial heterogeneity for perennial and seasonal snow layers. Annals of Glaciology, 38(1), 253-260.

Sturm, M., Holmgren, J., & Liston, G. E. 1995. A seasonal snow cover classification system for local to global applications. Journal of Climate, 8(5), 1261-1283.

Sturm, M., Johnson, J. B., & Holmgren, J. (2004, April). Variations in the mechanical properties of arctic and subarctic snow at local (1-m) to regional (100-km) scales. In Proceedings ISSMA-2004, International Symposium on Snow Monitoring and Avalanches, Manali, India (Vol. 12, p. 16).

Subin Z.M., Koven C.D., Riley W.J., Torn M.S., Lawrence D.M., Swenson S.C. 2013. Effects of soil moisture on the responses of soil temperature to climate change in cold regions. Journal of Climate. 26(10): 3139-3158 doi: http://dx.doi.org/10.1175/JCLI-D-12-00305.1.

Formatted: Font: +Body (Cambria), 12 pt

Tang, J., & Zhuang, Q. 2011. Modeling soil thermal and hydrological dynamics and changes of growing season in Alaskan terrestrial ecosystems. *Climatic change*, *107*(3-4), 481-510.

Tape, K. D., Rutter, N., Marshall, H. P., Essery, R., & Sturm, M. 2010. Recording microscale variations in snowpack layering using near-infrared photography. Journal of Glaciology, 56(195), 75-80.

VAN WIJK, W. R. 1963. Physics of plant environment. Physics of Plant Environment.

~~Watanabe, K., & Wake, T. 2009. Measurement of unfrozen water content and relative permittivity of frozen unsaturated soil using NMR and TDR. Cold Regions Science and Technology, 59(1), 34-41.~~

Weller, G., Holmgren, B. 1974. The microclimates of the arctic tundra. Journal of Applied Meteorology 13, 854-862.

Wieringa, J. and E. Rudel, 2002. Station exposure metadata needed for judging and improving quality of observations of wind, temperature and other parameters. Paper 2.2 in WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2002).

Williams, P. J., and Smith M. W. 1991. The Frozen Earth. Cambridge University Press, Cambridge, UK

Yang, D., Goodison, B. E., Ishida, S., and Benson, C.: 1998 Adjustment of daily precipitation data of 10 climate stations in Alaska: Applications of world meteorological organization intercomparison results, Water Resour. Res., 34, 241–256.

Yi, S., Wischnewski, K., Langer, M., Muster, S., & Boike, J. 2014. Freeze/thaw processes in complex permafrost landscapes of northern Siberia simulated using the TEM ecosystem model: impact of thermokarst ponds and lakes. *Geoscientific Model Development*, *7*(4), 1671-1689, doi:10.5194/gmd-7-1671-2014.

Zhang, T., Osterkamp, T.E., Stamnes, K. 1996. Influence of the depth hoar layer of the seasonal snow cover on the ground thermal regime. Water Resources Research 32(7): 2075-2086.

Zhang, T. 2005. Influence of the seasonal snow cover on the ground thermal regime: an overview. Reviews of Geophysics. 43: RG4002, doi:10.1029/2004RG000157.

Zhang Y., Carey S.K., Quinton W.L., Janowicz J.R., Pomeroy J.W., Flerchinger G.N. 2010. Comparison of algorithms and parameterizations for infiltration into organic-

covered permafrost soils. Hydrology and Earth Systems Sciences. 14: 729-750. doi:10.5194/hess-14-729-2010.

Zona, D., D. A. Lipson, J. H. Richards, G. K. Phoenix, A. K. Liljedahl, M. Ueyama, C. S. Sturtevant, and W. C. Oechel. 2013. Delayed responses of an Arctic ecosystem to an extremely dry summer: impacts on net ecosystem exchange and vegetation functioning. Biogeosciences Discussions, 10(12), 19189-19217.

Table 1. Valid parameter range for calibration sets

Notation/Units	Moss-Range	Peat-Range	Mineral-Range
Porosity [-]	0.88 -- 0.95	0.7 -- 0.93	0.2 -- 0.75
VG Alpha [1/Pa]	1×10^{-5} -- 2.35×10^{-3}	3.1×10^{-7} -- 1.2×10^{-3}	2.9×10^{-4} -- 1×10^{-3}
VG n [-]	1.3 -- 2.82	1.3 -- 1.9	0.1 -- 0.33
Residual VWC [-]	0.02 -- 0.18	0.04 -- 0.22	0.05 -- 0.18
$K_{dry, Bulk}$ [W/m K]	0.007 -- 0.3	0.05 -- 0.38	0.2 -- 1.6
$K_{unfrozen, Bulk\ Sat}$ [W/m K]	0.5 -- 0.59	0.43 -- 2.9	0.96 -- 3.1
$K_{frozen, Bulk\ Sat}$ [W/m K]	0.81 -- 2.8	0.81 -- 2.3	1.31 -- 2.8
$K_{dry, material}$ [W/m K]	0.022 -- 0.20	0.05 -- 0.38	0.2 -- 4.0
$\alpha_{T,uf}$ [-]	--	--	--
$\alpha_{T,f}$ [-]	--	--	--

** $K_{dry, material}$ [W/m K] is back calculated from $K_{dry, Bulk}$

Table 2. The calibration error from the measured values reported as the RMSE °C (phi) increased between the 1) BPC model to the 2) MC saturated model. Thus there was greater error in the model results, but the calibrated parameters were more realistic. Phi then decreased between the 2) MC saturated model and 3) the MC unsaturated model.

	BPC			MC			MC - Freed Pressure	
Calibration Start	Center	Trough	Rim	Center	Trough	Rim	Center	Trough
1	0.461	0.616	0.642	0.646	0.834	0.831	0.503	0.781
2	0.444	0.586	0.649	0.898	1.347	0.796	0.880	1.186
3	0.433	0.654	0.653	0.523	0.764	0.775	0.372	0.586
4	0.410	0.671	0.689	0.625	0.879	0.658	0.633	0.619
5	0.414	0.771	0.707	0.566	0.900	0.665	0.399	0.612
6	0.455	0.588	0.674	1.275	1.212	1.666	0.544	0.770

7	0.414	0.609	0.682	0.751	1.247	0.754	0.465	1.162
8	1.406	0.531	0.678	0.846	0.927	0.919	0.472	0.787
Average	0.555	0.628	0.672	0.766	1.014	0.883	0.533	0.813

Table 3. Measured snow depth ranges where gathered from a compilation of 258 snow depth measurements taken May 2nd 2013 in the area encompassing all three borehole temperature measurements. Utm coordinates: Northing 7910330-7910350, Easting 585900-585930. Measured snow water equivalence (SWE) ranges were calculated from measured snow depth and the measured average snowpack density of 326 [kg/m^3]. All simulated values were taken on simulation day May 2nd, 2013.

	Snow Depth [cm]		Snow Density [kg/m^3		Snow Water Eqv. [cm]	
	Measured Range	Simulated	Measured Ave.	Simulated	Measured Range	Simulated
Center	20 - 40	24.6	326	349.3	6.5 – 13	9.5
Rim	10 – 20	14.6		320.2	3.25 – 6.5	5.2
Trough	40 - 60	40.3		370.4	13 – 19.5	16.25

Table 4. The ALT for all three columns are listed for each iteration of the calibration process, also with the range of possible ALT from the observed data. The observed ALT range was made by finding the deepest borehole measurement for center rim and trough with a temperature above 0 C^o for at least a day and the shallowest borehole measurement with all temperatures below 0 C^o.

	Center	Rim	Trough
Calibrated Subsurface	48.2	44.2	48.1
Surface Energy Balance	37.7	41.0	33.7
Snow Distribution	40.5	41.3	38.4
Observed ALT	50 - 60	40 - 50	35 - 40

Table 5. Final Calibrated Parameter Table (referred to throughout the text)

Notation/Units	Calibrated Moss	Calibrated Peat	Calibrated Mineral (Silty Loam)
Porosity [-]	0.9	0.876	0.596
VG Alpha [1/Pa]	2.3×10^{-3}	9.5×10^{-4}	3.3×10^{-4}
VG n [-]	1.38	1.44	1.33
Residual VWC [-]	0.05	0.34	0.199
$K_{dry, Bulk}$ [W/m K]	0.024	0.025	0.104
$K_{unfrozen, Bulk\ Sat}$ [W/m K]	0.446	0.427	0.788
$K_{frozen, Bulk\ Sat}$ [W/m K]	1.81	1.73	3.2
$K_{dry, material}$ [W/m K]	0.1	0.11	2.23
$\alpha_{T,uf}$ [-]	0.5	0.4	0.8
$\alpha_{T,f}$ [-]	1	2	0.73

** $K_{dry, Bulk}$, $K_{frozen, Bulk}$, and $K_{unfrozen, Bulk}$ [W/m K] are back calculated from $K_{material, Bulk}$

1434

1435

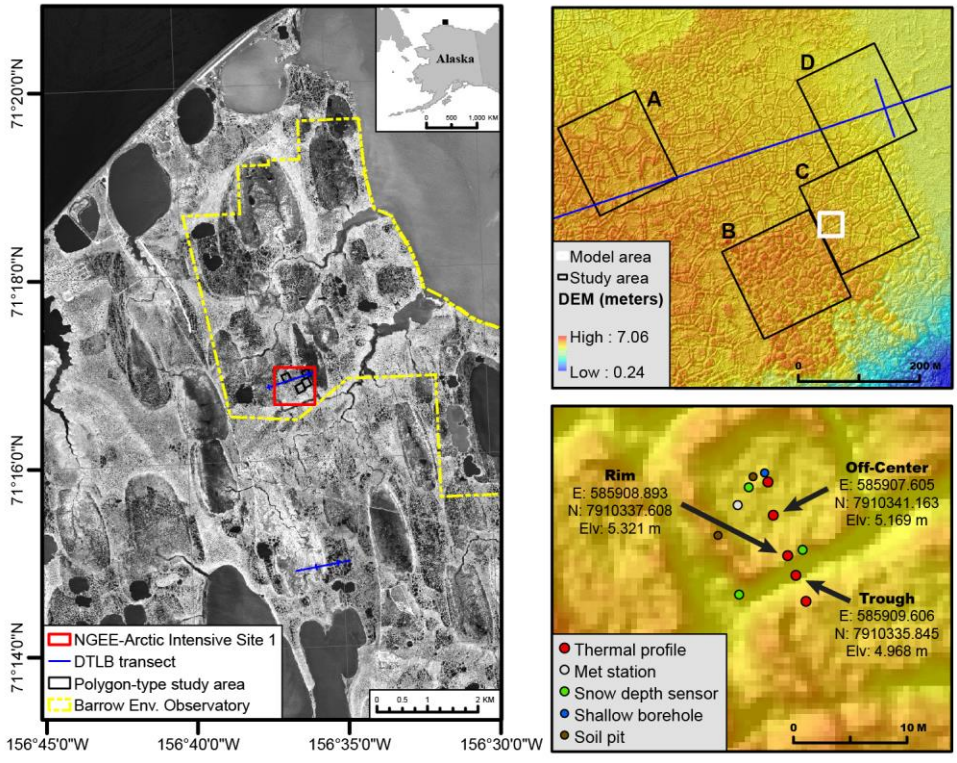
1436

1437

1438

1439

1440 ***Figures:***



1441

1442 Figure 1. LIDAR of site-C with the three observation locations mapped and greater
1443 Barrow, AK area. (Credit Garrett Altmann).
1444
1445
1446
1447

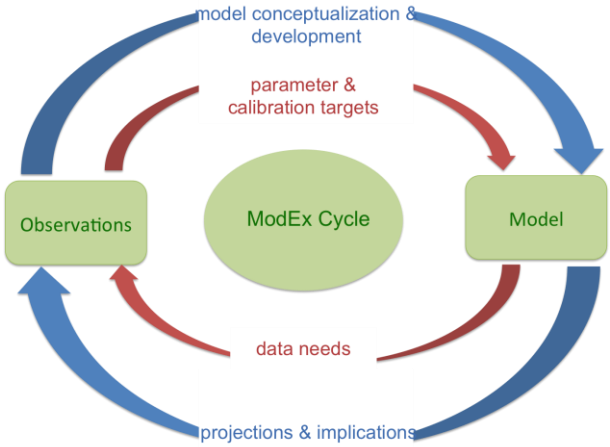


Figure 2. Schematic representation of a Model Observation/Experiment (ModEx) process involving traditional parameter estimation/calibration (inner loop) and model structural/conceptual refinement (outer loop). Observations inform simulation input and provide a starting point for a conceptual model. Both the conceptual and numerical model is then tested against observations. In successive ModEx iterations the model is then refined and at times re-drawn in order to elicit governing processes that shape model outcome to match observed and measured phenomena. Finally model experiments and the identification of governing processes inform future observations as to which measurements are needed to assess the state of the system.

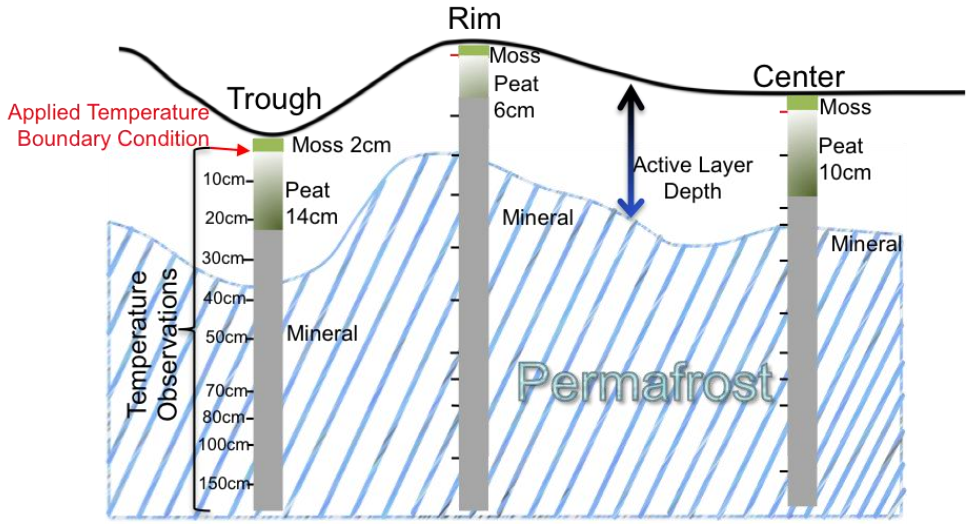


Figure 3. Diagram of the three 1-D columns and the associated measured soil temperature depths.

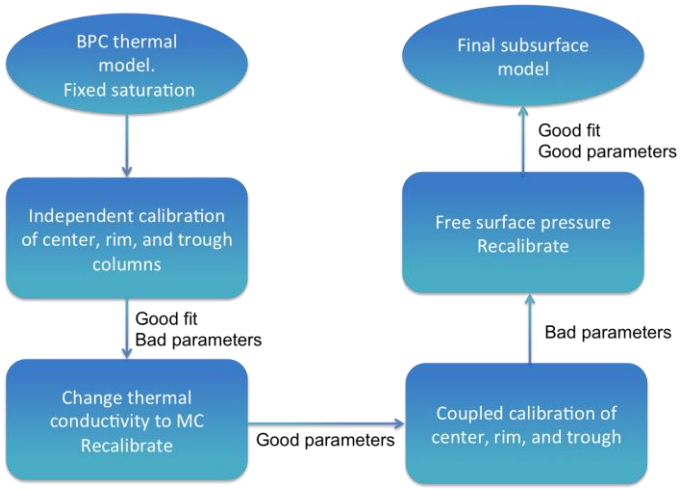


Figure 4 The ModEx cycle as applied here to subsurface thermal hydrologic system in freezing/thawing soils.

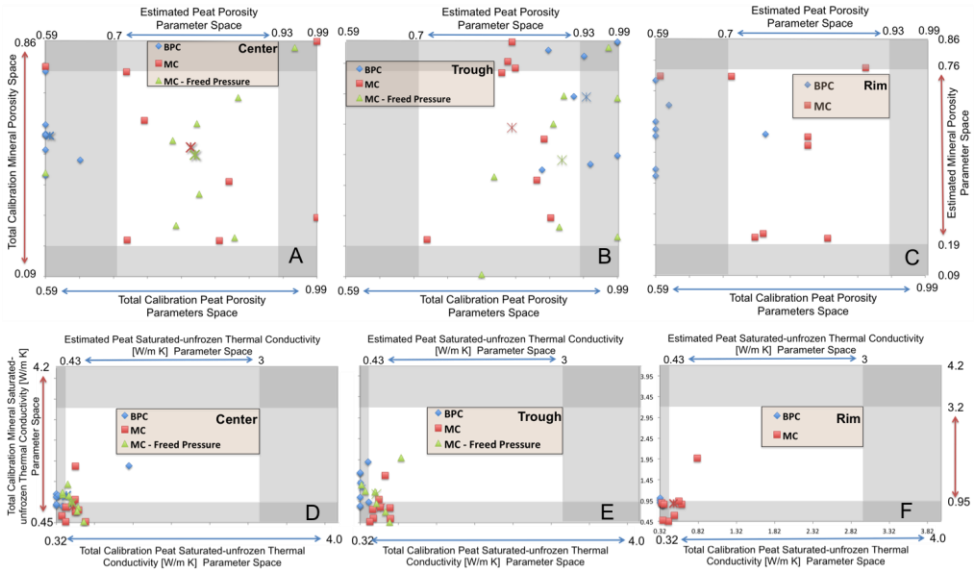


Figure 5. Plots A, B, and C show Center, trough and rim respective calibrated peat and mineral porosities from 8 calibrations starts. Plots D, E, and F show calibrated saturated unfrozen thermal conductivities ($K_{sat,uf}$) for peat and mineral soil layers from the same 8 calibrations starts. $K_{sat,uf}$ values from the MC calibration are

calculated from equation 3. Blue diamonds used the BPC model for soil thermal conductivity, red squares used the MC model for soil thermal conductivity, and green triangles added surface pressures as a free calibration parameter to the MC model for soil thermal conductivity. Color-coded asterisks represent the average calibrated parameter for each model tested for the 8 calibration starts, but are not actual calibrated results. Accepted parameter space delineated from literature and site observations in all cases are mapped as clear areas. Shaded areas are the calibration space outside of the acceptable parameter space. This figure shows how the calibration response surface changes as the model changed from 1) BPC to 2) MC to 3) unsaturated.

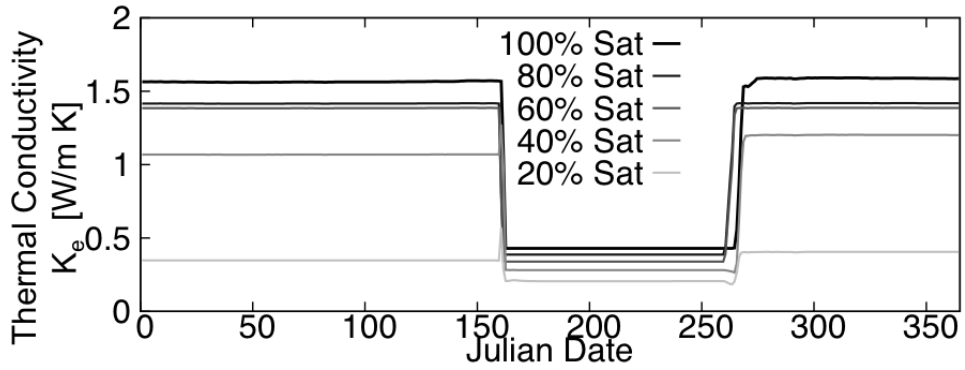


Figure 6. Thermal conductivity of peat throughout a year with different surface pressures. Percent liquid saturation is based off of summer time water liquid saturation, which changes during winter due to an increase in ice saturation. The change in thermal conductivity coincides with spring thaw, approximately Julian Day 160 or early-June, and fall freeze-up near Julian Day 265 or late September.

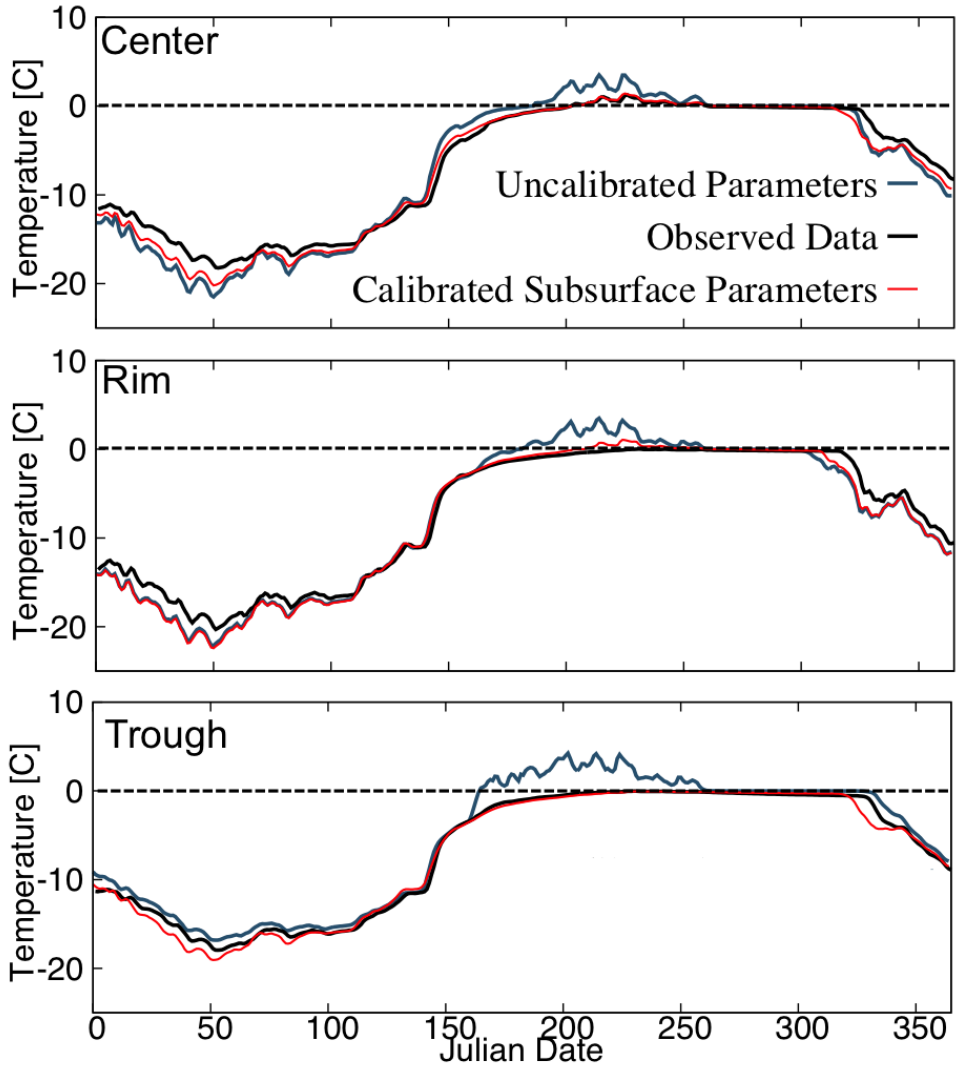


Figure 7. The subsurface un-calibrated and calibrated temperature time-series is compared to measured soil temperature time-series to showcase the improvement from the calibration process at 40cm depth for the center, trough, and rim. The initial un-calibrated parameters were selected from the literature search described in section 2.4 and Appendix C. Calibration fit to observation varies from the three columns, but shows marked improvement from initial un-calibrated time-series and are most accurate for all three during the summer at depth where active layer thickness is delineated.

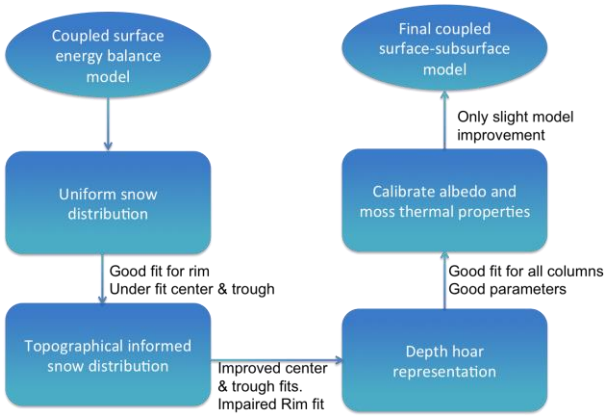


Figure 8. The ModEx cycle applied to the surface energy balance and moss parameters.

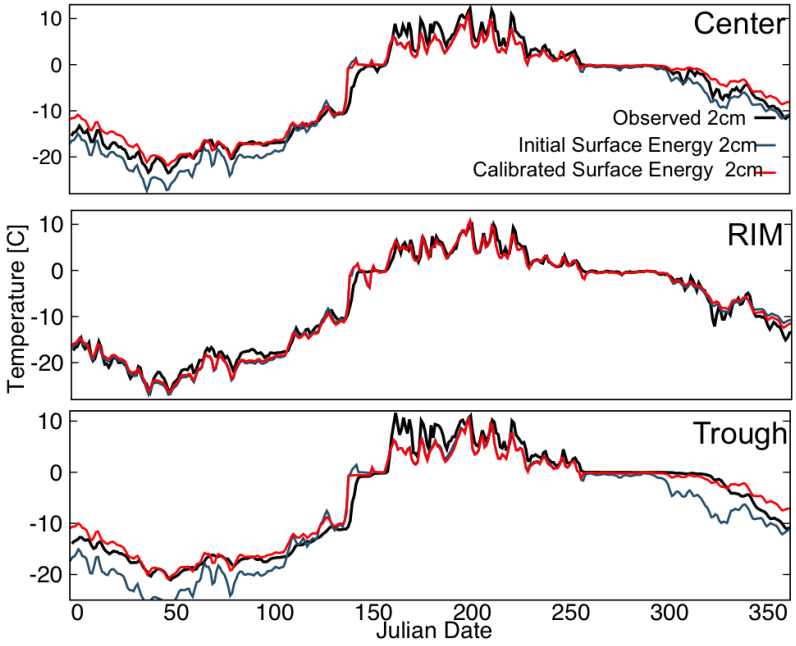


Figure 9. Temperature profiles for a 2cm depth are shown for the Center (plot A), Rim (plot B), and trough (plot C), using the initial surface energy balance parameters (blue), calibrated surface energy balance (red), and measured soil temperature profile (black). The biggest difference between initial temperature profiles and the calibrated profiles is the wintertime temperature for each column and is a result if distributing snow on the center, rim, and trough and depth hoar representation. Snow distribution also had the greatest control in the ALT (Table 4).

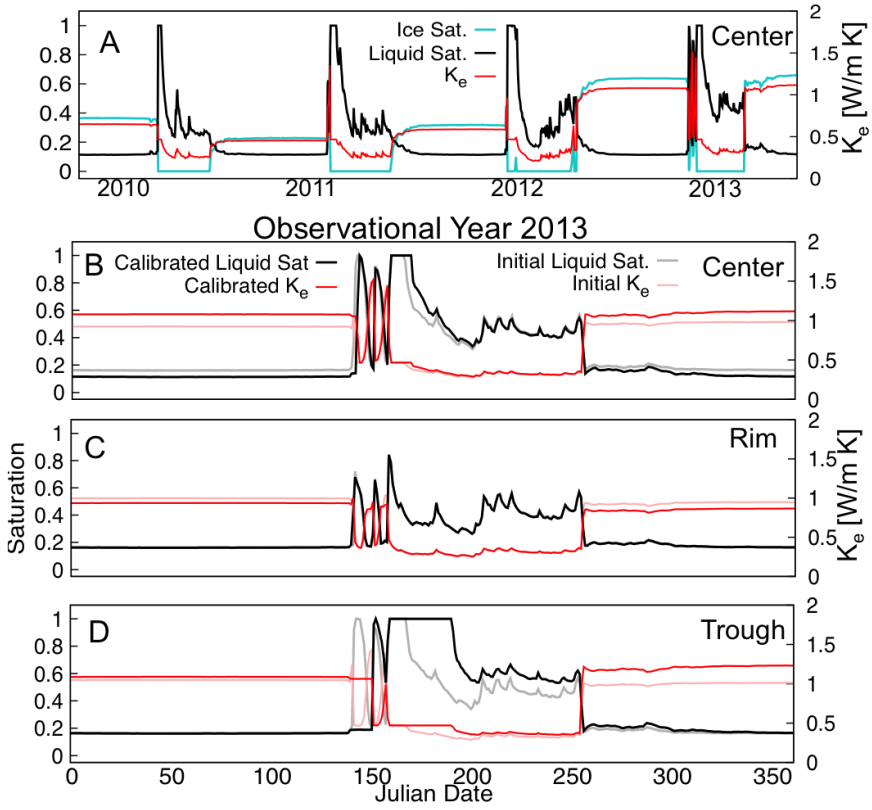


Figure 10. Ice and liquid saturation are shown in plot A for the simulated years of 2010-2013 at 2cm depth along with bulk thermal conductivity for a center column. Notice that ice saturation and thermal conductivity during the winters are unique for each simulation year. Plot B is a zoomed in view to year 2013 of ice and liquid saturation and the bulk thermal conductivity for the center. Plot C and D show the corresponding ice and liquid saturations for the trough and rim, along with the respective thermal conductivities for the 2 cm soil depth for the year 2013. Plots B-D have unique ice and liquid saturation and therefore bulk thermal conductivity for each column, which is a result of both the maximum ponded depth for each column and the snow distribution that mimics wind scouring of the snow surface at Barrow, AK.