**Response to comments of reviewer #1:**

**General comments**

**Snow and its insulation effects are critical for accurately simulating soil temperature and permafrost in high latitudes. This paper assessed the skills of nine land surface models based on the response patterns of Tsoil and the difference of Tsoil-Tair to snow depth in winter in high latitudes. The observed patterns at 268 climate stations in Russia were used as ground truth. Such an assessment is better than direct point-by-point comparison with station observations. It reveals some structural issues of the models in simulating snow depths and its insulation effects on soil temperature. The results from the observation stations are interesting as well. The data source is solid, the results and analysis are detailed and well presented in most parts. It is worthy to be published.**

We thank Reviewer 1 for the positive feedback on our paper and the useful specific comments that helped improving our manuscript. Please find **the reviewers comments in bold**, our point-by-point answers without formatting, and *changes to the initial manuscript in Italics*.

**Specific comments**

**1. The authors put several lines in abstract about near-surface permafrost. However permafrost results were not described in results and discussion sections, and it is only mentioned in summary and conclusion section. A somewhat proportional amount of description (in terms of length or importance) should be given in results and discussion sections so that it can be included in the summary and abstract. You need to add at least one paragraph about permafrost in the result (as suggested below) or in discussion sections.**

Done. We agree and accordingly we add an additional section “*5. Permafrost area*” in the results section in the paper which include a new table *Table 4*. This new table shows the simulated permafrost area for the nine models.

The new section “5. Permafrost area” reads: “*Snow cover plays an important role in modulating the variations of soil thermodynamics, and hence near-surface permafrost extent (e.g., Park et al., 2015). Here we evaluate if there is a simple relationship between the simulated Northern hemisphere permafrost area and the sophistication and ability of the snow insulation component in the LSM to match observed snow packs. The simulated near-surface permafrost area varies greatly across the nine models in the hindcast simulation (1960-2009; Table 4). Some of the better performing snow insulation effect models (CLM4.5, JULES) simulate a near-surface permafrost area of 13.19 to 15.77 million km2, which is comparable with the IPA map estimate (16.2 million km2) (Brown et al., 1997; Slater and Lawrence, 2013). CoLM and ORCHIDEE, identified as reasonable models with respect to snow insulation, simulate much lower (7.62 million km2) and higher (20.01 million km2) areas, respectively. The main deficiency of CoLM is its too small soil depth (3.4 m) compared with CLM4.5 (45.1 m) despite having very similar snow modules (Table 1). However, ISBA, one of the two models that showed rather limited skill in representing snow insulation effects, also simulates the highest permafrost area (20.86 million km2). This is inconsistent with previous studies (e.g., Vavrus, 2007; Koven et al., 2013) which concluded that the**first-order control on modelled near-surface permafrost distribution is the representation of the air-to-surface soil temperature difference. Table 4 shows that the situation is more complex and that snow insulation simulation is not the dominant factor in a good permafrost extent simulation. When the land surface models having poor snow models are eliminated, the remaining models’ simulated permafrost area show little or no relationship with the performance of the snow insulation component, because several other factors such as differences in the treatment of soil organic matter, soil hydrology, surface energy calculations, model soil depth, and vegetation also provide important controls on simulated permafrost distribution (e.g., Marchenko and Etzelmüller, 2013).****”***

Accordingly, we shortened the permafrost part in the “6. Summary and conclusions” section. The according paragraph reads: “*Snow and its insulation effects are critical for accurately simulating soil temperature and permafrost in high latitudes. The simulated near-surface permafrost area varies greatly across the nine models (from 7.62 to 20.86 million km2). However, it is hard to find a clear relationship between the performance of the snow insulation in the models and the simulated area of permafrost, because several other factors e.g. related to soil depth and properties and vegetation cover also provide important controls on simulated permafrost distribution.”*

**2. P.9: Before analyzing the Tair-dsnow-Tsoil relationship, it would be interesting to briefly describe the modeled distribution and errors in snow depth and soil temperature comparing with observations in Russia. The section 4.1.2 about snow depth can be moved to here (table 3 in supplementary can be moved to here as well), and add something similar about the soil temperature. With the soil temperature results, you may add the results of permafrost extent and distribution as you mentioned in the summary and abstract. You may well aware and it is worthwhile to emphasize that the simulated snow depth and soil temperature could be influenced by inputs of the model, and the station observations have limitations in spatial coverage (covers only part of Russia, and may not well represent the grids). However, the response patterns of Tsoil and Tsoil- Tair to dsnow should be consistent and can reveal deeper structural issues of the models.**

We appreciate the effort and thought of the referee to suggest some reorganization of the paper structure. However, we do want to keep it as is principally because the main scope of the paper is to evaluate the relationship between air and soil temperatures and its modulation by snow depth and climate regimes. Therefore, we present in the “Results” (section 3) the results for the relationships. This is then followed by the in-depth discussion of the indicated across-model differences in the relationships and its influencing factors (section 4). Here, snow depth comes into play and is discussed in detail (section 4.1.2). There we also discuss the relationship between the input precipitation/snowfall and simulated snow depth. We agree that snow depth is important for the snow insulation effect; therefore we followed your suggestion and show now for the snow depth both the spatial patterns (Fig. 6) and the station-based bias statistics. This means we moved earlier SI Tab. 3 into the new *Table 3*. In accordance with this, we moved also the ΔT bias statistics into the text (means we moved the whole earlier SI Tab. 3 to the new *Tab. 3*), because ΔT is also in detail discussed in section 4. We also followed your suggestion to present the permafrost extent (new *section 5* and new *Tab. 4*); see our above answer to your related comment #1). We do present information about soil temperature biases. Actually, the presented bias statistics for both air temperature (SI Tab.2) and for ΔT (*Tab.3*) give the information about soil temperature bias. An additional table in the manuscript would be redundant information. For your convenience, we add here explicitly the Tsoil bias table:

|  |  |  |  |
| --- | --- | --- | --- |
| Tsoil bias statistics (n=479); StDevObs=1.5 K | | | |
|  | BIAS (K) | StDev (K) | RMSE (K) |
| CLM45 | -2,4 | 1,2 | 3,5 |
| CoLM | 1,8 | 1,0 | 2,9 |
| ISBA | -9,4 | 1,8 | 9,9 |
| JULES | -3,1 | 2,6 | 4,5 |
| LPJ-GUESS | -1,0 | 0,9 | 3,2 |
| Miroc-ESM | -2,5 | 2,1 | 4,8 |
| ORCHIDEE | -6,1 | 1,8 | 6,6 |
| UVic | -6,2 | 1,8 | 7,0 |
| UW-VIC | -1,8 | 1,4 | 4,0 |

Further, Figs. 4 and 5 present also information of the simulated Tsoil in comparison to station observations. For example, both figures clearly show the strongest underestimation of Tsoil in ISBA, as shown in the table above too. Finally, we emphasize that the focus of the paper is on the relationships or functional behaviors. Even if Tsoil is biased, the relationship between Tsoil and Tair can be well represented compared with observations. For example, CLM45 and JULES have a cold bias in Tsoil (too cold by ca. 3 K), but can represent the dependency of the Tsoil-Tair relationship on snow depth regime well (Fig.4; Tab. 2).

Yes, the station data set covers the Russian Arctic. We focus on this because this data set was compiled within PCN project which initiated this model intercomparison study. However it is important to emphasize that the spatial coverage of the 579 stations reporting snow depth, 268 stations reporting simultaneously air and soil temperatures and snow depth, and 518 stations reporting air temperature is quite good (see Figs. 3, 6, SI figures) to cover the model grid boxes of 0.5x0.5 deg. And indeed, the presented response patterns allow a much better assessment than direct point-by-point comparison with station observations. Therefore we have chosen this approach. We follow your suggestion and highlight this better by adding a sentence in the “Summary and conclusions” section: “*The presented relation diagrams of Tsoil and the difference of Tsoil-Tair to snow depth allow a much better assessment to reveal structural issues of the models than a direct point-by-point comparison with station observations*.”

**3. P.8, L.11-13: “We assume that …in winter”. I feel such an assumption is not necessary. The effects of soil moisture and texture do have effects but is much smaller than that of snow. You may revise it to “The effects of other factors on ∆T are much smaller than that of snow” or delete the sentence.**

We agree and have revised this to: “*In winter, the effects of other factors (e.g. soil moisture, texture) on ∆T are much smaller than that of snow*.”

**4. P.12, L.2-5: This sentence does not connect well with the previous one (why LPJ-GUESS produces very low correlation coefficients). In addition, the meaning of the sentence is problematic. The correlation between the snowfall and its simulated snow depth and soil temperature should be somewhat consistent. As you indicated in section 4, the effects of inputs are limited.**

We agree and deleted this sentence.

**P.12, L.21-24: “the average … of Fig. 4.” 1) The authors seem like to provide a single criterion (one ratio) to assess the behavior of the models. Observations show clearly the difference between deep and shallow snow conditions. It would be better to assess the models for both deep and snow conditions, and Fig. 4 already show such results. 2)In this paragraph, the “stronger relationship” means “higher correlation coefficient” or “larger slope in the regression equations”? The term “gradient” used in the abstract and here actually means the slope of the regression between Tsoil and Tair. Gradient between Tsoil and Tair can be misunderstood as changes of temperature from soil to air. Probably it is better to indicate its true meaning (slope of the regression, or the ratio between Tsoil and Tair in winter). 3) It is very similar to the freezing season n-factor used in permafrost modeling. You may compare to the winter n-factors used by others.**

1) It seems to us that there is probably a misunderstanding. Indeed, this whole second paragraph in section 3.2 evaluates the behavior of the model’s Tair vs.Tsoil relation under both thick and thin snow conditions. And, we assess the simulated relationships by different measures. First, we compare the simulated slopes of the Tair vs.Tsoi relationship under thin and thick snow with the according slopes from the observed relationships. This quantification is given in Table 2. Second, we calculate the RMSE between the observed and modeled relationship. These numbers are given in each model panel in Fig. 4 for both thin and thick snow. And third, we calculate the ratio of the slopes under these two snow regimes (ratio of slope under shallow snow divided by that of thick snow). All three criteria give a solid evaluation of the models relationships and they quantitatively confirm each other by arriving at the same conclusion: some specific model’s behavior under thick and thin snow is in agreement, other models cannot reproduce the observation.

It seems we were not enough clear and improved this paragraph accordingly. It reads now:

*Figure 4 clearly shows that some models (CoLM, CLM45, JULES) can capture this modification of the Tair-Tsoil relation by snow depth regime well. Their regression slopes for thick and thin snow are well separated and in agreement with those from the observed relationship (Table 2). The RMSE of their modeled Tsoil vs. Tair relationships from observations is smaller than 4°C. These models better reproduce the observed ΔT vs. dsnow relationship. Other models (LPJ-GUESS, MIROC-ESM, ORCHIDEE) strongly underestimate the increase of the Tsoil vs. Tair regression slope for decreasing snow depth. They also produce a regression slope for thick snow more than twice as large as observations. Two models (ISBA, UVic) fail here and do not show any sensitivity in the Tair-Tsoil relation to snow conditions (Fig.4, Tab.2). Another measure quantitatively confirms the same models behavior: The observed average dsnow in the shallow snow regime is 13.7 cm and that for the thick snow regime is 58.5 cm, so we would expect, if near-surface air temperature and conductivities were equal in both snow depth classes, a ratio between the slopes for shallow and thick snow of 4.3. CLM4.5, CoLM, and JULES reproduce this observed variation in the Tsoil vs Tair relation better than others (Table2). JULES and CoLM indicate a factor of 4 change, while CLM4.5 indicates a factor of 2 change. Other models (LPJ-GUESS, MIROC-ESM, ORCHIDEE) strongly underestimate the increase of the regression slope for decreasing snow depth; they simulate only a factor change of about 1.5. The two models that had also unrealistic ΔT vs dsnow relationships (ISBA, UVic) also fail in this evaluation of their Tsoil vs Tair relationship. They simulate a too strong sensitivity of Tsoil to Tair (regression slopes larger than 0.9°C/°C, R2>0.7; Table 2) that are almost completely independent of the snow depth regimes, particularly in ISBA, which is not consistent with observations. These models' spatial correlation patterns between Tsoil and Tair also differ greatly from the observations and the other models (SI Fig. 3) and show very high positive correlation (r > 0.8) in most regions, as may be expected from the large regression slope shown in Fig. 4. The RMSE of their modeled Tsoil vs Tair relationships from observations reaches ca. 10°C.*

2) We agree and we changed the wording through all the manuscript. We use either “larger slope in the regression between Tsoil and Tair” or “larger regression slope”, or “stronger sensitivity of Tsoil to Tair”.

3) Yes, we agree and included one paragraph “*This is consistent with observations that the mean freezing n-factor (the ratio of freezing degree days at the ground surface to air freezing degree days) is high at sites where the snow cover is thin or absent, and low at sites where the snow cover is thick (e.g., for Yukon Territory in Canada; Karunaratne and Burn, 2003)*.”

**Minor points**

**P.3, L.2: revise “modelling” to “modeling”**

Done.

**P.3, L.6: replace “as expressed by” to “in the”, delete the two commas around “(∆T)”.**

Done.

**P.4, L.14: references are needed at the end of “… soil temperature” to support the treatment.**

Done. This has been clarified; this sentence is related to the sentence and references before.

**P.5, L.24-25: “these simulated relationships”: it is not clear what do you mean about “these relationships” without read the entire paper.**

Done. We revised it to: “*What is the range of the simulated air-soil temperature relationship across the model ensemble?*”

**P.6, L.10: “divided in 14 layers”, revise “in” to “into”**

Done.

**P.8, L.31-33: “the sentences “We illustrate … 3 regimes.” seems can be simplified as “We illustrate the dependence of Tsoil on Tair for three Tair ranges”.**

Done. It reads now “*We illustrate the dependence of ΔT and Tsoil on dsnow for three Tair ranges*.”

**You used “Larger snow depth”, “higher snow depth”. Probably can be revised as “thicker snow”, or “when the snow is deep”, or “with increase in snow depth” etc.**

Done.

**P.9, L.10: You do not need to redefine the symbols of ∆T and dsnow here. Actually, I feel you can replace the word descriptions by the symbols in many places, at least do not need to mention both the word description and symbols.**

We would like to keep it here at the beginning of the “Results” section, just for reminder the reader. We also keep it in first paragraph of “Summary and conclusions”. But, except this, we followed your suggestion and replaced the word descriptions by the symbols in the manuscript.

**P.9, L.29, L.31: ∆T/dsnow do mean a ratio as shown in Table 2. Revise “∆T/dsnow relationship”, to “∆T-dsnow relationship” here and many other places.**

Done. We do agree. We replaced “/” by “vs.” at all the respective places in the whole manuscript.

**P.9, L.31: “Figure 2 views the ∆T/dsnow relationship in the complementary form of the PDFS of …”, revised as “Figure 2 shows the ∆T-dsnow relationship in a complementary form using the PDFS of ……”**

Done. We also change “∆T-dsnow” to “∆T vs. dsnow”. See above answer.

**P.10, L.6: “the better models”, revise to “the five successful models”**

Done.

**P.10, L.11: “that affect the air soil temperature difference”, revise to “that affect the thermal conductivity of the snow”.**

Done.

**P.11, L.25: “reasonable pattern correlation coefficient with observations”, probably means “reasonable spatial pattern of correlation coefficient comparing to that of the observations”.**

Done. Actually, here we refer to the similarity between the simulated spatial patterns and the spatial pattern from observations. For this, we calculated the spatial pattern correlation coefficient. To be more precise, we improved this sentence to “… *show a reasonable spatial pattern correlation coefficient* …”.

**P.11, L.34: “a reverse pattern correlation than observations” revise to “a reverse spatial pattern comparing to that of the observations”**

Done.

**P.12, L.6: “emphasizing the weakening role of snow depth for Tsoil under thick snow**

**conditions”. Probably should be “emphasizing the reduced sensitivity of Tsoil to snow**

**depth under thick snow conditions”**

Done.

**Figures: revise “AirT” to “Tair”**

Done.