**Anonymous Referee #2**

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

**Major points:**

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

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

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

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

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

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

**We agree that upscale transfer of knowledge gained through fine-scale simulations to inform and improve simulations at larger spatial scales is a critical issue. While we can not comprehensively address that research problem in the current study, we have added text to indicate how the fine-scale simulation capability explored in this study contributes to a broader scaling strategy for the NGEE Arctic project. The following text has been added at Page 27 Lines 11-25:**
**"While the knowledge gained by developing and evaluating fine-scale 3D simulations is valuable from the perspective of increased understanding of complex process interactions, the explicit long-term goal of the NGEE Arctic project is to improve predictions of Arctic ecosystem processes at scales relevant to coupled climate and Earth system simulation. One element of our strategy to migrate knowledge across scales is to improve the grid and sub-grid representations in the land model component of our Earth system models to capture observed modes of variability in physical, biological, and biogeochemical processes. For example, our new top-level grid topology for global-scale land modeling follows watershed boundaries instead of the typical and arbitrary rectangular gridcell arrangement (Tesfa et al. 2015). Sub-grid schemes are being developed that represent topographic variation within basins, and our goal is to apply those methods in the micro-topographic setting of polygonal tundra to capture the variation in thermal, hydrologic, and biogeochemical regimes, and interactions with vegetation communities. The current study is one step toward identifying the relevant modes of variation among diverse landforms in the polygonal tundra region. Another element of our scaling strategy is to use, to the full extent possible, a common set of modeling tools to construct simulations at various spatial scales. Even though many processes that can be represented explicitly at the finest scales (such as lateral flows of energy and water) must be parameterized for efficiency in a larger-scale simulation, having a common underlying set of equations helps to reduce unintentional loss of information across scales due, for example, to aggregation and disaggregation operators."**

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

**Minor points:**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

11. Page 14, Line 17: How long is the spin up?
Spin up simulations were conducted for a period of 10 years. A statement clarifying that has been added to Page 13, Line 17-21.
“Spin up simulations were conducted for a period of by cycling annual time series of forcing. Spin up simulations were continued until a periodic steady state was achieved (i.e. close to zero interannual variability in annual thermal regime). Spin up duration of 10 years was used at all the sites and was determined to be sufficient.”

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

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

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

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

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

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

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

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

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

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

We have moved the discussion to the conclusion section.

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

**References:**

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