

General assessment

This study uses a pedon-scale soil physical model - called DynSoM-2D – to describe how small-scale heterogeneity affects soil thermodynamics and surface heat fluxes in a specific type of permafrost topography (non-sorted circles), with a particular focus on the effect of heterogeneity-induced lateral heat fluxes. This is an important topic, since most land surface models fail to represent subgrid-scale heterogeneity in permafrost-affected areas, thereby leading to an inaccurate representation of permafrost dynamics and biogeochemistry. The approach, tools and simulation design employed in this study are highly relevant for addressing these questions. The discussion is particularly interesting and provides an insightful interpretation of the results. The code is clear and well commented, making it easy to read. Unfortunately, there are some major problems.

We are grateful to the reviewer for emphasizing the importance of this study and the constructive feedback. We addressed the mentioned problems point by point.

First, the quantitative analysis is based on a single year, which is insufficient for deriving robust conclusions about the effect of soil heterogeneity on soil temperature. While the authors briefly explain their decision to conduct a single-year analysis (l.193), the results are highly dependent on the specific atmospheric conditions of that year. More robust results could be obtained by including several years in the analysis (e.g. a 10-year climatology or a longer timeseries), especially given that this modelling study is not limited by a lack of data.

We agree with the reviewer that a single year data is not enough to describe the soil temperature of a site. However, the aim of our paper is to identify the differences on the temperature induced by heterogeneity of the soil and related lateral heat fluxes and avoid errors linked to aggregation. As stated on the manuscript, the simulations are preceded by a 50 year spin up that provides the conditions for the final run with aims to be a “snap-shot” of the differences on the soil. However, we added the full simulation timeseries (fig. 2) to the SI to show that the observed pattern are consistent across years. For figure-clarity, we removed the intra-model ranges of M1 and M2.

Second, (inter- and intra-) model ranges are highly sensitive to extreme values and are not absolute metrics. They lack a reference point, such as a mean or median, against which they can be compared (for example, an increase in the intra-model range does not have the same impact on permafrost physics and biogeochemistry at high temperatures or close to 0°C). Although extreme values can be informative, other statistical metrics (e.g. mean/standard deviation or median/interquartile range) are a more robust choice for describing the results.

We assume that the reviewer might have misunderstood, what we have done, and in that case, we will clarify this in the manuscript. We have used mean simulations among columns of the simulations M0 – M2 to derive the inter-model differences, meaning that the inter-model difference between M1 and M0 is the difference between the mean over all columns of M0 and M1, respectively. The intra-model differences refer to the largest differences between columns of the same model, which is only applicable for M1 and M2, because M0 has a homogeneous soil input and consequently no differences among columns. These ranges (for M1 and M2) are highlighted as shadowed areas in our timeseries figures and include the minimum and maximum range of simulations.

Finally, as this is a new model that has not been described elsewhere, a much more detailed description is required, particularly with regard to snow, hydrology and the calculation of soil thermal properties. In particular, the description lacks the key equations that underpin the results presented in this study. Some specific points are detailed in the “Specific Comments”.

We added the specific chapter and/or equation from Bonan (2019), and added Beer et al. (2018) as reference for our snow model, where necessary. We also added a more detailed description of the snow dynamics to the SI.

Overall, this is an interesting study that contributes to a growing body of literature on an important and challenging topic. It would definitely be a valuable contribution for the land surface modelling community. However, major revisions are required in a number of areas.

Specific Comments

- The title is quite technical, making it difficult to guess what the article is about at first glance (for instance, it does not mention “soil heterogeneity”). I do not have any other suggestions, but the article would benefit from a clearer title.

We changed the title to “Impact of Soil Heterogeneity and Lateral Heat Fluxes on Soil Temperature Simulations in a Permafrost-Affected Soil” and hope that this title is now better understandable.

- l.101 : The principles to which the authors are referring to are unclear. Maybe just keep the following sentence (l.102) or combine both ?

We will rephrase the sentences as following: The vertical scheme of the model essentially couples soil thermal and hydrological processes through (i) phase change of soil water (Ekici et al., 2014) and (ii) changes in soil thermal properties due to soil moisture (Bonan, 2019, chapters 5 & 8).

- l.108-109 : Could you give the soil layer thickness for all layers ?

The layer thickness of DynSoM is user-adjustable. We added the following for better understanding: For our simulations, we chose a set-up of 10 columns with a width of 10cm each, adding to a total horizontal extension of 1m, and layer thicknesses of 10cm within the uppermost meter, followed by layers with increasing thicknesses of 1m, 3m, 5m, 10m, and 30m, adding to a total depth of 50m.

- l.111-112 : Bulk density is missing from the list (needed in pedotransfer functions of Wösten et al. (1999)).

DynSoM calculates bulk density based on the given information about soil texture, i.e. the fractions of clay and silt and organic matter content, based on the equations published by Martín et al. (2017) based on Leonaviciute (2000). We clarified this in the text. The precise equations are included in the published code.

- l.115 : What about the other soil parameters besides soil porosity ?

In the pedotransfer functions, the following parameters are considered: van Genuchten parameters (alpha, m and n), residual porosity, saturated conductivity and from Cosby parameters (Cosby et al. 1984, saturated matric potential and the b parameter). In addition, the pedotransfer function of the code includes the calculation of the field capacity matric potential, field capacity effective saturation and field capacity moisture (all based in van Genuchten parameters), as well as initial porosity water and ice based on the calculated porosity and input fractions. We clarified this in the text and added a table with mean values of the most relevant soil parameters to the SI.

- l.118 : The variables driving the model should be specified. Also specify how rainfall and snowfall are separated from total precipitation.

We extended this section as following and added a description of the snow dynamics to the SI.

In addition to this constant site information, DynSoM-2D requires a site-specific atmospheric forcing as an upper boundary condition. Specifically, DynSoM-2D uses precipitation, which assumed to be snow, if the air temperature is below 0°C, and serves as upper boundary for the soil hydrological scheme (cf. A2), air temperature (only for the rain-snow separation), incoming long-wave and short-wave radiation (to calculate the energy balance at the surface, which is used as upper boundary for the soil heat conduction scheme, which is described in the following section), surface pressure, specific humidity, water vapor pressure and wind velocity. This.

- I.140 : Please provide the equations used to calculate soil thermal conductivity and capacity as these parameters largely control soil temperature dynamics.

We added this to the SI.

- Given the crucial role of soil water content and frozen/unfrozen fractions on soil temperature, the soil hydrological scheme should be described (including boundary conditions).

We added this to the SI

- I.156 “The heat is recalculated to a skin layer/surface temperature” is not very precise and should be clarified.

We added the specific sections of Bonan (2019).

- I.160-169 : The equations governing snow cover dynamics should be provided, along with the values of snow conductivity, capacity and any other parameters controlling snow dynamics.

We added this to the SI.

- I.174 : For clarity, please recall that the soil column extends below 1m. If I have understood the code correctly, the soil column is 50m deep, but this should be clearly explained in the manuscript.

We clarified this.

- I.175 : Which soil properties are used at depths below 1m ?

We clarified this.

- Fig.1 : The vertical axis needs a unit/label.

We added them.

- I.178 : Vegetation growth is excluded but it is unclear whether vegetation is present in the simulations. If so, please describe the type of vegetation present.

We clarified this.

- Table 2 : The percentages on the left do not add up to 100% for A-B.

Sorry, that's a typo. The A-B has 57.4% of soil. We changed this. We also changed the sand content of the OA in the right column to 3.25%.

- I.188 : The authors should provide evidence that the soil temperatures and soil moisture are close to equilibrium after the 50-year spinup (with no long-term drift).

We have chosen a 50-years spin-up period, as we found this as a typical spin-up time for geophysical soil models that do not consider vegetation/a vegetated ecosystem, which would have needed to grow and find its equilibrium. In our preliminary analysis for the spin up length we have not observed sensible long-term remaining trends after the 20 first years spin up. Unfortunately, we haven't stored

the spin-up simulations as part of the provided dataset, as we considered the information of the complete spin-ups is not relevant enough to justify the storage that it requires. Consequently we are unable to prove this with our current simulations that are provided at zenodo. If the editor considers this information to be critical, we can create an new dataset containing the full spin-up result series.

- l.191-193 : Analysing a single year does not produce robust quantitative results, as these depend on the atmospheric forcing of that particular year (e.g. the amount of precipitation influencing both snow cover and soil thermal properties via soil moisture). The aggregation error could be different if a different year had been chosen. Why not use a longer time period ? I do not understand the need to “preserve the current atmospheric signal” (l.192). The results would be more robust if a longer period was included in the analysis, for example daily or monthly averages over 10 years, or the analysis of a longer timeseries if the authors prefer not to use time averages.

As described above, we did not want to describe the soil temperature at a specific site and validate the model, where a single year would be definitely not sufficient (besides the fact that we do not have measurement data), but to investigate the effect of soil heterogeneity and heterogeneity induced lateral heat fluxes on simulated soil temperatures (given the same atmospheric forcing). For our analysis, we used a single year to represent a “snap-shot” of the differences in soil simulations, which are solely caused by either soil heterogeneity or soil heterogeneity and heterogeneity induced lateral heat fluxes. We added the full timeseries to the SI. We also extended the OM vs soilT analysis (sec. 3.4) to the full simulation period to make our results more robust.

- l.201 : Please mention how θ_{het} is calculated. Horizon-wise average ?

The values in table 3 are spatially averaged for topsoil (0-20cm), subsoil (20-80cm) and deep soil (80-100cm), which means that they show the mean over all ten columns and within the given depths, and for the entire soil (0-100cm), again for all ten columns. See figure 1 for clarification.

- l.203 “In general, the horizon-wise averaged heat conductivity is lower in the heterogeneous soil than in the homogeneous soil” → This statement is not supported by the data in Table 3 where $k_{het} > k_{hom}$ for both the subsoil and the total soil.

We rephrased the sentence as following: In general, the horizon-wise averaged heat conductivities (k_{het}) differ between the heterogeneous and the homogeneous soil...

- l.208-210 : I do not think it can be concluded from Fig.A1c&e that “the calculated albedo is hardly affected”. The authors should provide data to support this statement.

We added numbers to support the statement. We also added the equations to calculate albedo.

- l.234 : What does “per day” mean ? I suppose it refers to daily values but this should be clarified.

We clarified this.

- l.255 : “at the simulated distance of 1m” → unclear what this refers to.

We will rephrase this as following: ... , which exceeds 4°C on a daily basis and 2°C on a monthly basis across the simulated distance of 1m,...

- l.257 : “After freezing” → unclear, please precise.

We will clarify this as following: After freezing, i.e. when the soil is fully frozen up to 1m depth,...

- Using DOY in Fig.2 and months in Fig.3 makes it difficult to compare the two figures. Adding months to Fig.2 would improve understanding.

We agree that the figures are hard to compare. We added DOY information to figure 3.

- l.276 : “reflecting differences in soil moisture content that affect surface albedo” → albedo data are missing to support this statement.

We added numbers in sec. 3.1. And also the equations to calculate albedo based on relative soil water content.

- l.300 : What does “from below” refer to ? The freezing front ?

We will rephrase this as following: During this week the soil freezes from both sides (fig. 3), i.e. from the surface, which is strongly cooled by the atmosphere, and from colder frozen layers below.

- Fig.4 : A statistical test, such as a t-test, is required in order to assess the significance of soil temperature differences.

If necessary, we can add the outcome of a t test, but the differences are small, and indeed not significant, which we also did not say here. We observed a clear pattern, linking OM content to larger differences in soil temperatures between model simulations that are modified by lateral fluxes. But neither the initial heterogeneity induced differences (M1-M0), nor the reduction (O horizon grid cells)/increase (OA, A, B, C horizon grid cells) of differences due to lateral fluxes (M2-M0 vs M2-M0) are significant.

- Fig.4 : The difference in soil temperature is likely to be dominated by snow and ice in winter, resulting in values close to zero from February to June. Maybe focusing on the period when soil temperature differences are significant (July to January) would help to assess the effect of OM distribution on these differences, and would avoid pulling the median down artificially due to the near-zero ΔT_{soil} during half of the year ?

The reviewer is right, that the observed full-year differences are resulting from close to zero differences within the first half of the year, but we think that the fact, that second-half year differences are still notable in full-year differences is noteworthy. We could artificially boost the effect by filtering the data, but we decided not to do this, because (1) OM content has also an effect on frozen soil properties, which we did not want to take out, and (2) the boosted effect would not be representative for the full year. However, we could aggregate the data seasonally (DJF, MAM, JJA, SON) and provide the figures in the SI.

- Section 3.4 lacks quantification. Statistical metrics and tests should be used.

As said before, the differences, and also not the change in differences due to lateral fluxes is not significant, and we never said it is. We observed a relationship between OM content and differences in simulated soil temperatures between a homogeneous and a heterogeneous soil, which is different for a model without and with lateral fluxes. But given our very small domain, the differences are small and not significant. We also show both, box whisker plots (incl. median, quartile ranges) of differences and mean differences, to be totally clear about the fact that (1) medians are close to zero and inner-quartile ranges largely overlap, which means that differences are not significant, and (2) mean values, which deviate stronger, are largely influenced by few outliers, which would also not show any significance, but need to be acknowledged.

- l.315-317 : “By aggregating...lower OM content” → not sure if this is significant. A statistical test is needed.

We rephrased this as following: By aggregating the grid cells by their difference in OM content in the soil input (fig. 1), which can be higher in the heterogeneous soil compared to the homogeneous soil, equal, or lower, depending on the certain grid cell within its layer...

- I.317-318 : “this relationship...from M2 to M0” → based on which metrics ?

We clarified this by adding: However, differences are not significant, and we cannot derive any clear mathematical relation between soil temperature shifts and OM differences.

- I.356-357 : This is counter-intuitive as OM generally insulates the ground against warmer air temperatures in permafrost regions (Zhu 2019, Lorant 2018). However this can be explained by the higher thermal conductivity in columns with a higher OM content, due to higher soil moisture, as the authors explain. A figure showing the 2D evolution of soil moisture (vertical and horizontal, similar to Fig.3) would support this statement.

Since DynSoM tends to be wet and (almost) saturated (see fig. A1), a figure showing soil moisture (water+ice) differences similar to figure 3 (or figure A2) indeed basically shows the difference in soil porosity, i.e. soil moisture is higher in the heterogeneous soil, where the heterogeneous soil contains more OM, and lower, where the heterogeneous soil contains less OM than the homogeneous soil (see figure below this text). Differences are constant in frozen soil (Nov-June), and only differ slightly, where soil is thawed in summer. The differences in the active layer thickness are visible by the small negative deviation in around 0.5m depth in summer, where the model with homogeneous soil is frozen, but the models with heterogeneous soil thaw 10-20cm deeper. We can add this figure to the SI, but we do not see a major benefit here.



- I.375-376 : “The only time when snow actually has an effect on simulated surface heat fluxes is at the end of the snow melt period” → rather on the difference of surface heat fluxes between homogeneous and heterogeneous configurations than on surface heat fluxes themselves (which are always impacted by snow) ?

We rephrased this as following: The only time when snow impacts surface heat fluxes, which are simulated by the three different models, differently is at the end of the snow melt period.

- I.409 : “which will be consistent for any aggregation error.” → I do not think this conclusion can be drawn from a single-site study. Such a conclusion would require simulations for other sites, as the

amplification of the aggregation error by lateral fluxes probably depends on the spatial distribution of soil texture and OM content.

We rephrased this as following: “which will likely be consistent for any aggregation error.”

- l.443-446 : I do not understand why the amplification of the aggregation error by lateral fluxes provides more confidence that these results would remain for other sites.

We explained this by adding: When transferring the setup to another site, it would still imply a (linear) averaging over soil properties, which will cause the observed aggregation error due to the non-linearity of relationships. However, the direction of the aggregation error may change with site.

- l.448 : In the context of land surface modelling, there are no global maps of soil texture and OM at a 10 cm resolution. Could you briefly describe how these results could help improving land surface models ?

We added: Finally, we would like to turn our perspective from our very extreme example with a cm-scale resolution back to LSMs, which usually run on scales of (hundreds of) kilometers. Our results clearly show that with decreasing model resolution, heterogeneity becomes more obvious, and heterogeneity induced lateral fluxes will become more important. Thus, we would like to emphasize that as soon as model developers think about decreasing model resolution in order to improve their simulation results, they have also to think about emerging heterogeneity and induced lateral fluxes and acknowledge their potential influence on simulation results, which may increase simulation uncertainty, especially in combination with the lack of resolution-suitable sub-surface soil data.

- l.453 : “we assume that it will remain at other sites and under changed environmental conditions”
→ I do not think this assumption can be made based on a single-site study.

We will rephrase this as following: “we assume that it will likely remain at other sites and under changed environmental conditions”

- Fig. A3 and Fig.A4 are not referenced in the main text. They should either be included or removed.

We added a reference to the first figure and removed the second one.

Technical corrections

- Bonan 2019 is cited for various aspects of the model and it would be helpful to provide more specific references for each citation (e.g. refer to the relevant chapter in the book).

We added chapter/section numbers and/or referred to equations

- l.105 : Please give the section number.

Done

- l.123 : “phase change” → water phase change

Done

- l.126 : Please refer to the specific chapter/equations from Bonan (2019).

Done

- l.131-134 : Please precise the units.

Done

- l.136 : Please give the section number.

Done

- l.138 : Please precise the units.

Done

- l.154 : Please refer to the specific chapter/equations from Bonan (2019).

Done

- l.171 : Please specify which figure from Gentsch et al. (2015) you are referring to.

Done

- Table 2 : Please define O, A, B and C in the table legend.

Done

- Figure 1 : Please precise in the legend what the red dash-dot lines refer to (separation topsoil, subsoil and deep soil ?).

Done. Added to figure description.

- Table 3 : Please add units.

We will add this.

- Fig.2 (b) : The peak of maximum difference looks cropped.

The exact numbers can be found in the text.

- l.234 : "In general" should be removed.

Done

- deVrese et al. Paper is now published and is no longer a preprint.

Done.

References

Loranty, M. M., Abbott, B. W., Blok, D., Douglas, T. A., Epstein, H. E., Forbes, B. C., Jones, B. M., Kholodov, A. L., Kropp, H., Malhotra, A., Mamet, S. D., Myers-Smith, I. H., Natali, S. M., O'Donnell, J. A., Phoenix, G. K., Rocha, A. V., Sonnentag, O., Tape, K. D., and Walker, D. A.: Reviews and syntheses: Changing ecosystem influences on soil thermal regimes in northern high-latitude permafrost regions, *Biogeosciences*, 15, 5287–5313, <https://doi.org/10.5194/bg-15-5287-2018>, 2018.

Wösten, J., Lilly, A., Nemes, A., and Le Bas, C.: Development and use of a database of hydraulic properties of European soils, *Geoderma*, 90, 169–185, 1999.

Zhu, D., Ciais, P., Krinner, G. et al. Controls of soil organic matter on soil thermal dynamics in the northern high latitudes. *Nat Commun* 10, 3172 (2019). <https://doi.org/10.1038/s41467-019-11103-1>

Beer, C., Porada, P., Ekici, A., & Brakebusch, M. (2018). Effects of short-term variability of meteorological variables on soil temperature in permafrost regions. *The Cryosphere*, 12(2), 741-757. <https://doi.org/10.5194/tc-12-741-2018>

Leonaviciute, N. (2000). Predicting soil bulk and particle densities by pedotransfer functions from existing soil data in Lithuania. *Geografijos metraštis*, 33, 7-330.

Martín, M. Á., Reyes, M., & Taguas, F. J. (2017). Estimating soil bulk density with information metrics of soil texture. *Geoderma*, 287, 66-70. <https://doi.org/10.1016/j.geoderma.2016.09.008>

The manuscript presents an attempt to quantify the significance of the treatment of small-scale variations in soil properties on the simulated soil and surface temperatures, as well as the surface heat fluxes. The effects of heterogeneity and their impact on larger-scale averages are undoubtedly intriguing topics that warrant the attention of the modeling community. However, I think the manuscript would benefit from some revisions of the description of the model and analysis, and expanding of analysis to include the effects of interannual variability.

We are grateful to the reviewer for emphasizing the importance of this study and the constructive feedback. We addressed the mentioned problems point by point.

To start, I would suggest changing the title of the manuscript. First, for readers not intimately familiar with the permafrost features and process, it is not at all obvious what “non-sorted circles” in the title refers to. As a very minimum, the title should make it clear that the study is specific to the permafrost processes. Second, the focus of the study is on the differences in results of several modeling approaches, but to the casual reader the term “aggregation error” implies comparison with observations (or the perfect representation of the processes), which is not the focus of the manuscript. I recommend avoiding this term, at least in the title.

We changed the title to “Impact of Soil Heterogeneity and Lateral Heat Fluxes on Soil Temperature Simulations in a Permafrost-Affected Soil” and hope that this title is now better understandable.

I think another problem is that the description of the model misses an essential part: the method used to calculate surface turbulent fluxes. Clearly, on such a small horizontal scale of ~10 cm that the described model uses, the Monin-Obukhov Similarity Theory (MOST) approach commonly employed in large-scale mosaic schemes would not work because a number of assumptions important for the MOST applicability are violated. Therefore, it is essential to describe what alternative approach was used to calculate surface fluxes, especially given that a significant portion of the manuscript is devoted to the analysis of differences in turbulent fluxes and energy balance. Without that, it is very hard to judge the validity of the results.

The method used to calculate the surface turbulent fluxes is indeed Monin-Obukhov Similarity Theory, but from our perspective, there are three reasons, which we can use MOST for our model: (1) All MOST assumptions are limited to the “atmospheric” part of our model that is only applied vertically (1D). Derived energy fluxes are then used for all columns individually. (2) For this specific study, we used only “bare soil” and neglected any topographical differences, which means no differences in height and/or roughness between columns. Snow is added to the soil scheme, which does not affect surface height in this model configuration. Consequently, soil roughness versus the reference height of 2m (lowest level taken from CRUNCEP data) is well enough the recommended value of 50. However, (3) if roughness differences (due to vegetation or topographical differences) are present, DynSoM couples MOST with a roughness sublayer parameterization following Harman and Finnigan (2007, 2008). We will clarify this in the model description.

The description also seems to contradict itself, saying in section 2.1.1 that “a prescribed skin temperature (see following section) serves as the upper boundary condition”, while equation (3) implies that the surface skin temperature is calculated given the atmospheric meteorological forcing and prognostic equations of water and energy balance in the soil. I think providing more details about the calculation of surface fluxes and energy balance would help to resolve this confusion.

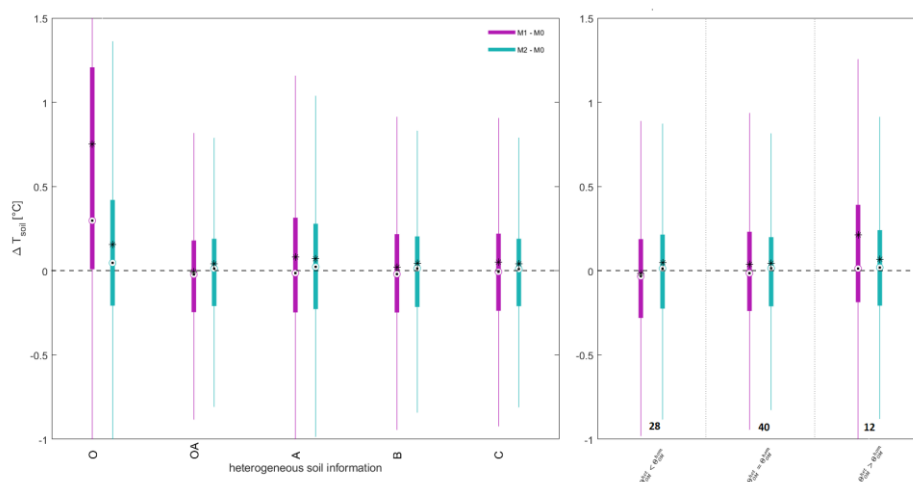
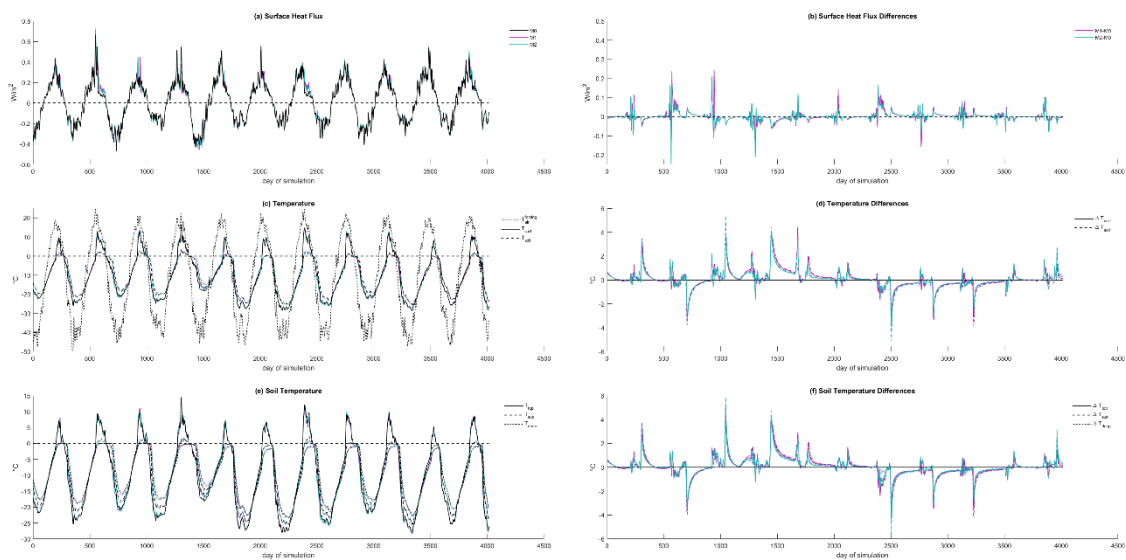
We took this method from Bonan (2019) and referred to the specific chapter/specific sections here.

In the section describing the results of the simulations, the authors chose to limit the analysis of the results to just one year, but the motivation is not entirely clear to me: the text in lines 191-192 says

“to preserve the current atmospheric forcing signal, which is strongly superimposed when another averaging method is used.” I am not sure about the meaning of this phrase; especially since it somewhat contradicts the one-week smoothing applied to the results to “avoid overly fuzzy near-surface” values. I think analyzing and presenting the statistics across all available years of forcing is essential to take into account the interannual variability and to have confidence in the robustness of the results.

This does not, of course, preclude using the results from one year (or one season) as an example illustrating physical processes, if necessary.

We agree that founding our study on a single year limits its significance. We added the full timeseries of figure 2 to the SI (see upper figure below this text) to show that the general observed pattern, i.e. simulation differences that are solely caused by soil heterogeneity and their increase by lateral heat fluxes, is not dependent on single years, whereby of course the absolute differences between models (inter-model ranges) differ across years. For figure clarity, we left out the intra-model ranges here, whereby these obviously also differ across years. To keep this information, we replaced figure 4 (see lower figure below this text as example; as well as related figures in the SI) by figures showing boxplots for the entire period and rewrote section 3.4, where necessary.



In the description of the energy-related fluxes and balances, the authors frequently use the confusing phrase “kW/m² per day”: the fluxes are typically measured in W/m², and it is absolutely not clear what “per day” refers to. Similarly confusing is the phrase “°C per day” in the description of the range of temperatures.

We changed this.

Assuming that my understanding of the units used in the analysis is correct, some of the energy balance numbers seem to be unreasonably large. For example, on lines 221-223, the manuscript says “the total simulated annual heat budget ... (M0: 13.7kW/m², M1: 14.1kW/m², M2: 13.5kW/m²)”. Of course, the total long-term average energy balance at the surface of the well-spun-up land model should be close to zero, so it is not clear what these numbers represent.

I can only assume that these results represent the annual average sum of sensible and latent heat fluxes (which should compensate radiative fluxes), but even then the numbers seem excessively high. For comparison, the solar radiation incident on the area perpendicular to the sun rays at the top of the Earth atmosphere is ~1360 W/m². It is not clear how it is possible that the annual heat balance at the site (~69N latitude) can be so large, given attenuating factors due to site latitude, annual averaging, and absorption/reflection/scattering by the atmosphere and clouds. Unless this is a typo, an explanation must be provided. Likewise, the intra-model differences in heat fluxes are on the order of hundreds of W/m²: that of course is not impossible on the short time scale, but would strongly depend on the way the turbulent fluxes are calculated, and needs to be discussed.

We changed this.

In figure A1, soil water content is measured in kg/kg; this is kilogram of water per kilogram of what dry soil or wet soil? Or per dry soil+ice? Why the commonly accepted definition of volumetric water content is not used?

We added the information to the figure legend.

Technical comments:

Line 38, and elsewhere replace “snow height” with “snow depth”

Done

L 171: Provide coordinates for Cherskii site

Done

L 178: Typo: “growthto” should be “growth to”

Done

L 203: Replace “horizon-wise averaged” with “horizontally averaged”

We disagree here, because “horizontally averaged” would only imply the horizontal averaging over single rows, which have (in our model configuration) a vertical extension of 10cm, whereas the “horizon-wise averaging” that we applied implies a larger vertical extension (in our model configuration).

Caption of figure 2, and elsewhere: does T_surf refer to the temperature of the soil surface, or the surface interacting with the atmosphere (i.e. surface of the snowpack if present and soil surface T otherwise)?

We added a clarification note to the text.

Caption of figure 2: "T_soil, all depths" — does it mean averaged over entire soil column?

Averaged until the depth of 1m. We clarified this.