Response to 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 small discussion of the FUSE and FLEX modeling approach at 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 calibration and model structure reduction, and provides other similar methods to diagnose structural model needs. The updated text will read: "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 different models and evaluating model performance (Clark et al., 2008; Kavetski and Fenicia, 2011; Fenicia et al., 2011). We implement ModEx model refinement by focusing on the plausibility of calibrated parameters in addition to the mismatch between field measurements and simulated responses."*

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 preferentially target larger 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, we don't see how it would apply to the single-objective calibrations that we performed. Ranges of calibration parameters such as porosity or thermal conductivity could have been limited to physical values, 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 identified 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.

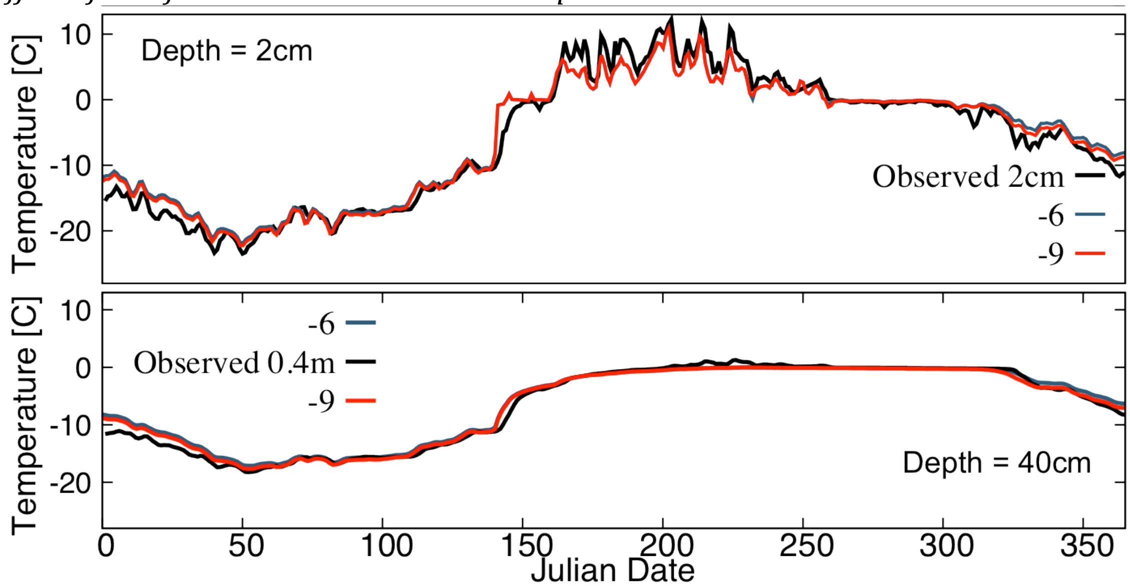
It is somewhat interesting that there is no consideration of the impact of uncertainty in the parameter definitions on the modeling performance. Clearly, this is a complex model with various interactions (hence the ModEx approach adopted). With that, it would be interesting to understand better the role 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 by updating the conclusion section to: "Thus, field and laboratory work to better constrain hydrothermal representation and the governing model parameters would help reduce uncertainty in model projections. Likewise modeling efforts *that focus on uncertainty analysis and environmental parameters sensitivity can identify which parameters govern model outcome. This information then can be used to direct new observational efforts that monitor key environmental variables."*

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℃ 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 that, "A far field bottom boundary condition was held constant at -6℃ to provide a low temperature gradient."

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 revisit 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 in section 3.4 at 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.